



Statistik

Lehr- und Handbuch
der angewandten Statistik

von

o. Prof. Dr. Joachim Hartung
Thomas-L.-Saaty-Preisträger

Dr. Bärbel Elpelt

Fakultät Statistik der
Technischen Universität Dortmund

Prof. Dr. Karl-Heinz Klösener
Fachhochschule Trier

15., überarbeitete und wesentlich erweiterte Auflage

Oldenbourg Verlag München

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <<http://dnb.d-nb.de>> abrufbar.

© 2009 Oldenbourg Wissenschaftsverlag GmbH
Rosenheimer Straße 145, D-81671 München
Telefon: (089) 45051-0
oldenbourg.de

Das Werk einschließlich aller Abbildungen ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Lektorat: Wirtschafts- und Sozialwissenschaften, wiso@oldenbourg.de
Herstellung: Anna Grosser
Coverentwurf: Kochan & Partner, München
Gedruckt auf säure- und chlорfreiem Papier
Gesamtherstellung: Druckhaus „Thomas Müntzer“ GmbH, Bad Langensalza

ISBN 978-3-486-59028-9

Kapitelverzeichnis

Einführung und Grundlagen	1
KAPITEL I: Aufbereitung und Darstellung von Datenmaterial – Deskriptive Statistik	15
KAPITEL II: Grundbegriffe der Wahrscheinlichkeitsrechnung	91
KAPITEL III: Statistische Schlußweisen	123
KAPITEL IV: Spezielle Verteilungen und statistische Schlüsse über Kenngrößen von Verteilungen mittels einer Meßreihe (Stichprobe)	143
KAPITEL V: Aspekte der Datengewinnung – Stichprobentheorie, Meßfehler, Ausreißertests, Datentransformationen, Versuchsplanung, Klinische Versuche, Skalierung	269
KAPITEL VI: Qualitätskontrolle	381
KAPITEL VII: Analyse diskreten Datenmaterials in Form von Kontingenztafeln	407
KAPITEL VIII: Vergleich zweier Meßreihen (Stichproben)	505
KAPITEL IX: Die Korrelation von Merkmalen	545
KAPITEL X: Regressionsanalyse	569
KAPITEL XI: Varianzanalyse	609
KAPITEL XII: Zeitreihenanalyse	637
KAPITEL XIII: Analyse von Lebensdauern und Zuverlässigkeit von Systemen	745
KAPITEL XIV: Explorative Datenanalyse (EDA) und Robuste Verfahren	825
KAPITEL XV: Meta-Analyse zur Kombination von Studien, Experimenten und Prognosen	887
ANHANG	1023

Inhaltsverzeichnis

<i>Vorworte</i>	XXVII
<i>Ein kurzer Überblick</i>	XXXI
Einführung und Grundlagen	1
1. Was ist Statistik	1
2. Das Experiment	3
3. Die Erhebung	5
4. Zur Statistik und ihren philosophischen Voraussetzungen	7
5. Zur Geschichte der Statistik	10
Kapitel I: Aufbereitung und Darstellung von Datenmaterial – Descriptive Statistics	15
1. <i>Grundlegende Begriffe und Überblick</i>	15
1.1. Untersuchungseinheiten, Merkmale und Merkmalsausprägungen	15
1.2. Charakterisierung von Merkmalen	16
1.3. Grundgesamtheit und Stichprobe	18
1.4. Überblick über die Methoden der deskriptiven Statistik	19
2. <i>Der Häufigkeitsbegriff</i>	20
2.1. Absolute und relative Häufigkeiten	20
2.2. Die graphische Darstellung von Häufigkeiten	21
2.3. Die empirische Verteilungsfunktion	23
3. <i>Der Häufigkeitsbegriff bei Klassenbildung</i>	24
3.1. Die Klassenbildung	26
3.2. Absolute und relative Häufigkeiten bei Klassenbildung	26
3.3. Die graphische Darstellung von Häufigkeiten bei Klassenbildung	27
3.4. Die empirische Verteilungsfunktion bei Klassenbildung	28
4. <i>Lagemaße von Häufigkeitsverteilungen</i>	31
4.1. Das arithmetische Mittel	31
4.2. Der Median und das α -Quantil	32
4.2.1. Der Median einer Beobachtungsreihe	32
4.2.2. Das α -Quantil einer Beobachtungsreihe	34
4.3. Der Modalwert	35
4.4. Das geometrische Mittel und das harmonische Mittel	35
4.5. Einige Bemerkungen zu den Lagemaßen	37
5. <i>Streuungsmaße von Häufigkeitsverteilungen</i>	40
5.1. Die Spannweite	40
5.2. Der Quartilsabstand	41
5.3. Die mittlere absolute Abweichung vom Median	42
5.4. Varianz, Standardabweichung und Variationskoeffizient	43
5.4.1. Die Varianz	44
5.4.2. Die Standardabweichung	46
5.4.3. Der Variationskoeffizient	47

5.5. Die Schiefe und der Exzeß	47
5.5.1. Die Schiefe einer Häufigkeitsverteilung	47
5.5.2. Der Exzeß einer Häufigkeitsverteilung	49
6. Konzentrationsmaße für Häufigkeitsverteilungen	50
6.1. Die Lorenzkurve	50
6.2. Das Lorenzsche Konzentrationsmaß; der Gini-Koeffizient	52
7. Verhältniszahlen	55
7.1. Gliederungszahlen	55
7.2. Beziehungszahlen	56
7.2.1. Verursachungszahlen	56
7.2.2. Entsprechungszahlen	56
7.3. Indexzahlen	57
7.3.1. Meßzahlen	57
7.3.2. Standardisierung von Meßzahlen, Sterbeziffern	60
7.3.3. Zusammengesetzte Indexzahlen	62
A. Der Wertindex	62
B. Preisindizes nach Laspeyres und nach Paasche	63
C. Ein Beispiel	63
D. Mengenindizes nach Laspeyres und nach Paasche	65
E. Preisbereinigung; Deflationierung	65
F. Preis- und Mengenindizes als gewogene Mittel von Meßzahlen: Subindizes	66
7.3.4. Vergleich von Preisindizes nach Laspeyres und nach Paasche; der Fishersche Idealindex; der Preisindex nach Lowe	70
8. Die empirische Ausfallrate	70
9. Darstellung zweidimensionalen Zahlenmaterials und deskriptive Korrelationsrechnung	72
9.1. Die Kontingenztafel	72
9.2. Der Korrelationskoeffizient nach Bravais-Pearson	73
9.3. Der Fechnersche Korrelationskoeffizient	78
9.4. Der Spearmansche Rangkorrelationskoeffizient	79
9.5. Der Kendallsche Rangkorrelationskoeffizient	81
9.6. Der Yulesche Assoziationskoeffizient für die Vierfeldertafel	82
10. Praktische Berechnung einiger Kenngrößen	83
10.1. Berechnung des arithmetischen Mittels und der Standardabweichung	83
10.2. Berechnung der mittleren absoluten Abweichung vom Median	87
Kapitel II: Grundbegriffe der Wahrscheinlichkeitsrechnung	91
1. Ereignisse und Zufallsexperimente	91
2. Wahrscheinlichkeiten	93
3. Kombinatorik und Beispiele für die Berechnung von Laplace-Wahrscheinlichkeiten	96
3.1. Permutationen	96
3.2. Kombinationen	96
3.3. Beispiele zur Berechnung von Laplace-Wahrscheinlichkeiten	97

4. Bedingte Wahrscheinlichkeiten und Unabhängigkeit	98
5. Die Bayessche Formel	102
6. Zufallsvariable und Verteilungen	103
7. Unabhängigkeit und Funktionen von Zufallsvariablen	108
7.1. Unabhängigkeit von Zufallsvariablen	108
7.2. Funktionen von Zufallsvariablen	108
A. Lineare Transformation; Normalverteilung	109
B. Summe; Faltung; Binomialverteilung	109
C. Maximum	111
D. Minimum	111
8. Kenngrößen von Zufallsvariablen	112
8.1. Lageparameter	112
A. Der Erwartungswert	112
B. Der Median und andere Quantile	114
C. Der Modalwert	116
8.2. Streuungsparameter	116
A. Die Varianz und die Standardabweichung; standardisierte Zufallsvariable; Tschebyscheffsche Ungleichung	116
B. Der Variationskoeffizient	117
C. Der Quartilsabstand	118
8.3. Momente von Zufallsvariablen; Schiefe; Exzeß	118
8.4. Kovarianz und Korrelation von Zufallsvariablen	119
9. Grenzwertsätze	121
Kapitel III: Statistische Schlussweisen	123
1. Schätzen von Parametern	124
A. Momentenmethode	126
B. Maximum-Likelihood-Methode	126
C. Methode der kleinsten Quadrate	128
2. Konfidenzintervalle	129
3. Prognose- und Toleranzintervalle	132
4. Statistische Tests	133
5. Beurteilungskriterien für statistische Tests	137
6. Arten von Hypothesen und allgemeine Bemerkungen	138
7. Nichtparametrische (verteilungsfreie) Verfahren	139
8. Zufällige Auswahl, Randomisation	141
9. Notation von Zufallsvariablen	142

Kapitel IV: Spezielle Verteilungen und Statistische Schlüsse über Kenngrößen von Verteilungen mittels einer Meßreihe (Stichprobe)	143
<i>1. Die Normalverteilung und daraus abgeleitete Verteilungen</i>	143
1.1. Die Normalverteilung und ihre Bedeutung	143
1.2. Einige in enger Beziehung zur Normalverteilung stehende Verteilungen	148
1.2.1. Aus der Normalverteilung abgeleitete Verteilungen	148
A. Die gestutzte Normalverteilung	148
B. Die Lognormalverteilung	151
1.2.2. Prüfverteilungen	152
A. Die χ^2 -Verteilung	152
B. Die t-Verteilung	154
C. Die F-Verteilung	156
1.3. Punktschätzungen und Konfidenz-, Prognose- und Toleranzintervalle bei normalverteilter Grundgesamtheit	157
1.3.1. Schätzen der Parameter μ und σ^2	157
1.3.2. Konfidenzintervalle für μ , σ^2 und σ	160
1.3.3. Prognose- und Toleranzintervalle	163
1.4. Bestimmung von benötigten Stichprobenumfängen bei Intervallschätzungen	166
1.4.1. Einhaltung absoluter Genauigkeiten	166
1.4.2. Einhaltung prozentualer Genauigkeiten	173
A. Das Variationszahlverfahren	173
B. Das Streuzahlverfahren	175
1.5. Testen von Parameter-Hypothesen und Bestimmung des benötigten Stichprobenumfangs	178
1.5.1. Testen von Hypothesen über die Parameter einer normalverteilten Grundgesamtheit	178
A. Hypothesen über den Mittelwert μ	178
B. Hypothesen über die Varianz σ^2	179
1.5.2. Bestimmung des Stichprobenumfangs n beim Testen von Hypothesen über den Erwartungswert μ einer normalverteilten Grundgesamtheit bei vorgegebenem Fehler 1. Art α und Fehler 2. Art β	181
1.6. Anpassungstests an die Normalverteilung	182
A. Der χ^2 -Anpassungstest	182
B. Der Kolmogoroff-Smirnov-Anpassungstest	183
C. Ein Beispiel	186
C1. χ^2 -Anpassungstest	186
C2. Kolmogoroff-Smirnov-Anpassungstest	187
1.7. Weitere Verfahren zum Testen von Normalverteilungshypothesen	189
1.7.1. Test auf Schiefe und Exzeß	189
1.7.2. Überprüfung der Normalverteilungsannahme mit Hilfe von Wahrscheinlichkeitspapier	190
<i>2. Die Gleichverteilung und die Dreiecksverteilung</i>	192
2.1. Die stetige Gleichverteilung	192
2.1.1. Die eindimensionale Gleichverteilung und ihre Anwendung in der Computersimulation	192
2.1.2. Die zweidimensionale Gleichverteilung	194
2.2. Die Dreiecksverteilung	195

2.3. Punkt- und Intervallschätzungen für die Gleichverteilung	197
2.4. Der χ^2 -Anpassungstest für die Gleichverteilung	198
3. <i>Einige diskrete Verteilungen</i>	199
3.1. Die Binomialverteilung	199
3.1.1. Punkt- und Intervallschätzung des Parameters p	202
3.1.2. Testen von Hypothesen über den Parameter p	205
3.1.3. Bestimmung des Stichprobenumfangs n beim Testen von Hypothesen über den Parameter p einer Binomialverteilung bei vorgegebenen Fehlern 1. und 2. Art	206
3.2. Die hypergeometrische Verteilung	207
3.2.1. Punktschätzungen für die Hypergeometrische Verteilung	208
3.3. Die Multinomialverteilung	209
3.3.1. Konfidenzbereich für die Multinomialverteilung	211
3.4. Die Poissonverteilung	212
3.4.1. Punkt- und Intervallschätzung für den Parameter λ einer $Po(\lambda)$ -Verteilung	214
3.4.2. Test über den Parameter λ einer $Po(\lambda)$ -Verteilung	214
3.4.3. Der χ^2 -Anpassungstest für die Poissonverteilung	216
4. <i>Einige Lebensdauerverteilungen</i>	218
4.1. Die Exponentialverteilung	219
4.1.1. Punkt- und Intervallschätzung für den Parameter einer $Ex(\lambda)$ -Verteilung	220
4.1.2. Tests von Hypothesen über den Parameter einer $Ex(\lambda)$ -Verteilung	223
4.1.3. Der χ^2 -Anpassungstest für die Exponentialverteilung	225
4.1.4. Der Kolmogoroff-Smirnov-Anpassungstest für die Exponentialverteilung	226
4.2. Die Weibullverteilung	230
4.2.1. Schätzen der Parameter α und β der Weibullverteilung	232
4.3. Die IDB-Verteilung (Hjorth-Verteilung)	232
4.3.1. Schätzen der Parameter α , β und γ der IDB-Verteilung	234
4.4. Die Erlang-n-Verteilung	234
5. <i>Nichtparametrische Test- und Schätzmethoden im Ein-Stichproben-Fall</i>	235
5.1. Konfidenzintervalle und Tests für Quantile	235
5.1.1. Ein Konfidenzintervall für Quantile	236
5.1.2. Tests für Quantile	237
5.2. Nichtparametrische Toleranzintervalle	238
5.3. Konfidenzstreifen für eine unbekannte Verteilungsfunktion	240
5.4. Nichtparametrische Einstichproben-Lokationsvergleiche und Tests auf Trend	242
5.4.1. Der Zeichentest	242
5.4.2. Der Vorzeichenrangtest nach Wilcoxon	243
5.4.3. Tests auf Trend	247
A. Der Test von Cox und Stuart	247
B. Der Test nach Mann	249
6. <i>Sequentielle Quotiententests</i>	251
6.1. Der sequentielle Quotiententest für die Binomialverteilung	252
6.2. Sequentieller Quotiententest für den Erwartungswert einer Normalverteilung	259
6.3. Sequentieller Quotiententest für eine Exponentialverteilung	261

**Kapitel V: Aspekte der Datengewinnung – Stichprobentheorie,
Meßfehler, Ausreißertests, Datentransformationen,
Versuchsplanung, Klinische Versuche, Skalierung269**

<i>1. Abriß der klassischen Stichprobentheorie am Beispiel der Inventur auf Stichprobenbasis</i>	269
1.1. Die Stichprobe	270
1.2. Überlegungen und Vorgehensweisen bei Stichprobenerhebungen	271
1.3. Verteilungsannahmen bei Stichprobenerhebungen	272
1.4. Die einfache Zufallsauswahl	273
1.5. Geschichtete Zufallsauswahl	278
1.5.1. Die optimale Aufteilung (Neyman-Tschuprow-Aufteilung)	282
1.5.2. Die proportionale Aufteilung	285
1.5.3. Aufteilung nach Auswahl der Stichprobe	286
1.5.4. Genauigkeitsvergleiche	287
1.6. Klumpenstichprobenverfahren	288
1.6.1. Einstufige Auswahlverfahren	289
1.6.2. Mehrstufige Auswahlverfahren	291
2. Weitere Verfahren der Stichprobentheorie	292
2.1. Ziehen mit und ohne Zurücklegen	292
2.2. Schätzen von Anteilen	293
2.3. Die systematische Stichprobe	295
2.4. Stichproben mit ungleichen Auswahlwahrscheinlichkeiten	297
2.5. Die Formel von Horwitz-Thompson	299
2.6. Verhältnis-, Differenzen- und Regressionsschätzung, gebundene und freie Hochrechnung	300
2.6.1. Die Verhältnisschätzung	300
2.6.2. Die Differenzenschätzung	303
2.6.3. Die Regressionsschätzung	303
2.7. Zweiphasige Problemstellungen	304
3. Probleme bei der praktischen Durchführung einer Erhebung	305
3.1. Die Abgrenzung der Grundgesamtheit	305
3.2. Endliche und unendliche sowie fiktive Grundgesamtheiten	306
3.3. Auswahltechniken und Erhebungsprobleme	307
3.4. Probleme im Zusammenhang mit Befragungen	309
3.4.1. Fragestellung und Fragebogen	309
3.4.2. Typen von Befragungen	310
3.4.3. Das Problem der Nichtbeantwortung	311
3.5. Vergleich zwischen den Schichten	313
3.6. Stichprobenverfahren in der Marktforschung	314
3.6.1. Marktforschung – Zielsetzungen und Problemstellungen	314
3.6.2. Beurteilungsstichproben in der Marktforschung	316
A. Typische Auswahl	316
B. Auswahl nach dem Konzentrationsprinzip	317
C. Quotenauswahl	318
3.7. Die Bedeutung der Stichprobenverfahren	320
4. Theorie der Meßfehler, Ausreißertests, Datentransformationen	320
4.1. Der Meßfehler bei der Datengewinnung	321
4.2. Das Gaußsche Fehlerfortpflanzungsgesetz	326

4.3. Kontrolle und Erfassung von Meßfehlern	332
4.3.1. Kontrolle und Ermittlung systematischer Fehler	332
4.3.2. Die Verwendung von Kontrollkarten	335
4.3.3. Ringversuche: Inter- und Intralaboratorische Vergleiche	337
4.3.4. Präzision, Spezifität, Richtigkeit und Sensibilität von Meß- und Analyseverfahren	341
4.4. Das Ausreißerproblem	343
A. Der David-Hartley-Pearson-Test	344
B. Der Grubbs-Test	345
C. Dixon's r-Statistiken	346
D. Test auf ein Ausreißerpaaar	347
4.5. Transformationen	349
A. Die reziproke Transformation	349
B. Die Wurzel-Transformation	349
C. Die Logarithmische Transformation	351
D. Die Box-Cox-Transformation	352
E. Die Arcus-Sinus-Transformation	352
F. Die Fishersche Z-Transformation	354
5. Allgemeine Aspekte der Planung von Versuchen	354
6. Anlage von klinischen Versuchen	363
6.1. Ethische Probleme bei klinischen Versuchen	366
6.2. Auswahl und Zuordnung von Versuchspersonen	367
6.2.1. Die retrospektive Zuordnung	368
6.2.2. Zuordnung auf freiwilliger Basis	369
6.2.3. Pseudaleatorische und aleatorische Zuordnung	370
6.2.4. Einige weitere Zuordnungsverfahren	370
6.3. Die Vergleichbarkeit der Versuchsergebnisse	370
6.4. Auto- und Heterosuggestion, Blindversuche	371
6.5. Sequentielle Studien	372
6.6. Ein Beispiel	373
7. Skalierung von Merkmalsausprägungen und Testergebnissen	374
Kapitel VI: Qualitätskontrolle	381
1. Stichprobenpläne in der Eingangs- und Endkontrolle	381
1.1. Einfache Stichprobenpläne für qualitative Merkmale	383
A. Vorgabe zweier Punkte der Operationscharakteristik	384
B. Vorgabe des Indifferenzpunktes und der Steilheit	387
1.2. Mehrfache und sequentielle Stichprobenpläne für qualitative Merkmale	389
A. Doppel Stichprobenpläne	390
B. Sequentielle Stichprobenpläne	392
1.3. Stichprobenpläne für quantitative Merkmale	395
2. Laufende Kontrolle der Produktion (Kontrollkarten)	401
2.1. Laufende Kontrolle bei quantitativen Merkmalen	401
2.2. Laufende Kontrolle bei qualitativen Merkmalen	404
3. Kontinuierliche Stichprobenpläne	406

Kapitel VII: Analyse diskreten Datenmaterials in Form von Kontingenztafeln	407
<i>1. Die 2 x 2-Felder-Tafel</i>	411
1.1. Hypothesen für die 2 x 2-Felder-Tafel	412
A. Die Unabhängigkeitshypothese	412
B. Die Homogenitätshypothese	412
C. Beziehungen zwischen den Hypothesen	412
1.2. Tests auf Unabhängigkeit in der 2 x 2-Tafel	413
1.2.1. Der χ^2 -Test	413
1.2.2. Exakte Tests	414
A. Ein exakter Test, der auf größere Kontingenztafeln übertragbar ist ..	414
B. Der exakte Test von Fisher	416
1.2.3. Einseitige Hypothesen in 2 x 2-Tafeln	416
1.3. Tests auf Homogenität in der 2 x 2-Tafel	418
1.3.1. Die Durchführung der Tests	418
1.3.2. Der erforderliche Stichprobenumfang	419
A. Die Formel mittels Approximation der Gütfunktion des χ^2 -Tests ..	419
B. Die Arcus-Sinus-Formel	420
C. Die Formel nach Casagrande/Pike/Smith	420
D. Exakte Stichprobenumfänge	421
1.4. Tests auf Symmetrie in der 2 x 2-Tafel	422
1.4.1. Der McNemar-Test	423
1.4.2. Cochran's Q	423
<i>2. Loglineare Modelle und Tests für r x s-Tafeln</i>	425
2.1. Das loglineare Modell für die r x s-Tafel	425
2.1.1. Entwicklung des Modells am Beispiel der 2 x 2-Tafel	425
A. Aufstellung des Modells	425
B. Schätzen der Parameter	427
C. Die approximative Varianz der Schätzungen	428
D. Die Interpretation der Parameter	428
2.1.2. Das Modell für die allgemeine r x s-Tafel	429
2.2. Hypothesen und Tests in r x s-Tafeln	432
2.2.1. Einige Hypothesen für r x s-Tafeln	433
A. Die Unabhängigkeits- bzw. Homogenitätshypothese	433
B. Die bedingte Gleichverteilungshypothese	434
C. Die totale Gleichverteilungshypothese	434
2.2.2. Einige Testverfahren für r x s-Tafeln	435
A. χ^2 - und Likelihood-Quotienten-Test	435
B. Die Statistiken τ_A und τ_B	439
C. Der Test auf Symmetrie nach Bowker	440
<i>3. Assoziationsmaße für 2 x 2 und r x s-Tafeln</i>	442
3.1. Assoziationsmaße in der 2 x 2-Kontingenztafel	442
3.1.1. Assoziationsmaße, die in Beziehung zum cross-product ratio q stehen ..	442
A. Die Eigenschaften des cross-product ratio	442
B. Der Q-Koeffizient von Yule	443
C. Der Verbundenheitskoeffizient von Yule	444

D. Punkt- und Intervallschätzungen für die Yuleschen Assoziationsmaße	444
E. Eigenschaften der Yuleschen Assoziationsmaße	446
3.1.2. Assoziationsmaße, die in Beziehung zum Korrelationskoeffizienten Q stehen	446
A. Der Korrelationskoeffizient Q (Phikoeffizient) und seine Eigenschaften	446
B. Der Pearsonsche Kontingenzkoeffizient	449
3.2. Assoziationsmaße in allgemeinen 2-dimensionalen Kontingenztafeln	450
3.2.1. Assoziationsmaße, die in Beziehung zur χ^2-Statistik stehen	451
A. Der Pearsonsche Kontingenzkoeffizient für die $r \times s$ -Tafel	451
B. Der korrigierte Pearsonsche Kontingenzkoeffizient	451
C. Das Assoziationsmaß von Tschuprow	451
D. Das Assoziationsmaß von Cramer	452
E. Schätzung der Varianzen der Assoziationsmaße	452
F. Ein Beispiel	452
3.2.2. Die λ- und die τ-Maße	455
A. Die λ -Maße λ_A , λ_B und λ	456
B. Die τ -Maße τ_A , τ_B und τ	459
4. Loglineare Modelle und Tests für mehrdimensionale Kontingenztafeln	464
4.1. Die Parameter des saturierten Modells	465
A. Schätzen der Parameter des saturierten Modells	466
B. Varianz- und Intervallschätzungen für die Parameter des saturierten Modells	471
4.2. Testen von Hypothesen über die Parameter des saturierten Modells	477
4.2.1. Ein iteratives Verfahren zur Schätzung erwarteter Häufigkeiten unter einer Hypothese	477
4.2.2. Die Bestimmung der Freiheitsgrade	485
4.2.3. Die Partitionierung der Teststatistiken	488
5. Verteilungsannahmen, Logit-Modell und Adjustieren bei Kontingenztafeln	492
5.1. Kontingenztafeln und Verteilungen	492
5.1.1. Verteilungsannahmen bei Kontingenztafeln	492
5.1.2. Vergleich der Parameter mehrerer diskreter Verteilungen	495
A. Vergleich der Parameter von s Poissonverteilungen	495
B. Vergleich der Parameter verschiedener Binomialverteilungen	496
C. Vergleich der Parameter mehrerer Multinomialverteilungen	498
5.2. Das Logit-Modell bei Kontingenztafeln	498
5.3. Adjustieren von Kontingenztafeln	501
Kapitel VIII: Vergleich zweier Meßreihen (Stichproben)	505
1. Vergleich zweier unabhängiger Meßreihen	505
1.1. Lokationsvergleiche bei normalverteilter Grundgesamtheit	505
1.1.1. Tests und Konfidenzintervalle bei bekannten Varianzen σ_1^2 und σ_2^2 der Grundgesamtheiten	505
1.1.2. Tests und Konfidenzintervalle bei unbekannten aber gleichen Varianzen σ_1^2 und σ_2^2 der beiden Grundgesamtheiten	508

1.1.3. Tests und Konfidenzintervalle bei unbekannten und ungleichen Varianzen σ_1^2 und σ_2^2 der beiden Grundgesamtheiten	510
1.1.4. Bestimmung von Stichprobenumfängen bei Tests und Konfidenzintervallen	511
1.2. Verteilungsfreie Lokationsvergleiche	513
1.2.1. Der Wilcoxon-Rangsummentest, der U-Test von Mann-Whitney	513
1.2.2. Der Kolmogoroff-Smirnov-Test	520
1.3. Dispersionsvergleiche bei normalverteilten Grundgesamtheiten — Tests und Konfidenzintervalle	524
1.4. Verteilungsfreie Dispersionsvergleiche	526
1.4.1. Der Test von Ansari-Bradley-Freund, der Siegel-Tukey-Test	526
1.4.2. Der Test von Moses	529
1.5. Test auf Trend	531
2. <i>Vergleich zweier abhängiger Meßreihen</i>	533
2.1. Lokationsvergleiche bei normalverteilten Grundgesamtheiten — Tests, Konfidenzintervalle und Bestimmung der Stichprobenumfänge	534
A. Die Varianz σ_d^2 ist bekannt	534
B. Die Varianz σ_d^2 ist unbekannt	536
2.2. Dispersionsvergleiche bei normalverteilten Grundgesamtheiten	538
2.3. Verteilungsfreie Lokationsvergleiche	539
Kapitel IX: Die Korrelation von Merkmalen	545
1. <i>Die Korrelation zweier normalverteilter Merkmale</i>	546
2. <i>Die Rangkorrelation zweier Merkmale</i>	553
2.1. Der Spearmansche Rangkorrelationskoeffizient	553
2.2. Der Kendallsche Rangkorrelationskoeffizient	559
3. <i>Die partielle Korrelation</i>	561
3.1. Die partielle Korrelation zwischen normalverteilten Merkmalen	561
3.2. Der partielle Rangkorrelationskoeffizient nach Kendall	563
4. <i>Die bi-partielle Korrelation</i>	563
5. <i>Die multiple Korrelation</i>	564
6. <i>Ein Test auf Unabhängigkeit von p Meßreihen</i>	567
Kapitel X: Regressionsanalyse	569
1. <i>Lineare Regression</i>	573
1.1. Die Methode der kleinsten Quadrate	574
1.2. Schätzen der Fehlervarianz σ^2	578
1.3. Zusammenhang zwischen Regressions- und Korrelationsrechnung; das Bestimmtheitsmaß	578
1.4. Konfidenzintervalle und Testen von Hypothesen über die unbekannten Parameter α , β und σ^2	580
1.5. Konfidenz- und Prognosestreifen	582

<i>1.6. Regression durch einen vorgegebenen Punkt, Regression ohne Absolutglied (eigentlich-lineare Regression)</i>	584
<i>2. Residualanalyse</i>	585
<i>3. Transformationen auf Linearität</i>	587
<i>4. Nichtlineare Regression und Schätzen des Maximums (Minimums) einer quadratischen Regressionsfunktion</i>	589
<i>5. Multiple Regression</i>	595
<i>6. Regression bei Fehlern in den Variablen</i>	601
A. Die Varianz σ_1^2 ist bekannt	601
B. Die Varianz σ_2^2 ist bekannt	602
C. Der Quotient der Varianzen σ_1^2 und σ_2^2 ist bekannt	603
D. Das Berkson-Modell	604
<i>7. Regressionsgerade nach Wald</i>	605
Kapitel XI: Varianzanalyse	609
<i>1. Vergleich von p unabhängigen Meßreihen (Stichproben) — einfache Varianzanalyse, vollständig randomisierter Versuchsplan</i>	610
1.1. Testen auf signifikante Lokationsunterschiede	611
A. Der F-Test (normalverteilte Grundgesamtheit)	611
B. Der Test von Kruskal und Wallis	613
1.2. Simultane Vergleiche von p Mittelwerten	614
A. Die Tests von Scheffé und Tukey	616
A1. Scheffe-Test	616
A2. Tukey-Test	616
B. Der Test von Steel und Dwass	616
1.3. Dispersionsvergleiche	617
A. Der Bartlett-Test	617
B. Der Levene-Test	617
C. Scheffe's χ^2 -Test	617
1.4. Modellbetrachtung	619
2. Das einfache Blockexperiment	619
2.1. Verfahren bei normalverteilter Grundgesamtheit	620
2.2. Verteilungsfreie Verfahren	622
3. Zweifache Varianzanalyse mit mehreren Beobachtungen pro Faktorstufenkombination (pro Zelle)	624
3.1. Modell mit Wechselwirkungen zwischen den Faktoren A und B	625
3.2. Modell ohne Wechselwirkungen zwischen den Faktoren	627
4. Die Komponenten der Streuung — Modelle der Varianzanalyse mit zufälligen Effekten (Modell II)	629
4.1. Die einfach hierarchische Klassifikation	630
4.2. Ein nicht-klassisches Varianzanalysemmodell aus der Geodäsie	634

Kapitel XII: Zeitreihenanalyse	637
1. <i>Deskriptive Methoden der Zeitreihenanalyse</i>	640
1.1. Die Komponenten einer Zeitreihe	640
1.2. Nichtlineare Trendmodelle — Trendschätzung mittels nicht-linearer Regression	642
1.2.1. Die logistische Funktion	642
1.2.2. Die Mitscherlich-Funktion und die Gompertz-Kurve	648
1.2.3. Die allometrische Funktion	654
1.3. Trend- und Saison-Schätzung bzw. -Elimination durch Glättung bzw. Filterung	660
1.3.1. Gleitende Durchschnitte	660
1.3.2. Polynome und Splines	666
1.3.3. Die Differenzenmethode	668
1.3.4. Exponentielles Glätten	672
1.3.5. Lineare Filter	673
1.4. Autokovarianzen, Autokorrelationen und partielle Autokorrelationen	675
2. <i>Selbsterklärende Zeitreihenmodelle und die Methode von Box und Jenkins</i>	678
2.1. Autoregressive Prozesse (AR-Prozesse)	678
2.2. Moving average Prozesse (MA-Prozesse)	681
2.3. Gemischte Prozesse (ARMA-Prozesse)	682
2.4. Instationäre stochastische Prozesse, ARIMA- und SARIMA-Prozesse, Box-Cox-Transformation	684
2.5. Die Methode von Box und Jenkins	686
2.5.1. Modellidentifikation	686
2.5.2. Schätzen der Modellparameter	688
2.5.3. Modellüberprüfung	690
2.5.4. Prognose und Prognosegüte	691
2.6. Ein Beispiel zur Methode von Box und Jenkins	694
3. <i>Die Spektralanalyse</i>	699
3.1. Komplexe Zahlen	700
3.2. Spektrum und Spektraldichte	701
3.3. Spektraldichten gefilterter Prozesse	703
3.4. Schätzen der Spektraldichte	709
3.4.1. Das Periodogramm und das Stichprobenspektrum	709
3.4.2. Geglättete Spektraldichteschätzungen, Spektral- und Lag-Fenster	711
3.4.3. Ein Beispiel zum Stichprobenspektrum	715
3.5. Die harmonische Analyse einer Zeitreihe	717
3.6. Das Berliner Verfahren zur Saisonbereinigung	722
4. <i>Analyse des Zusammenhangs zweier Zeitreihen</i>	727
4.1. Analyse im Zeitbereich	728
4.2. Analyse im Frequenzbereich — Kreuzspektralanalyse	728
4.2.1. Das Kreuzspektrum	729
4.2.2. Schätzung kreuzspektraler Größen	734
5. <i>Gemischte Regressions-Zeitreihen-Modelle</i>	735

5.1. Regressionsmodelle mit korrelierten Fehlern	736
5.1.1. Allgemeine Vorgehensweise	736
5.1.2. Regressionsmodelle mit AR(1)-Fehler-Prozeß	737
A. Die Cochrane-Orcutt-Methode	738
B. Der Durbin-Watson-Test	740
C. Prognose und Prognosegüte	741
D. Ein Beispiel	741
5.2. Autoregressive Regressionsmodelle	744
6. Tests auf Stationarität einer Zeitreihe [Dickey-Fuller Unit Root Test] und eines kreuzkorrelierten Zeitreihenpanels	744

Kapitel XIII: Analyse von Lebensdauern und Zuverlässigkeit von Systemen 745

1. Die Zuverlässigkeit von Komponenten und Systemen	746
1.1. Die Zuverlässigkeit elementarer Systeme	746
1.2. Zuverlässigkeitsschaltbilder	748
1.3. Darstellung monotoner Systeme mittels minimaler Pfade und minimaler Schnitte	750
1.4. Systemstrukturfunktion und Systemzuverlässigkeitsfunktion	751
1.4.1. Die Strukturfunktion monotoner Systeme	751
A. Bestimmung einer Strukturfunktion mittels disjunktiver Normalform	752
B. Bestimmung einer Strukturfunktion mittels minimaler Pfad- oder Schnittmengen	754
C. Gewinnung des Zuverlässigkeitsschaltbildes zu einer Strukturfunktion	755
1.4.2. Die Zuverlässigkeitsfunktion monotoner Systeme	756
1.5. Stochastische Assoziiertheit von Systemkomponenten	759
1.6. Klassifizierung der Zuverlässigkeitsfunktionen monotoner Systeme mit Komponenten gleicher Zuverlässigkeit	759
1.7. Methoden zur Erhöhung der Zuverlässigkeit – Redundanz bei Komponenten und Systemen, Systeme mit heißer und kalter Reserve	760
1.8. Systeme mit mehr als zwei Zuständen (Multi-State-Systems)	763
1.8.1. Die Bestimmung des Systemzustandes mittels minimaler Pfad- oder Schnittmengen	763
1.8.2. Kritische Pfadvektoren bei Multi-State-Systemen	764
1.9. Die Fehlerbaumanalyse	765
1.10. Systembetrachtungen bei mehrphasigen Missionen	769
1.10.1. Mehrphasige Missionen	769
1.10.2. Phasenzuverlässigkeiten und Missionszuverlässigkeit	770
1.10.3. Die Transformation mehrphasiger Missionen	772
2. Klassen von Lebensdauerverteilungen	774
2.1. IFR- und DFR-Verteilungen	776
2.1.1. Die Verteilungsklassen IFR und DFR	776
2.1.2. IFR- und DFR-Tests	777
A. Der Proschan-Pyke-Test	777
B. Der cttot-Test nach Epstein	778

2.2. NBU- und NWU-Verteilungen	779
2.2.1. Die Verteilungsklassen NBU und NWU	779
2.2.2. NBU- und NWU-Tests: Der Hollander-Proschan-Test	780
2.3. IFRA- und DFRA-Verteilungen	782
2.3.1. Die Verteilungsklassen IFRA und DFRA	782
2.3.2. IFRA- und DFRA-Tests: Der cttot-Test	784
2.4. NBUE- und NWUE-Verteilungen	784
2.4.1. Die Verteilungsklassen NBUE und NWUE	784
2.4.2. NBUE- und NWUE-Tests: Der Hollander-Proschan-Test	784
2.5. Beziehungen zwischen den Verteilungsklassen	786
2.6. Zensierte Lebensdauerprüfungen	787
3. <i>Punkt- und Intervallschätzungen für die Parameter einiger spezieller Lebensdauerverteilungen</i>	788
3.1. Das Modell exponentialverteilter Lebensdauern	788
3.1.1. Punkt- und Intervallschätzungen bei fest vorgegebener Beobachtungsdauer T_0	788
A. Experimente mit Ersetzung ausgefallener Objekte	788
B. Experimente ohne Ersetzung ausgefallener Objekte	789
3.1.2. Punkt- und Intervallschätzungen bei fest vorgegebener Zahl von Ausfällen	789
A. Experimente mit Ersetzung ausgefallener Objekte	789
B. Experimente ohne Ersetzung ausgefallener Objekte	790
3.2. Das Modell Weibull-verteilter Lebensdauern	791
3.2.1. Schätzen der Parameter α und β bei fest vorgegebener Anzahl von Ausfällen ohne Ersetzung	791
3.2.2. Schätzen der Parameter α und β bei unzensierten Lebensdauerprüfungen	793
3.3. Das Modell der Lognormalverteilung	793
3.4. Das Modell der Hjorth-Verteilung (IDB-Verteilung)	794
4. <i>Zur Problematik zeitraffender Zuverlässigkeitstests</i>	794
4.1. Die Extrapolation der zeitraffenden Prüfung an Normalbedingungen	796
4.1.1. Das Eyring-Modell	796
4.1.2. Das Arrhenius-Modell	796
4.1.3. Das verallgemeinerte Eyring-Modell	799
4.2. Screening-Tests, Burn-Ins	803
4.3. Labor- und Einsatzbedingungen	803
4.3.1. Der Einfluß von Strahlungen auf den Alterungsprozeß	803
4.3.2. Wertungsfaktoren für im Labor ermittelte Ausfallraten	804
5. <i>Wartungs- und Erneuerungsüberlegungen</i>	806
5.1. Wartung und Wartbarkeit, Erneuerung	806
A. Die Wartbarkeit von Systemen	806
B. Die Wartung von Systemen	806
C. Die Erneuerung von Systemen	807
5.2. Erneuerungsstrategien, Schranken der Erneuerungsfunktion	807
A. Altersabhängige Erneuerungsstrategie und Gruppenerneuerungsstrategie ..	807
B. Der Erneuerungsprozeß und die Erneuerungsfunktion	807
C. Schranken der Erneuerungsfunktion	808

5.3. Die Zuverlässigkeit von Straßenverkehrssignalanlagen – Ein Beispiel	808
5.3.1. Straßenverkehrssignalanlagen-Systeme	809
A. Das System ohne Reserve	809
B. Das System mit heißer Reserve	810
C. Systeme mit kalter Reserve	811
C1. Das System mit kalter Macro-Reserve	811
C2. Das System mit kalter Micro-Reserve	812
5.3.2. Vergleich der Straßenverkehrssignalanlagen-Systeme	812
5.3.3. Die Wirtschaftlichkeit der Systeme	815
A. Inspektionszeiträume und Mindestzuverlässigkeit der Systeme	817
B. Systemkosten bei Mindestzuverlässigkeit der Systeme	818
6. Verfügbarkeit von Systemen und Instandhaltungsstrategien	820
6.1. Die Verfügbarkeit von Systemen	820
A. Momentane Verfügbarkeit und Dauerverfügbarkeit	820
B. Punkt- und Intervallschätzungen, Testen von Hypothesen über die Dauerverfügbarkeit	820
6.2. Methoden zur Erhöhung der Verfügbarkeit	822
A. Die Redundanzplanung	822
B. Die vorbeugende Instandsetzung	823
B1. Bereitschafts- und Präventivstrategien	823
B2. Periodische, sequentielle und optionale Strategien	823
7. Vergleichende Lebensdauerprüfungen: Log-Rank Test, Wilcoxon-Gehan Test, Tarone-Ware Tests und Cox Proportional Hazards Regression Model	824
8. Schätzen der [Über-] Lebensdauerverteilung: Kaplan-Meier Survival Analysis	824

Kapitel XIV: Explorative Datenanalyse (EDA) und Robuste Verfahren 825

1. Verfahren für einzelne Merkmale in der EDA	827
1.1. Empirische Kenngrößen	827
1.2. Empirische Kenngrößen bei gruppierten Daten	831
1.3. Datentransformationen	832
1.4. Box-Plots	835
1.5. Stamm-und-Blätter-Darstellungen	838
1.6. Histogramm und empirische Verteilungsfunktion	839
1.7. Empirische Dichten	840
1.8. Wurzeldiagramme	844
1.9. Q-Q-Plots zur Überprüfung von Verteilungsannahmen	847
2. Verfahren für zwei Merkmale in der EDA	849
2.1. Glätten zweidimensionaler Punktescharen	850
2.2. Explorative Regressionsgeraden	852
2.3. Linearisieren zweidimensionaler Punktescharen	854
3. Verfahren für mehrdimensionale Daten in der EDA	857
3.1. Der Scatter-Plot	858
3.2. Weitere explorative Verfahren für mehrdimensionale Daten	860

4. Robuste Schätzungen	861
4.1. Charakterisierung von Robustheitseigenschaften	862
4.1.1. Die Sensitivitätskurve	862
4.1.2. Die Einflußkurve	863
4.1.3. Der Bruchpunkt.	864
4.2. Robuste Skalenschätzer	864
4.2.1. Der Median der absoluten Abweichungen vom Median	865
4.2.2. Der Quartilsabstand	866
4.3. M-Schätzer für die Lokation	866
4.3.1. Huber-k-Schätzer	869
4.3.2. Andrews' wave und Tukey's biweight	872
4.3.3. Die Berechnung von M-Schätzern	874
4.3.4. M-Schätzer für einige typische Beispiele	878
4.4. L-Schätzer für die Lokation	880
4.4.1. Das α -getrimmte Mittel	880
4.4.2. Das α -winsorisierte Mittel	881
4.4.3. Das α -Gastwirth-Cohen-Mittel	882
4.5. R-Schätzer für die Lokation	883

**Kapitel XV: Meta-Analyse zur Kombination von Studien,
Experimenten und Prognosen887**

1. Kombination von Wissen — Combining Information	887
1.1. Astronomie	887
1.2. Evidence Based Medicine	888
1.3. Verhaltensforschung — Kombination unabhängiger p-Werte und Tests	891
1.4. Landwirtschaft — Genetik — Heterogenitätsmaße	892
1.5. Gesamtwirtschaftliche Vorhersagen	893
1.6. Betriebswirtschaftliche Prognosen	894
1.7. Stationarität eines kreuz-korrelierten Zeitreihenpanels — Kombination abhängig p-Werte	895
1.8. Umwelt — Physik — Technik	895
1.9. Pareto-Optimale Portfolioselektion im Wertpapiermanagement	897
1.10. Normalverteilungs-Annahmen	899
2. Fixed Effects Model (FE) der Meta-Analyse	899
2.1. Darstellung des Modells und Gauß-Markov-Schätzer	900
2.2. Kombinierte Schätzung des Gesamteffektes	900
2.3. Konfidenzintervalle für den Gesamteffekt	901
2.4. Homogenitätstest von Cochran	901
2.5. Beispiel: Klinische Prüfungen zu Multipler Sklerose MS [Zweiarmige Studien]	902
2.6. Korrelationen	903
2.6.1. Kombination von Korrelationskoeffizienten	904
2.6.2. Beispiel: Korrelation zwischen Rückenspeckdicke und Fleisch-Fett- Verhältnis im Schlachtkörper von Schweinen	904
3. Random Effects Model (RE) der Meta-Analyse	906
3.1. Schätzung des Heterogenitätsparameters	906
3.2. Gesamtergebnis und Konfidenzintervalle	907

3.3. Beispiel: MS-Studien im RE-Model	908
3.4. Beispiel: Landwirtschaft — Besamungsquote von Bullen [Einarmige Studien]	909
3.5. Studien-Cluster [Hartung-Knapp]	910
3.6. Beispiel: Fortsetzung Landwirtschaft	912
4. <i>Diskrete Beobachtungen</i>	913
4.1. Risiko Differenz RD	914
4.2. Odds Ratio OR	915
4.3. Relatives Risiko RR und Noninferiority	915
4.4. Beispiel: Klinische Prüfungen zu Funktioneller Dyspepsie [Zweiarmige Studien]	917
4.5. Odds Ratio Schätzer von Mantel-Haenszel und Peto oder Yusuf-Peto	918
4.6. Poissonverteilte Daten	920
4.7. Beispiel Umwelt — Technik: Asbest-Faser Zählung [Einarmige Studien]	921
5. <i>Überlebenszeiten und Time-to-Event-Daten</i>	924
5.1. Tests auf Gleichheit zweier Survival-Funktionen	925
5.1.1. Der Log-Rank Test	926
5.1.2. Der Wilcoxon-Gehan Test und Tarone-Ware Tests	928
5.1.3. Beispiel: Progressionsfreies Überleben bei Tumorerkrankung	928
5.2. Kombination von vergleichenden Überlebenszeitstudien	930
5.2.1. Kombination der Hazard Ratios	931
5.2.2. Beispiel Progressionsfreies Überleben PFS [3 Studien]: Kombiniertes Konfidenzintervall für den Hazard Ratio	931
5.3. Schätzen der Überlebenswahrscheinlichkeiten und Konfidenzbereiche	932
5.3.1. Kaplan-Meier Schätzer der Survival-Funktion und Bestimmung eines Konfidenzstreifens	933
5.3.2. Schätzen der medianen Überlebenszeit und Konfidenzintervall	933
5.3.3. Beispiel Progressionsfreies Überleben PFS: Bestimmung der Kaplan-Meier PFS-Kurven und der medianen Überlebenszeiten	934
5.4. Kombination von Überlebenszeiten	935
5.4.1. Kombination der Kaplan-Meier Überlebenskurven	936
5.4.2. Kombination der medianen Überlebenszeiten	937
5.4.3. Kombination der Konfidenzintervalle für Überlebenswahrscheinlichkeiten zu markanten Zeitpunkten	938
5.5. Beispiel Progressionsfreies Überleben PFS: Kombination der Kaplan-Meier Überlebenskurven, der medianen Überlebenszeiten und der 6-Monats-Konfidenzintervalle	938
5.6. Cox's Proportional Hazards Regression Model: Vergleich mehrerer Gruppen, Kovariablen und Kombination der Effekte	940
6. <i>Generalized Common Mean (GCM) Problem und Varianzschätzung bei Unkenntnis der Einzelvarianzen</i>	941
6.1. Das GCM-Problem	942
6.2. Direkter Varianzschätzer von Hartung-Böckenhoff-Knapp	942
7. <i>Wirtschafts-Prognosen mit Scoringssystem</i>	943
7.1. Entwicklung eines Scoringssystems	943
7.1.1. Bekannte Referenzwerte	943
7.1.2. Referenzbezug Gruppenmittel	944

7.2. Kombinierte Prognosen mit Scoringpunkten	944
7.3. Beispiel: Prognosen des BIP von sieben Wirtschaftsinstituten	945
<i>8. Combining Forecasts mit geschätzten Zuverlässigen und Korrelationen</i>	947
8.1. Beispiel: Kombinierte Prognose in der Investitionsberatung	947
8.2. Beispiel: Investitionsberatung im RE-Model	948
8.3. Kombination von Prognosen bei geschätzten Korrelationen aber homogenen Prognosen [EFE-Model]	949
8.4. Beispiel: Investitionsberatung im EFE-Model und „Borrowing of Strength“	950
8.5. Kombination von Prognosen bei geschätzten Korrelationen und heterogenen Prognosen [ERE-Model]	951
8.6. Beispiel: Investitionsberatung im ERE-Model	951
8.7. Ein stabiler Hartung-Varianzschätzer für die kombinierte Prognose im ERE-Model	952
8.8. Vergleich der vier Prognose-Modelle im Beispiel zur Investitionsberatung	953
8.8.1. Vergleich der kombinierten Prognosen	953
8.8.2. Vergleich der kombinierten Konfidenzintervalle	953
<i>9. Kombination unabhängiger p-Werte oder unabhängiger Tests</i>	954
9.1. Inverse Normal- und Stouffer-Methode, Probits	954
9.2. Inverse χ^2 - und Fisher-Methode, Fisher-Probits	955
9.3. Heterogenität: Hartung-Knapp-Methode	956
9.4. Beispiel Verhaltensforschung: Konformität bei Männern und Frauen	956
<i>10. Combining Forecasts bei unbekannten Korrelationen und Kombination abhängiger p-Werte oder Tests</i>	958
10.1. Strategie zu Combined Forecasting	959
10.2. Hartung-Methode zur Kombination von abhängigen p-Werten oder Tests	959
10.3. Hartung-Methode für Fisher-Probits	960
10.4. Beispiel: Investitionsprognosen mit unbekannter Korrelation zwischen Betrieb und Berater	960
<i>11. Kombination von kreuz-korrelierten Zeitreihen zu einem Panel</i>	962
11.1. Dickey-Fuller Unit Root Test auf Stationarität einer Zeitreihe	962
11.2. Prüfung eines Panels kreuz-korrelierter Zeitreihen auf Stationarität	964
11.2.1. Kombination der Dickey-Fuller p-Werte nach Hartung	964
11.2.2. Kombination der Dickey-Fuller p-Werte nach Hartung mit den Fisher-Probits	964
11.3. Interpretation des kombinierten Tests auf Stationarität	965
<i>12. Delphi Methode für Vorhersagen und zur Gewinnung Subjektiver Wahrscheinlichkeiten</i>	965
12.1. Anwendungsbereiche der Delphi Methode	965
12.2. Kombination der Einzelprognosen oder subjektiven Risikobeurteilungen mittels der Delphi Methode	966
12.3. Beispiel Investitionsberatung: Cash Flow Vorhersagen ohne Zuverlässigkeitssangaben	967
12.4. Iterative Delphi Methode	967
<i>13. Heterogenitätsmaß für Meta-Analysen: Verallgemeinerter Intraklass- Korrelationskoeffizient</i>	968
13.1. Hartung-Konfidenzintervalle für die Heterogenitätsmaße	969
13.2. Beispiel Landwirtschaft – Genetik: Erblichkeit der Besamungsfähigkeit	970

13.3. Ein weiterer expliziter Hartung-Schätzer für den Heterogenitätsparameter	972
13.4. Median Unverzerrte Hartung-Schätzer für die Heterogenitätsmaße	973
13.5. Zusammenfassendes Beispiel: Zweiarmige Studien zu Multipler Sklerose MS	974
<i>14. Normalverteilte Daten: Exakte Verfahren und spezielle Effektmaße</i>	976
14.1. Effektmaß — Differenz der Mittelwerte	978
14.1.1. Exakte Punkt- und Intervallschätzer im Fixed Effects Model [Hartung-Knapp-Verfahren]	978
14.1.2. Beispiel: Zusammenfassung zweier Noninferiority Studien zu Asthma Bronchiale [Differenz der Mittelwerte]	979
14.2. Effektmaß — Quotient der Mittelwerte [Hartung-Verfahren]	981
14.2.1. Exakte Methoden im Fixed Effects Model für Quotienten	981
14.2.2. Beispiel Asthma Bronchiale: Zusammenfassung zweier Noninferiority Studien [Quotient der Mittelwerte]	982
14.2.3. Random Effects Model für „Quotient der Mittelwerte“	982
14.2.4. Beispiel MS-Studien: RE-Model für „Quotient der Mittelwerte“	984
14.2.5. Alternatives FE-Model für Quotienten mit Beispiel Asthma Bronchiale	986
14.3. Effektmaß — Standardisierte Mittlere Differenzen	991
14.3.1. Exakte Methoden im FE-Model für Standardisierte Mittlere Differenzen [Hartung-Verfahren]	991
14.3.2. Beispiel: Zusammenfassung zweier kleiner Superiority Studien zu Acne Papulopustulosa [Standardisierte Mittlere Differenzen] — Klinische Relevanz	993
14.3.3. Bisherige Methoden im FE- und RE-Model für Standardisierte Mittlere Differenzen	994
14.3.4. Beispiel: Verbesserung der Belastbarkeit bei Angina Pectoris [Standardisierte Mittlere Differenzen] — Klinische Relevanz	995
<i>15. Konfidenzintervalle und implizite Schätzer für den Heterogenitätsparameter ...</i>	996
15.1. Hartung-Knapp-Konfidenzintervall für den Heterogenitätsparameter	996
15.2. Beispiel: Fortsetzung MS-Studien	997
15.3. Weitere Konfidenzintervalle und Tests für den Heterogenitätsparameter	998
15.4. Implizit definierte Schätzer für den Heterogenitätsparameter	999
15.4.1. Mandel-Paule Schätzer	999
15.4.2. Median Unverzerrte Schätzer nach Hartung	1000
15.4.3. Pseudo Maximum Likelihood Schätzer nach Hartung	1000
15.4.4. Beispiel: Fortsetzung MS-Studien	1000
15.4.5. Maximum Likelihood Schätzer	1000
15.4.6. Restringierte Maximum Likelihood Schätzer	1001
15.4.7. Beispiel: Fortsetzung MS-Studien	1001
15.5. Zusammenfassendes Beispiel: Fortsetzung Landwirtschaft — Genetik	1002
<i>16. Portfolioselektion im Wertpapiermanagement</i>	1004
16.1. Rendite und Volatilität, mit Beispiel	1004
16.2. Pareto-Optimale Portfolioselektion	1005
16.3. Die Rendite-Volatilitäts-Projektionsebene, mit Beispiel	1007
16.4. Rendite-Schwelle und Volatilitäts-Schranke, mit Beispiel	1008
16.5. Korrelation der Wertpapierkurse, mit Beispiel	1010

<i>17. Weitere Aspekte der Meta-Analyse</i>	1013
17.1. Ergänzende Tests auf Homogenität	1013
17.2. „Partial Weighting“ nach Yates-Cochran bei Dominanz einer Studie oder Prognose und Verallgemeinertes RE-Model	1013
17.3. Meta-Regression, mit Beispiel „Tuberkulose Studien“	1013
17.4. Prospektive Meta-Analyse	1014
17.4.1. Adaptive und „Self-designing“ Versuchsanlagen zur Prospektiven Meta-Analyse	1016
17.4.2. Beispiel: Adaptive Placebo-kontrollierte prospektive Nichtunterlegenheits-Studien zu Asthma Bronchiale	1016
17.5. Eigenschaften des Gauß-Markov-Schätzers und der Varianzschätzer sowie verwandter Schätzer	1018
Anhang	1023
1. Tabellenverzeichnis	1023
1.1. Kritische Werte, Quantile	1023
1.2. Weitere allgemeine Tabellen	1024
2. Tabellenanhang	1025
3. Griechisches Alphabet	1043
4. Symbolverzeichnis	1044
5. Literaturverzeichnis	1048
6. Literaturhinweise zu den einzelnen Kapiteln	1075
7. Sach- und Namensregister	1077
8. Weiterführende und ergänzende Hinweise	1144

Vorwort zur 15. Auflage

In der nun schon 15. Auflage, überarbeitet und stark erweitert, dieses Lehr- und Handbuches der angewandten Statistik konnten wir modernsten Entwicklungen und vielfach geäußerten Leserwünschen nachkommen: Das Buch enthält jetzt das **umfangreiche neue Kapitel XV :**

Meta-Analyse zur Kombination
von
Studien, Experimenten und Prognosen

mit zahlreichen detailliert durchgerechneten realen Beispielen, dokumentiert in 39 Tabellen und illustriert in 36 Abbildungen, die das jeweilige Geschehen visuell verdeutlichen und somit helfen, das Verständnis zu festigen.

Die **Meta-Analyse** gehört zu unseren aktuellen Arbeitsgebieten und so konnten wir neueste Forschungsergebnisse direkt in ihrer relevanten praktischen Bedeutung berücksichtigen. Die Meta-Analyse ist eine moderne Disziplin und von rasant zunehmender Bedeutung. Mit dem Anstieg verfügbarer Informationen gewinnt die Meta-Analyse an Wichtigkeit. **Combining Information** oder **Kombination von Wissen** bzw. **Combining Forecasts** – wenn es sich um die Kombination von Prognosen oder Vorhersagen handelt – sind weitere Bezeichnungen für diese Disziplin; speziell die Kombination von Wertpapieren zu einem Portfolio wird als **Portfolioauswahl** bezeichnet.

Wir behandeln in Kapitel XV konkrete und wichtige Anwendungen u. a. in den Bereichen Gesamtwirtschaft, Betriebswirtschaft, Wertpapiermanagement, Ernährungswirtschaft, Landwirtschaft, Genetik, Verhaltensforschung, Umwelt, Technik, Pharmazie und Medizin, wobei hier speziell auch Überlebenszeit- oder Time-to-Event-Studien mit zensierten Daten analysiert und zu einem relevanten Gesamtergebnis kombiniert werden. Dank der **Evidenz Basierten Medizin [EBM]** erlangt in diesem Bereich die Meta-Analyse außer der medizinischen auch eine zunehmende wirtschaftliche Bedeutung.

Wir sind davon überzeugt, dass das Buch in der nun vorliegenden überarbeiteten und wesentlich erweiterten Form, auch durch das aktualisierte und stark ausgedehnte Literaturverzeichnis, in entscheidend stärkerem Maße sowohl Lernenden als auch praktisch Arbeitenden in ihrem Bedürfnis nach aktuellen und relevanten statistischen Methoden gerecht wird.

Die Entscheidung, die Meta-Analyse neu in das Buch aufzunehmen, wurde beeinflusst zum einen durch an uns herangetragene Anregungen seitens der Leser und zum anderen auch durch die Verleihung des

Thomas L. Saaty Prize
vom
American Journal of
Mathematical and Management Sciences

im Jahre 2006 für „outstanding“ anwendungsorientierte Forschungsleistungen auf dem Gebiet der Meta-Analyse an den erstgenannten Autor dieses Buches.

An dieser Stelle möchten wir **Stefan Tewes** aus Iserlohn für die komplette Erstellung des Typoskripts und **Jan Edgar Hartung** aus Königswinter für seine Assistenz bei numerischen Berechnungen sowie für die komplette Erstellung der 36 Abbildungen für das neue Kapitel XV danken. J.E. Hartung haben wir auch zu danken für anregende Diskussionen und sein sorgfältiges Korrekturlesen.

Wir danken **Prof. Dr. Peter Pflaumer** für intensive Gespräche, welche die Ausrichtung des neuen Kapitels XV wesentlich beeinflusst haben. Ferner danken wir den Lesern, für deren Anregungen wir auch künftig ein offenes Ohr haben werden. Unser Dank gilt natürlich auch dem Oldenbourg-Verlag, Lektorat Wirtschafts- und Sozialwissenschaften, Steuern, Recht, und da insbesondere Dr. Jürgen Schechler sowie Rainer Berger, die uns nach Kräften bei der Neugestaltung des Buches unterstützten.

Joachim Hartung
Bärbel Elpelt
Karl-Heinz Klösener

Vorwort zur 7. Auflage

Die unvermindert starke Nachfrage hat nunmehr das Erscheinen der siebten Auflage dieses Lehr- und Handbuches der angewandten Statistik zur Folge. Nachdem in Vorauflagen eine vollständige Überarbeitung und starke Erweiterung vorgenommen wurde, konnte — nach einer nochmaligen kritischen Durchsicht — der Text hier weitgehend unverändert bleiben.

Um den an uns herangetragenen Leserwünschen gerecht zu werden, sind mittlerweile
Elpelt / Hartung: Grundkurs Statistik

sowie

Hartung / Heine: Statistik-Übungen — Deskriptive Statistik

bzw.

Hartung / Heine: Statistik-Übungen — Induktive Statistik

im selben Verlag erschienen. Diese drei Bücher sowie das nun in 3. Auflage erschienene Buch Multivariate Statistik (Hartung / Elpelt) werden am Schluß dieses Buches in ihrer Konzeption vorgestellt.

Vorwort zur 4. Auflage

In der nun vorliegenden, überarbeiteten und stark erweiterten vierten Auflage dieses Lehr- und Handbuches der angewandten Statistik haben wir moderne Entwicklungen und vielfach geäußerte Leserwünsche berücksichtigt: Das Buch enthält stark erweiterte bzw. neue Kapitel über

Zeitreihenanalyse,
Robuste Verfahren,
Explorative Datenanalyse (EDA).

Zudem konnten wir die bestehenden Kapitel vollständig überarbeiten, wobei auch Anregungen der Leser einbezogen wurden. Wir sind davon überzeugt, daß das Buch in der nun vorliegenden Form, nicht zuletzt auch durch das wesentlich erweiterte Literaturverzeichnis, in noch stärkerem Maße sowohl Lernenden als auch praktisch Arbeitenden gleichermaßen gerecht wird.

An dieser Stelle möchten wir cand. stat. Rolf Dietl, cand. stat. Thomas Goddemeier und cand. stat. Manfred Jutzi danken, die uns bei der technischen Herstellung der neuen Kapitel behilflich waren. Weiterhin gilt unser Dank den aufmerksamen Lesern, für deren Anmerkungen wir auch weiterhin dankbar sind, sowie dem R. Oldenbourg-Verlag und insbesondere Herrn Martin Weigert, die uns bei der Überarbeitung freie Hand ließen.

Aus dem Vorwort zur 3. Auflage

Der von Lesern mehrfach an uns herangetragene Wunsch einer stärkeren Berücksichtigung multivariater Verfahren ließ sich wegen der ohnehin schon sehr hohen Seitenzahl nicht realisieren. Jedoch ist bereits

Hartung / Elpelt: MULTIVARIATE STATISTIK
Lehr- und Handbuch der angewandten Statistik

im selben Verlag erschienen.

Aus dem Vorwort zur 1. Auflage

Statistische Methoden werden überall dort in zunehmendem Maße angewandt, wo durch Experiment oder Erhebung gewonnene Daten ausgewertet werden müssen. Solche Bereiche sind die Wirtschaftswissenschaften, die Ingenieurwissenschaften, die Medizin, die Landwirtschaft, die Physik, die Biologie, die Chemie, die Pharmazie, die Geodäsie, die Raumplanung, die Soziologie, die Psychologie, die Pädagogik, die Linguistik, die Archäologie, die Geologie, die Geographie, die Sportwissenschaft, die Umweltforschung usw.

Dieses Buch eignet sich nun zum einen als Handbuch für alle diejenigen, die Statistik bzw. statistische Methoden benötigen, zum anderen dient es dem Studenten in allen Anwendungsgebieten als vorlesungsbegleitendes, vertiefendes Lehrbuch. Nicht zuletzt wegen der Vielzahl vollständig durchgerechneter Zahlenbeispiele eignet es sich aber auch in hervorragender Weise zum Selbststudium.

Um allen angesprochenen Personenkreisen gerecht zu werden, haben wir die benötigten mathematischen Vorkenntnisse auf ein Minimum reduziert, das durch reine Schulmathematik abgedeckt wird, und zudem haben wir nicht nur solche Methoden aufgenommen, die in Grundvorlesungen über Statistik behandelt werden.

Dem Spezialisten wird ein schnelles Nachschlagen, dem Studenten und dem Praktiker wird die Wiederholung, Vertiefung und Erweiterung seiner Kenntnisse sowie ein fruchtbares Selbststudium ermöglicht.

An dieser Stelle möchten wir cand. stat. Johanna Behrens, cand. stat. Jürgen Böchynek, cand. stat. Johannes Ecke, cand. stat. Bärbel Heine, Dipl.-Stat. Manfred Korter, cand. stat. Lothar Kreienbrock, Dipl.-Math. Heinz Krüger, cand. stat. Heike Küntzel, Dipl.-Stat. Harald Kunitz, cand. stat. Gerd Mettendorf, cand. stat. Rüdiger Ostermann, Frau Bärbel Skopp, cand. math. Bärbel Tjakraatmadja, cand. stat. Elisabeth Vieth, Dipl.-Math. Bernard Voet, cand. stat. Peter Woboril und Dipl.-Stat. Bernd Wolter danken, die zum Teil bei der technischen Herstellung des Manuskriptes und zum Teil auch beim Sammeln des Materials und der Beispiele — für das endgültige Manuskript oder eines seiner verschiedenen Vorläufer — behilflich waren. Weiterhin gilt unser Dank Herrn Prof. Dr. Franz Hering für Ratschläge und Diskussionen, sowie dem R. Oldenbourg-Verlag, insbesondere Herrn Martin Weigert, für die gute Zusammenarbeit und weitestgehende Freiheit bei der Gestaltung des Buches.

Ein kurzer Überblick

An dieser Stelle soll ein kurzer Überblick über den Inhalt dieses Buches gegeben werden; einen ausführlicheren Einblick bieten das detaillierte Inhaltsverzeichnis und nicht zuletzt auch die Einleitungen zu den einzelnen Kapiteln und Abschnitten.

Die beschreibende, deskriptive Statistik dient der Aufbereitung und tabellarischen bzw. graphischen Darstellung von Datenmaterial. Als recht elementares Gebiet der Statistik und Grundlage vieler komplizierterer Methoden wird sie in **Kapitel I** ausführlich behandelt. Hier werden unter anderem empirische Kenngrößen von Häufigkeitsverteilungen, Konzentrationsmaße, Indexzahlen sowie empirische Zusammenhangsmaße vorgestellt.

Ausgangspunkt der induktiven Statistik ist die in **Kapitel II** in einem kurzen Abriß dargestellte Wahrscheinlichkeitsrechnung.

Die Schlußweisen der induktiven Statistik, basierend auf Punkt- und Intervallschätzungen sowie statistischen Tests, beinhaltet das sich anschließende **Kapitel III**.

Auch das **Kapitel IV** kann in gewisser Weise noch als einführend angesehen werden. Hier werden die wichtigsten statistischen Verteilungen und ihre Eigenschaften dargestellt. Zugleich werden aber in diesem Kapitel viele spezielle Methoden bzw. Verfahren betrachtet; zu den verschiedenen Verteilungen werden direkt entsprechende Punktschätzer, Intervallschätzer und Tests über die Parameter sowie Anpassungstests vorgestellt. Außerdem beinhaltet das Kapitel IV nichtparametrische (verteilungsfreie) Test- und Schätzmethoden im Einstichprobenproblem sowie einen ausführlichen Abschnitt über sequentielle Tests.

Das **Kapitel V** befaßt sich dann mit Problemen der Datengewinnung. Neben der Behandlung klassischer Methoden der Stichprobentheorie wird auch auf spezielle Stichprobenprobleme, z.B. in der Marktforschung, eingegangen. In einem Abschnitt über Meßfehler wird die Berücksichtigung, Erfassung und Kontrolle derselben angeprochen; hier sind nicht zuletzt auch die vorgestellten Ausreißertests und Datentransformationen von Interesse. Außerdem enthält das Kapitel V ausführliche Hinweise zur Planung und Auswertung von Versuchen, auch speziell von klinischen Versuchen, sowie Methoden der Skalierung von Merkmalsausprägungen.

Kapitel VI ist dann der in allen Bereichen der Fabrikation wichtigen Qualitätskontrolle auf Stichprobenbasis gewidmet. Neben einfachen, mehrfachen und sequentiellen Stichprobenplänen für quantitative und qualitative Merkmale sowie kontinuierlichen Stichprobenplänen wird auch auf die laufende Kontrolle der Produktion mittels Kontrollkarten eingegangen.

Vor allem bei der Auswertung nominalen Datenmaterials ist die Kontingenztafelanalyse von hervorragender Wichtigkeit. Diese wird in **Kapitel VII** ausführlich dargestellt. Neben klassischen Methoden der Analyse von zweidimensionalen Kontingenztafeln ist hier die Vielzahl vorgestellter Assoziationsmaße, die Analyse mittels loglinearem Modell — auch in mehrdimensionalen Kontingenztafeln — und nicht zuletzt das Logit-Modell, das Adjustieren von Kontingenztafeln und die Behandlung von Tests zum Vergleich der Parameter mehrerer diskreter Verteilungen herzuheben.

Kapitel VIII beschäftigt sich dann mit dem Vergleich zweier Meßreihen bzw. Stichproben. Behandelt werden hier Tests und Konfidenzintervalle für Lokations- und Dispersionsparameter einschließlich verteilungsfreier Verfahren.

Die Untersuchung der Stärke eines Zusammenhangs zwischen verschiedenen Merkmalen erfolgt mittels der in **Kapitel IX** dargestellten Korrelationsanalyse; für nominal skalierte Merkmale wird dies jedoch schon in Kapitel VII behandelt. Für einfache, partielle, bi-partielle und multiple Korrelationskoeffizienten werden neben Schätzern auch entsprechende Testverfahren betrachtet. Weiterhin findet man in **Kapitel IX** einen Test auf Unabhängigkeit mehrerer Meßreihen.

Das Kapitel X ist der Regressionsanalyse gewidmet, die der funktionalen Spezifikation des Zusammenhangs verschiedener Merkmale dient; hier werden verschiedene Probleme der linearen und nichtlinearen (insbesondere quadratischen) Regression sowie der multiplen Regression angesprochen. Ebenfalls wird Wert gelegt auf die Regression bei Fehlern in den Variablen.

Modelle der Varianzanalyse zur Beurteilung gemeinsam wirkender Einflußfaktoren werden in **Kapitel XI** behandelt; speziell finden sich hier auch verschiedene Methoden zum Vergleich von mehreren unabhängigen Meßreihen bzw. Stichproben.

Mit den Methoden der Zeitreihenanalyse beschäftigen wir uns im **Kapitel XII**. Neben klassischen Verfahren der Trend- und Saison-Schätzung bzw. -Elimination werden hier Zeitreihen als Realisationen stationärer und instationärer Prozesse (Box-Jenkins-Methode) behandelt. Weiterhin enthält dieses Kapitel die wesentlichen Aspekte der Spektral- und Kreuzspektralanalyse, also der Analyse von Zeitreihen im Frequenzbereich, sowie einen Abschnitt über gemischte Regressions-Zeitreihen-Modelle. Tests auf Stationarität einer Zeitreihe [Dickey-Fuller] und eines [kreuzkorrelierten] Zeitreihenpanels sind wegen des methodischen Zusammenhangs in den Abschnitt 11 von Kapitel XV verlagert.

Das **Kapitel XIII** beschäftigt sich mit der Zuverlässigkeitsanalyse von Systemen (z. B. technischer oder biologischer Natur) und der Analyse von Lebensdauern. Neben klassischen Systembetrachtungen werden auch Multi-State-Systeme, Fehlerbaumanalyse und die Analyse mehrphasiger Missionen miteinbezogen. Weiterhin betrachten wir hier verschiedene Klassen von Lebensdauerverteilungen, speziell Tests über die Zugehörigkeit zu diesen Klassen, und Parameterschätzungen bei den wichtigsten Lebensdauerverteilungen. Besondere Aufmerksamkeit verdienen auch die Ausführungen zu beschleunigten Lebensdauerprüfungen (Analyse im Eyring- oder Arrhenius-Modell), Wartungs- und Instandhaltungsstrategien sowie Überlegungen zur Verfügbarkeit von Systemen. Vergleichende Lebensdauerprüfungen [Log-Rank Test, Wilcoxon-Gehan Test, Tarone-Ware Tests], Cox's Proportional Hazards Regression Model und Kaplan-Meier Survival Analysis sind methodisch bedingt in Abschnitt 5 von Kapitel XV dargestellt.

In **Kapitel XIV** werden die Methoden der Explorativen Datenanalyse (EDA) dargestellt, die i. w. eine Ergänzung zur deskriptiven Statistik in Kapitel I sind, aber auf Terminologien späterer Kapitel zurückgreifen. Außerdem behandeln wir die grundsätzlichen Vorgehensweisen im Bereich der Robusten Statistik: M-, L- und R-Schätzer sowie Kenngrößen der Robustheit werden vorgestellt.

Die in **Kapitel XV** dargestellte **Meta-Analyse** verarbeitet unterschiedliche Ergebnisse verschiedener Studien, Experimente, Erhebungen oder Prognosen bzw. Vorher-

sagen und kombiniert sie auf geeignete Weise zu einem Gesamtergebnis. Das Methodenspektrum für diese aktuelle und zunehmend wichtiger werdende **Kombination von Wissen** oder **Combining Information** wird vorgestellt und ausführlich für eine Reihe von Anwendungsbereichen an Hand realer Beispiele demonstriert, in 39 Tabellen dokumentiert und mit 36 Abbildungen graphisch veranschaulicht. Dies geschieht für qualitative, quantitative oder Lebensdauer- bzw. Time-to-Event-Daten mit Zensierungen in den beteiligten Studien. Insbesondere wird auch die Kombination wirtschaftlicher Vorhersagen real diskutiert [**Combining Forecasts**]. Etwa im Bereich Umwelt – Technik werden widersprüchliche Ergebnisse konkurrierender Meß-Labore zur Einschätzung der Asbestbelastung eines Objekts zu einem Gesamtergebnis kombiniert mit dem Anspruch, eine gerechte Beurteilung zu erreichen. Im Bereich der **Evidenz Basierten Medizin [EBM]** hängt sowohl die medizinische als auch die wirtschaftliche Zukunft eines Medikamentes entscheidend von der richtigen Meta-Analyse ab, d. h. von der adäquaten Kombination der Resultate aus aufwendig durchgeführten Klinischen Prüfungen oder Studien. Im Wertpapiermanagement umfaßt die Meta-Analyse etwa die **Portfolioauswahl**, in der ein Portfolio aus Wertpapieren nach den Kriterien Rendite und Risiko wohl abgestimmt zusammenzustellen ist.

Natürlich konnten in diesem Buch nicht alle statistischen Methoden bis in die letzte Einzelheit berücksichtigt werden, da dies den Rahmen des ohnehin schon umfangreichen Buches gesprengt hätte. Wir haben aber die wichtigsten, mit elementaren Kenntnissen der Mathematik zu bewältigenden und für die Anwendungen relevanten Methoden erfaßt. Die Auswahl der Beispiele zu den einzelnen Verfahren, die aus verschiedensten Anwendungsbereichen entlehnt sind, streicht die weitgestreuten und somit für jeden Leser interessanten Anwendungsmöglichkeiten der vorgestellten Methoden und Verfahren heraus. Für den speziell an einem Gebiet weitergehend interessierten Leser haben wir zudem ausführliche Hinweise auf Literatur angegeben.

Diejenigen, die insbesondere auch an den Verfahren der sogenannten Multivariaten Statistik (z.B. Skalierungsverfahren, Faktorenanalyse, Clusteranalyse, multivariate Regressions- und Varianzanalyse), welche hier nur angerissen werden können, interessiert sind, seien hingewiesen auf das im selben Verlag erschienene Buch

Hartung / Elpelt: **MULTIVARIATE STATISTIK**
Lehr- und Handbuch der angewandten Statistik ,

in dem derartige Verfahren auf über 800 Seiten dargestellt werden.

Für die erstmals mit Statistik konfrontierten Leser möchten wir abschließend noch einige Hinweise zum Arbeiten mit diesem Buch geben. Es ist keinesfalls notwendig, das Buch fortlaufend Seite für Seite durchzuarbeiten. Vielmehr sollte man zunächst die ersten Teile der Abschnitte 1 bis 5 von Kapitel I und die Kapitel II und III lesen. Hier sind vor allem die Begriffe Zufallsvariable, Verteilungsfunktion, Dichte, (Mittel- oder) Erwartungswert, Varianz, Standardabweichung, Punktschätzer, Intervallschätzer, Konfidenzintervall und statistischer Test wichtig für das weitere Verständnis. Auch die in Kapitel IV angegebenen grundlegenden Eigenschaften von Verteilungen, insbesondere von Normal-, Binomial- und Exponentialverteilung, sind für den Anfänger wichtig. Beim weiteren Studium des Buches ist die Einhaltung einer bestimmten Reihenfolge nicht mehr nötig; alle Kapitel können weitgehend unabhängig voneinander gelesen und verstanden werden.

Einführung und Grundlagen

1. Was ist Statistik

Das Wort Statistik hat wohl jeder schon oft gehört und benutzt: Man spricht von Verkaufsstatistiken, Statistik bei Wahlen, Unfallstatistik, Bevölkerungsstatistik usw. In diesem Zusammenhang assoziiert man mit dem Begriff Statistik zunächst einmal Zahlenkolonnen, Tabellen, in denen große Datenmengen mehr oder weniger übersichtlich zusammengestellt sind. In eben diesen Rahmen paßt wohl die Herkunft des Wortes Statistik: Abgeleitet vom neulateinischen Begriff „status“, der etwa „Staat“ oder auch „Zustand“ bedeutet, verstand man im achtzehnten und auch im neunzehnten Jahrhundert unter Statistik vorwiegend die „Lehre von der Zustandsbeschreibung des Staates“, also das Sammeln von Daten über Heer, Gewerbe und Bevölkerung (Tauf-, Heirats-, Sterberegister).

Eine andere Assoziation zum Begriff Statistik, die einem sofort in den Sinn kommt, ist der Mittelwert, der Durchschnittswert. So spricht man oft vom mittleren Einkommen, von der durchschnittlichen Zuschauerzahl, vom durchschnittlichen Benzinerbrauch, von der mittleren Temperatur oder etwa vom Durchschnittsalter. Jemand, der schon einmal den Prospekt eines Reiseveranstalters studiert hat, wird auf Angaben von Regentagen, Sonnentagen usw. pro Jahr in bestimmten Gebieten gestoßen sein. Obwohl hier nicht explizit vermerkt ist, daß diese Angaben nur Durchschnittswerte der vergangenen Jahre sind, handelt es sich natürlich auch hier um Mittelwerte. Was ein Mittelwert eigentlich ist, welche Aussagekraft er hat, kommt in folgendem Gedicht, das wir Herrn P. H. List, Professor für pharmazeutische Technologie in Marburg, verdanken, vgl. auch Krafft (1977), treffend zum Ausdruck.

Ein Mensch, der von Statistik hört,
denkt dabei nur an Mittelwert.
Er glaubt nicht dran und ist dagegen,
ein Beispiel soll es gleich belegen:

Ein Jäger auf der Entenjagd
hat einen ersten Schuß gewagt.
Der Schuß, zu hastig aus dem Rohr,
lag eine gute Handbreit vor.

Der zweite Schuß mit lautem Krach
lag eine gute Handbreit nach.
Der Jäger spricht ganz unbeschwert
voll Glauben an den Mittelwert:
Statistisch ist die Ente tot.

Doch wär' er klug und nähme Schrot
– dies sei gesagt, ihn zu bekehren –
er würde seine Chancen mehren:
Der Schuß geht ab, die Ente stürzt,
weil Streuung ihr das Leben kürzt.

Wie dem Jäger in diesem Gedicht geht es vielen: Man sagt Mittelwert und interpretiert diese Kenngröße als den wahren, den wirklichen Wert. Weiß man etwa, daß eine Sorte von Fernsehgeräten eine mittlere Lebensdauer von sieben Jahren hat, so heißt dies aber noch lange nicht, daß ein bestimmtes Gerät, das man kauft, wirklich

sieben Jahre lang hält. Wenn man Pech hat, kann es schon nach wenigen Tagen kaputtgehen, wenn man Glück hat, vielleicht auch erst nach fünfzehn Jahren. Im Einzelfall ist also der Mittelwert nicht identisch mit dem wahren Wert. Daraus darf man aber nicht schließen, daß die Statistik ungenaue oder gar falsche Angaben hervorbringt. Man muß sich aber vor einer falschen Interpretation von Kenngrößen wie dem Mittelwert hüten: Er ist nicht gleichzusetzen mit einem wahren Wert; er charakterisiert zwar die Verhältnisse, die Charakterisierung ist aber nicht vollständig, denn er beschreibt nur ein Zentrum, einen Durchschnitt aller vorkommenden Werte. Genauso verhält es sich mit allen anderen statistischen Kenngrößen: Die Angabe des minimal und des maximal möglichen oder vorkommenden Wertes, der Index der Verbraucherpreise, der Reallohnindex usw. geben eine unvollständige, aber durchaus sinnvolle und hilfreiche Vorstellung von den wirklichen Verhältnissen. Durch die Angabe von statistischen Kenngrößen hofft man also, das Wesentliche oder Typische der wirklichen Gegebenheiten zu erfassen und erhebt nicht, wie oft fälschlich angenommen, den Anspruch auf Gültigkeit im Einzelfall. Bei der Erfassung des Wesentlichen ist nicht nur die Angabe von Kenngrößen, sondern vielmehr auch die übersichtliche tabellarische oder auch graphische Darstellung der wahren Gegebenheiten oft sehr hilfreich.

Die bisher angesprochene Charakterisierung und Darstellung der realen Verhältnisse ist der wohl jedem bekannte und geläufige Teil der Statistik. Er ist jedoch nur ein kleiner Teil, den man auch die beschreibende oder deskriptive Statistik nennt. Wesentlicher ist die schließende, die induktive Statistik. Aufgrund vorliegender Daten werden mit Hilfe der Methoden der induktiven Statistik Schlüsse gezogen und Aussagen gemacht:

Wie wird sich etwa der Absatz eines Produkts oder die Zahl der Arbeitslosen in Zukunft entwickeln? Kann man aus der bisherigen Entwicklung Prognosen gewinnen? Kann man aus dem Wahlverhalten einiger Personen auf das Verhalten aller Wähler schließen (Wahlprognose)? In diesen Beispielen kann man die Prognosen, die Schlüsse aus den bekannten Gegebenheiten auf unbekannte Gegebenheiten im nachhinein überprüfen. Es stellt sich aber die Frage, ob man nicht schon vorher Aussagen über die Zuverlässigkeit solcher Prognosen oder Schlußfolgerungen machen kann, insbesondere dann, wenn eine Überprüfung nicht vorgesehen oder gar nicht möglich ist. Befragt man etwa einige Personen in der Bundesrepublik Deutschland über ihren Gesundheitszustand, um auf den Gesundheitszustand der Gesamtbevölkerung zu schließen, so kann man die Richtigkeit, die Genauigkeit der Schlußfolgerung kaum überprüfen: Es wäre viel zu aufwendig im nachhinein den Schluß zu überprüfen, indem man alle interessierenden Personen medizinisch untersucht. Es stellen sich dann die Fragen, wie man die zu befragenden Personen auswählt (Frage nach dem Stichprobenplan), wieviele Personen man auswählt (Frage nach dem Stichprobenumfang) usw., um mit vertretbarem Aufwand eine möglichst hohe Genauigkeit zu erreichen.

Natürlich darf man nicht darüber hinwegsehen, daß Stichprobenergebnisse und auch einzelne Meßergebnisse stets Ungenauigkeiten beinhalten. Dies ist auch in einer so exakten Wissenschaft wie z. B. der Physik der Fall: Hier besagt etwa die Heisenbergsche Unschärferelation, daß es unmöglich ist, Ort und Geschwindigkeit eines Teilchens gleichzeitig exakt zu messen.

Sicherlich werden aber ganz genaue, ganz exakte Angaben auch gar nicht benötigt. Die Fehler müssen jedoch in angemessenen, vernünftigen Größen gehalten werden: Bei der Bestimmung der Entfernung zwischen zwei Städten ist ein Meßfeh-

ler von einem Meter i. a. sicherlich irrelevant. Dagegen wird man bei der Messung der Dicke einer Metallfolie einen Fehler von 0,5 mm schon als zu groß bezeichnen; hier wäre die mit einer Mikrometerschraube erreichbare Genauigkeit von 0,01 mm als angemessen zu betrachten.

Das benötigte Instrumentarium zur Abschätzung von Art und Größe von Meßfehlern, Fehlern bei Schlüssen von Stichproben auf Gesamtheiten usw. wird von der Statistik bereitgestellt. Sie ist überall dort einzusetzen, wo der Zufall eine Rolle spielt oder wo Situationen so komplex sind, daß sie sich einer deterministischen Beschreibung entziehen. Dies soll noch an einem einfachen Beispiel erläutert werden: Wirft man eine Münze, so sagt man, es ist zufällig, ob auf der Oberseite „Kopf“ oder „Zahl“ erscheint. In Wirklichkeit aber ist der Münzwurf deterministisch: Ob Kopf oder Zahl fällt, ist durch physikalische Gesetzmäßigkeiten (Abwurfwinkel, Fallgesetze usw.) eindeutig bestimmt. Die Zahl der Faktoren, die das Ergebnis eines Münzwurfs beeinflussen, ist jedoch so groß, daß man vom Zufall spricht.

Die induktive Statistik könnte man also als ein Modell des Zufalls verstehen. Sie versucht, „den Zufall in den Griff zu bekommen“, „Sicherheit über Unsicherheit zu gewinnen“.

2. Das Experiment

Jede Beobachtung in der Natur oder in der Umgebung zieht bestimmte Verhaltensweisen nach sich. Hat man beispielsweise einmal auf eine heiße Herdplatte gefaßt und sich dabei die Finger verbrannt, so wird man das aufgrund dieser schlechten Erfahrung nicht noch einmal tun. Wie hier, zieht einmalige schlechte Erfahrung schon eine bestimmte Verhaltensweise nach sich, was auch im Sprichwort „Gebranntes Kind scheut das Feuer“ zum Ausdruck kommt.

Andere Beobachtungen werden erst zur Erfahrung, wenn man einen bestimmten Vorgang vielmals beobachtet hat. So wundert sich niemand darüber, daß ein Gegenstand, den man fallen läßt, nach unten und nicht nach oben fällt. Aber auch, wenn im Herbst die Blätter von den Bäumen fallen, erstaunt es nicht, wenn ein Blatt „nach oben fällt“. Ganz selbstverständlich schließt man, daß es durch einen Windstoß nach oben getragen wird. Niemand wird deshalb Zweifel an der Tatsache hegen, daß Gegenstände ohne Einwirkung von außen stets nach unten fallen; dies ist einfach eine Erfahrung.

Das Experiment ist nun eine Formalisierung von Beobachtungen in der Natur: Aus Beobachtungen in der Natur (Erfahrungen) wird ein Prinzip abgeleitet, d. h. eine Hypothese wird aufgestellt und dann experimentell überprüft. Eine der Voraussetzungen zum Betreiben der Naturwissenschaften, der Wissenschaften des Messens und Vergleichens, zu denen auch die Statistik gehört, ist das Experimentieren.

Naturgesetze z. B. basieren auf Beobachtungen aus denen ein Prinzip abgeleitet und experimentell überprüft wird. Sie gelten dann als bewiesen (durch das Experiment), solange kein Widerspruch auftritt bzw. kein besseres Prinzip (Modell) vorhanden ist. Dies sei an einem Beispiel ausführlich dargestellt (vgl. H. Seitz (1953): Das Weltbild der Physik, in: Die moderne Volkshochschule, Schwab, Stuttgart).

Aufgrund der subjektiven Erfahrung vieler Personen hat sich der objektive Befund herauskristallisiert, daß alle Körper unabhängig von Beobachtungsort und

-zeit senkrecht nach unten fallen, wenn man sie losläßt oder ihnen die tragende Unterlage entzieht. Diese Aussage an sich ist aber noch nicht befriedigend: Man möchte auch wissen, wie die Körper fallen. Daher werden Experimente durchgeführt, in denen der Vorgang des Falls immer wieder unter den verschiedensten Bedingungen wiederholt wird: Man läßt gleich schwere, aber verschiedenen große, gleich große, aber verschiedenen schweren Körper immer wieder aus derselben Höhe fallen und stellt dabei fest, daß alle Körper – ganz gleich wie groß oder wie schwer sie sind – gleich schnell fallen, sofern kein Luftwiderstand den Fall stört. Zum Gesetz des freien Falls kommt man experimentell schließlich dadurch, daß man auch die Fallhöhe variiert. Nachdem man experimentell erforscht hat, wie ein Körper fällt, stellt sich natürlich die Frage, warum er so und nicht anders fällt. Diese Frage ist ungleich schwerer zu ergründen, da hier ein induktiver Schluß von der Wirkung auf die Ursache notwendig ist. Dieser Schluß ist nie eindeutig, so wie es der Schluß von der Ursache auf die Wirkung ist: Tritt etwa eine Warenverknappung auf, so werden die Preise natürlich steigen; sind dagegen die Preise gestiegen, so heißt das noch lange nicht, daß eine Verknappung eingetreten ist.

Zur Beantwortung der Frage „Warum ist etwas so und nicht anders?“ stellt man eine (Arbeits-)Hypothese auf und vergleicht die Erfahrungen mit den Annahmen in der Hypothese. Zur Erklärung der Fallgesetze etwa wurde die Hypothese aufgestellt, daß die Erde eine Anziehungskraft auf die fallenden Körper ausübt. Diese Anziehungskraft nennt man dann Schwerkraft. Als Isaac Newton (1642–1727) nachweisen konnte, daß dieselbe Kraft, die einen Körper nach unten fallen läßt, auch den Mond zu seiner Bahn um die Erde zwingt, wurde die Vorstellung von der Schwerkraft auf das ganze Universum erweitert: Aus dem Fallgesetz entstand das Gravitationsgesetz. Aufgrund des Gravitationsgesetzes gelang es, die Existenz bisher unbekannter Planeten (z. B. Pluto) vorauszusagen, was zur Folge hatte, daß nicht mehr an der Richtigkeit der Hypothese über die Gravitationskraft gezweifelt wurde. Die Hypothese wurde zur Theorie. Eine Theorie ist aber nie eine unumstößliche Wahrheit, sie ist lediglich eine zweckmäßige Darstellung des Zusammenhangs aller bisherigen Erfahrungen und behält ihre Gültigkeit nur solange keine Widersprüche zur Erfahrung entstehen.

Das statistische Experiment oder das Zufallsexperiment ist nun dort besonders wichtig, wo Objekte (Untersuchungseinheiten) zur Verfügung stehen, an denen Effekte von Behandlungen meßbar, aber nicht exakt vorhersagbar sind, d. h. von Versuch zu Versuch verschieden sind. Anhand der gemessenen Effekte möchte man dann Schlüsse über die unbekannten, wahren Effekte ziehen, die Ergebnisse zukünftiger Experimente (Versuche) vorhersagen (bei Einräumung von spezifizierten Fehlerwahrscheinlichkeiten), d. h. man möchte „Sicherheit über Unsicherheit“ gewinnen. Besonders wichtig ist das statistische Experiment also überall dort, wo Unsicherheiten (Zufälligkeiten) auftreten. Um nun wirklich „Sicherheit über Unsicherheit“ gewinnen zu können, muß das Experiment genau geplant werden: Die Versuchsergebnisse müssen ja in möglichst hohem Maße vergleichbar sein, die Versuchsbedingungen müssen möglichst konstant gehalten werden, das Experiment muß kontrollierbar und jeder Zeit wiederholbar sein. Die Planung eines Experiments muß natürlich stets vor der Durchführung erfolgen und muß sich an den praktischen Gegebenheiten orientieren. (Natürlich sind die Planungsmöglichkeiten durch äußere Umstände eingeschränkt.)

Eine besonders große Rolle spielt das statistische Experiment naturgemäß dort, wo die Unsicherheit am größten ist. Dies ist in Bereichen wie z. B. der Medizin, der

Psychologie, der Pädagogik, der Biologie, der Landwirtschaft und den Ingenieurwissenschaften der Fall. Die Ergebnisse eines Experiments in solchen Bereichen, die O. Kempthorne die "noisy sciences" nennt, sind nie exakt: Störungen (Rauschen, noise) machen exaktes Messen unmöglich. Typisch ist hier, daß ein Objekt nach einmaliger Versuchsdurchführung nicht wieder in den ursprünglichen Zustand zurückgesetzt werden kann. Hat man etwa die Lebensdauer eines Transistors festgestellt, so ist dieser dann defekt, und man kann nicht noch einmal seine Lebensdauer bestimmen. Hat man den Erfolg der Behandlung eines Krebspatienten mit einem bestimmten Medikament festgestellt, so kann man im selben Experiment das Medikament diesem Patienten i. a. nicht noch einmal verabreichen, um den Behandlungserfolg zu überprüfen, denn der Ausgangszustand des Patienten ist dann ein anderer. Eine Versuchswiederholung ist also nur mit einem anderen Objekt möglich. Dann stellt sich aber die Frage, ob sich die Objekte genügend ähnlich sind, d. h. die Frage nach der Konstanz der Versuchsbedingungen.

Am ehesten kontrollierbar sind wohl Experimente im physikalischen und chemischen Bereich. Aber auch hier treten Unsicherheiten auf; man kann nicht von absoluten Wahrheiten sprechen. Dies wurde schon bei den Ausführungen zum Fallgesetz und zum Gravitationsgesetz deutlich: Die Theorie hat nur so lange Gültigkeit, bis sie durch Erfahrung widerlegt wird. Die Unsicherheit bei solchen gut planbaren, kontrollierbaren Experimenten, die zwar relativ gering ist, kommt auch in einem Brief Maxwells an Lord Rayleigh zum Ausdruck; Maxwell schreibt dort: „Das zweite Gesetz der Thermodynamik hat denselben Wahrheitsgehalt wie die Aussage, daß man, nachdem man einen Eimer voll Wasser ins Meer geschüttet hat, nicht denselben Eimer Wasser wieder dem Meer entnehmen kann.“ Maxwell hält also die Gültigkeit des zweiten Gesetzes der Thermodynamik für sehr wahrscheinlich, erkennt jedoch, daß es nicht völlig gesichert ist.

Im wesentlichen sind also in allen Bereichen mehr oder weniger große Unsicherheiten vorhanden; ein „Rauschen“ kann quasi niemals völlig ausgeschlossen werden. Die Statistik (das statistische Experiment) dient nun der Extraktion der vom Rauschen überlagerten Information, d. h. man möchte diejenige Information aus dem Experiment ziehen, die auf andere Umstände, andere Objekte usw. übertragbar ist; man vergleiche hierzu auch die Ausführungen von Krafft (1977).

3. Die Erhebung

Wir haben gesehen, daß sich Störeinflüsse auch in sogenannten exakten Wissenschaften nicht gänzlich ausschließen lassen; auch ein noch so gut kontrollierbares Experiment birgt gewisse Unsicherheiten. In wesentlich stärkerem Umfang treten unkontrollierbare Störungen z. B. im sozial- und wirtschaftswissenschaftlichen Bereich auf. Die Preisentwicklung etwa wird von saisonalen Einflüssen überlagert; oft aber röhren höhere Preise auch daher, daß Qualitätsverbesserungen mit ihnen einhergehen oder aus Angst Panikkäufe (Hamsterkäufe) getätigter werden. Hat man die Preisentwicklung eines Produktes beobachtet, so lassen sich diese Beobachtungen nicht in einem kontrollierten Experiment überprüfen: Die Zeit kann nicht zurückgedreht werden, Störeinflüsse können nicht ausgeschaltet werden. Wie will man etwa Saisonschwankungen ausschalten und beobachten, wie sich das Preisniveau entwickelt hätte, wenn keine saisonalen Effekte vorhanden gewesen wären?

Man ist hier gezwungen, seine Schlußfolgerungen direkt aus den einmal erhobenen (beobachteten) Daten zu ziehen. Dabei kann man natürlich Vergleichsdaten

(etwa aus anderen Perioden) zu Hilfe nehmen. Ist etwa die momentane Preisentwicklung „wesentlich“ von der in anderen Jahren verschieden, so wird man dann sicherlich annehmen können, daß nicht nur saisonale Schwankungen das Preisniveau beeinflußt haben, denn diese dürften in allen Jahren etwa gleich sein.

Eine (nichtkontrollierbare) Erhebung, ein Sammeln von nicht im nachhinein überprüfbaren Daten kann also nicht dem experimentellen Nachweis einer Hypothese, sondern nur der Erfassung eines Sachverhaltes dienen.

Auch eine Befragung birgt ein gehöriges Maß an Unsicherheit. Befragt man etwa einige Personen nach ihrem Einkommen und will daraus auf die Einkommensstruktur z. B. in der Bundesrepublik Deutschland schließen, so macht man unter Umständen einen großen Fehler. Befragt man etwa nur Frauen, so ist ein Schluß auf die Gesamtheit sicher mit einem enormen Fehler behaftet, denn es ist bekannt, daß die Einkommensstruktur bei Frauen anders ist als bei Männern.

Auch eine Befragung ist natürlich i. a. kein Experiment zum Nachweis einer Hypothese. Man kann nicht die Einkommensstruktur beobachten und dann eine aus der Beobachtung gewonnene Hypothese experimentell überprüfen: Da die Zeit fortschreitet, kommt es ständig zu Verschiebungen in der Einkommensstruktur (z. B. durch Inflationsausgleiche oder einseitige Einkommenskürzungen im Öffentlichen Dienst).

Die Feststellung der momentanen Einkommensstruktur in der Bundesrepublik Deutschland, die natürlich nur eine Erhebung (und kein Experiment) ist, läßt sich aber planen: Man kann z. B. im vorhinein überlegen, welche Personen man befragt und wieviele Personen man befragt. Solche Fragen der Planung von Erhebungen werden von der Stichprobentheorie beantwortet. Man muß überlegen, auf welche Gesamtheit ein Erhebungsergebnis übertragen werden soll. Will man wissen, was der durchschnittliche momentane Verdienst aller Personen in der Bundesrepublik Deutschland ist, oder will man den Durchschnittsverdienst aller Berufstätigen ermitteln, d. h. welche Grundgesamtheit sollen die befragten Personen repräsentieren? Wie genau möchte man den Durchschnittsverdienst ermitteln? Je genauer das Ergebnis sein soll, desto größer muß in der Regel die Zahl der befragten Personen sein. Die geplante Erhebung unterscheidet sich also von der nichtgeplanten dadurch, daß ihr ein Stichprobenplan zugrundeliegt.

Man könnte nun natürlich einwenden, daß doch stets erreicht werden kann, daß bei einer Befragung kein Fehler beim Schluß auf eine Grundgesamtheit vorhanden ist. Theoretisch braucht man ja nur die gesamte Grundgesamtheit zu erfassen, d. h. etwa alle in Frage kommenden Personen zu befragen. In einem solchen Fall spricht man dann nicht mehr von einer Stichproben- oder Teilerhebung, sondern von einer Voll- oder Totalerhebung. Meistens sind aber Vollerhebungen viel zu teuer und zu aufwendig. Auch sind sie nur theoretisch fehlerfrei: Bei den meisten Erhebungen, die Personenbefragungen als Grundlage haben, kommen psychologische Effekte ins Spiel. Befragt man etwa alle Bewohner der Bundesrepublik Deutschland danach, ob sie arbeitslos sind oder nicht und ob sie „schwarzarbeiten“, um den Prozentsatz der „arbeitslosen Schwarzarbeiter“ festzustellen, so werden einige gar keine Antwort, andere eine falsche Antwort geben, da ihnen die Wahrheit unangenehm ist; wer gibt schon gerne zu, daß er arbeitslos ist und schwarzarbeitet? Praktisch enthalten also auch Totalerhebungen Fehlerquellen. Manchmal ist es sogar so, daß der Fehler bei einer Stichprobenerhebung geringer ist als bei einer Vollerhebung: Befragt man nur wenige Personen, so kann die Befragung durch ein Interview sehr intensiv durchgeführt werden. Will man dagegen eine Totalerhebung

durchführen, so ist man wohl auf das Versenden von Fragebögen angewiesen. Die Antworten sind dann oft unklar; ein Nachfragen ist nicht möglich. Auch werden viele Fragebögen erst gar nicht zurückgeschickt.

Bisher sind wir davon ausgegangen, daß eine Erhebung durchgeführt wird, um eine bestimmte Frage zu beantworten, einen Sachverhalt zu erfassen, z. B.: Wie entwickelt sich das Preisniveau? Wieviel Prozent aller Personen sind arbeitslos? Wie hoch ist der mittlere Verdienst in der Bundesrepublik Deutschland? Eine solche Erhebung nennt man auch eine primärstatistische Erhebung, denn die Erhebung wird eigens zur Beantwortung einer konkreten Frage durchgeführt.

Oft werden aber Daten sowieso erhoben. Beispielsweise wird, ohne daß eine bestimmte Frage verfolgt wird, in der Flensburger Verkehrskartei jede Neuzulassung von Personenkraftwagen registriert. Will man also die Anzahl der durchschnittlich pro Monat neu zugelassenen PKW berechnen, so braucht man nicht eigens zu diesem Zweck eine Erhebung durchzuführen; man kann die existierende Kartei zu Hilfe nehmen. Aber auch die Bestimmung der durchschnittlichen Neuzulassungen ist eine Erhebung, nämlich eine sogenannte sekundärstatistische Erhebung. Bei einer sekundärstatistischen Erhebung werden also die Daten nicht extra zwecks Beantwortung einer Fragestellung erhoben; vielmehr werden ohnehin vorhandene Originaldaten zur Klärung der Fragen herangezogen. Ist das Material einer Erhebung nicht nur zu einem anderen Zwecke erhoben worden, sondern handelt es sich auch noch um verfälschtes Material, d. h. um bereits verformte, manipulierte Daten, so spricht man schließlich von einer tertiärstatistischen Erhebung. Ist man bei primärstatistischen Erhebungen oft noch in der Lage, die Erhebung der zu beantwortenden Frage entsprechend optimal zu planen, d. h. einen möglichst gut angemessenen Stichprobenplan aufzustellen, so ist dies bei sekundär- und erst recht bei tertiärstatistischen Erhebungen natürlich nicht mehr möglich. Schlüsse, die aus nichtprimärstatistischen Erhebungen gezogen werden, sind daher mit einem noch größeren Fehler behaftet als bei primärstatistischen Erhebungen; die Aussagekraft von Schlußfolgerungen – ihr Wahrheitsgehalt – ist also mitunter nicht sehr groß. Dies gilt natürlich für tertiärstatistische Erhebungen in hohem Maße insbesondere dann, wenn man nicht weiß, in welcher Weise Verformungen und Manipulationen der Originaldaten vorgenommen wurden.

4. Zur Statistik und ihren philosophischen Voraussetzungen

Die Statistik ist als methodisches Instrumentarium an allen empirisch arbeitenden Wissenschaften bzw. Wissenschaftsbereichen orientiert. Philosophische Grundlagen/Voraussetzungen der Statistik sind daher mit den Grundlagen dieser Wissenschaften gegeben. Einige Hinweise auf solche philosophischen Voraussetzungen können dies aufzeigen. Charakteristisch für alle statistischen Methoden ist der Verzicht auf eine Vielzahl von Kenntnissen über die Untersuchungseinheiten einer Erhebung bzw. über die Objekte, die in einem Experiment berücksichtigt werden. Stellt man in einer Erhebung z. B. die mittlere Größe der Nutzfläche aller landwirtschaftlichen Betriebe in der Bundesrepublik Deutschland fest, so verzichtet man bewußt auf die Registrierung des Viehbestandes, der Region, in der der Betrieb liegt, usw. Bestimmt man in einem Experiment die mittlere Gewichtsabnahme in einem Zeitraum bei Verabreichung einer bestimmten Diät, so wird bewußt z. B. das Alter, das Anfangsgewicht und der Gesundheitszustand der Objekte (Personen, Tiere) außer acht gelassen.

Man sagt sich von vielen Kenntnissen los und begnügt sich mit einigen wenigen: „An die Stelle der tatsächlich existierenden Einheit, des tatsächlichen Geschehens, tritt der statistische Schatten, in dem nur Teile der Wirklichkeit registriert sind“ (vgl. O. Anderson (1965): Probleme der statistischen Methodenlehre in den Sozialwissenschaften, Physica, Würzburg).

Real gesehen sind alle Objekte zufallsbedingt verschieden; durch den Verzicht auf Kenntnisse über das einzelne Objekt verliert es seine Individualität, an deren Stelle sein statistischer Schatten tritt: Objekte mit identischen Schatten sind statistisch gleich. Alle landwirtschaftlichen Betriebe mit gleich großer Nutzfläche sind in dieser Hinsicht gleich, obwohl sie in völlig verschiedenen Regionen liegen können, völlig unterschiedliche Viehbestände haben können usw. Alle Objekte, bei denen durch Verabreichung einer Diät die gleiche Gewichtsabnahme erreicht wird, werden als gleich angesehen, obwohl sie natürlich individuelle Unterschiede wie z. B. unterschiedliches Alter aufweisen.

Entscheidend für die Aussagekraft der Statistik (der Erhebung, des statistischen Experiments) ist es, inwieweit alle irgendwie relevanten Umstände im statistischen Schatten berücksichtigt werden können. Wesentlich für die Beurteilung der Ertragssteigerung durch Verwendung eines Düngemittels ist die Berücksichtigung der Bodengüte, der Witterungsbedingungen, der Bewässerungsmöglichkeiten, des Saatgutes usw. Sicherlich können nicht alle irgendwie relevanten Umstände im Schatten berücksichtigt werden, so daß Aussagen aufgrund von Erhebungen oder Experimenten nie den Anspruch auf absolute Wahrheit erheben können; wohl aber kann der Wahrheitsgehalt einer Aussage sehr hoch sein.

Andersons Charakterisierung des Wesens der Statistik (Ersetzen des Individuums durch seinen Schatten, bewußter Verzicht auf Kenntnisse über das Individuum) ist der Philosophie Immanuel Kants (1724–1804), der „kritisch“ lehrt, daß man nicht ein „Ding an sich“, sondern nur „Erscheinungen“ von ihm beobachten kann, entnommen; demzufolge kann kein Anspruch auf absolute Gültigkeit erhoben werden: Nichts ist eben absolut wahr. Daher versucht man durch Beobachten, Erheben und Experimentieren im „Raum der Erscheinungen“ Aussagen zu machen. Die Gewinnung von Erkenntnissen unter den Bedingungen von „Raum und Zeit“ ist heute Grundlage aller modernen Wissenschaften. Daß dies nicht immer so war und diese Auffassung nicht immer vertreten wurde, wird bei Durchsicht der Geschichte der Philosophie deutlich.

Descartes (1596–1650) formulierte den bekannten Satz „Ich denke, also bin ich“ („cogito, ergo sum“), d. h. nur das eigene Denken (Existenz) ist absolut sicher. Dieser Satz ist der Kern einer Philosophie, die sehr richtig bemerkt, daß absolute Wahrheit nicht durch Beobachtung, sondern nur durch logisches Denken deduktiv gewonnen werden kann, und die es deshalb generell ablehnt, sich auf empirische Beobachtungen als Grundlage zu stützen, um zu neuen Erkenntnissen zu gelangen. In der Antike wurde diese Philosophie von Plato (427–347 v. Chr.) vertreten. Aber auch damals gab es im Ansatz schon die heute vertretene Ansicht, daß man aus Beobachtungen sehr wohl induktiv zu neuen Erkenntnissen gelangen kann; dies kommt in der Philosophie des Aristoteles (384–322 v. Chr.) zum Ausdruck. Er akzeptiert das Vorhandensein von Variabilität, glaubt aber auch an die Immanenz des Parameters (des wahren Wertes, der nicht von Störungen überlagert ist) im Individuum. Die Unterschiede zwischen den Mitgliedern einer Gesamtheit sind nur vom Zufall bedingt, entstehen nur durch „Unfälle“. Dieser philosophische Ansatz des Aristoteles wurde in der Folgezeit aber gelegentlich übersehen. Zwischenzeitlich

ging das philosophische Interesse auch weg von der Natur und der Wissenschaft und konzentrierte sich mehr auf die Theologie mit Ethik und Moral sowie Ästhetik. Unter dem Einfluß eines Philosophen wie z. B. Augustinus (354–430), der als höchste Erkenntnisstufe die göttliche Eingebung und als die niedrigste aller Erkenntnisstufen die Gewinnung von Erkenntnissen aus Beobachtungen und Experimenten betrachtet, kann sich keine philosophische Grundlage entwickeln, auf der Naturwissenschaften – und damit auch die Statistik – aufbauen können.

Die platonische Philosophie wird neben der aristotelischen auch weiter vertreten. So bekannte Philosophen wie Descartes und Leibniz (1646–1716) vertreten in der Konsequenz die Ansicht, daß Experimentieren kein wahres Wissen bringt. Descartes etwa vertritt den Standpunkt, daß die Axiome der Mathematik deduktiv, also nicht durch Beobachtungen (Erfahrungen), entstanden sind.

Die Lehren des Galileo Galilei (1564–1642) bereiten so etwas wie eine Wende im Denken vor. Er beschäftigte sich mit der Berechnung von Chancen bei Glücksspielen, wobei er natürlich Beobachtungen zugrundelegte. Dies findet etwa ein Jahrhundert später seinen Niederschlag u. a. in der Philosophie von Hume (1711–1776) und Kant. Hume vertritt einen extremen empiristischen Standpunkt. Er vertritt die philosophische Ansicht, daß man Dinge, die man nicht experimentell erarbeiten kann, auch nicht weiß. Nach Hume sind also nur Experimente geeignet, das Wissen zu vermehren; man kann nur induktiv, also durch den Schluß von der Wirkung auf die Ursache, vom Teil auf das Ganze, zu neuen Erkenntnissen gelangen. Der „kritische“ Standpunkt Kants zeigt dagegen positive Möglichkeiten und Grenzen und kommt damit dem heutigen Wissenschaftsverständnis wohl am nächsten. Deshalb kann man die Kantsche Philosophie, die teilweise bei Aristoteles anknüpft, wohl als philosophische Grundlage der Naturwissenschaften, wie sie heute betrieben werden, betrachten. Kant versucht die „Begriffe“ und die „empirische Anschauung“ zu vereinen: Durch Konstruktion im „Gemüt“ werden die vielfachen sensorischen Informationen, die das Gehirn empfängt, synthetisiert. Kant stellt fest, daß man nicht das „Ding an sich“ (die absolute Realität), sondern nur Erscheinungen von ihm beobachtet, und daß man daher die Natur durch „Schemata“ (Modelle) nicht vollständig erklären kann. Ein „Schema“ (versinnlichter Begriff oder Hypothese) kann nach Kant sehr wohl dabei helfen, die empirische Wirklichkeit anhand von Beobachtungen „analytisch“ und „synthetisch“ zu erfassen.

Wie wichtig für die heutige Zeit die Wende im Denken durch die Werke Galileo Galileis und die daran anschließende philosophische Besinnung auf Aristoteles war, läßt sich vielleicht ermessen, wenn man weiß, daß zuvor praktisch nur die Gesetzmäßigkeiten der Sternenbewegung bekannt waren.

Faßt man die philosophischen Grundlagen der modernen Wissenschaften, und damit der Statistik noch einmal zusammen, so ergibt sich:

Das Experiment ist ein Verifikationsinstrumentarium, d. h. es dient zur Überprüfung von Hypothesen, die aufgrund von Beobachtungen aufgestellt werden. Die Erhebung hingegen ist nur ein Erkenntnisinstrumentarium: Aufgrund von Beobachtungen versucht man, reale Gesetzmäßigkeiten, reale Zustände zu erfassen. Ganz gleich, ob Experimente oder Erhebungen durchgeführt werden, sieht sich die Wissenschaft heute im Kantschen Verständnis.

Man muß Variabilität und Veränderung als wirklich und nichtillusorisch akzeptieren; man kann Teile (Einheiten, Objekte) nicht einzeln sehen, denn jedes Teil gibt nur ein unvollständiges Bild des Ganzen, zu dem es gehört; es gibt keinen speziellen

Grund für Variabilität, denn sie ist zufällig; Wissen und Erkenntnis können aus Beobachtungen abstrahiert werden; der induktive Prozeß, der es erlaubt, von Teilen auf das Ganze zu schließen, muß akzeptiert werden. (Deduktiv kann man nur vom Ganzen auf Teile schließen.)

5. Zur Geschichte der Statistik

Wir wollen hier einen kurzen Überblick über die wichtigsten Stationen der Entwicklung der Statistik geben, der natürlich keinesfalls den Anspruch auf Vollständigkeit erheben kann.

Die Grundlage aller statistischen Methoden ist die Wahrscheinlichkeitstheorie, die ihren Ursprung wohl in der Berechnung von Chancen bei Glücksspielen hat.

Schon sehr früh wird in einem Gedicht („de vetula“), das vermutlich Richard de Fournival (1200–1250) zuzuschreiben ist, die Anzahl der möglichen Wege mit Berücksichtigung der Reihenfolge angegeben, in denen drei Würfel fallen können, ohne sich dabei um irgendwelche Wahrscheinlichkeiten zu kümmern. Erst im 16. Jahrhundert legt der italienische Mathematiker, Physiker und Philosoph Gerolimino Cardano (1501–1576) in seinem ca. 1526 erschienenen Buch „Liber de Ludo Aleae“ empirisch fest, daß beim Würfeln mit einem „idealen“ Würfel als Ergebnis alle Seiten gleichwahrscheinlich sind. Er verallgemeinert dieses Ergebnis fälschlicherweise auf einen aus dem Sprungbein von Schafen gewonnenen vierseitigen Würfel: Hier trifft die Gleichwahrscheinlichkeit wegen der unregelmäßigen Form nicht zu.

Exakte Berechnungen der Wahrscheinlichkeiten beim Würfeln mit 3 Würfeln finden sich dann in dem Werk „Sopra le Scoperte dei Dadi“ des Italieners Galileo Galilei (1564–1642), das erstmals im Jahre 1718 unter dem Titel „Considerazione sopra il Gioco dei Dadi“ veröffentlicht worden ist.

Die ersten wirklich mathematisch orientierten Betrachtungen zur Berechnung von Chancen bei Glücksspielen werden durch folgende Anfrage aus dem Jahre 1654 des Chevalier de Méré beim französischen Mathematiker Blaise Pascal (1623–1662) angeregt:

„Stimmt es, daß von den beiden folgenden Wetten

- I. Ein Würfel wird viermal geworfen; gesetzt wird darauf, daß mindestens eine Sechs auftritt;
- II. Zwei Würfel werden (gleichzeitig) vierundzwanzigmal geworfen; gesetzt wird darauf, daß mindestens ein Sechserpasch auftritt;

man bei der ersten Wette öfter gewinnt als bei der zweiten?“

Aus dieser Anfrage resultierte die Korrespondenz zwischen Blaise Pascal und Pierre de Fermat (1601–1665), die die Mérés Vermutung aufgrund häufigen Spielens bestätigte: Im ersten Fall gewinnt man in 51,77% aller Fälle, im zweiten nur in 49,14% aller Fälle. Die Ideen der französischen Mathematiker wie etwa Pascal und Fermat wurden von Christian Huygens (1629–1695) in „de Ratiociniis in Alea Ludo“, einem Anhang zu seinem 1657 erschienenen „Exercitationes Mathematicae“, herauskristallisiert und erweitert. Im 18. Jahrhundert untersuchte dann der englische Geistliche Thomas Bayes (1702–1761), welche Aussage man bei Kenntnis der Ausgänge eines Spiels über die zugrundeliegenden Wahrscheinlichkeiten ma-

chen kann. Schließlich erscheint im Jahre 1812 die „Théorie Analytique des Probabilités“ von Pierre Simon Marquis de Laplace (1749–1827), in der er einen Überblick über den Stand der Wahrscheinlichkeitsrechnung gibt. Auch hier beschäftigt man sich fast ausschließlich noch mit solchen Problemen, wo nur endlich viele Ereignisse mit gleicher Wahrscheinlichkeit auftreten, also mit Würfel-, Karten- und Glücksspielen.

Die Wahrscheinlichkeitstheorie, wie sie heute betrieben wird, erlebte ihren Aufschwung erst mit ihrer Axiomatisierung durch A. N. Kolmogoroff im Jahre 1933.

Die eigentliche Statistik – bzw. was man heute unter Statistik versteht – entwickelte sich erst vom Beginn dieses Jahrhunderts an sehr stürmisch. Einige Ansätze und Grundlagen gehen jedoch schon auf frühere Zeiten zurück. So wurde im 16. Jahrhundert vom Astronom Tycho Brahe Dani (1546–1601) schon das arithmetische Mittel verwandt. Er beobachtete vom Februar 1582 bis April 1588 in unregelmäßigen Abständen den Stand von α Arietis. Dabei faßte er stets zwei Beobachtungen zusammen und bildete deren arithmetisches Mittel, um so Parallaxen und Refraktionen auszugleichen. Die so erhaltenen Ergebnisse waren erstaunlich exakt und wurden auch von Kepler (1571–1630) verwandt. Die Hypothese, daß Mädchen- und Knabengeburten gleichverteilt sind, wurde vom schottischen Arzt und Satiriker John Arbuthnot (1667–1735) durch statistisches Zahlenmaterial widerlegt. Im Jahre 1713 wird das Buch „Ars conjectandi“ von Jakob Bernoulli (1654–1705) veröffentlicht. Dort taucht der Begriff „Stochastisch“ auf, die Binomialverteilung und die Gesetze der großen Zahlen werden wohl erstmals erwähnt.

Von Abraham de Moivre (1667–1754) wird dann der Zusammenhang zwischen Binomialverteilung und Normalverteilung erkannt und 1718 in seinem Werk “The Doctrine of Chances” veröffentlicht. Ein frühes Beispiel für das Vorkommen der Normalverteilung gab der belgische Astronom und Physiker Adolphe Quetelet (1796–1874) mit der Messung des Brustumfangs von 5738 schottischen Soldaten. Der Name Normalverteilung wird allerdings erst von Francis Galton (1822–1911) um 1880 verwandt; heute nennt man sie zu Ehren von Carl Friedrich Gauß (1777–1855) manchmal auch Gaußverteilung. Die Nützlichkeit der Vertafelung der Verteilungsfunktion der Standardnormalverteilung wird schon im Jahre 1783 von Pierre Simon Marquis de Laplace erkannt. Zu Beginn des 19. Jahrhunderts erscheinen dann kurz hintereinander 2 Arbeiten über die Methode der kleinsten Quadrate und zwar (in dieser Reihenfolge) von Adrien Marie Legendre (1752–1833) und Carl Friedrich Gauß; letzterer baute darauf seine Theorie der Beobachtungsfehler auf. Die Methode der kleinsten Quadrate ist z. B. die Grundlage der modernen Regressionsrechnung, die schon in ersten Ansätzen (Minimierung der Abstandssumme) im Jahre 1755 bei Boscovich zu finden ist. Die u. a. für Gauß’ Theorie der Beobachtungsfehler wichtige χ^2 -Verteilung taucht erstmals im Jahre 1863 in der Arbeit „Über die Gesetzmäßigkeit in der Verteilung bei Beobachtungsreihen“ von Ernst Abbe (1840–1905) auf. Unabhängig davon wurde sie von dem Mathematiker, Astronomen und Geodäten Friedrich Robert Helmert (1843–1917) im Jahre 1876 entdeckt, geriet aber dann zunächst in Vergessenheit. Dann beginnt eine recht stürmische Phase in der Entwicklung der modernen Statistik.

1888 wird von Sir Francis Galton der Korrelationskoeffizient, der heute nach Karl Pearson benannt wird, angegeben. Karl Pearson (1857–1936) entdeckt im Jahre 1900 die χ^2 -Verteilung wieder, entwickelt den χ^2 -Anpassungstest, treibt die Korrelations- und Regressionsrechnung voran, vertafelt die vollständige Betafunktion, entwickelt die Gamma- und die Betaverteilung usw. Im Jahre 1901 gründet er

bereits die Zeitschrift Biometrika, die das wichtigste Organ der angelsächsischen Statistiker war. In den Jahren 1900, 1901 und 1903 erscheinen die Arbeiten von G. U. Yule (1871–1951) über Assoziationskoeffizienten in Vierfeldertafeln. We sentliche Fortschritte auf dem Gebiet der induktiven Statistik werden dann dadurch erzielt, daß der Mathematiker und Bierbrauer William Sealy Gosset (1876–1937) die t-Verteilung entdeckte und dies im Jahre 1908 – unter dem Pseudonym Student – veröffentlichte. Die Entwicklung der Methoden der Varianzanalyse von Sir Ronald Aylmer Fisher (1890–1962), der erstmals im Jahre 1935 mit Erscheinen seines Buches „The Design of Experiments“ eine ausführliche Darstellung von Varianzanalyse und Versuchsplanung veröffentlichte, wurde erst durch die Ableitung der F-Verteilung durch George W. Snedecor möglich; den Namen F-Verteilung erhielt diese Verteilung zu Ehren Fishers. Am Aufschwung der statistischen Testtheorie seit Beginn des Zweiten Weltkrieges waren vor allem Jerzy Neyman und Egon S. Pearson maßgeblich beteiligt. Schließlich wurden im Jahre 1950 die Schlußweisen der Statistik von Abraham Wald (1902–1950) in seinem Buch „Statistical Decision Functions“ einer entscheidungstheoretischen Betrachtungsweise zugänglich gemacht.

Der Ursprung des Wortes „Statistik“ selbst liegt wohl im Rahmen der bisher historisch kaum angesprochenen beschreibenden Statistik, genauer in der Beschreibung des Staates (seines Heeres, Gewerbes, seiner Bevölkerung usw.). Auf die geschichtliche Entwicklung in diesem Bereich wollen wir nun noch kurz eingehen.

Schon in der Bibel nimmt die Volkszählung eine wichtige Stellung ein: Im vierten Buch Moses, das ja „Numeri“ heißt, wird von der „Musterung des Volkes“ gesprochen. Diese Volkszählung wurde von Gott angeordnet; dagegen lastete auf der von David (2. Buch Samuel) eigenmächtig durchgeführten Zählung ein Fluch; David wurde dafür sogar von Gott bestraft. Noch im Mittelalter zeigte diese biblische Warnung ihre Wirkung: Die Erhebungen beschränken sich in dieser Zeit quasi auf Vermögenserhebungen. Karl der Große etwa ordnete die Bestandsaufnahme all seiner Güter, Domänen usw. an, und 250 Jahre später wurde ein Buch, das man als Grundkataster des Okzidents bezeichnen könnte, nämlich das Doomsday Book, im Auftrag von Wilhelm dem Eroberer angelegt.

In den Kirchenbüchern Englands beginnt dann Anfang des 16. Jahrhunderts die systematische Aufzeichnung von Geburts- und Todesfällen. Dies geschieht auf Anweisung des Lordkanzlers Thomas Cromwell (eines entfernten Verwandten von Oliver Cromwell), der (vielleicht auch aus zusätzlichen Gründen) des Hochverrats beschuldigt und hingerichtet wird. Im Jahre 1693 erscheint dann ein Werk des britischen Astronomen Edmond Halley (1656–1742) über Sterbewahrscheinlichkeiten (anhand der Kirchenbücher der Stadt Breslau) und daraus resultierenden Prämien für Lebensversicherungen. Im Jahre 1741, mit Erscheinen seines Werkes „Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen“, verhilft der preußische Feldprediger Johann Peter Süßmilch (1707–1767) der Bevölkerungsstatistik im deutschsprachigen Raum zum Durchbruch.

Die z. B. von Halley und Süßmilch vertretene Art der Bevölkerungsstatistik, der Staatskunde, ist auch unter dem Namen der politischen Arithmetik bekannt. In krassem Gegensatz hierzu steht eine andere, gleichzeitig betriebene Art der Staatskunde: die Staatsbeschreibung. Eines der ersten Werke dieser Richtung ist Francesco Sansovisos 1562 erschienenes „Del governo et amministrazione di diversi regni“. Im Jahre 1656 erscheint „Teutscher Fürstenstaat“ von Veit Ludwig von

Seckendorf. Dieses Werk ist ein Vorläufer der Staatsbeschreibungen Gottfried Achenwalls (1719–1772), die dieser selbst 1748 als „Statistik“ bezeichnet. Vermutlich bezieht er das Wort Statistik auf das neulateinische „status“, das soviel wie Zustand oder auch Staat heißt; Statistik ist im Sinne Achenwalls also die Zustandsbeschreibung des Staates. Auf diese Prägung des Wortes „statistique“ von Achenwall verweisen auch die französischen Wissenschaftler des 18. Jahrhunderts ausdrücklich.

Verstand Achenwall unter der Statistik noch die Vollerhebung zur Beschreibung des Staates, so ging man in der Folgezeit immer mehr zu den heute üblichen Stichprobenerhebungen über.

Kapitel I:

Aufbereitung und Darstellung von Datenmaterial

– Deskriptive Statistik

In vielen Bereichen wie z. B. den Wirtschaftswissenschaften, den Naturwissenschaften, der Medizin oder der Psychologie wird man häufig mit der Aufgabe konfrontiert, umfangreiches und unübersichtliches Datenmaterial zu beschreiben und eventuell zu analysieren.

Die **deskriptive Statistik** stellt nun zum einen **deskriptive also beschreibende Methoden** bereit, d. h. sie ist das direkte Instrumentarium zur Beschreibung von Datenmaterial; zum anderen ist jede deskriptive Methode aber auch eine Vorstufe der Analyse von Datenmaterial, eine Vorstufe der **schließenden Statistik**, die wir in den weiteren Kapiteln dieses Buches noch behandeln werden: Einerseits müssen Hypothesen aufgestellt werden, die aus den Resultaten der Deskription abgeleitet werden, andererseits müssen deskriptive Kenngrößen zur Hypothesenprüfung bestimmt werden.

Bevor wir uns mit den Methoden der deskriptiven Statistik konkret beschäftigen, soll nun im 1. Abschnitt ein Überblick über diese Methoden gegeben werden. Weiterhin werden in Abschnitt 1 einige der grundlegenden Begriffe, die uns nicht nur in diesem Kapitel, sondern auch in allen folgenden Kapiteln begegnen, erläutert.

1. Grundlegende Begriffe und Überblick

Die deskriptive Statistik dient der Beschreibung, der Strukturierung und der Verdeutlichung unübersichtlichen und umfangreichen Datenmaterials: Wo kommt dieses Material her, um was für Daten handelt es sich? Diese und ähnliche Fragen müssen zunächst einmal untersucht werden.

Daten kann man z. B. durch Befragung von Personen oder auch durch Messungen gewinnen.

Beispiel:

- (a) Indem man eine Reihe von Personen nach ihrem Alter, ihrem Familienstand, ihrer Kinderzahl usw. fragt, erhält man eine Vielzahl von Daten.
- (b) Befragt man Besitzer von landwirtschaftlichen Betrieben nach der Größe der ihnen zur Verfügung stehenden landwirtschaftlichen Nutzflächen, nach der Güte dieser Nutzflächen, nach dem Viehbestand ihres Betriebes und ähnlichem, so gewinnt man ebenfalls eine Vielzahl verschiedener Daten.
- (c) Mißt man zu verschiedenen Zeitpunkten Temperatur, Luftdruck, Niederschlagsmengen, Luftfeuchtigkeit usw., so stehen umfangreiche Datenmengen, die Auskunft über das Klima geben sollen, zur Verfügung.
- (d) Eine Vielzahl von Daten erhält man etwa auch dann, wenn mittels verschiedener Tests Musikalität, sprachliche Begabung, mathematische Begabung usw. von Personen ermittelt werden.

1.1. Untersuchungseinheiten, Merkmale und Merkmalsausprägungen

Die befragten Personen bzw. die Objekte, an denen Messungen vorgenommen wer-

den, nennt man auch **Untersuchungseinheiten**; die Größen, auf die sich Fragen oder Messungen beziehen, heißen **Merkmale**.

Beispiel:

- (e) Im Beispiel (a) sind die Personen die Untersuchungseinheiten und etwa Alter und Familienstand die beobachteten Merkmale. Im Beispiel (b) ist der Besitzer eines landwirtschaftlichen Betriebes bzw. der Betrieb selbst eine Untersuchungseinheit. Die interessierenden Merkmale sind hier z. B. Größe und Güte der landwirtschaftlichen Nutzfläche.
- (f) Im Beispiel (c) ist die Untersuchungseinheit das „Wetter“ zu einem bestimmten Zeitpunkt, und die beobachteten Merkmale sind Klimafaktoren wie z. B. die Temperatur. Im Beispiel (d) schließlich sind die Personen wiederum Untersuchungseinheiten, an denen Merkmale wie Musikalität beobachtet und geeignet gemessen werden.

Bei der Beobachtung von Merkmalen an Untersuchungseinheiten stellt man die **Merkmalsausprägung** (den **Merkmalswert**) jedes Merkmals an jeder Untersuchungseinheit fest; die Merkmalsausprägungen insgesamt sind dann identisch mit dem gewonnenen Datenmaterial.

Beispiel:

- (g) Im Beispiel (a) sind etwa die Ausprägungen des Merkmals „Alter“ verschiedene Zahlen von Jahren, wie z. B. 20 Jahre oder 75 Jahre, und die möglichen Ausprägungen des Merkmals „Familienstand“ sind: ledig, verheiratet, geschieden, verwitwet.

1.2. Charakterisierung von Merkmalen

Wie wir in Beispiel (g) schon sehen, können Merkmalsausprägungen sehr unterschiedlich sein: Die Ausprägungen sind nicht immer Zahlen, sie können z. B. auch Zustände sein. (Aus diesem Grund sprechen wir auch von Daten- und nicht von Zahlenmaterial.) Ausgehend von solchen Unterschieden möglicher Ausprägungen von Merkmalen werden Merkmale in Klassen eingeteilt. Eine Art der Klassifizierung ist die Unterscheidung **quantitativer und qualitativer Merkmale**: Die Ausprägungen quantitativer Merkmale unterscheiden sich durch ihre Größe, die Ausprägungen qualitativer Merkmale durch ihre Art.

Beispiel:

- (h) Quantitative Merkmale sind z. B. Alter, Gewicht, Körpergröße und Einkommen von Personen, Wasser- und Lufttemperaturen, Lebensdauern von Maschinen oder Viehbestände von landwirtschaftlichen Betrieben. Dagegen sind Merkmale wie Geschlecht, Beruf, Rasse, Haarfarbe, Fabrikat qualitativ. Auch Schulnoten, Güteklassen und z. B. Härtegrade von Wasser sind qualitative Merkmale, die mitunter jedoch auch quantitativ aufgefaßt werden.

Um die Ausprägung eines Merkmals messen oder erfragen zu können, muß man natürlich zunächst eine **Skala** (diese ist oft schon vorgegeben) festlegen, die alle möglichen Ausprägungen eines Merkmals beinhaltet. Auch die Art, das **Niveau** von solchen Skalen kann zur Klassifizierung von Merkmalen herangezogen werden.

Die Skala mit dem niedrigsten Niveau ist die sogenannte **Nominalskala**: Die Werte einer Nominalskala unterliegen keiner Rangfolge und sind nicht vergleich-

bar. Merkmale, deren Ausprägungen nur einer solchen Skala genügen, nennt man auch **nominale Merkmale**.

Beispiel:

- (i) Die Ausprägungen von Merkmalen wie Farbe, Beruf oder Rasse sind nicht vergleichbar und unterliegen keiner Rangfolge: Man kann z.B. nicht sagen, daß Schlosser besser als Bäcker sind oder umgekehrt. Es handelt sich hier also um nominale Merkmale.

Ein etwas höheres Niveau hat schon die **Ordinalskala**, deren Werte sich in ihrer Intensität unterscheiden und nach der Stärke der Intensität ordnen lassen. Ausprägungen **ordinaler Merkmale** unterliegen also einer bestimmten Rangfolge; Abstände zwischen verschiedenen Ausprägungen (Werten der Skala) lassen sich jedoch nicht interpretieren.

Beispiel:

- (j) Merkmale mit Ausprägungen wie gut, mittel und schlecht, also etwa Zensuren oder Härtgrade, sind ordinale Merkmale: die Ausprägungen unterliegen einer Rangfolge, Abstände sind nicht interpretierbar. Bei Schulnoten wird man etwa sagen können 1 (sehr gut) ist besser als 3 (befriedigend); der Abstand der Noten $3 - 1 = 2$ läßt sich jedoch nicht interpretieren, da 1 und 3 willkürlich gewählte Zahlen, die gewissen Leistungen zugeordnet werden, sind. Man könnte die Leistung „befriedigend“ etwa auch mit 6 und die Leistung „sehr gut“ mit 10 beziffern; der Abstand wäre dann $10 - 6 = 4$.

Die Skala mit dem höchsten Niveau ist die **metrische Skala**, die dadurch gekennzeichnet ist, daß ihre Werte nicht nur einer Reihenfolge unterliegen, sondern die Abstände zwischen den Werten auch interpretierbar sind. **Metrische** oder **metrisch skalierte Merkmale** sind natürlich gerade solche, deren Ausprägungsskala metrisch ist.

Beispiel:

- (k) Lebensdauern, Gewichte, Größen, Stromstärken u.ä. sind natürlich metrische Merkmale. Man kann etwa nicht nur sagen, daß eine Stromstärke von 10 Ampère größer ist als eine von 1 Ampère; vielmehr läßt sich auch der Abstand der Stromstärken interpretieren: 10 Ampère sind $10 - 1 = 9$ Ampère mehr als 1 Ampère.

Wir wollen nun noch eine weitere Art der Klassifizierung von Merkmalen kennenlernen, nämlich die Einteilung in **stetige** und **diskrete Merkmale**. Ein diskretes Merkmal kann nur endlich viele oder höchstens abzählbar unendlich viele Ausprägungen besitzen; dagegen nennen wir ein Merkmal stetig, wenn es als Ausprägung jeden beliebigen Wert in einem Bereich haben kann.

Beispiel:

- (l) Ein diskretes Merkmal ist etwa das „Geschlecht“ von Personen. Es hat nur zwei, also endlich viele Ausprägungen, nämlich männlich und weiblich. Auch die „Anzahl der Einwohner einer Stadt“ ist ein diskretes Merkmal, dessen Ausprägungen prinzipiell alle positiven ganzen Zahlen (also 1,2,3,...) sind; dies sind abzählbar unendlich viele Ausprägungen. Das Merkmal „Temperatur“ dagegen ist stetig, denn es kann beliebige Werte in einem Bereich als Ausprägungen besitzen. Auch das Merkmal „Lebensdauer“ ist stetig, denn die

Lebensdauer einer Untersuchungseinheit kann eine beliebige positive Zahl sein.

Es sei noch erwähnt, daß häufig stetige Merkmale nur diskret beobachtet werden, was z. B. daran liegen kann, daß ein Meßverfahren nicht beliebig genau ist: Mißt man etwa Temperaturen, so werden die Meßergebnisse zumeist nur in vollen Gradzahlen angegeben.

1.3. Grundgesamtheit und Stichprobe

Bisher sind wir davon ausgegangen, daß an Untersuchungseinheiten Merkmale beobachtet, erhoben werden. Was sind nun aber die Untersuchungseinheiten einer **Erhebung**? Wie wählt man diese aus? Für eine reine Beschreibung von Datenmaterial sind solche Fragen zweitrangig, denn Aussagen werden hier nur über die konkret beobachteten Untersuchungseinheiten gemacht. Sieht man jedoch die deskriptive Statistik als Vorstufe der schließenden Statistik, so sind diese Fragen grundlegend.

Beispiel:

(m) Befragt man 1000 Personen danach, ob sie arbeitslos sind oder nicht, so erlaubt die rein deskriptive Statistik nur Aussagen der Form „Von den 1000 Personen sind 50 Personen oder 5% arbeitslos“. Möchte man jedoch ausgehend von einer solchen Aussage schließen können, wieviel Prozent aller Personen (z. B. in der Bundesrepublik Deutschland) arbeitslos sind, so muß im vorhinein überlegt werden, welche Personen grundsätzlich befragt werden sollen und wie man aus allen möglichen Untersuchungseinheiten 1000 Personen für die Erhebung auswählt.

Die Frage nach den möglichen Untersuchungseinheiten erfordert eine **Abgrenzung der Grundgesamtheit** aller Untersuchungseinheiten.

Beispiel:

(m) Interessiert man sich für den prozentualen Anteil der Arbeitslosen an den arbeitsfähigen Personen zwischen 17 und 65 Jahren, die in der Bundesrepublik Deutschland leben, oder für den Prozentsatz der Arbeitslosen an der Gesamtbevölkerung? Im ersten Fall sind alle in der Bundesrepublik Deutschland lebenden, arbeitsfähigen Personen zwischen 17 und 65 Jahren mögliche Untersuchungseinheiten (sie bilden die Grundgesamtheit), im zweiten Fall besteht die Grundgesamtheit aus allen in der Bundesrepublik Deutschland lebenden Personen.

Welche Personen nun wirklich in die Erhebung einbezogen werden, ist ein Problem der Stichprobentheorie. Diese wird erst im Kapitel V in den Abschnitten 1 bis 3 ausführlich behandelt und ist hier auch nicht so wesentlich.

Im Beispiel (m) haben wir postuliert, daß nur 1000 Personen befragt werden. Prinzipiell könnte man auch alle zur Grundgesamtheit gehörigen Personen befragen. Schöpft man eine Grundgesamtheit voll aus, so spricht man von einer **Vollerhebung** oder **Totalerhebung**; tut man dies nicht, so führt man eine **Stichprobenerhebung** oder **Teilerhebung** durch.

Beispiel:

(n) Die bekanntesten und grundlegendsten Erhebungen zur Bestandsaufnahme der Bevölkerung sind die Volkszählungen. Dies sind Totalerhebungen mit

Haushalten als Untersuchungseinheiten; sie sollen in der Bundesrepublik Deutschland in der Regel alle zehn Jahre durchgeführt werden. Der lange Abstand zwischen den Volkszählungen wird durch die 1957 eingeführte „Repräsentativerhebung der Bevölkerung und des Erwerbslebens“ (**Mikrozensus**) überbrückt. Der Mikrozensus ist natürlich eine Teilerhebung.

Die Gründe für die Durchführung von Stichprobenerhebungen anstelle von Totalerhebungen sind vielfältig: Stichprobenerhebungen sind billiger und manchmal auch genauer als Totalerhebungen, da sie im Detail sorgfältiger durchgeführt werden können. Auch ist eine Vollerhebung mitunter sinnlos und zu langwierig: Will man etwa die Lebensdauern einer bestimmten Serie von Kühlschränken analysieren, so läge ein Ergebnis der Totalerhebung ja erst dann vor, wenn alle Kühlschränke defekt sind; eine Qualitätsbeurteilung ist für eventuelle Käufer dann uninteressant. Auch wäre die Dauer eines entsprechenden Versuchs viel zu lang.

1.4. Überblick über die Methoden der deskriptiven Statistik

Um das Datenmaterial einer Erhebung übersichtlich zu gestalten, bietet sich zunächst eine **tabellarische** und **graphische Darstellung**, die in den Abschnitten 2 und 3 behandelt wird, an. Bei metrischen und ordinalen Merkmalen empfiehlt es sich etwa, die beobachteten Merkmalsausprägungen der Größe nach zu ordnen. Weiterhin bietet es sich an, zu jeder Ausprägung eines Merkmals die Anzahl der Untersuchungseinheiten, bei denen sie beobachtet worden ist, zu berechnen und zum Beispiel durch unterschiedlich hohe Säulen graphisch darzustellen.

Oft ist es auch nützlich, große Datenmengen durch wenige charakteristische Größen (**Kenngrößen**) grob zu beschreiben. Zum einen kann man Zentren – sogenannte **Lageparameter** –, vergleiche auch Abschnitt 4, der beobachteten Ausprägungen eines Merkmals (z. B. die am häufigsten beobachtete Ausprägung) bestimmen, zum anderen können Maßzahlen der Variation der Ausprägungen – sogenannte **Streuungsparameter** –, vergleiche Abschnitt 5, berechnet werden (z. B. die Größe des Bereichs, in dem alle Ausprägungen liegen).

In Abschnitt 6 beschäftigen wir uns mit der graphischen Darstellung und der Bestimmung der **Konzentration** von Merkmalsausprägungen. Zwei Fragen, die mittels dieser Betrachtungen behandelt werden können, sind etwa die folgenden: Wieviel Prozent des gesamten Nettoeinkommens aller Beschäftigten entfallen auf wieviel Prozent der Beschäftigten? Entfällt vielleicht der größte Teil des Einkommens auf eine nur sehr kleine Zahl von Beschäftigten?

Im Beispiel (m) haben wir uns mit dem prozentualen Anteil der Arbeitslosen an der Gesamtbevölkerung beschäftigt. Hier wird die Zahl der Arbeitslosen zur Gesamtbevölkerung ins Verhältnis gesetzt. Solche sogenannten **Verhältniszahlen**, zu denen auch die **Indizes** gehören, werden in Abschnitt 7 ausführlich behandelt.

In Abschnitt 8 wird eine spezielle Kenngröße bei der Erhebung von Lebensdauern, nämlich die **empirische Ausfallrate**, vorgestellt.

Erhebt man an den Untersuchungseinheiten einer Erhebung mehrere Merkmale zugleich (z. B. Alter, Körpergröße und Gewicht von Personen), so kann man die einzelnen Merkmale (**eindimensionale Merkmale**) auch zu einem **mehrdimensionalen Merkmal** zusammenfassen. Auf die Darstellung und Beschreibung der Zusammenhänge bei speziell zwei Merkmalen wird in Abschnitt 9 eingegangen.

Schließlich werden in Abschnitt 10 noch einige Möglichkeiten zur Vereinfachung der Berechnung von Kenngrößen angegeben.

2. Der Häufigkeitsbegriff

Bei einer Erhebung wird nun an n Untersuchungseinheiten ein Merkmal X beobachtet, d.h. an jeder Einheit wird die Ausprägung dieses Merkmals festgestellt: Sind a_1, \dots, a_k die möglichen Ausprägungen des Merkmals X , so wird also der i -ten Untersuchungseinheit ($i = 1, \dots, n$) seine Ausprägung a_j als **Merkmalswert** x_i zugeordnet:

$$x_i = a_j(i).$$

Beispiel: Die Arbeiter (Untersuchungseinheiten) einer Fabrik werden gefragt, ob sie mit öffentlichen Verkehrsmitteln (a_1), privaten Verkehrsmitteln (a_2) oder zu Fuß (a_3) zu ihrer Arbeitsstelle gelangen. Kommt etwa der sechste befragte Arbeiter zu Fuß, so ist der zugehörige Merkmalswert gerade

$$x_6 = a_3(6) = „zu Fuß“.$$

Die insgesamt n Merkmalswerte x_1, \dots, x_n heißen auch **Beobachtungsreihe** oder **Urliste**.

Beispiel: Bei $n = 100$ Schulkindern einer Jahrgangsstufe sollen die Zähne auf Karies untersucht werden: Bei jedem Kind wird die Anzahl der kariösen Zähne registriert, wobei sich die Urliste x_1, \dots, x_{100} aus **Tab. 1** ergibt.

Tab. 1: Urliste der Anzahl kariöser Zähne bei 100 Schulkindern

1,0,0,3,1,5,1,2,2,0,1,0,5,2,1,0,1,0,0,4,0,1,1,3,0,
1,1,1,3,1,0,1,4,2,0,3,1,1,7,2,0,2,1,3,0,0,0,0,6,1,
1,2,1,0,1,0,3,0,1,3,0,5,2,1,0,2,4,0,1,1,3,0,1,2,1,
1,1,1,2,2,0,3,0,1,0,1,0,0,0,5,0,4,1,2,2,7,1,3,1,5

Wie man am Beispiel erkennt, ist das Material einer Urliste oft sehr unübersichtlich, so daß man – zumal häufig nur interessiert, welche Ausprägung eines Merkmals wie oft auftritt – zu einer anderen Darstellungsform übergehen sollte.

2.1. Absolute und relative Häufigkeiten

Beobachtet man an n Untersuchungseinheiten ein Merkmal X , das in k Ausprägungen a_1, \dots, a_k vorkommt, so ist

$$H_n(a_j) = „Anzahl der Fälle, in denen a_j auftritt“ für $j = 1, \dots, k$$$

die **absolute Häufigkeit** der Ausprägung a_j ; natürlich ist

$$H_n(a_1) + H_n(a_2) + \dots + H_n(a_k) = \sum_{j=1}^k H_n(a_j) = n.$$

Die absoluten Häufigkeiten hängen stark vom Umfang n der Erhebung ab, so daß sie sich nicht zum Vergleich von Erhebungen unterschiedlichen Umfangs eignen. Daher bietet es sich an, eine Größe zu suchen, die unabhängig vom Umfang n ist. Eine solche ist die **relative Häufigkeit**

$$h_n(a_j) = \frac{1}{n} \cdot H_n(a_j) \quad \text{für } j = 1, \dots, k,$$

die den prozentualen Anteil der Untersuchungseinheiten angibt, die die Ausprä-

gung a_j tragen: $h_n(a_j)$ nimmt unabhängig von n stets Werte zwischen Null und Eins an, und es gilt

$$h_n(a_1) + \dots + h_n(a_k) = \frac{1}{n} \cdot \sum_{j=1}^k H_n(a_j) = \frac{n}{n} = 1.$$

Wie wir im Beispiel noch sehen werden, gewinnt man die absoluten und relativen Häufigkeiten der Ausprägungen eines Merkmals aus der Urliste am besten durch Anfertigen einer **Strichliste**.

Beispiel: Zur Urliste aus Tab. 1 sollen nun die absoluten und relativen Häufigkeiten der einzelnen Ausprägungen des beobachteten Merkmals „Anzahl kariöser Zähne“ mittels einer Strichliste bestimmt werden, vgl. **Tab. 2**.

Tab. 2: Absolute und relative Häufigkeiten der verschiedenen Anzahlen kariöser Zähne von 100 Schulkindern, vgl. auch Tab. 1

Anzahl der kariösen Zähne a_j	Strichliste	$H_n(a_j)$	$h_n(a_j)$
0		30	0,30
1		34	0,34
2		14	0,14
3		10	0,10
4		4	0,04
5		5	0,05
6		1	0,01
7		2	0,02
Gesamt		100	1,00

Die Folge der Zahlen $H_n(a_1), \dots, H_n(a_k)$ nennt man auch die **absolute Häufigkeitsverteilung** und die Folge $h_n(a_1), \dots, h_n(a_k)$ auch die **relative Häufigkeitsverteilung** des beobachteten Merkmals. Dazu sei noch angemerkt, daß es sich bei metrischen und ordinalen Merkmalen empfiehlt, die Ausprägungen a_1, \dots, a_k der Größe nach zu ordnen, wie dies im Beispiel bereits geschehen ist.

2.2. Die graphische Darstellung von Häufigkeiten

Eine Folge von Zahlen, wie dies die Häufigkeitsverteilungen sind, läßt sich besonders gut übersehen, wenn man sie graphisch darstellt. Hierzu eignen sich zunächst Stabdiagramme, Häufigkeitspolygone oder Histogramme.

Beim **Stabdiagramm** wird über jeder Ausprägung auf der Abszisse die zugehörige Häufigkeit als senkrechte Strecke abgetragen; es eignet sich besonders dann, wenn das beobachtete Merkmal diskret ist. In **Abb. 1** sind die relativen Häufigkeiten des Beispiels „Anzahl kariöser Zähne“, vgl. auch Tab. 2, in Form eines Stabdiagramms dargestellt.

Das **Häufigkeitspolygon** erhält man, wenn man die Endpunkte der Strecken des Stabdiagramms miteinander verbindet, wie dies für die relativen Häufigkeiten unseres Beispiels in **Abb. 2** geschehen ist. Besonders geeignet ist ein Häufigkeitspolygon zur Darstellung zeitlicher Verläufe.

Beispiel: Drei Jahre lang wird an jedem Monatsende die Anzahl der Arbeitslosen registriert. Die insgesamt 36 Anzahlen werden dann durch ein Häufigkeitspolygon

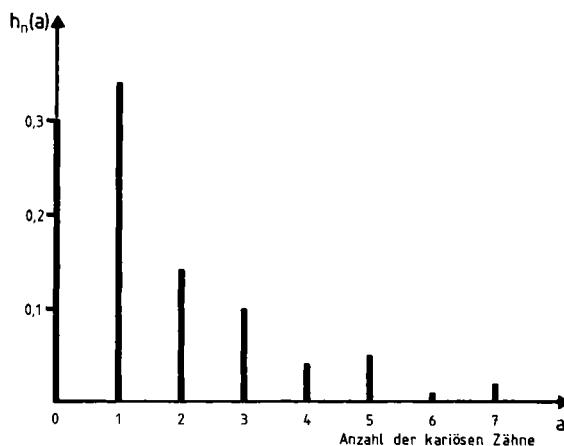


Abb. 1: Stabdiagramm

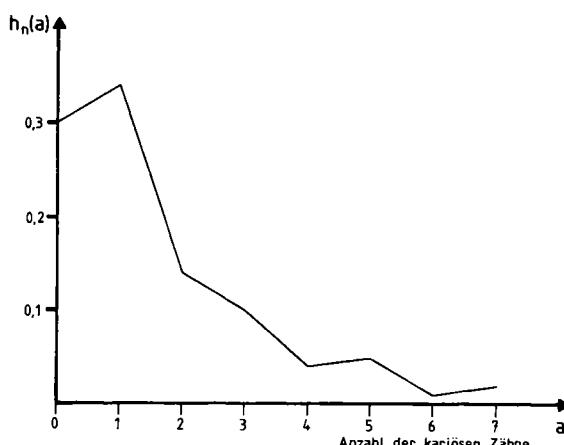


Abb. 2: Häufigkeitspolygon

beschrieben, wenn auf der Abzisse (= „Zeitachse“) über jedem Monat die Anzahl der Arbeitslosen abgetragen wird und diese Anzahlen verbunden werden.

Ein **Histogramm** erhält man, wenn man die beobachteten Häufigkeiten als aneinanderstoßende Rechtecke, deren **Flächen proportional den Häufigkeiten** sind, darstellt. Diese graphische Darstellungsform, die in Abb. 3 für unser Beispiel „Anzahl kariöser Zähne“ veranschaulicht wird, eignet sich vor allem dann, wenn das beobachtete Merkmal stetig ist.

Will man speziell eine Struktur darstellen, so bedient man sich sogenannter **Flächendiagramme**, bei denen die Häufigkeiten durch Flächen repräsentiert werden.

Beispiel: In Abb. 4 ist ein spezielles Flächendiagramm, nämlich ein **Kreisdiagramm** gewählt worden, um die Anzahlen der Beschäftigten in verschiedenen Wirtschafts-

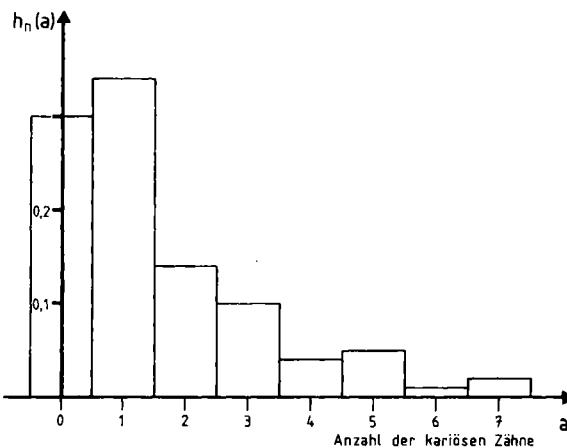


Abb. 3: Histogramm

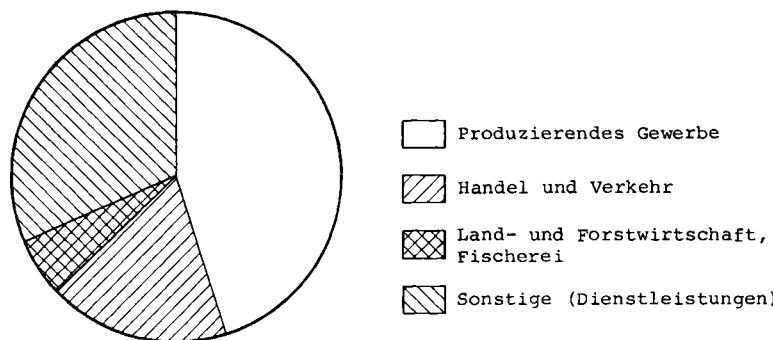


Abb. 4: Kreisdiagramm; Erwerbstätige nach Wirtschaftsbereichen 1979

zweigen darzustellen: Der Kreis wurde in Flächen zerlegt, die proportional zu den beobachteten Anzahlen sind.

2.3. Die empirische Verteilungsfunktion

Hat man bei ordinalen oder metrischen Merkmalen die möglichen Ausprägungen a_1, \dots, a_k der Größe nach geordnet, so bietet sich eine weitere Möglichkeit zur Beschreibung einer Beobachtungsreihe an. Statt wie bisher die Häufigkeiten einer jeden Ausprägung zu betrachten, bildet man sogenannte Summenhäufigkeiten:

Die **absolute Summenhäufigkeit** der j-ten Ausprägung a_j ist die Zahl der Untersuchungseinheiten, bei denen eine der Ausprägungen a_1, a_2, \dots, a_j beobachtet wurde, d.h. die absolute Summenhäufigkeit von a_j ist gegeben als

$$H_n(a_1) + H_n(a_2) + \dots + H_n(a_j) = \sum_{i=1}^j H_n(a_i) \quad \text{für } j = 1, \dots, k.$$

Entsprechend versteht man unter der **relativen Summenhäufigkeit** von a_j die Größe

$$h_n(a_1) + \dots + h_n(a_j) = \frac{1}{n} \cdot \sum_{i=1}^j H_n(a_i),$$

also die Summe der relativen Häufigkeiten von a_1, \dots, a_j .

Beispiel: Im Beispiel der kariösen Zähne bei Schulkindern, vgl. Tab. 2, gibt die absolute Summenhäufigkeit von a_j an, wieviele der 100 untersuchten Kinder höchstens a_j kariöse Zähne haben, und die relative Summenhäufigkeit von a_j ist der Anteil der Kinder mit höchstens a_j kariösen Zähnen.

Durch die Folge der relativen Summenhäufigkeiten wird nun die sogenannte **empirische Verteilungsfunktion** oder **Summenhäufigkeitsfunktion** $S_n(x)$ des beobachteten Merkmals X bestimmt:

Für alle Zahlen x mit $a_j \leq x < a_{j+1}$ ist

$$S_n(x) = \frac{1}{n} \cdot \sum_{i=1}^j H_n(a_i) = \sum_{i=1}^j h_n(a_i).$$

Die empirische Verteilungsfunktion ist also auf jedem Intervall $[a_j, a_{j+1})$ konstant und springt bei a_{j+1} um den Wert $h_n(a_{j+1})$ nach oben. Die erste Sprungstelle liegt bei der kleinsten, die letzte bei der größten beobachteten Merkmalsausprägung. Natürlich ist .

$$S_n(x) = 0 \quad \text{für } x < a_1 \quad \text{und} \quad S_n(x) = 1 \quad \text{für } x \geq a_k.$$

Beispiel: Ausgehend von Tab. 2 wollen wir die empirische Verteilungsfunktion der Anzahl kariöser Zähne bei 100 Schulkindern bestimmen. In Tab. 3 sind die relativen Summenhäufigkeiten angegeben, und in Abb. 5 ist die zugehörige empirische Verteilungsfunktion graphisch dargestellt.

Tab. 3: Relative Summenhäufigkeiten im Beispiel, vgl. auch Tab. 2

Anzahl der kariösen Zähne a_j	relative Häufigkeit $h_n(a_j)$	relative Summenhäufigkeit $\sum_{i=1}^j h_n(a_i)$
0	0,30	0,30
1	0,34	0,64
2	0,14	0,78
3	0,10	0,88
4	0,04	0,92
5	0,05	0,97
6	0,01	0,98
7	0,02	1,00

3. Der Häufigkeitsbegriff bei Klassenbildung

Die Häufigkeitsverteilung wird dann unübersichtlich, wenn ein Merkmal in sehr vielen Ausprägungen auftritt. In solchen Fällen betrachtet man nicht mehr die einzelnen Ausprägungen, sondern faßt verschiedene Ausprägungen in Klassen zusammen. Dies ist natürlich insbesondere dann der Fall, wenn das beobachtete Merkmal stetig ist. Man geht dann zumeist von vornherein von Klassen aus, denn

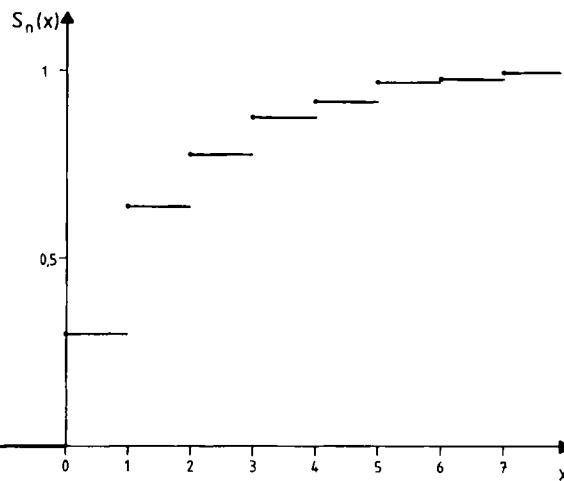


Abb. 5: Empirische Verteilungsfunktion im Beispiel

- selbst wenn man formal keine Klassen bildet, kann eine Merkmalsausprägung in Wirklichkeit einer Klasse entsprechen, weil man nicht beliebig genau messen kann;

Beispiel: An n Hühnereiern wird jeweils der größte Abstand x_i zwischen zwei Schalenpunkten gemessen. Wählt man als Maßeinheit „Millimeter“, so entspricht etwa ein Meßwert $x_i = 57$ mm einem wahren Abstand, der im Intervall [56,5 mm, 57,5 mm) liegt. In Wirklichkeit betrachtet man also Klassen; der Meßwert x_i entsteht also durch Rundung oder weil nicht ganz exakt gemessen werden kann.

- bei stetigen Merkmalen kann eine Ausprägung jede beliebige Zahl in einem bestimmten Bereich sein, so daß es selten vorkommt, daß mehrere gleiche Ausprägungen beobachtet werden. Häufig kommt es jedoch vor, daß einige Ausprägungen dicht beieinander liegen. Es bietet sich dann an, diese in einer Klasse zusammenzufassen.

Beispiel: In Tab. 4 sind die Lebensdauern von 87 Kühlaggregaten (der Größe nach geordnet) angegeben. Es kommt kaum einmal vor, daß mehrere Lebensdauern gleich sind; jedoch sind die Unterschiede oft relativ gering.

Tab. 4: Lebensdauern von 87 Kühlaggregaten (der Größe nach geordnet) in Jahren

0,05	0,06	0,06	0,08	0,11	0,13	0,15	0,16	0,20
0,22	0,24	0,25	0,25	0,28	0,31	0,34	0,37	0,42
0,43	0,47	0,51	0,51	0,53	0,59	0,60	0,61	0,63
0,68	0,75	0,76	0,76	0,79	0,87	0,88	0,88	0,92
0,99	1,12	1,16	1,18	1,22	1,27	1,35	1,38	1,39
1,42	1,45	1,49	1,53	1,69	1,74	1,81	1,83	1,87
1,92	1,93	2,07	2,09	2,15	2,22	2,24	2,36	2,39
2,41	2,47	2,49	2,53	2,64	2,69	2,83	2,90	3,21
3,25	3,49	3,61	3,80	3,88	4,37	4,58	4,62	5,29
5,68	6,02	6,23	6,71	7,82	9,93			

3.1. Die Klassenbildung

Wie geht man nun konkret bei der Klassenbildung vor? Wir betrachten hier nur den Fall ordinaler oder metrischer Merkmale, denn die Ausprägungen nominaler Merkmale sind meist nicht so zahlreich.

Zunächst zerlegt man das Intervall, in dem alle Beobachtungswerte liegen, in mehrere Teilintervalle, die sogenannten **Klassen** oder **Gruppen**.

Eine solche Klasse ist dann eindeutig bestimmt durch **Klassenmitte** und **Klassenbreite** oder durch **untere** und **obere Klassengrenze**. Natürlich muß im vorhinein festgelegt werden, zu welcher Klasse die Klassengrenzen gehören:

Zu jeder Klasse gehört die untere Klassengrenze bzw.

zu jeder Klasse gehört die obere Klassengrenze bzw.

man zählt einen Wert, der auf die Klassengrenze fällt, je zur Hälfte zu den beiden benachbarten Klassen.

Um Schwierigkeiten aus dem Weg zu gehen, sollte man, wenn möglich, die Klassengrenzen so wählen, daß kein Beobachtungswert dort liegt.

Obwohl sich rechnerische Vereinfachungen ergeben, wenn man als Klassenmitten möglichst einfache Zahlen und alle Klassen gleich breit wählt, legt man aus sachlichen Gründen oft verschiedene Klassen fest.

Beispiel: Sind die beobachteten Merkmalsausprägungen die in einer Klausur erreichten Punktzahlen, so bietet es sich an, die Klasseneinteilung durch zugehörige Noten festzulegen, obwohl die zu den Noten gehörigen Punktintervalle nicht alle gleich groß sind.

3.2. Absolute und relative Häufigkeiten bei Klassenbildung

Genauso, wie wir in Abschnitt 2 absolute und relative Häufigkeiten von Merkmalsausprägungen bestimmt haben, können wir Klassenhäufigkeiten bestimmen.

Ist p die Anzahl der gebildeten Klassen, so ist die **absolute Klassenhäufigkeit** der i -ten Klasse K_i für $i = 1, \dots, p$ gerade

$$H_i = \text{„Anzahl der Beobachtungswerte in der Klasse } K_i\text{“},$$

und die **relative Klassenhäufigkeit** dieser Klasse ist

$$h_i = \frac{1}{n} H_i.$$

Beispiel: (a) In Abschnitt 2 haben wir als ein Beispiel die Anzahl kariöser Zähne bei 100 Schulkindern betrachtet. Die Merkmalsausprägungen waren $0,1, \dots, 7$ kariöse Zähne, und die zugehörigen absoluten und relativen Häufigkeiten sind in Tab. 2 angegeben. Wir wollen nun in diesem Beispiel zur Klassenbildung übergehen:

geringer Kariesbefall $\hat{=} 0, 1$ oder 2 kariöse Zähne,

mittlerer Kariesbefall $\hat{=} 3$ oder 4 kariöse Zähne,

starker Kariesbefall $\hat{=} 5, 6$ oder 7 kariöse Zähne.

Da es für die graphische Darstellung, wie wir noch sehen werden, günstig ist, daß die Klassen aneinanderstoßen, setzen wir $K_1 = [-0,5; 2,5)$, $K_2 = [2,5; 4,5)$ und $K_3 = [4,5; 7,5)$, obwohl etwa die Ausprägung $2,5$ kariöse Zähne gar nicht auftreten kann.

Ausgehend von dieser Klasseneinteilung in 3 Klassen sind in Tab. 5 die absoluten und relativen Klassenhäufigkeiten angegeben.

Tab. 5: Absolute und relative Klassenhäufigkeiten im Beispiel „Anzahl kariöser Zähne“; vgl. auch Tab. 2

Klasse K_i	kariöse Zähne	absolute Klassenhäufigkeit	relative Klassenhäufigkeit
$K_1 = [-0,5; 2,5)$	geringer Kariesbefall	78	0,78
$K_2 = [2,5; 4,5)$	mittlerer Kariesbefall	14	0,14
$K_3 = [4,5; 7,5)$	starker Kariesbefall	8	0,08
Gesamt		100	1

(b) Auch im Beispiel der Lebensdauern von 87 Kühlaggregaten, vgl. Tab. 4, wollen wir eine Klassenbildung vornehmen. Wir bilden 10 Klassen der Breite 1 (= 1 Jahr), d. h. es ist

- $K_1 = \{\text{Lebensdauer von weniger als einem Jahr}\} = [0,00; 1,00),$
 $K_2 = \{\text{Lebensdauer von mindestens einem und weniger als zwei Jahren}\}$
 $= [1,00; 2,00), \dots,$
 $K_{10} = \{\text{Lebensdauer von mindestens neun und weniger als zehn Jahren}\}$
 $= [9,00; 10,00).$

Die zugehörigen absoluten und relativen Klassenhäufigkeiten sind in **Tab. 6** angegeben.

Tab. 6: Klasseneinteilung für die Lebensdauern der 87 Kühlaggregate, vgl. auch Tab. 4

Klasse K_i	absolute Klassenhäufigkeiten H_i	relative Klassenhäufigkeiten h_i (gerundet)
$K_1 = [0; 1)$	37	0,425
$K_2 = [1; 2)$	19	0,218
$K_3 = [2; 3)$	15	0,172
$K_4 = [3; 4)$	6	0,069
$K_5 = [4; 5)$	3	0,034
$K_6 = [5; 6)$	2	0,022
$K_7 = [6; 7)$	3	0,034
$K_8 = [7; 8)$	1	0,011
$K_9 = [8; 9)$	0	0,000
$K_{10} = [9; 10)$	1	0,011

3.3. Die graphische Darstellung von Häufigkeiten bei Klassenbildung

Will man Häufigkeiten von in Klassen zusammengefaßten Beobachtungswerten graphisch darstellen, so bedient man sich zumeist eines **Histogramms**. Hier wird es wichtig, daß die Klassen aneinanderstoßend gewählt wurden, denn nur dann lassen sich die Häufigkeiten sinnvoll durch aneinanderstoßende Rechtecke darstellen. Die Flächen dieser Rechtecke beschreiben die Häufigkeit einer Klasse, d. h. nicht die Höhe des Rechtecks über einer Klasse ist proportional zur Klassenhäufigkeit, sondern die Fläche selbst ist es. Somit ist das Produkt von Rechteckhöhe und Klassenbreite einer Klasse proportional zur beobachteten Klassenhäufigkeit.

Wir wollen uns im folgenden auf die Darstellung relativer Klassenhäufigkeiten beschränken. Setzen wir die Fläche des Rechtecks über der i-ten Klasse K_i gleich der relativen Klassenhäufigkeit h_i dieser Klasse, so ist es recht einfach, die Höhe r_i des Rechtecks zu bestimmen, wenn wir die Klassenbreite mit b_i bezeichnen:

$$h_i = b_i \cdot r_i, \quad \text{d.h.} \quad r_i = \frac{h_i}{b_i}.$$

Beispiel: (a) In Tab. 5 sind die relativen und absoluten Klassenhäufigkeiten für das Beispiel „Anzahl kariöser Zähne“ bei 3 Klassen angegeben. Wir wollen nun die relativen Klassenhäufigkeiten graphisch in Form eines Histogramms darstellen. Dazu müssen wir zunächst die Klassenbreiten angeben und daraus die Höhen r_i der Rechtecke über den Klassen bestimmen, wie dies in **Tab. 7** erfolgt ist. Nun können wir das Histogramm erstellen, vgl. **Abb. 6**.

Tab. 7 Berechnungen für das Histogramm in Abb. 6

Klasse K_i	relative Klassenhäufigkeit h_i	Klassenbreite b_i	Rechteckshöhe r_i
K_1	0,78	3	0,26
K_2	0,14	2	0,07
K_3	0,08	3	0,027

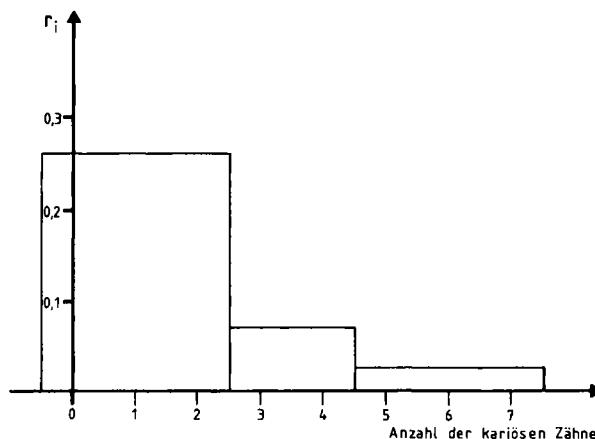


Abb. 6: Histogramm zum Beispiel

(b) Auch zum Beispiel der 87 Kühlaggregate wollen wir ein Histogramm ausgehend von Tab. 6 erstellen. Da hier die Klassenbreiten alle 1 Jahr sind, stimmen die relativen Klassenhäufigkeiten h_i jeweils mit der Höhe r_i der Rechtecke über den Klassen überein, so daß sich das Histogramm aus **Abb. 7** ergibt.

3.4. Die empirische Verteilungsfunktion bei Klassenbildung

Hauptsächlich dann, wenn die einzelnen Beobachtungswerte x_1, \dots, x_n nicht zur Verfügung stehen und nur noch gewisse Klassenhäufigkeiten bekannt sind, wird

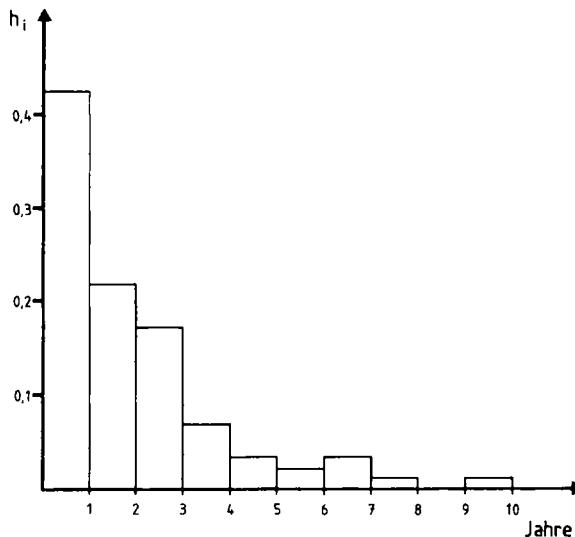


Abb. 7: Histogramm der Lebensdauer von 87 Kühlaggregaten bei Klassenbreite 1

man die **empirische Verteilungsfunktion** des beobachteten Merkmals X aus diesen Klassenhäufigkeiten berechnen. Hierbei gibt es verschiedene Vorgehensweisen, von denen zwei hier vorgestellt werden sollen, man vgl. aber auch Kap. XIV, 1.6.

Insbesondere im Fall symmetrisch um die Klassenmitten verteilter Beobachtungswerte bietet es sich an, den Wert der empirischen Verteilungsfunktion $S_p(x)$ (bei Bildung von p Klassen) an der Stelle x als die Summe der relativen Häufigkeiten aller Klassen, deren Mitten kleiner oder gleich x sind, festzulegen. Bezeichnet also m_i für $i = 1, \dots, p$ die Klassenmitte der i-ten Klasse K_i und ist h_i die relative Klassenhäufigkeit dieser Klasse, so kann man

$$S_p(x) = \sum_{j=1}^i h_j \quad \text{für } m_i \leq x < m_{i+1}$$

definieren. Die Sprungstellen der empirischen Verteilungsfunktion sind dann identisch mit den Klassenmittnen; für $x < m_1$ ist $S_p(x) = 0$ und für $x \geq m_p$ ist $S_p(x)$ natürlich gerade Eins.

Eine andere Möglichkeit der Definition der empirischen Verteilungsfunktion ist die folgende: Der Wert $S_p(x)$ ist für jede Zahl x gleich der Summe der relativen Klassenhäufigkeiten aller Klassen, deren Endpunkte kleiner oder gleich x sind. Ist also y_i für $i = 1, \dots, p$ der Endpunkt der Klasse K_i , die mit Häufigkeit h_i auftritt, so ist

$$S_p(x) = \sum_{j=1}^i h_j \quad \text{für } y_i \leq x < y_{i+1}.$$

Bei dieser Definition der empirischen Verteilungsfunktion fallen die Sprungstellen also mit den Klassenendpunkten zusammen und es ist $S_p(x) = 0$ für $x < y_1$ sowie $S_p(x) = 1$ für $x \geq y_p$.

| Beispiel: In Abb. 8 und Abb. 9 sind die empirischen Verteilungsfunktionen bei

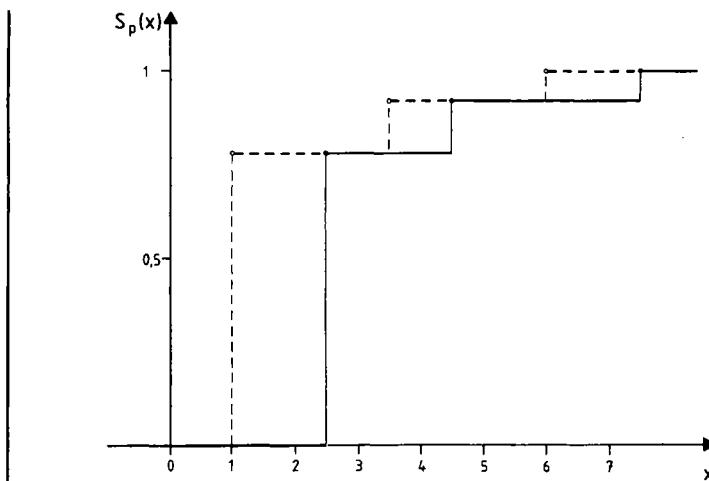


Abb. 8: Empirische Verteilungsfunktion der Anzahl kariöser Zähne bei Berechnung durch das Klassenende y_i (durchgehende Linie) und die Klassenmitte m_i (gestrichelte Linie)

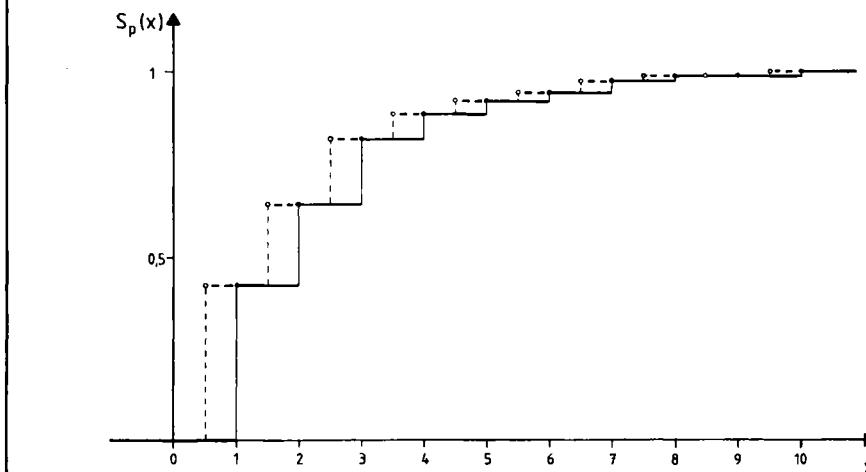


Abb. 9: Empirische Verteilungsfunktion der Lebensdauer von Kühlaggregaten bei Berechnung durch das Klassenende y_i (durchgehende Linie) und die Klassenmitte m_i (gestrichelte Linie)

Tab. 8: Berechnung der empirischen Verteilungsfunktion bei Klassenbildung im Beispiel „Anzahl kariöser Zähne“

Klasse K_i	Klassenmitte m_i	Klassenend- punkt y_i	relative Klassenhäufigkeit h_i	$S_p(x)$ für $m_i \leq x < m_{i+1}$ bzw. für $y_i \leq x < y_{i+1}$
K_1	1,0	2,5	0,78	0,78
K_2	3,5	4,5	0,14	0,92
K_3	6,0	7,5	0,08	1,00

Tab. 9: Berechnung der empirischen Verteilungsfunktion bei Klassenbildung im Beispiel „Lebensdauer von Kühlaggregaten“

Klasse K _i	Klassenmitte m _i	Klassenend- punkt y _i	relative Klassenhäufigkeit h _i	S _p (x) für m _i ≤ x < m _{i+1} bzw. für y _i ≤ x < y _{i+1}
K ₁	0,5	1	0,425	0,425
K ₂	1,5	2	0,218	0,644
K ₃	2,5	3	0,172	0,816
K ₄	3,5	4	0,069	0,885
K ₅	4,5	5	0,034	0,920
K ₆	5,5	6	0,022	0,943
K ₇	6,5	7	0,034	0,977
K ₈	7,5	8	0,011	0,989
K ₉	8,5	9	0,000	0,989
K ₁₀	9,5	10	0,011	1,000

Klassenbildung für die Beispiele „Anzahl kariöser Zähne“, vgl. auch Tab. 5, und „Lebensdauer von Kühlaggregaten“, vgl. Tab. 6, für beide Berechnungsmöglichkeiten dargestellt. Die nötigen Berechnungen findet man in **Tab. 8** bzw. **Tab. 9**.

4. Lagemaße von Häufigkeitsverteilungen

Oft lässt sich eine Beobachtungsreihe durch wenige charakteristische Größen kennzeichnen. Ein solches Charakteristikum ist z. B. eine Maßzahl, die in geeigneter Weise ein „Zentrum“ der Beobachtungswerte x₁, ..., x_n angibt. Einige solcher Maßzahlen, die man **Lagemaße** nennt, sollen hier vorgestellt werden.

4.1. Das arithmetische Mittel

Sind x₁, ..., x_n die beobachteten Merkmalswerte eines Merkmals X mit Ausprägungen a₁, ..., a_k, so ist das **arithmetische Mittel (Mittelwert, durchschnittlicher Wert)**

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{n} \cdot \sum_{j=1}^k a_j \cdot H_n(a_j)$$

ein sinnvolles Lagemaß vor allem dann, wenn das Merkmal X metrisch skaliert ist.

Sind die Beobachtungswerte in p Klassen eingeteilt, so ist natürlich

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^p \bar{x}_i \cdot H_i \quad \text{mit } \bar{x}_i = \frac{1}{H_i} \cdot \sum_{x_i \in K_i} x_i$$

das arithmetische Mittel aller Beobachtungswerte; \bar{x}_i ist gerade das arithmetische Mittel der Beobachtungswerte in der i-ten Klasse.

Mitunter kommt es vor, daß bei gruppierten Daten die Klassenmittel nicht bekannt sind. Dann berechnet man als arithmetisches Mittel, ausgehend von den Klassenmitteln m₁, ..., m_p,

$$\bar{x}_M = \frac{1}{n} \cdot \sum_{i=1}^p m_i \cdot H_i ;$$

natürlich gilt $\bar{x} = \bar{x}_M$ nur dann, wenn die Beobachtungswerte jeder Klasse symmetrisch um die Klassenmitte verteilt liegen.

Beispiel: Im Beispiel „Anzahl kariöser Zähne bei 100 Schulkindern“ aus Abschnitt 2, vgl. Tab. 2, ergibt sich als arithmetisches Mittel der ungruppierten Daten

$$\begin{aligned}\bar{x} &= \frac{1}{100} (0 \cdot 30 + 1 \cdot 34 + 2 \cdot 14 + 3 \cdot 10 + 4 \cdot 4 + 5 \cdot 5 + 6 \cdot 1 + 7 \cdot 2) = \\ &= 1,53,\end{aligned}$$

d.h.: Im Durchschnitt haben die 100 Kinder 1,53 kariöse Zähne. Gehen wir zu den gruppierten Daten aus Tab. 5 über, so ergibt sich natürlich ebenfalls

$$\bar{x} = 1,53,$$

denn es ist

$$\begin{aligned}\bar{x}_1 &= \frac{1}{78} (0 \cdot 30 + 1 \cdot 34 + 2 \cdot 14) = \frac{31}{39} = 0,79, \\ \bar{x}_2 &= \frac{1}{14} (3 \cdot 10 + 4 \cdot 4) = \frac{23}{7} = 3,29 \quad \text{und} \\ \bar{x}_3 &= \frac{1}{8} (5 \cdot 5 + 6 \cdot 1 + 7 \cdot 2) = \frac{45}{8} = 5,625, \quad \text{so daß sich} \\ \bar{x} &= \frac{1}{100} \left(\frac{31}{39} \cdot 78 + \frac{23}{7} \cdot 14 + \frac{45}{8} \cdot 8 \right) = 1,53\end{aligned}$$

ergibt.

Gehen wir einmal davon aus, daß wir die Klassenmittel nicht mehr kennen, und berechnen

$$\bar{x}_M = \frac{1}{100} (1 \cdot 78 + 3,5 \cdot 14 + 6 \cdot 8) = 1,75.$$

Der wirkliche Durchschnittswert \bar{x} wird hier also überschätzt, da alle Klassenmittel durch die Klassenmitten überschätzt werden.

Gibt man sich allgemein Zahlen (**Gewichte**) w_1, \dots, w_n mit $0 \leq w_i \leq 1$ und $\sum_{i=1}^n w_i = 1$ vor, so erhält man ein sogenanntes **gewogenes arithmetisches Mittel**

$$\bar{x}^w = \sum_{i=1}^n w_i x_i.$$

Mit $w_1 = \dots = w_n = 1/n$ ergibt sich wieder das gewöhnliche arithmetische Mittel \bar{x} .

4.2. Der Median und das α -Quantil

4.2.1. Der Median einer Beobachtungsreihe

Der **Median** oder **Zentralwert** ist dadurch charakterisiert, daß jeweils mindestens 50% der Beobachtungen x_1, \dots, x_n einen Wert größer oder gleich bzw. kleiner oder gleich dem Median annehmen. Sind also $x_{(1)} \leq \dots \leq x_{(n)}$ die der Größe nach geordneten Beobachtungswerte (sind mehrere Werte gleich, so spielt deren Reihenfolge keine Rolle), so läßt sich der Median $\tilde{x}_{0,5}$ als

$$\tilde{x}_{0,5} = \begin{cases} x_{((n+1)/2)}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(n/2)} + x_{((n+2)/2)}), & \text{falls } n \text{ gerade} \end{cases}$$

angeben. Bei einer ungeraden Anzahl von Merkmalsausprägungen ist der Median also der Wert in der Mitte der geordneten Reihe, bei einer geraden Anzahl ist jede Zahl, die zwischen den beiden mittleren Werten liegt, ein Median; wir aber wählen hier stets das arithmetische Mittel der beiden Werte als speziellen Median. Der Median hat gegenüber dem durchschnittlichen Wert \bar{x} als Lagemaß den Vorteil,

daß er auch bei ordinal skalierten Merkmalen verwandt werden kann und daß er weniger empfindlich gegenüber Ausreißern in der Beobachtungsreihe ist; d. h. Werte, die weit von allen übrigen entfernt liegen, beeinflussen den Wert des Medians kaum.

Beispiel: (a) Im Beispiel „Lebensdauern von 87 Kühlaggregaten“ aus Abschnitt 3 wollen wir den Median $\tilde{x}_{0,5}$ der Beobachtungsreihe bestimmen; in Tab. 4 ist die geordnete Reihe aller 87 Beobachtungswerte zu finden. Da $n = 87$ ungerade ist, entspricht der Median gerade dem $(n + 1)/2 = 88/2 = 44$ ten Wert in der Meßreihe:

$$\tilde{x}_{0,5} = x_{(44)} = 1,38.$$

(b) Der Median der Beobachtungsreihe $x_{(1)} = 3, x_{(2)} = 5, x_{(3)} = 6, x_{(4)} = 9$ ist $\tilde{x}_{0,5} = (5 + 6)/2 = 5,5$.

Auch wenn gruppierte Daten vorliegen, läßt sich natürlich der Median bestimmen: Sind K_1, \dots, K_p die verschiedenen Klassen, die mit den relativen Häufigkeiten h_1, \dots, h_p auftreten, so bestimmt man zunächst die Klasse K_i , die den Median enthält. Für sie muß gelten

$$\sum_{j=1}^{i-1} h_j < 0,5 \quad \text{und} \quad \sum_{j=1}^i h_j \geq 0,5, \quad \text{wobei} \quad \sum_{j=1}^0 h_j = 0 \quad \text{gesetzt wird.}$$

Hat man nun die Klasse K_i gefunden, so ermittelt man ihre untere Klassengrenze u_i und ihre Breite b_i . Der Median ist dann gerade

$$\tilde{x}_{0,5} = u_i + \frac{0,5 - \sum_{j=1}^{i-1} h_j}{h_i} \cdot b_i;$$

graphisch bedeutet dies, daß der Median die gesamte Histogrammfläche in zwei gleichgroße Teile zerlegt, vgl. Abb. 10.

Beispiel: Im Beispiel der Kühlaggregate aus Abschnitt 3 soll nun der Median aus den gruppierten Daten bei Klassenbreite 1 bestimmt werden, vgl. Tab. 6. Der Median liegt hier in der Klasse K_2 , denn es ist

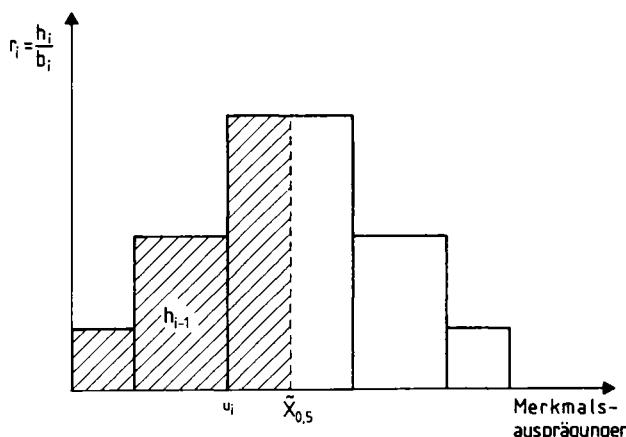


Abb. 10: Bestimmung des Median $\tilde{x}_{0,5}$

$$h_1 \simeq 0,425 < 0,5 \quad \text{und} \quad h_1 + h_2 \simeq 0,644 > 0,5.$$

Weiter ist

$$u_2 = 1 \quad \text{bzw.} \quad b_2 = 1$$

die untere Grenze bzw. die Breite der Klasse K_2 , so daß sich als Median der Wert

$$\tilde{x}_{0,5} = u_2 + \frac{0,5 - h_1}{h_2} \cdot b_2 \simeq 1 + \frac{0,5 - 0,425}{0,218} \cdot 1 = 1 + \frac{0,075}{0,218} \simeq 1,344$$

ergibt.

4.2.2. Das α -Quantil einer Beobachtungsreihe

Ist $x_{(1)} \leq \dots \leq x_{(n)}$ die geordnete Beobachtungsreihe zur Urliste x_1, \dots, x_n , so bezeichnet man als α -Quantil \tilde{x}_α ($0 < \alpha < 1$) dieser Reihe gerade den Wert

$$\tilde{x}_\alpha = \begin{cases} x_{(k)} & , \text{ falls } n \cdot \alpha \text{ keine ganze Zahl ist} \\ & (k \text{ ist dann die auf } n \cdot \alpha \text{ folgende ganze Zahl}) \\ \frac{1}{2}(x_{(k)} + x_{(k+1)}), & \text{falls } n \cdot \alpha \text{ eine ganze Zahl ist} \\ & (\text{Es ist dann } k = n \cdot \alpha) \end{cases}$$

Das α -Quantil \tilde{x}_α wird für ganze Zahlen $n \cdot \alpha$ so gewählt, daß $n \cdot \alpha$ Beobachtungswerte kleiner oder gleich und $n(1 - \alpha)$ Werte größer oder gleich \tilde{x}_α sind. Ist $n \cdot \alpha$ keine ganze Zahl, so soll für das α -Quantil gelten, daß $k - 1$ Werte kleiner und $n - k$ Werte größer als \tilde{x}_α sind, wobei k die auf $n \cdot \alpha$ folgende ganze Zahl ist.

Ist $n \cdot \alpha$ eine ganze Zahl, so würde diese Bedingung prinzipiell durch jede Zahl zwischen $x_{(n \cdot \alpha)}$ und $x_{(n \cdot \alpha + 1)}$ erfüllt; wir haben hier jedoch als spezielles α -Quantil den Durchschnitt dieser beiden Werte gewählt.

Wie man sieht ist der Median $\tilde{x}_{0,5}$ ein spezielles α -Quantil, nämlich gerade das 0,5-Quantil der Beobachtungsreihe. Weitere Quantile, die einen speziellen Namen tragen, sind das 0,25-Quantil und das 0,75-Quantil: Sie heißen **unteres** bzw. **oberes Quartil**.

Beispiel: Wie wollen das 0,1-Quantil der Lebensdauern von 87 Kühlaggregaten, vgl. Tab. 4 in Abschnitt 3, bestimmen. Es ist

$$n \cdot \alpha = 87 \cdot 0,1 = 8,7,$$

so daß, da 8,7 keine ganze Zahl ist, der neunte Wert $x_{(9)}$ in der Reihe der geordneten Beobachtungswerte dem 0,1-Quantil entspricht:

$$\tilde{x}_\alpha = \tilde{x}_{0,1} = x_{(9)} = 0,20.$$

Falls die Beobachtungsdaten nur gruppiert vorliegen, kann man bei der Bestimmung eines beliebigen α -Quantils analog zur Bestimmung des Medians vorgehen: Das α -Quantil liegt in der Klasse K_i , falls gilt

$$\sum_{j=1}^{i-1} h_j < \alpha \quad \text{und} \quad \sum_{j=1}^i h_j \geq \alpha.$$

Ist dann u_i die untere Grenze dieser Klasse und b_i die Klassenbreite, so ist

$$\tilde{x}_\alpha = u_i + \frac{\alpha - \sum_{j=1}^{i-1} h_j}{h_i} \cdot b_i.$$

Durch diese Vorschrift wird \tilde{x}_α gerade so gewählt, daß die Fläche unterhalb des Histogramms der gruppierten Daten links von \tilde{x}_α die Größe α hat. (Die Gesamtfläche unterhalb des Histogramms hat gerade die Größe Eins.)

Beispiel: Zu den Daten aus Tab. 6 wollen wir das obere Quartil, d. h. das 0,75-Quantil, bestimmen. Dieses liegt in der Klasse K_3 , deren untere Grenze bzw. Breite gerade

$$u_3 = 2 \quad \text{und} \quad b_3 = 1$$

sind, denn es ist

$$h_1 + h_2 \approx 0,644 < 0,75 \quad \text{und} \quad h_1 + h_2 + h_3 \approx 0,816 > 0,75.$$

Damit ergibt sich

$$\tilde{x}_\alpha = \tilde{x}_{0,75} = u_3 + \frac{\alpha - (h_1 + h_2)}{h_3} \cdot b_3 \approx 2 + \frac{0,75 - 0,644}{0,172} \cdot 1 \approx 2,616,$$

d. h. 75% der Kühlaggregate haben weniger als 2,616 Jahre gelebt.

4.3. Der Modalwert

Bei nominalskalierten Merkmalen wie Geschlecht oder Beruf kann man weder das arithmetische Mittel noch den Median als Lagemaß verwenden. Ein Lagemaß, das auch für diese Merkmale definiert ist, ist dagegen der **Modus (Modalwert, häufigster Wert)**. Er gibt die Ausprägung an, die die größte Häufigkeit in der Beobachtungsreihe besitzt.

Für den Modus x_{mod} gilt also: $h_n(x_{\text{mod}}) \geq h_n(a_j)$ für alle Merkmalsausprägungen a_1, \dots, a_k . Falls mehrere Ausprägungen diese Bedingung erfüllen, der Modus also nicht eindeutig ist, ist es nicht sinnvoll, ihn als Lagemaß zu verwenden.

Bei in Klassen eingeteilten Werten definiert man den Modus als Klassenmitte der am dichtesten besetzten Klasse, d. h. der Klasse K_i mit größter Histogrammhöhe r_i . Im Falle gleichbreiter Klassen bedeutet dies, daß der Modus die Klassenmitte der Klasse mit größter Klassenhäufigkeit h_i ist.

Beispiel: Im Beispiel mit den Kühlaggregaten besitzt, falls man Klassen der Breite Eins bildet, vgl. Tab. 6, die 1. Klasse K_1 die größte Klassenhäufigkeit. Daher ist $x_{\text{mod}} = 0,5$.

4.4. Das geometrische Mittel und das harmonische Mittel

Wenn die Merkmalsausprägungen relative Änderungen sind (z. B. Lohnerhöhung, Belegschaftsverminderung in %), ist es nicht sinnvoll, das arithmetische Mittel als Maß für einen Durchschnittswert zu verwenden, weil die Gesamtänderung nicht durch eine Summe, sondern durch ein Produkt beschrieben wird.

Bezeichnet x_1, \dots, x_n wiederum die Beobachtungsreihe und ist $x_i \geq 0$ für $i = 1, \dots, n$, so benutzt man als Lagemaß in solchen Fällen besser das **geometrische Mittel**

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}.$$

Beispiel: Vier Jahre lang wurden die Produktionssteigerungen eines Betriebes pro Jahr gemessen. Es ergaben sich die Werte aus **Tab. 10**.

Tab.10: Produktionssteigerungen eines Betriebes in vier Jahren

Jahr	1. Jahr	2. Jahr	3. Jahr	4. Jahr
Produktionssteigerung	2%	11%	4%	7%

Falls die Produktion im Jahr 0 (Anfangsproduktion) a beträgt, gilt für die nächsten Jahre:

- Produktion im 1. Jahr: $a \cdot 1,02$,
- Produktion im 2. Jahr: $a \cdot 1,02 \cdot 1,11$,
- Produktion im 3. Jahr: $a \cdot 1,02 \cdot 1,11 \cdot 1,04$,
- Produktion im 4. Jahr: $a \cdot 1,02 \cdot 1,11 \cdot 1,04 \cdot 1,07$.

Die Gesamtsteigerung der Produktion ist also durch das Produkt $1,02 \cdot 1,11 \cdot 1,04 \cdot 1,07$ bestimmt. Es ist deshalb nicht sinnvoll, die Summe der Einzeländerungen zu bilden und durch die Anzahl der Beobachtungen zu teilen, um die durchschnittliche Steigerung zu erhalten. Statt dessen bildet man das geometrische Mittel:

$$\bar{x}_g = \sqrt[4]{1,02 \cdot 1,11 \cdot 1,04 \cdot 1,07} = \sqrt[4]{1,26} = 1,06.$$

Die durchschnittliche Produktionssteigerung beträgt also ungefähr 6%. Dies bedeutet, daß, falls in jedem Jahr eine Produktionssteigerung von 6% stattgefunden hätte, im 4. Jahr diese konstante Steigerung zur selben Gesamtsteigerung wie wirklich vorhanden geführt hätte.

Das geometrische Mittel wird oft auch dort als Lagemaß verwandt, wo eine logarithmische Skala sinnvoll ist, denn wenn man den Logarithmus des geometrischen Mittels bildet, erhält man wieder ein arithmetisches Mittel:

$$\ln \bar{x}_g = \ln \sqrt[n]{x_1 \dots x_n} = \frac{1}{n} (\ln x_1 + \dots + \ln x_n).$$

Der Wert des geometrischen Mittels ist nie größer als der des entsprechenden arithmetischen Mittels. Vielmehr gilt

$$\bar{x}_g \leq \bar{x},$$

wobei die Gleichheit genau dann zutrifft, wenn

$$x_1 = x_2 = \dots = x_n$$

ist.

Gibt man sich Gewichte w_1, \dots, w_n mit $0 \leq w_i \leq 1$ und $\sum_{i=1}^n w_i = 1$ vor, so erhält man ein **gewogenes geometrisches Mittel**

$$\bar{x}_g^w = x_1^{w_1} \cdot x_2^{w_2} \cdot \dots \cdot x_n^{w_n}.$$

Mit $w_1 = \dots = w_n = 1/n$ ergibt sich wieder das gewöhnliche geometrische Mittel \bar{x}_g .

Hat man nur positive oder negative Beobachtungswerte x_1, \dots, x_n , so ist das **harmonische Mittel** \bar{x}_h gegeben durch

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}.$$

Genauso wie beim arithmetischen Mittel und beim geometrischen Mittel können natürlich auch beim harmonischen Mittel die Beobachtungen x_1, \dots, x_n mit Zahlen w_1, \dots, w_n , $0 \leq w_i \leq 1$ und $\sum_{i=1}^n w_i = 1$, gewichtet werden. Man erhält dann ein **gewogenes harmonisches Mittel**

$$\bar{x}_h^w = \frac{1}{\sum_{i=1}^n \frac{w_i}{x_i}}.$$

Für $w_1 = \dots = w_n = 1/n$ ergibt sich natürlich gerade wieder das gewöhnliche harmonische Mittel \bar{x}_h .

Beispiel: Hat man etwa Beziehungen $U = P \cdot M$ und $u_i = x_i \cdot m_i$ und ist $\sum u_i = U$, $\sum m_i = M$, so ergibt sich hieraus

$$P = \frac{U}{M} = \frac{\sum u_i}{\sum m_i} = \frac{\sum u_i}{\sum \frac{u_i}{x_i} \cdot x_i} = \frac{1}{\sum \frac{w_i}{x_i}} \quad \text{mit } w_i = u_i / \sum u_i.$$

P ergibt sich somit als mit w_i gewogenes harmonisches Mittel der x_i .

Obige Größen können z. B. sein:

U = Gesamtumsatz

M = Gesamtmenge

P = durchschnittlicher Preis pro Mengeneinheit

u_i = Einzelsatz des i-ten Gutes

x_i = Einzelpreis pro Mengeneinheit des i-ten Gutes

m_i = umgesetzte Menge des i-ten Gutes.

Der durchschnittliche Preis P ergibt sich als mit den Umsatzanteilen w_i gewogenes harmonisches Mittel der Einzelpreise x_i .

Analog berechnet man beispielsweise auch Durchschnittsgeschwindigkeiten aufgrund der Beziehung

Weg = Geschwindigkeit · Zeit

oder durchschnittliche Stückzahlen wegen der Beziehung

Gesamtstückzahl = (durchschnittliche Stückzahl/Zeiteinheit) · Zeit.

Hat man nur positive Beobachtungswerte x_1, \dots, x_n , so gilt stets die Beziehung

$$\bar{x}_h \leqq \bar{x}_g \leqq \bar{x};$$

bei identischen Beobachtungswerten $x_1 = \dots = x_n$ sind die Mittel gleich.

4.5. Einige Bemerkungen zu den Lagemaßen

Bei mehrgipfligen und insbesondere bei U-förmigen Häufigkeitsverteilungen sind – im Gegensatz zu eingipfligen Verteilungen –, vgl. Abb. 11, 12, 13, die Lagemaße oft nicht charakteristisch für die Häufigkeitsverteilung.

Beispiel: Die Lebensdauer der Menschen im Mittelalter hat eine U-förmige Häufigkeitsverteilung: Die Säuglingssterblichkeit war sehr hoch; hatte man jedoch die ersten Jahre überlebt, so war die Chance, ein recht hohes Alter zu erreichen, groß.

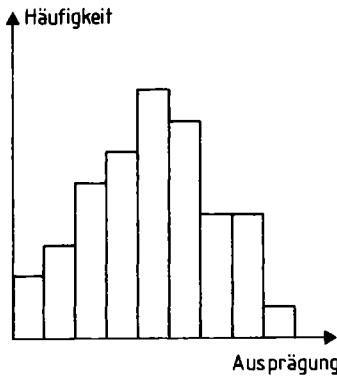


Abb. 11: Eingipflige Verteilung

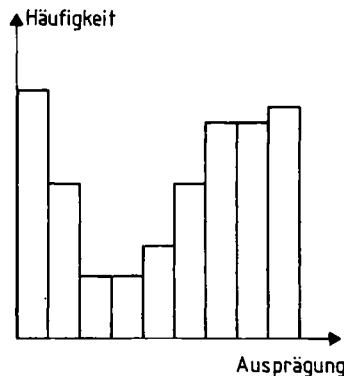


Abb. 12: U-förmige Verteilung

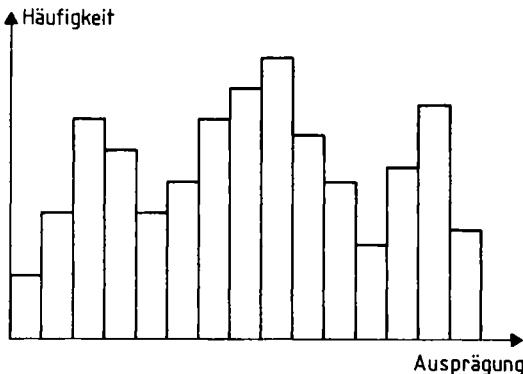


Abb. 13: Mehrgipflige Verteilung

Der Satz „Im Mittelalter wurden die Menschen durchschnittlich 25 Jahre alt“ gibt daher eine falsche Vorstellung von der damaligen Lebenserwartung.

Bevor man also zur Charakterisierung einer nicht eingipfligen Häufigkeitsverteilung ein Lagemaß bestimmt, muß man überprüfen, ob dieses sinnvoll zu interpretieren ist.

Abschließend wollen wir noch die eingipfligen Häufigkeitsverteilungen näher charakterisieren. Bezeichnet \bar{x} das arithmetische Mittel, $\tilde{x}_{0,5}$ den Median und x_{mod} den Modus einer eingipfligen Häufigkeitsverteilung, so heißt diese

rechtsschief oder **linkssteil**, falls $\bar{x} > \tilde{x}_{0,5} > x_{\text{mod}}$,

linksschief oder **rechtssteil**, falls $\bar{x} < \tilde{x}_{0,5} < x_{\text{mod}}$ und

symmetrisch, falls $\bar{x} = \tilde{x}_{0,5} = x_{\text{mod}}$,

vgl. auch Abb. 14, 15 und 16.

Beispiel: Die Verteilung der Anzahl kariöser Zähne bei 100 Schulkindern ist zwar mehrgipflig (Gipfel bei 1, 5 und 7 kariösen Zähnen), vgl. Abb. 3, aber nur der Gipfel

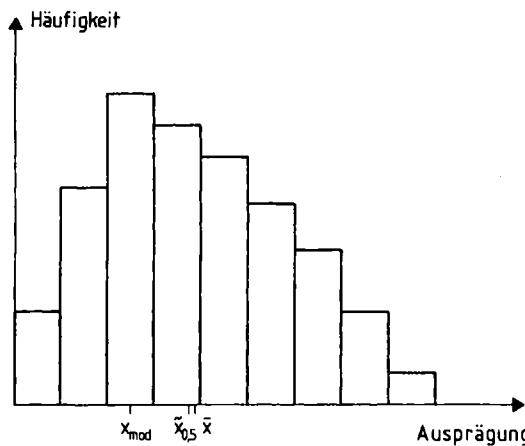


Abb. 14: Rechtsschiefe Verteilung

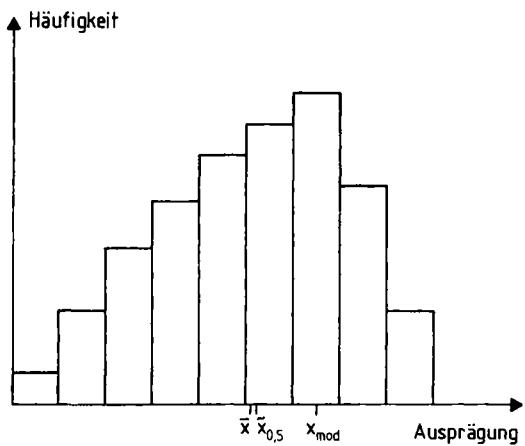


Abb. 15: Linksschiefe Verteilung

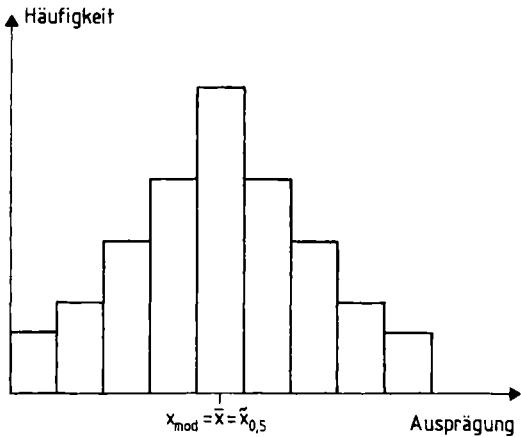


Abb. 16: Symmetrische Verteilung

bei Eins ist ausgeprägt, so daß man die Verteilung als annähernd eingipflig bezeichnen kann. Weiter ist

$$\bar{x} = 1,53 > \tilde{x}_{0,5} = 1 = x_{\text{mod}} = 1,$$

was auf eine rechtsschiefe Verteilung hindeutet. Bildet man zunächst Klassen wie in Abschnitt 3, Tab. 5 geschehen, so ist die Verteilung eingipflig, vgl. Abb. 6, und wegen

$$\bar{x} = 1,53 > \tilde{x}_{0,5} = 1,423 > x_{\text{mod}} = 1$$

rechtsschief.

5. Streuungsmaße von Häufigkeitsverteilungen

Lagemaße von Häufigkeitsverteilungen, wie wir sie in Abschnitt 4 kennengelernt haben, geben allein wenig Auskunft über eine Häufigkeitsverteilung. Sie beschreiben zwar ein Zentrum dieser Verteilung, liefern aber keinen Anhaltspunkt dafür, wie weit ein konkreter Merkmalswert von einem solchen Zentrum abweichen kann.

Beispiel: (a) Die Aussage, „Ein Beschäftigter der Firma Schulze verdient durchschnittlich 1347,- DM netto“, gibt keine Informationen darüber, wieviel der einzelne Beschäftigte wirklich verdient. Es könnte sein, daß alle etwa diesen durchschnittlichen Verdienst haben oder daß einige sehr viel weniger, andere dafür aber sehr viel mehr verdienen.

(b) Beträgt der durchschnittliche Preis für ein Kilogramm Schweinefleisch in einer bestimmten Gegend 17,- DM, so heißt dies nicht unbedingt, daß der Preis zwischen 16,- und 18,- DM liegt. Es könnte auch sein, daß man in einigen Geschäften nur 12,- DM, in anderen dagegen etwa 20,- DM bezahlt.

Maße, die die Abweichung von einem Zentrum einer Häufigkeitsverteilung beschreiben, sind nun die sogenannten **Streuungsmaße** oder **Dispersionsmaße**. Kennt man ein Lagemaß und ein Streuungsmaß einer Häufigkeitsverteilung, so hat man schon eine recht präzise Vorstellung von derselben. Einige solche Streuungsmaße für metrisch skalierte Merkmale wollen wir im folgenden vorstellen.

5.1. Die Spannweite

Der **Streubereich** einer Häufigkeitsverteilung ist derjenige Wertebereich, in dem alle Merkmalswerte einer Beobachtungsreihe liegen. Er ist somit bestimmt durch die Angabe des kleinsten und des größten beobachteten Merkmalswertes in einer Urliste. Ist also $x_{(1)} \leq \dots \leq x_{(n)}$ die geordnete Beobachtungsreihe, so ist das Intervall $[x_{(1)}; x_{(n)}]$ gerade der Streubereich der zugehörigen Häufigkeitsverteilung.

Beispiel: In $n = 40$ Geschäften wird der Preis für ein bestimmtes Farbfernsehgerät ermittelt. Das billigste Angebot lag bei $x_{(1)} = 1374,-$ DM und das teuerste bei $x_{(n)} = 1829,-$ DM. Diese beiden Zahlen geben den Streubereich der Häufigkeitsverteilung der Preise für das Gerät an.

Die Breite eines Streubereichs nennt man auch die **Spannweite (Range)** R einer Häufigkeitsverteilung:

$$R = x_{(n)} - x_{(1)}.$$

Beispiel: Die Spannweite für den Preis eines Fernsehgerätes war gerade $R = x_{(n)} - x_{(1)} = 1829 - 1374 = 455$.

Die Spannweiten verschiedener Beobachtungsreihen kann man nur dann vergleichen, wenn die Beobachtungsreihen die gleiche Länge n haben, denn natürlich kann die Verlängerung einer Reihe nur zur Folge haben, daß sich die Spannweite der Häufigkeitsverteilung entweder nicht verändert oder daß sie sich vergrößert. Eine Verkleinerung der Spannweite und damit des Streubereichs ist nicht möglich.

5.2. Der Quartilsabstand

Betrachten wir die Spannweite als Streuungsmaß einmal genauer, so fällt auf, daß sie lediglich vom kleinsten und vom größten Merkmalswert einer Beobachtungsreihe abhängt. Es kann nun vorkommen, daß z. B. der kleinste Wert sehr stark von allen anderen Werten abweicht, also ein sogenannter Ausreißer ist. Dann beeinflußt er den Wert R der Spannweite extrem stark.

Beispiel: Wäre im Beispiel aus Abschnitt 5.1 das zweitbilligste Angebot für den Farbfernseher etwa $x_{(2)} = 1509,-$ DM, so würde die Spannweite $R = 455$ eine schlechte Vorstellung von der Preisschwankung geben. Denn hätte man das Billigangebot $x_{(1)}$ nicht beobachtet, so wäre die Spannweite nur noch

$$1829 - 1509 = 320,$$

mithin sehr viel geringer gewesen.

Ein Streuungsmaß, das nicht so abhängig von Extremwerten und somit i. a. günstiger ist, ist der **Quartilsabstand**. Er gibt gerade die Größe des Bereichs zwischen oberem und unterem Quartil einer Beobachtungsreihe an; d. h. der Quartilsabstand ist

$$\tilde{x}_{0,75} - \tilde{x}_{0,25},$$

vgl. auch Abschnitt 4.2.2.

Zwischen unterem und oberem Quartil liegen mindestens 50% aller Merkmalswerte in einer Beobachtungsreihe, so daß der Quartilsabstand die Größe des Bereichs liefert, in dem etwa die Hälfte aller beobachteten Ausprägungen liegen.

Beispiel: $n = 30$ Glühbirnen werden einer Lebensdauerprüfung unterzogen; dabei stellt man die Lebensdauern x_i für $i = 1, \dots, 30$ aus **Tab. 11** fest.

Tab. 11: Lebensdauer x_i von 30 Glühbirnen in Stunden

i	x_i	i	x_i	i	x_i
1	375,3	11	772,5	21	918,3
2	392,5	12	799,4	22	935,4
3	467,9	13	799,6	23	951,1
4	503,1	14	803,9	24	964,9
5	591,2	15	808,7	25	968,8
6	657,8	16	810,5	26	1006,5
7	738,5	17	812,1	27	1014,7
8	749,6	18	848,6	28	1189,0
9	752,0	19	867,0	29	1215,6
10	765,8	20	904,3	30	1407,2

Um den Quartilsabstand dieser Häufigkeitsverteilung bestimmen zu können, müssen zunächst die Quartile $\tilde{x}_{0,25}$ und $\tilde{x}_{0,75}$ berechnet werden, vgl. Abschnitt 4.2.2. Es ist

$$n \cdot 0,25 = 30 \cdot 0,25 = 7,5$$

und somit

$$\tilde{x}_{0,25} = x_{(8)} = x_8 = 749,6.$$

Weiterhin ergibt sich mit

$$n \cdot 0,75 = 30 \cdot 0,75 = 22,5$$

nun

$$\tilde{x}_{0,75} = x_{(23)} = x_{23} = 951,1.$$

Damit ist der Quartilsabstand der Häufigkeitsverteilung gerade

$$\tilde{x}_{0,75} - \tilde{x}_{0,25} = 951,1 - 749,6 = 201,5.$$

5.3. Die mittlere absolute Abweichung vom Median

Als Streuungsmaße lassen sich auch solche Größen verwenden, die eine durchschnittliche Abweichung der Merkmalswerte in einer Beobachtungsreihe von einem festen Wert messen. Eine Möglichkeit besteht etwa darin, die mittlere betragliche Abweichung von einem festen Wert c zu betrachten:

$$\frac{1}{n} \sum_{i=1}^n |x_i - c| = \frac{1}{n} \sum_{j=1}^k |a_j - c| \cdot H_n(a_j);$$

hier bezeichnet x_i den Merkmalswert, der an der i -ten Beobachtungseinheit gemessen wurde und a_j die j -te Ausprägung des beobachteten Merkmals, die mit absoluter Häufigkeit $H_n(a_j)$ aufgetreten ist.

Es stellt sich nun die Frage, wie man den Wert c am besten wählt. Eine ausgezeichnete Stellung hat hier der Median $c = \tilde{x}_{0,5}$, denn es gilt die **Minimumseigenschaft des Medians**:

$$\sum_{i=1}^n |x_i - \tilde{x}_{0,5}| \leq \sum_{i=1}^n |x_i - c| \quad \text{für alle Zahlen } c.$$

Die Größe

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}_{0,5}|$$

nennt man auch die **mittlere absolute Abweichung vom Median**.

Beispiel: Für das Beispiel der Glühbirnen aus Abschnitt 5.2, vgl. auch Tab.11, wollen wir die mittlere absolute Abweichung vom Median berechnen. Der Median der Beobachtungsreihe selbst ist gerade das arithmetische Mittel der Werte

$$x_{(15)} = x_{15} = 808,7 \quad \text{und} \quad x_{(16)} = x_{16} = 810,5,$$

d.h. es ist

$$\tilde{x}_{0,5} = \frac{1}{2} \cdot (808,7 + 810,5) = \frac{1619,2}{2} = 809,6,$$

so daß sich die mittlere absolute Abweichung vom Median zu

$$\begin{aligned}
 d &= \frac{1}{30} \sum_{i=1}^{30} |x_i - 809,6| = \\
 &= \frac{1}{30} \cdot (|375,3 - 809,6| + |392,5 - 809,6| + \dots + |1407,2 - 809,6|) = \\
 &= \frac{1}{30} (434,3 + 417,1 + \dots + 597,6) = 161,207
 \end{aligned}$$

berechnet.

Auch wenn nur gruppierte Daten vorliegen, verwendet man die mittlere absolute Abweichung vom Median als Maß der Streuung einer Beobachtungsreihe.

Hat man p Klassen K_1, \dots, K_p gebildet, die mit absoluten Häufigkeiten H_1, \dots, H_p auftreten, so ist die mittlere absolute Abweichung vom Median gerade

$$\frac{1}{n} \sum_{j=1}^p |\bar{x}_j - \tilde{x}_{0,5}| \cdot H_j,$$

falls die Klassenmittel

$$\bar{x}_j = \frac{1}{H_j} \sum_{x_i \in K_j} x_i \quad \text{für } j = 1, \dots, p$$

bekannt sind, und

$$\frac{1}{n} \sum_{j=1}^p |m_j - \tilde{x}_{0,5}| \cdot H_j$$

falls nur die Klassenmitten m_1, \dots, m_p bekannt sind; anstelle der Klassenmittel kann man auch die Klassenmediane verwenden.

Beispiel: Im Beispiel aus Abschnitt 3 „Lebensdauern von 87 Kühlaggregaten“, vgl. auch Tab. 6, haben wir zehn Klassen der Breite 1 gebildet. Im Abschnitt 4.2.1 ist bereits der Median der so gruppierten Daten berechnet worden:

$$\tilde{x}_{0,5} = 1,344.$$

Nehmen wir einmal an, daß wir nur noch die Klassenmitten $m_1 = 0,5, m_2 = 1,5, \dots, m_{10} = 9,5$ kennen, so ergibt sich als mittlere absolute Abweichung vom Median der Wert

$$\begin{aligned}
 \frac{1}{n} \sum_{j=1}^p |m_j - \tilde{x}_{0,5}| \cdot H_j &= \frac{1}{87} \sum_{j=1}^{10} |m_j - 1,344| \cdot H_j \\
 &= \frac{1}{87} (|0,5 - 1,344| \cdot 37 + |1,5 - 1,344| \cdot 19 + \dots + |9,5 - 1,344| \cdot 1) \\
 &= 112,028/87 = 1,288.
 \end{aligned}$$

5.4. Varianz, Standardabweichung und Variationskoeffizient

Haben wir im vorigen Abschnitt als Bezugsgröße den Median einer Häufigkeitsverteilung gewählt, so beziehen sich die hier vorgestellten Maßzahlen auf das arithmetische Mittel einer Beobachtungsreihe.

Ein weiterer grundlegender Unterschied zwischen den hier betrachteten Maßen für die Streuung und der mittleren absoluten Abweichung vom Median besteht darin, daß nun als Abstandsmaß keine beträchtliche Differenz, sondern vielmehr ein quadratischer Abstand gewählt wird.

5.4.1. Die Varianz

Bezeichnen wiederum x_1, \dots, x_n die Merkmalswerte in der Urliste und a_1, \dots, a_k die Ausprägungen des betrachteten Merkmals, so ist die **Varianz** der Beobachtungsreihe, die wohl das am häufigsten verwandte Streuungsmaß ist, gerade

$$\begin{aligned}s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) = \\ &= \frac{1}{n-1} \sum_{j=1}^k (a_j - \bar{x})^2 H_n(a_j) = \frac{1}{n-1} \left(\sum_{j=1}^k a_j^2 H_n(a_j) - n\bar{x}^2 \right).\end{aligned}$$

Bei Verwendung dieses quadratischen Abstandsmaßes wählt man speziell das arithmetische Mittel \bar{x} als Bezugsgröße, da es – ähnlich wie der Median bei der absoluten Differenz – die **Minimumseigenschaft des arithmetischen Mittels** besitzt:

$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq \sum_{i=1}^n (x_i - c)^2 \quad \text{für jede beliebige Zahl } c.$$

Beispiel: Wir wollen die Varianz s^2 der Lebensdauer von $n = 30$ Glühbirnen, vgl. Tab. 11, berechnen. Mit

$$\bar{x} = \frac{1}{30} (375,3 + \dots + 1407,2) = 24791,8/30 = 826,39$$

ergibt sich

$$s^2 = \frac{1}{29} \sum_{i=1}^{30} (x_i - 826,39)^2 = 52025,361$$

als Varianz der Beobachtungsreihe.

Liegen die Beobachtungen nur in grupplerter Form vor, so lässt sich die Varianz mit Hilfe der Klassenmittel oder der Klassenmitten berechnen.

Ist p die Anzahl der Klassen, die mit absoluten Häufigkeiten H_1, \dots, H_p auftreten, und ist

$$\bar{x}_j = \frac{1}{H_j} \sum_{x_i \in K_j} x_i \quad \text{für } j = 1, \dots, p$$

das Klassenmittel, so berechnet sich die Varianz der gruppierten Daten zu

$$s_0^2 = \frac{1}{n-1} \sum_{j=1}^p (\bar{x}_j - \bar{x})^2 H_j.$$

Tab.11a: Arbeitstabelle zur Berechnung der Varianz der gruppierten Lebensdauern von 30 Glühbirnen, vgl. auch Tab. 11

Klasse K_j	H_j	\bar{x}_j	m_j
$K_1 = [300; 700)$	6	497,967	500
$K_2 = [700; 800)$	7	768,200	750
$K_3 = [800; 900)$	6	825,133	850
$K_4 = [900; 1000)$	6	940,467	950
$K_5 = [1000; 1500)$	5	1166,600	1250

Beispiel: Im Beispiel der Glühbirnen bilden wir 5 Klassen wie in Tab. 11a, wo auch die absoluten Häufigkeiten H_j und die Klassenmittel (sowie die Klassenmittnen) angegeben sind, und berechnen die Varianz s_0^2 der gruppierten Daten.

Es ergibt sich hier

$$\begin{aligned}s_0^2 &= \frac{1}{29} \sum_{j=1}^5 (\bar{x}_j - 826,39)^2 \cdot H_j = \\&= \frac{1}{29} ((497,967 - 826,39)^2 \cdot 6 + \dots + (1166,600 - 826,39)^2 \cdot 5) = \\&= 1327677,6/29 = 45781,986\end{aligned}$$

als Varianz der gruppierten Daten.

In diesem Beispiel sehen wir, daß die Varianz der Beobachtungsreihe durch die Varianz der gruppierten Daten unterschätzt wird. Das gilt auch allgemein, denn bei der Berechnung aus gruppierten Daten wird die Varianz innerhalb der Klassen vernachlässigt: Es ist

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

und weiter ist

$$\begin{aligned}\sum_{i=1}^n x_i^2 &= \sum_{j=1}^p \sum_{x_i \in K_j} x_i^2 \quad \text{sowie} \\- n\bar{x}^2 &= n\bar{x}^2 - 2n\bar{x}^2 = \bar{x}^2 \sum_{j=1}^p H_j - 2\bar{x} \sum_{j=1}^p \bar{x}_j H_j,\end{aligned}$$

so daß gilt

$$\begin{aligned}s^2 &= \frac{1}{n-1} \sum_{j=1}^p \left[\sum_{x_i \in K_j} x_i^2 + \bar{x}^2 H_j - 2\bar{x}\bar{x}_j H_j \right] \\&= \frac{1}{n-1} \sum_{j=1}^p \left[\sum_{x_i \in K_j} x_i^2 - \bar{x}_j^2 H_j + \bar{x}_j^2 H_j + \bar{x}^2 H_j - 2\bar{x}\bar{x}_j H_j \right] \\&= \frac{1}{n-1} \sum_{j=1}^p \left[\sum_{x_i \in K_j} (x_i - \bar{x}_j)^2 + (\bar{x}_j - \bar{x})^2 H_j \right] \\&= \frac{1}{n-1} \sum_{j=1}^p \sum_{x_i \in K_j} (x_i - \bar{x}_j)^2 + s_0^2 \\&\geq s_0^2\end{aligned}$$

Verwendet man nun zur Berechnung der Varianz der gruppierten Daten die Klassenmittnen, also

$$s_*^2 = \frac{1}{n-1} \sum_{j=1}^p (m_j - \bar{x})^2 H_j,$$

so wird man die Varianz der Beobachtungsreihe i. a. überschätzen, falls die Zahl der Klassen klein im Verhältnis zur Zahl der Beobachtungswerte ist. Das liegt daran, daß dann die Verteilung in den Klassen meist unsymmetrisch ist. vgl. Abb. 17. Je symmetrischer die Verteilung in den Klassen ist, umso kleiner ist auch s_*^2 im Verhältnis zu s_0^2 . Liegt völlige Symmetrie vor, so ist $s_*^2 = s_0^2 \leq s^2$.

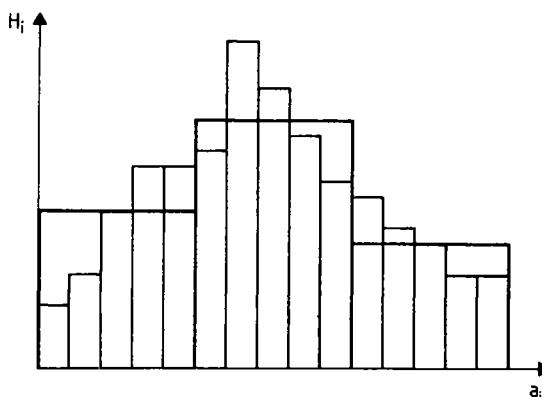


Abb. 17: Eine in den Klassen unsymmetrische Häufigkeitsverteilung

Beispiel: Bei Verwendung der Klassenmitten ergibt sich als Varianz der gruppierten Lebensdauern der Glühbirnen, vgl. Tab. 11a, der Wert

$$\begin{aligned}s_*^2 &= \frac{1}{29} \cdot \sum_{j=1}^5 (m_j - 826,39)^2 \cdot H_j \\&= \frac{1}{29} \cdot ((500 - 826,39)^2 \cdot 6 + \dots + (1250 - 826,39)^2 \cdot 5) \\&= 1672279/29 = 57664,792.\end{aligned}$$

Wir sehen also, daß die Varianz s^2 hier tatsächlich überschätzt wird, da die Verteilungen der Beobachtungswerte um die Klassenmitten unsymmetrisch sind.

Haben alle Klassen die gleiche Breite $b = b_1 = \dots = b_p$, so läßt sich der Wert s_*^2 korrigieren. Hierzu verwendet man die **Sheppardsche Korrektur** $b^2/12$:

$$s_{**}^2 = s_*^2 - \frac{b^2}{12}$$

und hofft, daß man so eine gute Näherung für die Varianz s^2 der ungruppierten Beobachtungswerte gefunden hat.

5.4.2. Die Standardabweichung

Die **Standardabweichung**

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

ist gerade die Wurzel aus der Varianz s^2 einer Beobachtungsreihe. Sie hat gegenüber der Varianz den Vorteil, daß sie die gleiche Dimension wie die Beobachtungswerte und damit wie \bar{x} besitzt.

Wird etwa \bar{x} in Stunden gemessen, so hat auch s die Dimension „Stunden“. Dagegen ist die Maßeinheit der Varianz (Stunde)².

Beispiel: Die Standardabweichung in der Beobachtungsreihe der Lebensdauern von 30 Glühbirnen, vgl. Tab. 11, ergibt sich wegen

$$s^2 = 52025,361$$

gerade zu

$$s = \sqrt{s^2} = \sqrt{52025,361} = 228,091.$$

5.4.3. Der Variationskoeffizient

Varianz und Standardabweichung benutzen zwar als Bezugspunkt das arithmetische Mittel \bar{x} , d. h. sie messen einen mittleren Abstand von \bar{x} , werden jedoch nicht ins Verhältnis zu \bar{x} gesetzt. Ist aber der Mittelwert einer Beobachtungsreihe z. B. 10000, so wird man sagen, daß eine Varianz von 10 recht klein ist, dagegen ist diese Varianz ziemlich groß zu nennen, wenn $\bar{x} = 1$ ist. Ein von \bar{x} bereinigtes Streuungsmaß, das also ohne direkte Nennung von \bar{x} interpretiert werden kann, ist nun der **Variationskoeffizient**

$$v = \frac{s}{\bar{x}}.$$

Der Variationskoeffizient mißt also das Verhältnis von Standardabweichung und arithmetischem Mittel. Er ist nur für solche Beobachtungsreihen ein sinnvolles Maß, die ausschließlich positive Merkmalswerte beinhalten und eignet sich natürlich insbesondere zum Vergleich der Streuungen verschiedener Meßreihen.

Beispiel: Der Variationskoeffizient der Lebensdauer von 30 Glühbirnen, vgl. Tab. 11, ergibt sich zu

$$v = \frac{s}{\bar{x}} = \frac{228,091}{826,39} = 0,276.$$

5.5. Die Schiefe und der Exzeß

Obwohl Schiefe und Exzeß keine direkten Streuungsparameter von Häufigkeitsverteilungen sind, wollen wir sie in diesem Zusammenhang erläutern. Beide Maßzahlen können nur dort sinnvoll verwandt werden, wo eingipflige Häufigkeitsverteilungen vorliegen.

5.5.1. Die Schiefe einer Häufigkeitsverteilung

Im Abschnitt 4.5 haben wir schon die Begriffe rechtsschiefe, linksschiefe und symmetrische eingipflige Häufigkeitsverteilung kennengelernt. Die Schiefe ist nun eine Maßzahl, die uns nicht nur sagt, in welcher Richtung eine Häufigkeitsverteilung schief ist, sondern vielmehr auch die Größenordnung dieser Schiefe angibt.

Sind x_1, \dots, x_n die Merkmalswerte in einer Beobachtungsreihe mit Mittelwert \bar{x} , so ist die Schiefe der zugehörigen Häufigkeitsverteilung gerade

$$g_1 = \frac{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right)}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^3}}.$$

Ist $g_1 = 0$, so heißt dies gerade, daß die Häufigkeitsverteilung symmetrisch ist. Je stärker negativ g_1 ist, desto linksschiefer ist die Verteilung, und sie ist umso rechtschiefer, je stärker positiv der Wert g_1 ist.

Natürlich läßt sich die Schiefe einer Häufigkeitsverteilung auch aus gruppiertem Datenmaterial berechnen:

Bei Verwendung der Klassenmittel $\bar{x}_1, \dots, \bar{x}_p$ ist

$$g_1 = \frac{\frac{1}{n} \sum_{i=1}^p (\bar{x}_i - \bar{x})^3 \cdot H_i}{\sqrt{\left[\frac{1}{n} \sum_{i=1}^p (\bar{x}_i - \bar{x})^2 \cdot H_i \right]^3}}$$

und bei Verwendung der Klassenmitten m_1, \dots, m_p ergibt sich

$$g_1 = \frac{\frac{1}{n} \sum_{i=1}^p (m_i - \bar{x})^3 \cdot H_i}{\sqrt{\left[\frac{1}{n} \sum_{i=1}^p (m_i - \bar{x})^2 \cdot H_i \right]^3}}.$$

Beispiel: Die Schiefe der Häufigkeitsverteilung der Lebensdauer von 30 Glühbirnen ergibt sich aus gruppierten Daten, vgl. Tab. 11a, zu

$$\begin{aligned} g_1 &= \left(\frac{1}{30} \sum_{i=1}^5 (\bar{x}_i - 826,39)^3 H_i \right) \Big/ \sqrt{\left(\frac{1}{30} \sum_{i=1}^5 (\bar{x}_i - 826,39)^3 H_i \right)^3} \\ &= -\frac{8133122,1}{30} \Big/ \sqrt{\left(\frac{1327677,6}{30} \right)^3} = -\frac{271104,07}{9310158,2} = -0,029, \end{aligned}$$

wenn man die Klassenmittel, und zu

$$\begin{aligned} g_1 &= \left(\frac{1}{30} \sum_{i=1}^5 (m_i - 826,39)^3 H_i \right) \Big/ \sqrt{\left(\frac{1}{30} \sum_{i=1}^5 (m_i - 826,39)^2 H_i \right)^3} \\ &= \frac{179742320}{30} \Big/ \sqrt{\left(\frac{1672279}{30} \right)^3} = \frac{5991410,7}{13160768} = 0,455, \end{aligned}$$

wenn man die Klassenmitten verwendet.

Bei Verwendung der Klassenmittel ergibt sich eine schwach linksschiefe und bei Verwendung der Klassenmitten eine rechtsschiefe Häufigkeitsverteilung.

Hier müßte man also wirklich von den ungruppierten Daten ausgehen; für diese ergibt sich

$$g_1 = \frac{52076124}{30} \Big/ \sqrt{\left(\frac{1508735,5}{30} \right)^3} = \frac{1735870,8}{11278147} = 0,154,$$

also eine leichte Rechtsschiefe.

Wir sehen, daß in diesem Beispiel der Wert g_1 der ungruppierten Daten durch Klassenbildung einmal unter- und einmal überschätzt wurde.

5.5.2. Der Exzeß einer Häufigkeitsverteilung

Eine andere traditionelle Maßzahl für eingipflige Häufigkeitsverteilungen ist der **Exzeß (Kurtosis, Wölbung)**

$$g_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} - 3,$$

der angibt, ob (bei gleicher Varianz) das absolute Maximum der Häufigkeitsverteilung größer als bei der Dichte der Normalverteilung, vgl. Abschnitt 1 in Kapitel IV, ist. Der theoretische Exzeß einer Normalverteilung ist gerade 0. Ist $g_2 > 0$, so ist das absolute Maximum der Häufigkeitsverteilung größer als das der zugehörigen Normalverteilung; ist $g_2 < 0$, so ist es kleiner.

Bei gruppierten Daten ergibt sich

$$g_2 = \frac{\frac{1}{n} \sum_{i=1}^p (\bar{x}_i - \bar{x})^4 \cdot H_i}{\left[\frac{1}{n} \sum_{i=1}^p (\bar{x}_i - \bar{x})^2 \cdot H_i \right]^2} - 3$$

bei Verwendung der Klassenmittel und

$$g_2 = \frac{\frac{1}{n} \sum_{i=1}^p (m_i - \bar{x})^4 \cdot H_i}{\left[\frac{1}{n} \sum_{i=1}^p (m_i - \bar{x})^2 \cdot H_i \right]^2} - 3$$

bei Verwendung der Klassenmittten.

Beispiel: Für die Häufigkeitsverteilung der ungruppierten Lebensdauern von $n = 30$ Glühbirnen, vgl. Tab. 11, ergibt sich ein Exzeß von

$$\begin{aligned} g_2 &= \frac{\frac{1}{30} \sum_{i=1}^{30} (x_i - 826,39)^4}{\left(\frac{1}{30} \sum_{i=1}^{30} (x_i - 826,39)^2 \right)^2} - 3 = \frac{2,6592080 \cdot 10^{11}/30}{(1508735,5/30)^2} - 3 \\ &= 3,504 - 3 = 0,504. \end{aligned}$$

Das absolute Maximum dieser Häufigkeitsverteilung ist also größer als das der zugehörigen Normalverteilung.

Sind Schiefe und Exzeß einer Häufigkeitsverteilung wesentlich von 0 verschieden, so ist das ein Hinweis darauf, daß die zugrundeliegende theoretische Verteilung der Gesamtheit, aus der die Untersuchungseinheiten stammen, vgl. Kap. II, wesentlich von einer Normalverteilung abweicht.

6. Konzentrationsmaße für Häufigkeitsverteilungen

Beobachtet man an n Untersuchungseinheiten ein metrisch skaliertes Merkmal X mit nur positiven Ausprägungen, so ist es oftmals von Interesse zu untersuchen, wie sich die Summe der Merkmalswerte x_1, \dots, x_n in der Beobachtungsreihe auf die Untersuchungseinheiten aufteilt. Ist diese Summe vielleicht auf wenige Einheiten konzentriert? Mit solchen und ähnlichen Fragen wollen wir uns in diesem Abschnitt beschäftigen.

Beispiel: In einer Region hat man an n Untersuchungseinheiten, nämlich den landwirtschaftlichen Betrieben, die Größe der landwirtschaftlichen Nutzflächen x_1, \dots, x_n beobachtet. Wie teilt sich die gesamte Nutzfläche der Region auf die einzelnen Betriebe auf?

Ist sie auf einige Betriebe konzentriert, d.h. besitzen wenige Betriebe einen Großteil der Nutzfläche?

Wir wollen hier davon ausgehen, daß die n beobachteten Merkmalswerte durch q verschiedene Merkmalsausprägungen $a_1 < \dots < a_q$, die mit absoluten Häufigkeiten H_1, \dots, H_q und relativen Häufigkeiten h_1, \dots, h_q auftreten, bestimmt werden; a_1, \dots, a_q können hier auch die Klassenmitten oder Klassenmittel von gruppierten Merkmalswerten sein. Dann ist natürlich der Gesamtbetrag der Merkmalswerte in der Beobachtungsreihe gerade

$$\sum_{j=1}^n x_j = \sum_{i=1}^q a_i H_i.$$

6.1. Die Lorenzkurve

Eine graphische Darstellung der Konzentration der Merkmalswerte liefert die sogenannte **Lorenzkurve**, die man, wie im folgenden beschrieben, erhält. Auf der Abszisse eines Koordinatenkreuzes werden die Größen

$$k_i = \sum_{j=1}^i \frac{H_j}{n} = \sum_{j=1}^i h_j \quad \text{für } i = 1, \dots, q$$

abgetragen; es werden also die nach der Größe der Merkmalsausprägungen bzw. Gruppen geordneten relativen Häufigkeiten aufsummiert (im Beispiel: Es wird erst der Prozentsatz der Betriebe mit der kleinsten Fläche bestimmt, dann wird der Prozentsatz der Betriebe mit zweitkleinster Fläche bestimmt und zum Prozentsatz der Betriebe mit der kleinsten Fläche addiert, usw.). Auf der Ordinate wird

$$l_i = \frac{\sum_{j=1}^i a_j H_j}{\sum_{j=1}^q a_j H_j} \quad \text{für } i = 1, \dots, q$$

abgetragen; die Ausprägungen werden also der Größe nach aufsummiert und auf die Summe aller Ausprägungen bezogen (im Beispiel: Die Flächenanteile der Betriebe bzgl. der Gesamtfläche werden der Flächengröße nach aufsummiert).

Ein Punkt $(k_i; l_i)$ der Lorenzkurve gibt dann an, daß auf $k_i \cdot 100\%$ der Untersuchungseinheiten $l_i \cdot 100\%$ des Gesamtbetrages aller Merkmalsausprägungen entfall-

len (im Beispiel: Auf $k_i \cdot 100\%$ der Betriebe entfallen $l_i \cdot 100\%$ der gesamten Nutzfläche).

Durch die Verbindung der Punkte (k_i, l_i) mit (k_{i+1}, l_{i+1}) für $i = 0, \dots, k-1$ entsteht die Lorenzkurve. Dabei wird $k_0 = l_0 = 0$ gesetzt, und es gilt stets $k_q = l_q = 1$.

Beispiel: In Tab. 12 ist die Aufteilung der landwirtschaftlich genutzten Fläche einer Region in Hektar (ha) abzulesen.

Tab. 12: Landwirtschaftlich genutzte Fläche einer Region

Flächengr. von ... bis ... ha	Anzahl der landw. Be- triebe	Flächengr. pro Gruppe in ha	Prozent- satz der landw. Betriebe h_i	Flächen- anteile $\frac{a_i \cdot H_i}{\sum_i a_i \cdot H_i}$	aufsumm. Betriebs- anteile k_i	aufsumm. Flächen- anteile l_i
bis 5 ha	21	63	0,42	0,063	0,42	0,063
5 bis 10 ha	9	72	0,18	0,072	0,60	0,135
10 bis 20 ha	9	135	0,18	0,135	0,78	0,270
20 bis 50 ha	8	280	0,16	0,280	0,94	0,550
> 50 ha	3	450	0,06	0,450	1,00	1,000
Gesamt	$n = 50$	1000	1,00	1,000		

Die zugehörige Lorenzkurve ist in Abb. 18 zu sehen.

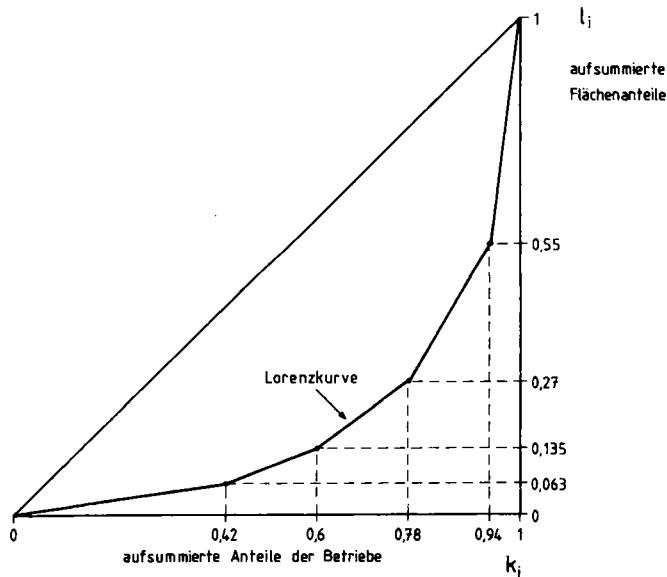


Abb. 18: Lorenzkurve im Beispiel

Der Punkt $(k_1; l_1) = (0,42; 0,063)$ auf der Lorenzkurve besagt z. B., daß auf 42% der landwirtschaftlichen Betriebe nur 6,3% der Fläche entfällt.

Es können zwei Extremfälle der Lorenzkurve auftreten:

- Es liegt keine Konzentration vor, d.h. alle Untersuchungseinheiten haben den gleichen Anteil am Gesamtbetrag (im Beispiel: Alle Betriebe besitzen gleich viel Fläche). In diesem Fall entfallen auf $z \cdot 100\%$ der Untersuchungseinheiten immer $z \cdot 100\%$ des Gesamtbetrages ($0 \leq z \leq 1$). Die Lorenzkurve ist also die Diagonale.
- Der Gesamtbetrag konzentriert sich fast vollständig auf eine Untersuchungseinheit, während der Anteil der anderen Untersuchungseinheiten verschwindend gering ist (im Beispiel: Ein landwirtschaftlicher Betrieb besitzt fast die ganze Fläche). Die Lorenzkurve nähert sich dann bei immer weiter fortschreitender Konzentration auf eine Untersuchungseinheit immer mehr der Abzisse und steigt immer näher am rechten Eckpunkt $(1; 0)$ fast senkrecht zum Eckpunkt $(1; 1)$ auf.

6.2. Das Lorenzsche Konzentrationsmaß; der Gini-Koeffizient

Mit Hilfe der Lorenzkurve möchte man eine Maßzahl für die Konzentration finden, die bei Nichtkonzentration den Wert 0 und bei vollständiger Konzentration den Wert 1 annimmt. Die Fläche zwischen Diagonale und Lorenzkurve wird mit zunehmender Konzentration immer größer.

Das **Lorenzsche Konzentrationsmaß (LKM)**, das man auch **Gini-Koeffizient** nennt, wird nun als das 2-fache der Fläche F zwischen Diagonale und Lorenzkurve definiert:

$$\text{LKM} = 2F$$

Die Fläche F kann durch die Summe der Trapezflächen F_i abzüglich der Fläche des oberen Dreiecks berechnet werden, vgl. auch Abb. 19:

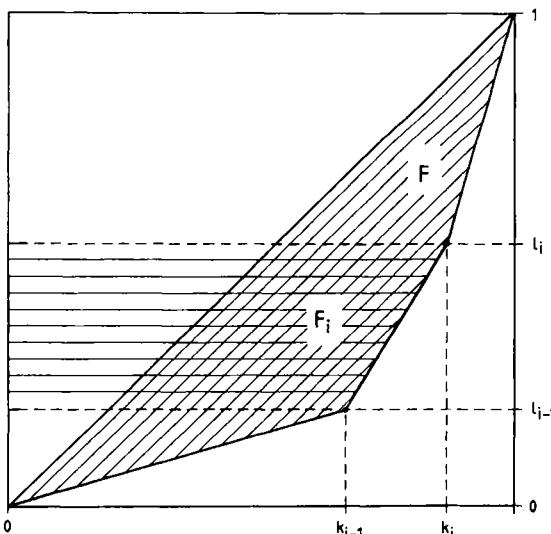


Abb. 19: Berechnung des Lorenzschen Konzentrationsmaßes

$$F = \sum_{i=1}^q F_i - 0,5 \quad \text{mit}$$

$$F_i = \frac{k_{i-1} + k_i}{2} (l_i - l_{i-1}) = \frac{k_{i-1} + k_i}{2} \frac{1}{\sum_{j=1}^q a_j \cdot H_j} \left(\sum_{j=1}^i a_j \cdot H_j - \sum_{j=1}^{i-1} a_j \cdot H_j \right)$$

$$= \frac{k_{i-1} + k_i}{2} \frac{a_i \cdot H_i}{\sum_{j=1}^q a_j \cdot H_j} \quad \text{für } i = 1, \dots, q.$$

Damit berechnet sich das Lorenzsche Konzentrationsmaß zu

$$LKM = \sum_{i=1}^q (k_{i-1} + k_i) \cdot \frac{a_i H_i}{\sum_{j=1}^q a_j H_j} - 1,$$

wobei gilt:

$$k_i = \sum_{j=1}^i \frac{H_j}{n} \quad \text{für } i = 1, \dots, q \quad \text{und} \quad k_0 = 0.$$

Für das Lorenzsche Konzentrationsmaß gilt stets

$$0 \leq LKM \leq \frac{n-1}{n};$$

es ist

$$LKM = 0,$$

falls alle Merkmalswerte in der Beobachtungsreihe gleich sind, und

$$LKM = \frac{n-1}{n},$$

falls eine einzige Untersuchungseinheit die gesamte Merkmalssumme trägt.

Damit man ein Konzentrationsmaß hat, das auch wirklich den Wert 1 annehmen kann (bei vollständiger Konzentration) normiert man den Gini-Koeffizienten oft wie folgt:

$$LKM_{nor} = \frac{n}{n-1} LKM.$$

Beispiel: Für die Aufteilung der landwirtschaftlichen Nutzflächen einer Region, vgl. Tab. 12, wollen wir das LKM berechnen. Aus Tab. 13 ergibt sich

$$LKM = \sum_{i=1}^5 2F_i - 1 = 1,6408 - 1 = 0,6408 \quad \text{bzw.}$$

$$LKM_{nor} = \frac{n}{n-1} LKM = \frac{50}{49} \cdot 0,6408 = 0,6539.$$

Das Lorenzsche Konzentrationsmaß kann trotz unterschiedlicher Konzentrationsstrukturen gleich sein.

Tab.13: Landwirtschaftlich genutzte Flächen, Arbeitstabelle zum Beispiel

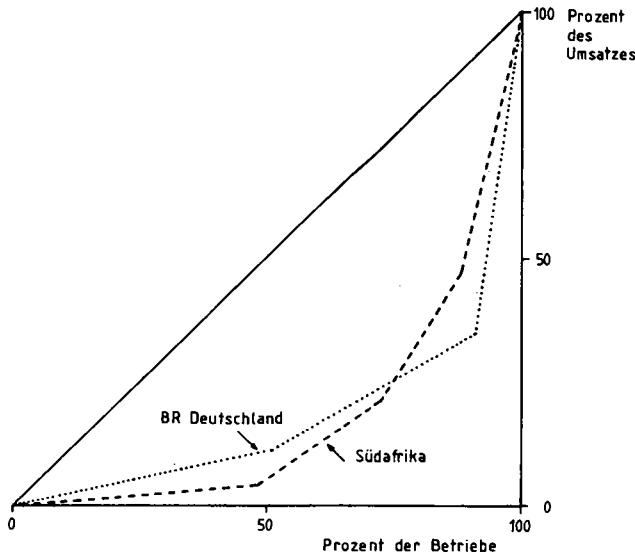
Flächengröße von... bis... ha	Flächen- anteile $a_i H_i$	aufsummierte Betriebsanteile k_i	$2 F_i$
	$\frac{1}{q} \sum_{j=1}^q a_j H_j$		
bis 5 ha	0,063	0,42	0,02646
5 bis 10 ha	0,072	0,60	0,07344
10 bis 20 ha	0,135	0,78	0,18630
20 bis 50 ha	0,280	0,94	0,48160
> 50 ha	0,450	1,00	0,87300
Gesamt	1,000		1,64080

Beispiel: In Abb. 20 ergibt sich für Südafrika: LKM = 0,63 und für die Bundesrepublik Deutschland: LKM = 0,64. Das LKM ist also in beiden Ländern fast gleich. Die Umsätze verteilen sich aber in Südafrika anders auf die Betriebe als in der Bundesrepublik Deutschland. Denn, während in Südafrika

weniger als 5% des Umsatzes auf fast 50% der Betriebe,
ungefähr 15% des Umsatzes auf ca. 20% der Betriebe,
ungefähr 25% des Umsatzes auf ca. 20% der Betriebe und
ungefähr 55% des Umsatzes auf ca. 10% der Betriebe entfallen,

konzentrieren sich in der Bundesrepublik Deutschland

ungefähr 10% des Umsatzes auf ca. 50% der Betriebe,
ungefähr 25% des Umsatzes auf ca. 40% der Betriebe und
ungefähr 65% des Umsatzes auf ca. 10% der Betriebe.

**Abb. 20:** Lorenzkurve der Umsatzkonzentration bei Ackerschleppern, Südafrika und Bundesrepublik Deutschland 1977

In Südafrika entfallen also sowohl auf Betriebe mit sehr geringem Anteil am Umsatz als auch auf Betriebe mit einem sehr hohen Anteil am Umsatz weniger Prozent des Umsatzes als in der Bundesrepublik Deutschland. Dabei ist der Anteil der Betriebe mit sehr geringem bzw. sehr hohem Umsatzanteil in beiden Ländern ungefähr gleich.

7. Verhältniszahlen

In der deskriptiven Statistik betrachtet man auch Zahlen, die einen bestimmten Sachverhalt quantitativ kennzeichnen. So beschreiben z. B. die Lagemaße ein „Zentrum“ einer Häufigkeitsverteilung, die Streungsmaße eine „Streuung“ der Merkmalsausprägungen und die Merkmalsausprägungen selbst eine Eigenschaft der Untersuchungseinheit. Diese werden als **statistische Maßzahlen** bezeichnet.

Bildet man den Quotienten zweier Maßzahlen, so erhält man eine wichtige Klasse von Maßzahlen: die sogenannten **Verhältniszahlen**.

Diese lassen sich, wie in der Abb. 21 gezeigt, untergliedern.

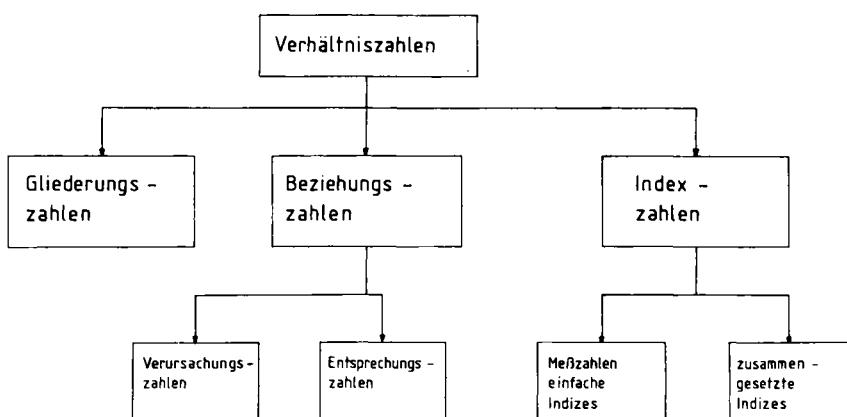


Abb. 21: Gliederung von Verhältniszahlen

7.1. Gliederungszahlen

Bezieht man eine Teilgröße auf eine ihr übergeordnete Gesamtgröße, so erhält man eine Gliederungszahl.

Besteht z.B. in einem Betrieb eine Tagesproduktion aus 1500 Maschinenteilen, wobei 300 Teile eine Qualitätskontrolle nicht überstehen, dann sind 20% der Tagesproduktion Ausschuß; dieser Ausschußanteil ist eine Gliederungszahl.

Weitere Beispiele findet man im Abschnitt 2: die relativen Häufigkeiten. Dort werden absolute Häufigkeiten eines Beobachtungswertes dividiert durch die Summe aller absoluten Häufigkeiten.

Man kann Gliederungszahlen in Anteilen, Prozenten oder in Promillen angeben.

7.2. Beziehungszahlen

Den Quotienten zweier sachlich sinnvoll in Verbindung stehender Maßzahlen, z. B. die Anzahl der Lebendgeborenen in einem Jahr bezogen auf die durchschnittliche Anzahl der Frauen in der Bundesrepublik Deutschland, bezeichnet man als **Beziehungszahl**.

Man unterscheidet zwischen **Verursachungs-** und **Entsprechungszahlen**.

7.2.1. Verursachungszahlen

Bei Verursachungszahlen bezieht man **Bewegungsmassen** auf die zugehörigen **Bestandsmassen**.

Dabei versteht man unter Bestandsmassen solche Massen, deren Elemente eine bestimmte Dauer haben; es wird zu einem festen Zeitpunkt der Bestand ermittelt. Dagegen sind die Elemente einer Bewegungsmasse Ereignisse, also einem festen Zeitpunkt zugeordnet. Eine zeitliche Abgrenzung erfolgt durch Festlegung eines Zeitraumes. Betrachten wir z. B. den Lagerbestand einer Brauerei am 15.4.1979 um 8⁰⁰ Uhr von 500 hl Bier: Werden an diesem Tag 200 hl Bier abgesetzt, so sind dies 40% des Bestandes; bei einer Tagesproduktion von 150 hl wird das Lager um 30% des alten Bestandes aufgefüllt.

Beispiel: Betrachten wir als weiteres Beispiel für eine Bestandsmasse die Wohnbevölkerung einer Stadt zu einem bestimmten Zeitpunkt. Die zugehörigen Bewegungsmassen sind Geburten, Todesfälle, Zu- und Fortzüge. Die sogenannte „**Geburtenziffer**“

$$G = \frac{\text{Zahl der Lebendgeborenen}}{\text{Einwohner}} \cdot 1000$$

sagt, wie viele Geburten auf 1000 Einwohner einer Stadt entfallen, vgl. **Tab. 14**.

Tab. 14: Geburten in zwei Städten

	Wohnbevölk.	Geburten	Geburtenziffer G	Frauen zwischen 15–45 Jahren	Fruchtbarkeitsrate F
Stadt A	178 503	1794	10,05	34 478	52,03
Stadt B	288 495	3054	10,59	62 728	48,69

Es ist zu beachten, daß der Aussagegehalt von Beziehungszahlen von der Wahl der Bezugsgröße abhängt.

Beispiel: Die Geburtenziffer für die Stadt B ist im obigen Beispiel größer als für die Stadt A, vgl. Tab. 14. Wenn man jedoch die Bestandsmasse von **sterilen Teilmassen** befreit und die „**Fruchtbarkeitsrate**“

$$F = \frac{\text{Zahl der Lebendgeborenen}}{\text{Frauen im gebärfähigen Alter (15–45 Jahre)}} \cdot 1000$$

für jede Stadt berechnet, so kehrt sich die Relation gerade um.

7.2.2. Entsprechungszahlen

Alle Beziehungszahlen, bei denen man Ereignisse nicht auf ihren Bestand beziehen kann, werden als **Entsprechungszahlen** bezeichnet.

Beispiel: Betrachten wir als Beispiel das Schüler-Lehrer-Verhältnis an Grund- und Hauptschulen in Nordrhein-Westfalen 1978:

$$\frac{\text{Zahl der Schüler}}{\text{Zahl der Lehrer}} = \frac{1581506}{66779} \simeq 23,7.$$

Es entfielen also auf einen Lehrer ungefähr 24 Schüler. Man beachte, daß diese Zahl i. a. nicht der durchschnittlichen Klassengröße entspricht.

Einige der Beziehungszahlen haben einen natürlichen Einzug in unseren Sprachgebrauch gefunden. Begriffe wie Hektarertrag oder Bevölkerungsdichte sind den meisten bekannt.

Doch bei der Interpretation von Beziehungszahlen ist, wie man bei der Konstruktion einer Fruchtbarkeitsrate in Abschnitt 7.2.1 gesehen hat, Vorsicht geboten.

7.3. Indexzahlen

Werden nicht mehr nur auch sachlich sinnvoll in Verbindung stehende Maßzahlen, sondern zwei Maßzahlen der gleichen Art in Beziehung gesetzt, so spricht man von **Indexzahlen**.

Man bezeichnet diese als **Meßzahlen (einfache Indizes)**, wenn die zugehörigen Maßzahlen einen realen Sachverhalt beschreiben.

Ist eine der Maßzahlen eine Zahl, die einen fiktiven Zustand beschreibt (z. B. das Ausgabeverhalten einer Person, wenn sie das Dreifache ihres Gehaltes verdienen würde), so erhält man eine **zusammengesetzte Indexzahl** (kurz **Indexzahl**).

7.3.1. Meßzahlen

In vielen Fällen hat man es nicht nur mit zwei Maßzahlen, sondern mit einer Reihe von Maßzahlen zu tun, welche man in Beziehung setzen möchte. Wir betrachten den Fall einer zeitlichen Reihe genauer:

Seien x_t Maßzahlen zu den Zeitpunkten (**Perioden**) t , wobei x_0 eine Maßzahl zum **Basiszeitpunkt** (Periode) 0 ist, dann heißt:

Tab. 15: Umsatzzahlen der beiden Firmen

t Quartal ab 1.1.1976	Umsatz x_t in 1000 DM	Maier & Co		Müller KG		
		Meßzahlen x_t/x_0	Meßzahlen x_t/x_4	Umsatz y_t in 1000 DM	Meßzahlen y_t/y_0	Meßzahlen y_t/y_4
0	1240	1,00	1,08	960	1,00	1,07
1	1210	0,98	1,05	980	1,02	1,09
2	1180	0,95	1,03	980	1,02	1,09
3	1190	0,96	1,03	920	0,96	1,03
4	1150	0,93	1,00	900	0,94	1,00
5	1080	0,87	0,94	860	0,90	0,96
6	1120	0,90	0,97	880	0,92	0,98
7	1130	0,91	0,98	920	0,96	1,02

$$I_{0t} = \frac{x_t}{x_0} \quad \text{für } t = 0, 1, 2, \dots$$

eine **Zeitreihe einfacher Indizes** (Meßzahlen).

Beispiel: Untersuchen wie z.B. die vierteljährlichen Umsatzzahlen der Firma Maier & Co und der Firma Müller KG über einen Zeitraum von 2 Jahren, vgl. Tab. 15.

Man erhält Meßzahlen I_{0t} zur Basisperiode 0 als: $I_{0t} = \frac{x_t}{x_0}$. Jede beliebige Periode $k = 0, 1, \dots, 7$ ist als Basisperiode wählbar. In diesem Beispiel wurde $I_{4t} = \frac{x_t}{x_4}$ berechnet.

In der graphischen Darstellung der Umsätze, vgl. Abb. 22, bedeutet ein paralleler Verlauf der Kurven, daß die Umsätze in diesem Quartal um den gleichen absoluten Betrag gestiegen oder gesunken sind. Dagegen bedeuten parallel verlaufende Strecken bei der graphischen Darstellung der Indizes, vgl. Abb. 23, 24, daß im Verhältnis zum Basisquartal die Umsätze in beiden Firmen sich gleich verhalten.

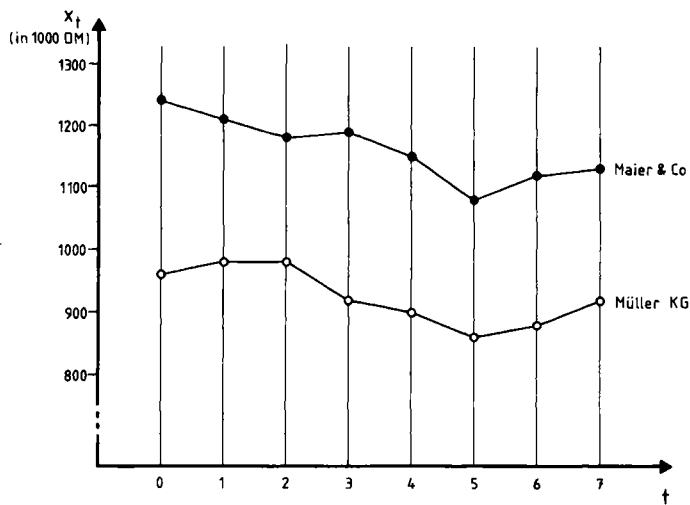


Abb. 22: Entwicklung des Umsatzes der beiden Firmen

Es ist üblich, die Meßzahlen mit 100 multipliziert zu betrachten. So besagt z.B. für Maier & Co

$$I_{05} \cdot 100 = \frac{x_5}{x_0} \cdot 100 = 87,$$

daß 87% des Umsatzes im **Basisquartal** im Quartal 5 umgesetzt werden, oder anders ausgedrückt, daß eine Minderung des Umsatzes um 13% vorliegt.

Vergleicht man z.B. die Meßzahlen

$$I_{05} \cdot 100 = 87 \quad \text{und} \quad I_{06} \cdot 100 = 90,$$

so sagt man, daß der Umsatz um 3 **Prozentpunkte** gestiegen ist.

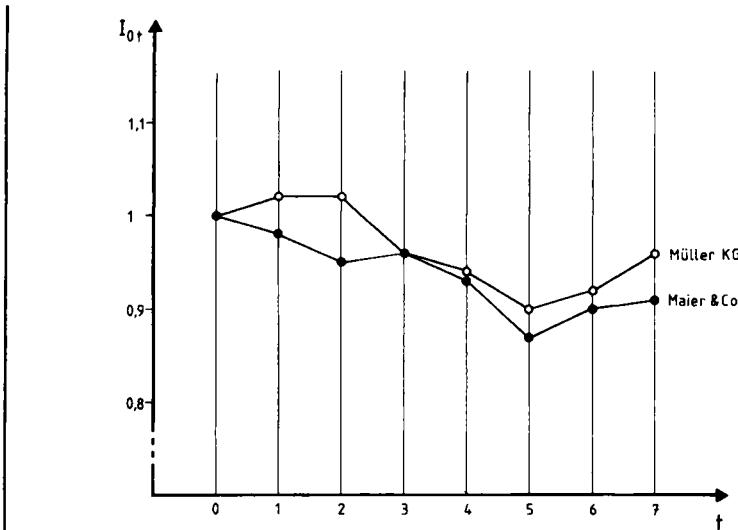


Abb. 23: Umsatzmeßzahlen mit Basisquartal 0

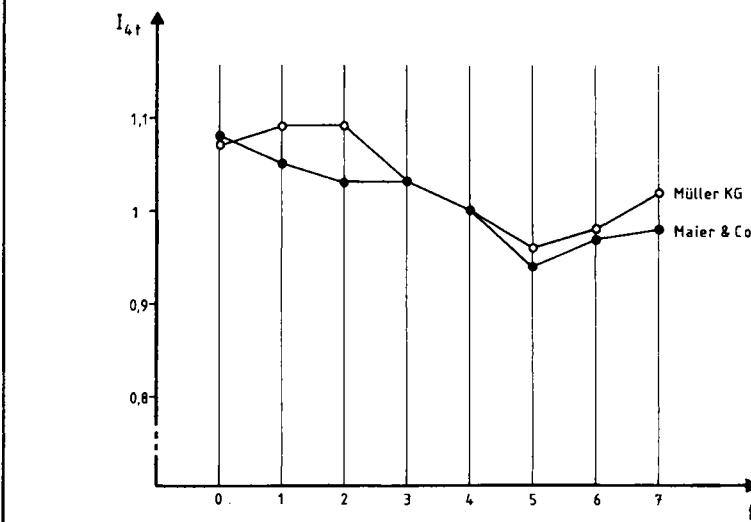


Abb. 24: Umsatzmeßzahlen mit Basisquartal 4

Im allgemeinen empfiehlt es sich nicht, lange zurückliegende Zeitperioden als Basisperiode zu wählen, und ebensowenig, Perioden mit außergewöhnlichen Zahlen zu wählen. Deshalb wird von Zeit zu Zeit eine **Umbasierung** der Zeitreihe notwendig. Dazu braucht man nicht die Ursprungsdaten zu kennen, da man die Indizes I_{kt} zur Basisperiode k wie folgt aus den Indizes zur Basisperiode 0 erhält:

$$I_{kt} = \frac{x_t}{x_k} = \frac{x_t/x_0}{x_k/x_0} = \frac{I_{0t}}{I_{0k}}.$$

Beispiel: In unserem Beispiel gilt für die Müller KG:

$$0,98 \simeq I_{46} = \frac{y_6}{y_4} = \frac{880}{900} \quad \text{und}$$

$$\frac{I_{06}}{I_{04}} = \frac{880/960}{900/960} \simeq \frac{0,92}{0,90} \simeq 0,98.$$

Direkt aus der Umbasierung läßt sich ihre inverse Operation, die **Verkettung** ableiten:

$$I_{0t} = I_{0k} \cdot I_{kt}.$$

Diese wird dann verwendet, wenn man für eine weitere Maßzahl x_t die Zeitreihe der Meßzahlen zur Basisperiode 0 ergänzen möchte, aber den Wert x_0 nicht mehr kennt; dabei muß man aber mit der Meßzahl I_{0k} auch die Maßzahl x_k kennen, um I_{kt} zu berechnen.

Beispiel: Im Beispiel sei für die Müller KG für $t = 8$ der Umsatz gleich 910 (in 1000 DM) gewesen. Dann ist

$$I_{08} = \frac{910}{960} \simeq 0,95 \quad \text{oder} \quad I_{08} = I_{04} \cdot I_{48} = 0,94 \cdot \frac{910}{900} \simeq 0,95.$$

Analog wie Zeitreihen von Meßzahlen lassen sich auch Reihen von Meßzahlen des örtlichen und sachlichen Verlaufs behandeln.

So ist z. B. der durchschnittliche Kartoffelpreis in verschiedenen Großstädten eine örtliche Reihe oder die Anzahl der verschiedenen Haustiere in 1000 Haushalten eine sachliche Reihe.

7.3.2. Standardisierung von Meßzahlen, Sterbeziffern

Beim Vergleich von gleichartigen Meßzahlen kann es zu Problemen bei der Interpretation kommen, wenn unterschiedliche Strukturen nicht berücksichtigt werden. So hat man z. B. bei der Konstruktion einer Fruchtbarkeitsrate, vgl. Abschnitt 7.2.1, versucht, diese Unterschiede weitgehend zu eliminieren, indem man die Bestandsmasse von sterilen Teilmassen befreit. Hier soll eine andere Methode, die der **Standardisierung**, anhand eines **Beispiels** erläutert werden.

Ziel ist es, für ein bestimmtes Jahr die Sterblichkeit der Wohnbevölkerung zweier Städte A und B zu vergleichen. Die Bevölkerung wird für jede Stadt zuerst in verschiedene Altersklassen – von 0 bis unter 1 Jahr, von 1 bis unter 10 Jahren usw. – eingeteilt, nun werden die Klassenhäufigkeiten H_{Ai} bzw. H_{Bi} und die Anzahl der Gestorbenen G_{Ai} bzw. G_{Bi} in den einzelnen Klassen bestimmt, vgl. auch Tab. 16.

Ein Vergleich der beiden globalen (mittleren) **Sterbeziffern** q_A und q_B – ermittelt aus der Gesamtzahl der Gestorbenen und der Wohnbevölkerung einer Stadt –

$$q_A = \frac{\sum_{i=1}^k G_{Ai}}{H_A} \quad \text{mit } H_A = \sum_{i=1}^k H_{Ai},$$

wobei k die Zahl der Altersklassen ist (q_B analog), scheint zu ergeben, daß die Sterblichkeit in der Stadt B um 4,4% höher liegt, als in der Stadt A:

Tab.16: Wohnbevölkerung in zwei Städten A und B, Gestorbene, Sterbeziffern und Arbeitstabelle zur Berechnung von standardisierten Meßzahlen

Alter von ... bis unter ... Jahren	Stadt A			Stadt B			Arbeitswerte			
	Wohn- bevölk. H _{Ai}	Gestor- bene G _{Ai}	Sterbe- ziffern q _{Ai} (· 1000)	Wohn- bevölk. H _{Bi}	Gestor- bene G _{Bi}	Sterbe- ziffern q _{Bi} (· 1000)	p _{Ai}	p _{Bi}	p _{Ai} q _{Bi}	p _{Bi} q _{Ai}
0– 1	4812	84	17,46	731	13	17,78	0,0097	0,0095	0,172	0,166
1–10	59 371	27	0,45	9 024	4	0,44	0,1198	0,1177	0,053	0,053
10–20	84 210	31	0,37	12 800	4	0,31	0,1700	0,1669	0,053	0,062
20–30	85 915	50	0,58	11 818	7	0,59	0,1734	0,1541	0,102	0,089
30–40	72 134	78	1,08	10 963	11	1,00	0,1456	0,1430	0,146	0,154
40–50	49 183	136	2,77	8 027	21	2,62	0,0993	0,1047	0,260	0,290
50–60	56 291	343	6,09	9 286	55	5,92	0,1136	0,1211	0,673	0,737
60–70	42 311	675	15,96	7 040	110	15,63	0,0854	0,0918	1,335	1,464
70–80	33 625	1677	49,87	5 704	278	48,74	0,0679	0,0744	3,309	3,710
80–90	6 715	955	142,22	1 133	154	135,92	0,0136	0,0148	1,849	2,105
90 und mehr	890	277	311,24	150	43	286,67	0,0018	0,0020	0,516	0,622
insgesamt	495 457	4333		76 676	700		1	1	8,47	9,45

$$I_{AB} = \frac{q_B}{q_A} = \frac{\sum_{i=1}^k G_{Bi}}{\sum_{i=1}^k G_{Ai}} \cdot \frac{H_A}{H_B} = \frac{700}{4333} \cdot \frac{495457}{76676} \simeq 1,044.$$

Betrachtet man die altersspezifischen Sterbeziffern q_{Ai} (bzw. q_{Bi}), definiert durch $q_{Ai} = \frac{G_{Ai}}{H_{Ai}}$ (bzw. $q_{Bi} = \frac{G_{Bi}}{H_{Bi}}$), so zeigt sich, daß die globalen Sterbeziffern nicht nur von den altersspezifischen Sterbeziffern abhängen, sondern auch vom Altersaufbau einer Stadt:

$$q_A = \frac{\sum_{i=1}^k G_{Ai}}{H_A} = \frac{\sum_{i=1}^k H_{Ai} q_{Ai}}{H_A} = \sum_{i=1}^k p_{Ai} q_{Ai} \quad \text{mit } p_{Ai} = \frac{H_{Ai}}{H_A}.$$

Um beim Vergleich der Sterblichkeit den Einfluß der Altersstruktur auszuschalten, muß man **standardisierte Sterbeziffern** verwenden, d.h. für beide Städte den gleichen „**Standardaltersaufbau**“ voraussetzen. So ist

$$q_B^{st} = \sum_{i=1}^k p_{Ai} q_{Bi}$$

die standardisierte Sterbeziffer der Stadt B, berechnet mit dem Altersaufbau p_{A1}, \dots, p_{Ak} der Stadt A, und entsprechend ist

$$q_A^{st} = \sum_{i=1}^k p_{Bi} q_{Ai} \quad \text{mit } p_{Bi} = \frac{H_{Bi}}{H_B}$$

die standardisierte Sterbeziffer für die Stadt A mit dem Altersaufbau p_{B1}, \dots, p_{Bk} der Stadt B.

| Für den **Vergleich der Sterbeziffern** bieten sich die Meßzahlen

$$I_{AB} = \frac{q_B^{st}}{q_A} = \frac{\sum_{i=1}^k p_{Ai} q_{Bi}}{q_A} \approx \frac{8,47}{8,75} = 0,968 \quad \text{bzw.}$$

$$I_{AB} = \frac{q_B}{q_A^{st}} = \frac{q_B}{\sum_{i=1}^k p_{Bi} q_{Ai}} \approx \frac{9,13}{9,45} \approx 0,966$$

an.

Hätte die Stadt B die gleiche Altersstruktur wie die Stadt A, dann hätte die Stadt B 3,2% weniger Gestorbene als die Stadt A. Oder im zweiten Fall wären es 3,4% weniger Gestorbene, falls in beiden Städten der Altersaufbau der Stadt B zugrundegelegt würde.

Bei den oben berechneten Indexzahlen handelt es sich bereits um **zusammengesetzte Indexzahlen**, da eine der beiden Maßzahlen einen nichtrealen Zustand beschreibt, in diesem Fall eine fiktive Sterbeziffer unter der Annahme einer anderen Altersstruktur.

Weitere Beispiele zur Standardisierung (Mengenindex, Preisindex) werden im nächsten Abschnitt folgen.

7.3.3. Zusammengesetzte Indexzahlen

Hat man Reihen gleichartiger Daten, so möchte man den Zusammenhang dieser Reihen oft durch eine Zahl beschreiben. Im Abschnitt 7.3.2 haben wir eine Methode kennengelernt, mit der man unterschiedliche Strukturen der Reihen ausschalten kann: die Standardisierung. Ein Vergleich einer standardisierten Meßzahl mit einer anderen Meßzahl führt uns zum Begriff der **zusammengesetzten Indexzahlen**, kurz **Indizes**. Am Beispiel eines **Warenkorbes** mit n verschiedenen Waren werden einige Indizes vorgestellt. Kennt man für die Waren bestimmte Mengen q_{t1}, \dots, q_{tn} und bestimmte Preise p_{t1}, \dots, p_{tn} zu einer bestimmten Periode t (für $t = 0$ **Basisperiode** oder für $t > 0$ **Berichtsperiode** genannt), so kann man z. B. für einen Warenkorb, bestehend aus den Getränken Milch, Limonade, Bier und Wein, den durchschnittlichen Verbrauch eines 4-Personenhaushaltes mit den durchschnittlichen Preisen pro Liter oder auch die Absätze einer Getränkefirma mit deren Getränkepreisen vergleichen. Der **Wert eines Warenkorbes** q_{t1}, \dots, q_{tn} in einer Periode t wird durch

$$\sum_{i=1}^n p_{ti} q_{ti}$$

bestimmt. Dies entspricht dann im Beispiel dem tatsächlichen Aufwand eines 4-Personenhaushaltes für diese Getränke.

A. Der Wertindex

Vergleicht man den Wert eines Warenkorbes zur Berichtsperiode t mit dem zur Basisperiode 0, so erhält man einen **Wertindex** (eine Meßzahl)

$$W_{0t} = \frac{\sum_{i=1}^n p_{ti} q_{ti}}{\sum_{i=1}^n p_{0i} q_{0i}}$$

Mit Hilfe des Wertindex $W_{0,t}$ lassen sich z. B. Aussagen darüber machen, wie sich der tatsächliche Aufwand eines 4-Personenhaushaltes für Getränke im Berichtsmonat t in Bezug zum Basismonat 0 verändert hat.

B. Preisindizes nach Laspeyres und nach Paasche

Möchte man jedoch Aussagen über Preisentwicklungen machen, so muß man für verschiedene Perioden das gleiche Mengenschema verwenden. Dies führt uns zu den **Preisindizes**

$$P_{0,t}^P = \frac{\sum_{i=1}^n p_{ti} q_{ti}}{\sum_{i=1}^n p_{0i} q_{ti}} \quad \text{nach Paasche und}$$

$$P_{0,t}^L = \frac{\sum_{i=1}^n p_{ti} q_{0i}}{\sum_{i=1}^n p_{0i} q_{0i}} \quad \text{nach Laspeyres.}$$

Beim Preisindex nach Paasche vergleicht man den Wert einer **Gütermengenkombination (Warenkorb)** q_{t1}, \dots, q_{tn} zur jeweiligen Berichtsperiode t mit dem Wert, den dieser unter der Preissituation zur Basisperiode gehabt hätte. Dagegen wird beim Preisindex nach Laspeyres die Gütermengenkombination q_{01}, \dots, q_{0n} der Basisperiode 0 für alle Berichtsperioden zugrundegelegt und ihr fiktiver Wert zur Berichtsperiode t mit ihrem Wert zur Basisperiode verglichen.

C. Ein Beispiel

Betrachten wir als Beispiel einen Preisindex für den durchschnittlichen Verbrauch an Fleisch aller privaten Haushalte in einer Gemeinde. Die durchschnittlich verbrauchten Mengen an Fleisch und die durchschnittlichen Fleischpreise wurden jeweils im Basismonat 0 und im Berichtsmonat 12 erfaßt, vgl. Tab. 17. Die Ergebnisse, die die Indizes liefern, werden in Prozent „auf der Basis 0“ angegeben. Der Wertindex

$$100 \cdot W_{0,12} = 100 \cdot \frac{\sum_{i=1}^7 p_{12i} q_{12i}}{\sum_{i=1}^7 p_{0i} q_{0i}} = 100 \cdot \frac{62750,5}{62012,5} \simeq 101,19$$

sagt uns, daß der tatsächliche Aufwand für Fleisch für die privaten Haushalte vom Basismonat bis zum Berichtsmonat um 1,19% gestiegen ist. Dabei ist nicht berücksichtigt, daß der durchschnittliche Verbrauch im Berichtsmonat um 205 g geringer ist.

Für die beiden Preisindizes erhält man:

$$100 \cdot P_{0,t}^P = \frac{\sum_{i=1}^7 p_{12i} q_{12i}}{\sum_{i=1}^7 p_{0i} q_{12i}} = 100 \cdot \frac{62750,5}{59962,5} \simeq 104,65$$

Tab.17: Durchschnittlicher Verbrauch an Fleisch aller privaten Haushalte und die durchschnittlichen Fleischpreise in einer Gemeinde im Basismonat und im Berichtsmonat; Arbeitstabelle

Fleisch	p_{0i} Preis im Basismonat (DM/kg)	q_{0i} verbrauchte Mengen (g)	$p_{1,2i}$ Preis im Berichtsmonat (DM/kg)	$q_{1,2i}$ verbrauchte Mengen (g)	$p_{0i} q_{0i}$	$p_{1,2i} q_{1,2i}$	$p_{0i} q_{1,2i}$	$p_{1,2i} q_{0i}$
Rindfleisch zum Kochen	5,75	1450	6,20	1410	8337,5	8742	8107,5	8990
Rindfleisch zum Schmoren und Braten	12,25	1260	13,55	1200	15435	16260	14700	17073
Rindfleisch, Lendenfilet	24,10	50	24,80	50	1205	1240	1205	1240
Roastbeef	17,80	210	18,10	215	3738	3891,5	3827	3801
Schweinebauchfleisch	4,55	2450	4,60	2470	11147,5	11362	11238,5	11270
Schweinebraten/Schluler	10,25	1310	10,40	1230	13427,5	12792	12607,5	13624
Schweinefleisch, Kotelett ohne Filet	8,90	980	9,10	930	8722	8463	8277	8918
insgesamt		7710		7505	62012,5	62750,5	59962,5	64916

und

$$100 \cdot P_{0,t}^L = \frac{\sum_{i=1}^7 p_{12i} q_{0i}}{\sum_{i=1}^7 p_{0i} q_{0i}} = 100 \cdot \frac{64916}{62012,5} \simeq 104,68.$$

Nimmt man für beide Monate den durchschnittlichen Verbrauch an Fleisch im Berichtsmonat als **Mengenschema (Warenkorb)** an, so sind die Preise in diesem Zeitraum um 4,65% gestiegen; oder: Für die im Basismonat verbrauchten Mengen an Fleisch muß man im Berichtsmonat 4,68% mehr Geld ausgeben.

D. Mengenindizes nach Laspeyres und nach Paasche

Ist man nun daran interessiert, wie sich die verbrauchten Mengen unabhängig von der Preisentwicklung verändert haben, so kann man einen **Mengenindex** betrachten. Standardisiert man nach den Preisen zur Berichtsperiode, so erhält man einen

Mengenindex nach Paasche:

$$Q_{0,t}^P = \frac{\sum_{i=1}^n p_{ti} q_{ti}}{\sum_{i=1}^n p_{0i} q_{0i}}$$

und entsprechend nach den Preisen zur Basisperiode einen

Mengenindex nach Laspeyres:

$$Q_{0,t}^L = \frac{\sum_{i=1}^n p_{0i} q_{ti}}{\sum_{i=1}^n p_{0i} q_{0i}}.$$

Beispiel: Für unser Beispiel, vgl. auch Abschnitt C, sagt

$$100 \cdot Q_{0,t}^P = 100 \cdot \frac{62750,5}{64916} \simeq 96,66,$$

dass der Verbrauch an Fleisch, gewichtet mit den Preisen im Berichtsmonat, um 3,34% gesunken ist.

Entsprechend besagt

$$100 \cdot Q_{0,t}^L = 100 \cdot \frac{59962,5}{62012,5} \simeq 96,69,$$

dass der Verbrauch an Fleisch, gewichtet mit den Preisen zum Basismonat, um 3,31% gesunken ist.

E. Preisbereinigung; Deflationierung

Die Umbasierung bzw. Verkettung, wie wir sie bei den Meßzahlen kennengelernt haben, lässt sich nicht auf zusammengesetzte Indizes übertragen, denn im allgemeinen ist $Q_{0,r} \cdot Q_{r,t}$ ungleich $Q_{0,t}$ bzw. $P_{0,r} \cdot P_{r,t}$ ungleich $P_{0,t}$.

Es gilt aber

$$Q_{0,t}^L = \frac{\sum_i p_{0i} q_{ti}}{\sum_i p_{0i} q_{0i}} \cdot \frac{\sum_i p_{ti} q_{ti}}{\sum_i p_{ti} q_{ti}} = \frac{W_{0t}}{P_{0,t}^P}$$

und

$$Q_{0,t}^P = \frac{\sum_i p_{ti} q_{ti}}{\sum_i p_{ti} q_{0i}} \cdot \frac{\sum_i p_{0i} q_{0i}}{\sum_i p_{0i} q_{0i}} = \frac{W_{0t}}{P_{0,t}^L}$$

Die Berechnung eines Mengenindex Q_{0t} aus einem Wertindex W_{0t} nennt man **Preisbereinigung** oder **Deflationierung**.

F. Preis- und Mengenindizes als gewogene Mittel von Meßzahlen: Subindizes

Eine Zahl, die das Preisverhältnis einer Ware zwischen Berichtsperiode und Basisperiode zum Ausdruck bringt, heißt **Preismeßzahl** und ist definiert durch $\frac{p_{ti}}{p_{0i}}$.

Die Preisindizes lassen sich dann als gewogenes Mittel von Preismeßzahlen darstellen: Den Preisindex nach Laspeyres erhält man als gewogenes arithmetisches Mittel:

$$P_{0,t}^L = \frac{\sum_i p_{ti} q_{0i}}{\sum_i p_{0i} q_{0i}} = \sum_i \frac{p_{ti} q_{0i}}{\sum_j p_{0j} q_{0j}} \cdot \frac{p_{0i} q_{0i}}{p_{0i} q_{0i}} = \sum_i g_{0i} \frac{p_{ti}}{p_{0i}}$$

mit $g_{0i} = \frac{p_{0i} q_{0i}}{\sum_j p_{0j} q_{0j}}$

(die Gewichte g_{0i} entsprechen den Ausgabenanteilen zur Basisperiode), den Preisindex nach Paasche erhält man als gewogenes harmonisches Mittel von Preismeßzahlen:

$$\frac{1}{P_{0,t}^P} = \frac{\sum_i p_{0i} q_{ti}}{\sum_j p_{0j} q_{tj}} = \sum_i \frac{p_{0i} q_{ti}}{\sum_j p_{0j} q_{tj}} \cdot \frac{p_{ti} q_{ti}}{p_{ti} q_{ti}} = \sum_i g_{ti} \frac{1}{p_{ti}/p_{0i}}$$

mit $g_{ti} = \frac{p_{ti} q_{ti}}{\sum_j p_{0j} q_{tj}}$.

(Dabei fungieren die Ausgabenanteile zur jeweiligen Berichtsperiode als Gewichte.)

Durch Vertauschen der Symbole p und q gelangt man zu entsprechenden Darstellungen für Mengenindizes. Der Grund dafür, daß in der amtlichen Statistik in der Regel die Wertgewichtsmethode zur Berechnung von Indizes verwendet wird, liegt in der Aufteilung von Waren in Untergruppen. So wird z.B. für den **Preisindex für die Lebenshaltung** das Güterverzeichnis für den privaten Verbrauch in 9 Hauptgruppen aufgeteilt; vgl. **Tab. 18**.

Für den Preisindex nach Laspeyres (Mengenindex analog) gilt folgende Aufspaltung in zwei **Subindizes** bei einem Warenkorb mit den Waren $i = 1, \dots, n$ unterteilt in zwei Warengruppen I und II mit den Waren $i = 1, \dots, k$ und $i = k + 1, \dots, n$:

Tab.18: Hauptgruppen des Güterverzeichnisses des Lebenshaltungsindex

-
- 1 Nahrungs- und Genußmittel
 - 2 Kleidung, Schuhe
 - 3 Wohnungsmieten u.ä.
 - 4 Elektrizität, Gas, Brennstoffe u.ä.
 - 5 Waren und Dienstleistungen für die Haushaltsführung
 - 6 Verkehrszwecke, Nachrichtenübermittlung
 - 7 Körper- und Gesundheitspflege
 - 8 Bildungs- und Unterhaltungszwecke
 - 9 Persönliche Ausstattung und sonstiges
-

$$\begin{aligned} P_{0,t}^L &= \sum_{i=1}^n g_{0i} \frac{p_{ti}}{p_{0i}} = \sum_{i=1}^k g_{0i} \frac{p_{ti}}{p_{0i}} + \sum_{i=k+1}^n g_{0i} \frac{p_{ti}}{p_{0i}} \\ &= g_0^I \sum_{i=1}^k \frac{g_{0i}}{g_0^I} \frac{p_{ti}}{p_{0i}} + g_0^{II} \sum_{i=k+1}^n \frac{g_{0i}}{g_0^{II}} \frac{p_{ti}}{p_{0i}} \end{aligned}$$

$$\text{mit den Gewichten } g_0^I = \frac{\sum_{i=1}^k p_{0i} q_{0i}}{\sum_{i=1}^n p_{0i} q_{0i}} \quad \text{und} \quad g_0^{II} = \frac{\sum_{i=k+1}^n p_{0i} q_{0i}}{\sum_{i=1}^n p_{0i} q_{0i}},$$

welche den Ausgabenanteilen der Warengruppen zur Basisperiode entsprechen.

$$\text{Da } g_{0i}^I = \frac{g_{0i}}{g_0^I} = \frac{p_{0i} q_{0i}}{\sum_{j=1}^n p_{0j} q_{0j}} \cdot \frac{\sum_{j=1}^n p_{0j} q_{0j}}{\sum_{j=1}^k p_{0j} q_{0j}} = \frac{p_{0i} q_{0i}}{\sum_{j=1}^k p_{0j} q_{0j}}$$

$$\text{und } g_{0i}^{II} = \frac{g_{0i}}{g_0^{II}} = \frac{p_{0i} q_{0i}}{\sum_{j=k+1}^n p_{0j} q_{0j}}$$

den Ausgabenanteilen einer Ware i an der jeweiligen Warengruppe zur Basisperiode entsprechen, erhält man die Preisindizes für diese Warengruppe; es ist also

$$P_{0,t}^L = g_0^I P_{0,t}^{L(I)} + g_0^{II} P_{0,t}^{L(II)}$$

mit den Subindizes

$$\begin{aligned} P_{0,t}^{L(I)} &= \sum_{i=1}^k g_{0i}^I \frac{p_{ti}}{p_{0i}} \quad \text{und} \\ P_{0,t}^{L(II)} &= \sum_{i=k+1}^n g_{0i}^{II} \frac{p_{ti}}{p_{0i}}. \end{aligned}$$

Analog lässt sich der Preisindex nach Paasche als gewichtetes harmonisches Mittel von Subindizes darstellen:

$$\begin{aligned}\frac{1}{P_{0,t}^P} &= \sum_{i=1}^n g_{ti} \frac{1}{p_{ti}/p_{0i}} = \sum_{i=1}^k g_{ti} \frac{1}{p_{ti}/p_{0i}} + \sum_{i=k+1}^n g_{ti} \frac{1}{p_{ti}/p_{0i}} \\ &= g_t^I \frac{1}{P_{0,t}^{P(I)}} + g_t^{II} \frac{1}{P_{0,t}^{P(II)}}\end{aligned}$$

mit den Gewichten $g_t^I = \frac{\sum_{i=1}^k p_{ti} q_{ti}}{\sum_{i=1}^n p_{ti} q_{ti}}$ und $g_t^{II} = \frac{\sum_{i=k+1}^n p_{ti} q_{ti}}{\sum_{i=1}^n p_{ti} q_{ti}}$

und den Subindizes

$$\begin{aligned}\frac{1}{P_{0,t}^{P(I)}} &= \sum_{i=1}^k \frac{g_{ti}}{g_t^I} \frac{1}{p_{ti}/p_{0i}} = \sum_{i=1}^k \frac{p_{ti} q_{ti}}{\sum_{j=1}^k p_{tj} q_{tj}} \frac{1}{p_{ti}/p_{0i}}, \\ \frac{1}{P_{0,t}^{P(II)}} &= \sum_{i=k+1}^n \frac{p_{ti} q_{ti}}{\sum_{j=k+1}^n p_{tj} q_{tj}} \frac{1}{p_{ti}/p_{0i}}\end{aligned}$$

Beispiel: Sehen wir uns die Berechnung von Subindizes in unserem Beispiel für die Unterwarengruppen Rind- und Schweinefleisch an, und beschränken wir uns auf die Preisindizes, vgl. Tab. 19.

Tab. 19: Berechnung der Subindizes mit der Gewichtungsmethode, mit Gewichten

$$g_{ti}^I = \frac{p_{ti} q_{ti}}{4} \text{ und } g_{ti}^{II} = \frac{p_{ti} q_{ti}}{7} \text{ für die Monate } t = 0 \text{ und } t = 12$$

Untergruppe	Preismeßzahlen		Gewichte		
	$\frac{p_{12i}}{p_{0i}} \cdot 100$	g_{0i}^I	g_{12i}^I	$g_{0i}^I \frac{p_{12i}}{p_{0i}} \cdot 100$	$g_{12i}^I \frac{1}{p_{12i}/p_{0i}}$
Rindfleisch					
zum Kochen	107,83	0,290	0,290	31,270	0,2690
zum Schmoren					
und Braten	110,61	0,538	0,540	59,509	0,4882
Lendenfilet	102,90	0,042	0,041	4,322	0,0398
Roastbeef	101,69	0,130	0,129	13,219	0,1269
gesamt		1	1	108,320	0,9239
Schweinefleisch					
Bauchfleisch	101,10	0,335	0,348	33,868	0,3442
Braten/Schulter	101,46	0,403	0,392	40,890	0,3863
Kotelett (o. Filet)	102,25	0,262	0,260	26,790	0,2543
gesamt		1	1	101,548	0,9848

Zu jeder Ware werden die Preismeßzahlen berechnet, welche die prozentualen Veränderungen gegenüber dem Basismonat angeben. Für Rindfleisch zum Kochen z. B. betrug der Preis im Berichtsmonat 107,83% des Preises im Basismonat, anders ausgedrückt ist der Preis um 7,83% gestiegen. Die in der Tabelle 19 stehenden Gewichte $g_0^{I(II)}$ bzw. $g_{12i}^{I(II)}$ geben an, welche Ausgaben für die i-te Ware im Verhältnis zu den Gesamtausgaben einer Unterwarengruppe, also Rindfleisch (I) oder Schweinefleisch (II), von allen privaten Haushalten der Gemeinde im Durchschnitt getätigt wurden (in beiden Monaten $t = 0$ und $t = 12$). Für die Preisindizes nach Laspeyres in den beiden Gruppen erhält man:

$$100 \cdot P_{0,12}^L = 108,32 \quad \text{bzw.} \quad 100 \cdot P_{0,12}^{I(II)} = 101,55.$$

Also ist der Warenkorb des Basismonats für die Warenguppe Rindfleisch um 8,32% und für die Warenguppe Schweinefleisch nur um 1,55% gestiegen.

Mit den Ausgabenanteilen der Warenguppen an den Gesamtausgaben (zum Basismonat) als Gewichte

$$g_0^I = \frac{\sum_{i=1}^4 p_{0i} q_{0i}}{\sum_{i=1}^7 p_{0i} q_{0i}} = \frac{28715,5}{62012,5} \approx 0,4631 \quad \text{und}$$

$$g_0^{II} = \frac{\sum_{i=5}^7 p_{0i} q_{0i}}{\sum_{i=1}^7 p_{0i} q_{0i}} = \frac{33297}{62012,5} \approx 0,5369,$$

erhält man den gemeinsamen Preisindex nach Laspeyres:

$$100 \cdot P_{0,12}^L = 100 \cdot (g_0^I P_{0,12}^{I(II)} + g_0^{II} P_{0,12}^{I(II)}) = 104,7.$$

Die Preisindizes nach Paasche für die beiden Warenguppen berechnet man aus den Gewichten g_{12i}^I bzw. g_{12i}^{II} und den Kehrwerten der Preismeßzahlen:

$$100 \cdot P_{0,12}^{P(I)} = \frac{100}{\sum_{i=1}^4 g_{12i}^I \frac{1}{p_{12i}/p_{0i}}} = \frac{100}{0,9239} \approx 108,24,$$

$$100 \cdot P_{0,12}^{P(II)} = \frac{100}{\sum_{i=5}^7 g_{12i}^{II} \frac{1}{p_{12i}/p_{0i}}} = \frac{100}{0,9848} \approx 101,54.$$

Den Warenkorb des Berichtsmonats zugrundelegend, ist der Preis für Rindfleisch um 8,24% und für Schweinefleisch um 1,54% gestiegen. Für den gemeinsamen Preisindex nach Paasche gilt:

$$100 \cdot P_{0,12}^P = \frac{100}{g_{12}^I \frac{1}{P_{0,12}^{P(I)}} + g_{12}^{II} \frac{1}{P_{0,12}^{P(II)}}}$$

$$= \frac{100}{0,4802 \cdot 0,9239 + 0,5198 \cdot 0,9848} \approx 104,65$$

mit den Gewichten

$$g_{12}^I = \frac{\sum_{i=1}^4 p_{12i} q_{12i}}{\sum_{i=1}^7 p_{12i} q_{12i}} = \frac{30133,5}{62750,5} \approx 0,4802 \quad \text{und}$$

$$g_{12}^{II} = \frac{\sum_{i=5}^7 p_{12i} q_{12i}}{\sum_{i=1}^7 p_{12i} q_{12i}} = \frac{32617}{62750,5} \approx 0,5198.$$

7.3.4. Vergleich von Preisindizes nach Laspeyres und nach Paasche; der Fishersche Idealindex; der Preisindex nach Lowe

Für den Preisindex nach Laspeyres spricht die Tatsache, daß der Warenkorb nur für die Basisperiode bestimmt werden muß und nicht für die weiteren Berichtsperioden wie beim Preisindex nach Paasche. Dabei ist nicht nur der Kostengesichtspunkt gemeint, sondern auch die Frage nach der Aktualität, da die umfangreichen Berechnungen eines Warenkorbes sich oft über einen längeren Zeitraum erstrecken. Sind diese Arbeiten einmal abgeschlossen, so läßt sich der Preisindex nach Laspeyres bei fortlaufender Preisnotierung schnell berechnen. Sein Nachteil besteht darin, daß die Ausgabenstruktur der Basisperiode veraltet, so daß nach einer bestimmten Zeit (ca. 4–8 Jahre) der Warenkorb neu festgesetzt werden muß. Der Preisindex nach Paasche paßt sich dagegen laufend der neuen Verbrauchssituation an. In den meisten Fällen wird daher der Preisindex nach Laspeyres benutzt und zu Kontrollzwecken der nach Paasche. Ergeben sich größere Abweichungen, so muß der Warenkorb neu festgelegt werden.

Häufig wird zur Überprüfung der Indizes auch noch der **Fishersche Idealindex** (nach Irving Fisher) berechnet, der sich gerade als geometrisches Mittel der Indizes nach Paasche und Laspeyres ergibt:

$$I_F = \sqrt{I_P \cdot I_L}.$$

Der Fishersche Idealindex nimmt also stets einen Wert an, der zwischen den Werten der Indizes von Paasche und Laspeyres liegt.

Manchmal wird nicht der Verbrauch der Basisperiode zur Berechnung der Indizes verwendet (z. B. wegen Saisonschwankungen), sondern ein anderer „geeigneter“ Verbrauch, z. B. beim Preisindex der Lebenshaltung der Durchschnittsverbrauch eines Jahres. Dies führt zum **Preisindex nach Lowe**, auf den wir hier nicht weiter eingehen wollen.

8. Die empirische Ausfallrate

Beschreiben die Daten aus einer Erhebung die Zeit bis zum Tod von Untersuchungsobjekten, so kann die **empirische Ausfallrate** zur Beschreibung der Daten herangezogen werden. Man unterteilt die Zeitachse in Intervalle $[0; t_1]$, $[t_1; t_2], \dots$ und berechnet die Anzahlen H_1, H_2, \dots der Versuchsstoffe mit Lebensdauern im Intervall $[0; t_1]$, $[t_1; t_2], \dots$. Bei insgesamt n betrachteten Lebensdauern ist die empirische Ausfallrate definiert als:

$$r_{\text{emp}}(t) = \frac{H_i}{(n - \sum_{j=1}^{i-1} H_j)(t_i - t_{i-1})} \quad \text{für } t \in [t_{i-1}; t_i), t_0 = 0.$$

Sie gibt – bezogen auf die Klassenbreite – den Anteil der im Intervall $[t_{i-1}; t_i)$ ausfallenden bzw. sterbenden Objekte H_i bezogen auf die Anzahl der zum Zeitpunkt t_{i-1} noch funktionierenden bzw. lebenden Objekte an. Sie hängt recht stark von der gewählten Klasseneinteilung ab, wie man im nachfolgenden Beispiel sieht.

Ist die empirische Ausfallrate ziemlich konstant, so heißt dies, daß die Objekte der Erhebung mit fortschreitender Zeit weder „altern“ noch sich „verjüngen“. Steigt die empirische Ausfallrate mit wachsender Zeit t , so „altern“ die Objekte bzw. zeigen Verschleißerscheinungen, fällt sie, so „verjüngen“ sich die Objekte mit fortschreitender Zeit, zeigen also mit zunehmender Zeit immer weniger Neigung zum Ausfallen; vgl. auch Abschnitt 4 in Kap. IV und Kap. XIII.

Beispiel: Wir wollen die empirische Ausfallrate $r_{\text{emp}}(t)$ am Beispiel der 87 Kühlaggregate, vgl. Tab. 4, berechnen und dafür zwei verschiedene Klasseneinteilungen wählen. Bei der Einteilung in Klassen der Breite 1 Jahr ergibt sich z. B.:

$$r_{\text{emp}}(t) = \frac{37}{(87 - 0)(1 - 0)} = 0,425 \quad \text{für } t \in [0; 1],$$

$$r_{\text{emp}}(t) = \frac{19}{(87 - 37)(2 - 1)} = 0,380 \quad \text{für } t \in [1; 2],$$

$$r_{\text{emp}}(t) = \frac{15}{(87 - 37 - 19)(3 - 2)} = 0,484 \quad \text{für } t \in [2; 3], \text{ usw.}$$

Diese empirische Ausfallrate ist in Abb. 25 abgetragen. Als zweites wählen wir die Einteilung in Klassen der Breite 0,5 Jahre und erhalten durch

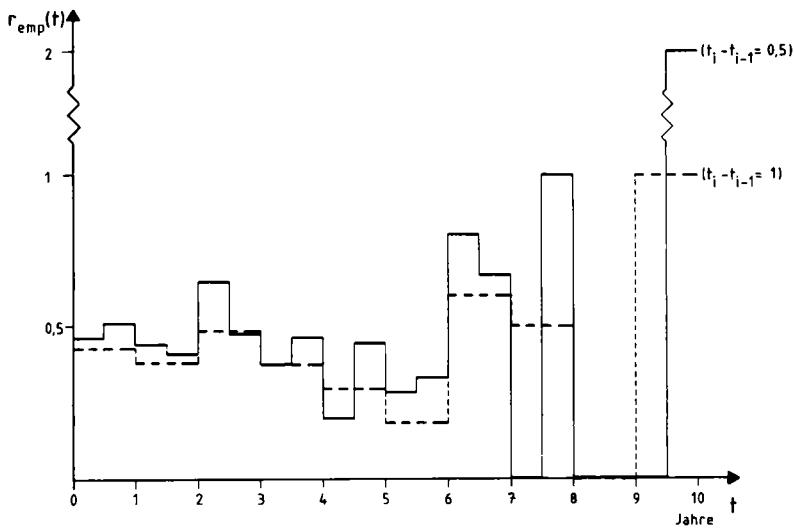


Abb. 25: Empirische Ausfallsrate der Kühlaggregate im Beispiel (bei verschiedenen Klasseneinteilungen)

$$r_{\text{emp}}(t) = \frac{20}{(87 - 0)(0,5 - 0)} = 0,460 \quad \text{für } t \in [0; 0,5),$$

$$r_{\text{emp}}(t) = \frac{17}{(87 - 20)(1 - 0,5)} = 0,508 \quad \text{für } t \in [0; 5, 1),$$

$$r_{\text{emp}}(t) = \frac{11}{(87 - 20 - 17)(1,5 - 1)} = 0,440 \quad \text{für } t \in [1; 1,5); \quad \text{usw.}$$

die andere in Abb. 25 dargestellte empirische Ausfallrate.

Der Zeichnung kann man entnehmen, daß 0,45 eine grobe konstante Approximation der empirischen Ausfallrate darstellt. Wir können für die in der Zeit beobachteten Kühlaggregate schließen, daß sie im Schnitt keine wesentlichen „Verschleißerscheinungen“ zeigten. Die relativen Ausfallhäufigkeiten zeigen keine gleichmäßig steigende oder fallende Tendenz.

9. Darstellung zweidimensionalen Zahlenmaterials und deskriptive Korrelationsrechnung

In Abschnitt 1 dieses Kapitels wurde der Begriff des mehrdimensionalen Merkmals eingeführt. Wir wollen uns hier auf zweidimensionale Merkmale beschränken und vor allem untersuchen, wie man die Abhängigkeit bzw. den Zusammenhang zwischen zwei Merkmalen messen kann.

Dabei können folgende Fragestellungen auftreten:

- Wie stark ist der Zusammenhang bzw. die Abhängigkeit?
- Läßt sich der Zusammenhang in einer bestimmten Form darstellen?

Wir wollen uns hier mit der ersten Fragestellung befassen, die unter dem Begriff **Korrelationsrechnung**, vgl. auch Kap. IX, bekannt ist. Die zweite Fragestellung wird in der **Regressionsrechnung** (vgl. Kap. X) behandelt.

Im folgenden betrachten wir n Untersuchungseinheiten, an denen die Merkmale X und Y beobachtet werden. Die verschiedenen Ausprägungen von X werden mit a_1, \dots, a_l , die von Y mit b_1, \dots, b_m bezeichnet. Somit erhält man das zweidimensionale Merkmal (X, Y) , dessen Ausprägungen (a_j, b_k) mit absoluter Häufigkeit H_{jk} und mit relativer Häufigkeit $h_{jk} = \frac{1}{n} H_{jk}$ auftreten.

9.1. Die Kontingenztafel

Die Häufigkeitsverteilung von (X, Y) wird meistens in Form einer **Kontingenztafel** dargestellt, vgl. Tab. 20 und Tab. 21.

$H_{j.} = \sum_{k=1}^m H_{jk}$ nennt man **absolute Randhäufigkeit** und

$h_{j.} = \sum_{k=1}^m h_{jk}$ **relative Randhäufigkeit** von a_j für $j = 1, \dots, l$.

Durch die Randhäufigkeiten erhält man die Häufigkeitsverteilung des Merkmals X , die dann **Randverteilung** genannt wird.

Tab.20 Kontingenztafel der absoluten Häufigkeiten H_{ij}

	b_1	b_2	\dots	b_m	\sum
a_1	H_{11}	H_{12}	\dots	H_{1m}	$H_{1\cdot}$
a_2	H_{21}	H_{22}	\dots	H_{2m}	$H_{2\cdot}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
a_l	H_{l1}	H_{l2}	\dots	H_{lm}	$H_{l\cdot}$
\sum	$H_{\cdot 1}$	$H_{\cdot 2}$	\dots	$H_{\cdot m}$	n

Tab.21: Kontingenztafel der relativen Häufigkeiten h_{ij}

	b_1	b_2	\dots	b_m	\sum
a_1	h_{11}	h_{12}	\dots	h_{1m}	$h_{1\cdot}$
a_2	h_{21}	h_{22}	\dots	h_{2m}	$h_{2\cdot}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
a_l	h_{l1}	h_{l2}	\dots	h_{lm}	$h_{l\cdot}$
\sum	$h_{\cdot 1}$	$h_{\cdot 2}$	\dots	$h_{\cdot m}$	n

Entsprechend liefern die Randhäufigkeiten von b_k , die durch

$$H_{\cdot k} = \sum_{j=1}^l H_{jk} \quad \text{bzw.} \quad h_{\cdot k} = \sum_{j=1}^l h_{jk} \quad \text{für } k = 1, \dots, m$$

gegeben sind, die Randverteilung von Y.

Es gilt:

$$h_{j\cdot} = \frac{1}{n} H_{j\cdot}, \quad h_{\cdot k} = \frac{1}{n} H_{\cdot k},$$

$$\sum_{j=1}^l \sum_{k=1}^m H_{jk} = \sum_{j=1}^l H_{j\cdot} = \sum_{k=1}^m H_{\cdot k} = n \quad \text{und}$$

$$\sum_{j=1}^l \sum_{k=1}^m h_{jk} = \sum_{j=1}^l h_{j\cdot} = \sum_{k=1}^m h_{\cdot k} = 1.$$

Auf die Analyse von Kontingenztafeln wird in Kap. VII eingegangen.

9.2. Der Korrelationskoeffizient nach Bravais-Pearson

Wir betrachten nun die Ausprägungen eines metrisch skalierten Merkmals (X, Y). Die verschiedenen Ausprägungen (a_j, b_k) treten mit Häufigkeiten H_{jk} für $j = 1, \dots, l, k = 1, \dots, m$ auf. Ein Maß für den Zusammenhang zwischen X und Y liefert dann der **Korrelationskoeffizient (nach Bravais-Pearson)** r_{XY} , der gegeben ist durch

$$r_{XY} = \frac{\sum_{j=1}^l \sum_{k=1}^m (a_j - \bar{a})(b_k - \bar{b}) H_{jk}}{\sqrt{\left(\sum_{j=1}^l (a_j - \bar{a})^2 H_{j\cdot} \right) \left(\sum_{k=1}^m (b_k - \bar{b})^2 H_{\cdot k} \right)}}$$

$$\begin{aligned}
 &= \frac{\sum_{j=1}^1 \sum_{k=1}^m a_j b_k H_{jk} - n \cdot \bar{a} \cdot \bar{b}}{\sqrt{\left(\sum_{j=1}^1 a_j^2 H_{j.} - n \bar{a}^2 \right) \left(\sum_{k=1}^m b_k^2 H_{.k} - n \bar{b}^2 \right)}} \\
 &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right)}}
 \end{aligned}$$

wenn (x_i, y_i) die Beobachtung an der i -ten Untersuchungseinheit bezeichnet.

Der Zähler von r_{XY} ist im wesentlichen die empirische **Kovarianz** zwischen X und Y, die als

$$\begin{aligned}
 s_{XY} &= \frac{1}{n-1} \sum_{j=1}^1 \sum_{k=1}^m (a_j - \bar{a})(b_k - \bar{b}) H_{jk} \\
 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right)
 \end{aligned}$$

definiert ist und die man auch als Maß des Zusammenhangs benutzen kann. Wenn man s_{XY} durch die Standardabweichungen s_X und s_Y von X bzw. Y dividiert, erhält man dann als ein normiertes Zusammenhangsmaß den Korrelationskoeffizienten r_{XY} , der nur Werte zwischen -1 und $+1$ annehmen kann.

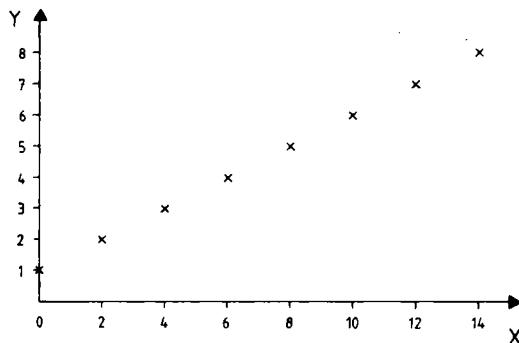
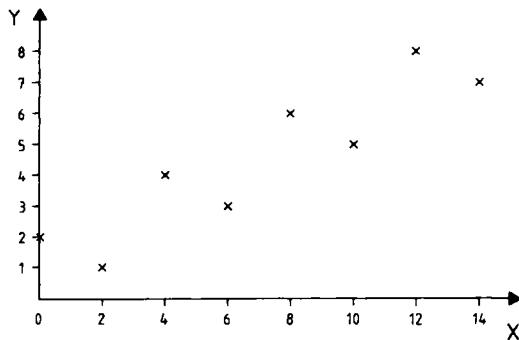
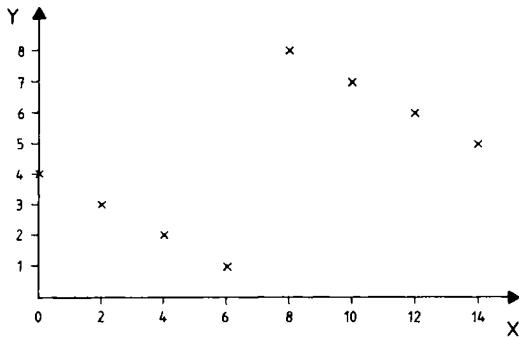
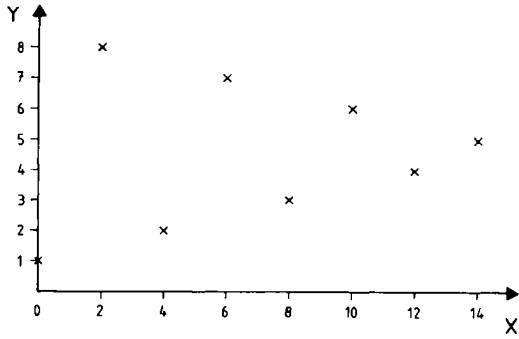
Um zu erkennen, wie sich der Korrelationskoeffizient r_{XY} bei verschiedenen Konstellationen von Ausprägungen des Merkmals (X, Y) verhält, werden im folgenden Beispiel jeweils für gleichbleibende Ausprägungen von X die Ausprägungen von Y so variiert, daß s_Y konstant bleibt.

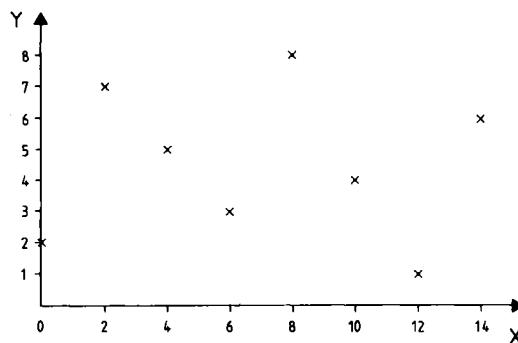
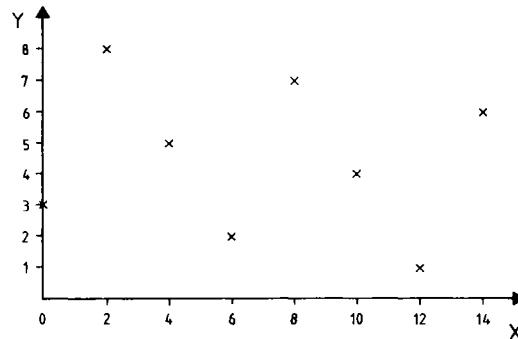
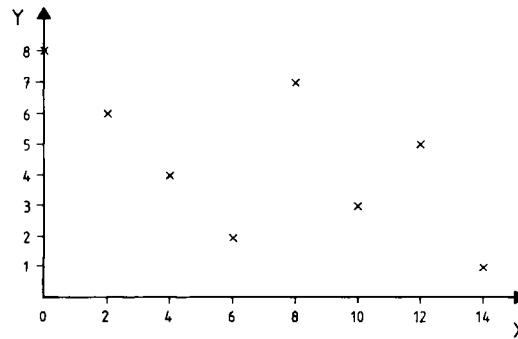
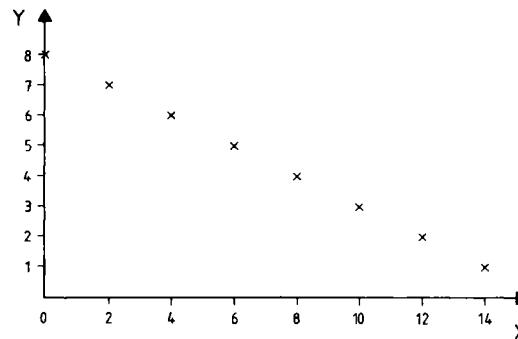
Beispiel: In Tab. 22 sind für je 8 Untersuchungseinheiten die Beobachtungsdaten (x_i, y_i) der Merkmale X und Y, wobei die Ausprägungen von X gleichbleiben und die von Y variieren, die Korrelationskoeffizienten nach Bravais-Pearson berechnet.

Tab. 22: Korrelationskoeffizient bei verschiedenen Konstellationen von Ausprägungen

x_i	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
0	1	2	4	1	2	3	8	8	
2	2	1	3	8	7	8	6	7	
4	3	4	2	2	5	5	4	6	
6	4	3	1	7	3	2	2	5	
8	5	6	8	3	8	7	7	4	
10	6	5	7	6	4	4	3	3	
12	7	8	6	4	1	1	5	2	
14	8	7	5	5	6	6	1	1	
$r_{XY} =$									
	1,00	0,905	0,52	0,19	0,00	-0,14	-0,62	-1,00	

Um eine Vorstellung von der Größe des Korrelationskoeffizienten r_{XY} zu bekommen, haben wir in den Abb. 26 bis 33 diese Kombinationen von Merkmalsausprägungen graphisch dargestellt.

Abb. 26: $r_{XY} = 1,00$ Abb. 27: $r_{XY} = 0,905$ Abb. 28: $r_{XY} = 0,52$ Abb. 29: $r_{XY} = 0,19$

**Abb. 30:** $r_{XY} = 0,00$ **Abb. 31:** $r_{XY} = -0,14$ **Abb. 32:** $r_{XY} = -0,62$ **Abb. 33:** $r_{XY} = -1,00$

Tab. 23: Studienanfänger im Jahr 1977 und Studenten insgesamt im Jahr 1977 an 10 Universitäten der Bundesrepublik Deutschland (vgl. Grund- und Strukturdaten 1979; der Bundesminister für Bildung und Wissenschaft)

	Univ. Bochum	Univ. Bremen	Univ. Claus- thal	Univ. Dort- mund	Univ. Frank- furt	Univ. Frei- burg	Univ. Ham- burg	Univ. Heidel- berg	Univ. Kiel	Univ. Mün- ster
Studien- anfänger	3970	732	499	1300	3463	2643	3630	3294	1931	5048
Studenten insgesamt	24273	5883	2847	5358	23442	17076	28360	19812	12379	31433

Man sieht am Beispiel, daß r_{XY} nahe 1 liegt, wenn der Zusammenhang zwischen X und Y annähernd positiv linear ist und nahe -1 liegt, wenn eine negativ lineare Abhängigkeit zu erkennen ist. Je „verstreuter“ die Ausprägungen (x_i, y_i) in der Ebene liegen, desto näher liegt r_{XY} bei Null.

Beispiel: Aus den Daten der Tab. 23 wollen wir die empirische Korrelation zwischen der Anzahl der Studienanfänger X und der Gesamtzahl von Studenten an einer Universität bestimmen. Es ergibt sich

$$\begin{aligned}
 r_{XY} &= \frac{\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^{10} (x_i - \bar{x})^2\right)\left(\sum_{i=1}^{10} (y_i - \bar{y})^2\right)}} \\
 &= \frac{\sum_{i=1}^{10} (x_i - 2651)(y_i - 17086,3)}{\sqrt{\left(\sum_{i=1}^{10} (x_i - 2651)^2\right)\left(\sum_{i=1}^{10} (y_i - 17086,3)^2\right)}} \\
 &= \frac{133195081}{\sqrt{20173924 \cdot 920380528}} = \frac{133195081}{136263300} = 0,9775.
 \end{aligned}$$

Es besteht also ein starker linearer Zusammenhang zwischen der Anzahl der Studienanfänger und der Anzahl aller Studenten an einer Universität.

Schon an dieser Stelle soll auf die Gefahr der **Scheinkorrelation** hingewiesen werden. Scheinkorrelation liegt dann vor, wenn die beiden Merkmale X und Y beide von einem dritten Merkmal Z abhängen: Die Gefahr eines Waldbrandes und schlechter Körnertrag zum Beispiel hängen von der Stärke der Sonnenbestrahlung ab. Ebenso sollte man darauf achten, daß die zu untersuchenden Größen in einem sachlogischen Zusammenhang stehen, damit keine **Nonsenskorrelation** entsteht, wie es z. B. die Korrelation zwischen der Anzahl der Störche und der Anzahl der Geburten in einem Land oder die Korrelation zwischen Apfelsinen – Einfuhrmenge und Anzahl der Krebserkrankungen ist. Gewarnt sei auch noch davor, durch geschickte Bastlei eine optimale Korrelation zu erhalten. Dies kann dadurch geschehen, daß man einige „unerwünschte“ Werte streicht, Werte einer Variablen einer zeitlichen Verschiebung unterwirft, bei einer Variablen zu logarithmischen Werten

übergeht etc. Ebenfalls sei noch einmal darauf hingewiesen, daß r_{XY} nur einen linearen Zusammenhang zwischen X und Y mißt und somit r_{XY} Null sein kann, obwohl ein starker nichtlinearer Zusammenhang vorliegt, vgl. Abb. 34.

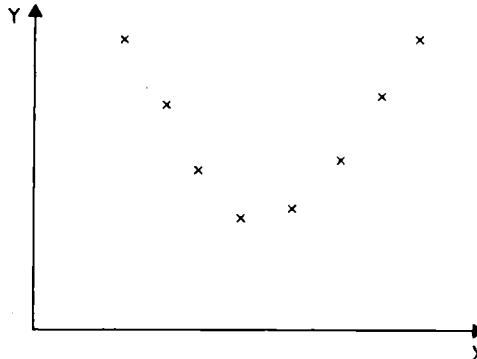


Abb. 34: Nichtlinearer Zusammenhang mit Korrelation $r_{XY} = 0$

9.3. Der Fechnersche Korrelationskoeffizient

Ein anderer – sehr leicht zu berechnender – Korrelationskoeffizient für metrisch skalierte Merkmalsausprägungen zweier Merkmale X und Y ist der **Fechnersche Korrelationskoeffizient** r_F . Dieser basiert nicht auf den ursprünglichen Merkmalsausprägungspaaren $(x_1, y_1), \dots, (x_n, y_n)$, sondern auf den Vorzeichen innerhalb der transformierten Paare $(x_i - \bar{x}, y_i - \bar{y}), \dots, (x_n - \bar{x}, y_n - \bar{y})$.

Wir bestimmen die Summe

$$V = \sum_{i=1}^n v_i, \quad \text{wobei } v_i = \begin{cases} 1, & \text{falls } x_i - \bar{x} \text{ und } y_i - \bar{y} \text{ das gleiche Vorzeichen haben oder beide Null sind} \\ 1/2, & \text{falls genau einer der Werte } x_i - \bar{x} \text{ bzw. } y_i - \bar{y} \text{ gleich Null ist} \\ 0, & \text{sonst} \end{cases}$$

und daraus den Fechnerschen Korrelationskoeffizienten

$$r_F = \frac{2V - n}{n},$$

der stets Werte zwischen -1 und $+1$ annimmt. Allerdings nimmt er z. B. den Wert $+1$ nicht nur bei positiv linearem Zusammenhang an, sondern schon dann, wenn für alle Paare (x_i, y_i) gilt:

$$(x_i < \bar{x} \text{ und } y_i < \bar{y}) \quad \text{oder} \quad (x_i > \bar{x} \text{ und } y_i > \bar{y}).$$

Beispiel: Bei $n = 15$ Hennen wurden die Merkmale $X \cong$ Körpergewicht und $Y \cong$ monatliche Legeleistung bestimmt. Als Maß für die Abhängigkeit zwischen diesen beiden Merkmalen wollen wir den Fechnerschen Korrelationskoeffizienten r_F bestimmen. Die festgestellten Merkmalsausprägungen sind in Tab. 24 zusammengestellt, die gleichzeitig auch unsere Arbeitstabelle ist.

Damit ergibt sich

Tab. 24: Legeleistung und Gewicht von 15 Hennen

Henne i	Körpergewicht x_i	Legeleistung y_i	$x_i - \bar{x} = x_i - 1847$	$y_i - \bar{y} = y_i - 21$	v_i
1	1763	19	-84	-2	1
2	1890	24	+43	+3	1
3	1872	23	+25	+2	1
4	1938	26	+91	+5	1
5	1791	22	-56	+1	0
6	1854	18	+7	-3	0
7	1960	21	+113	+0	1/2
8	1723	20	-124	-1	1
9	1898	21	+51	+0	1/2
10	1834	20	-13	-1	1
11	1946	24	+99	+3	1
12	1755	19	-92	-2	1
13	1846	21	+1	+0	1/2
14	1752	17	-95	-4	1
15	1884	20	+37	-1	0

$$r_F = \frac{2 \cdot 10,5 - 15}{15} = \frac{7}{15} = 0,4.$$

Der Fechnersche Korrelationskoeffizient ist also recht groß, da eindeutig eine positive Korrelation zwischen den beiden Merkmalen vorhanden ist.

Zum Vergleich wollen wir noch den Korrelationskoeffizienten nach Bravais-Pearson berechnen:

$$r_{XY} = \frac{s_{XY}}{\sqrt{s_X^2 \cdot s_Y^2}} = \frac{120,28492}{\sqrt{5755,0666 \cdot 6}} = \frac{120,28492}{185,82357} = 0,648.$$

Der Fechnersche Korrelationskoeffizient, der die Größe der Abstände ja nicht berücksichtigt, hat in diesem Fall sogar einen kleineren Wert als r_{XY} .

9.4. Der Spearmansche Rangkorrelationskoeffizient

Bei ordinal skalierten Merkmalen kann man den Korrelationskoeffizienten nach Bravais-Pearson nicht verwenden. Um ein Zusammenhangsmaß für ordinal skalierte Merkmale zu erhalten, nutzt man aus, daß die Ausprägungen solcher Merkmale eine eindeutige Rangfolge haben. Deshalb muß der Begriff der **Rangzahl** eingeführt werden. Dazu betrachtet man die Ausprägungen z_1, \dots, z_n eines Merkmals Z und ordnet sie der Größe nach an (hier von der größten zur kleinsten). Die geordnete Reihe bezeichnet man mit $z_{(1)}, z_{(2)}, \dots, z_{(n)}$. Wenn alle $z_{(1)}, \dots, z_{(n)}$ verschieden sind, numeriert man sie von 1 bis n durch und definiert die Rangzahl von $z_{(i)}$ als

$$R(z_{(i)}) = i \quad \text{für } i = 1, \dots, n.$$

Tritt eine Ausprägung mehrmals auf, so ordnet man diesen gleichen Ausprägungen als Rang das arithmetische Mittel der Ränge zu, die sie einnehmen: Man erhält z. B. für die Beobachtungsreihe $z_{(1)} = 8, z_{(2)} = 5, z_{(3)} = 5, z_{(4)} = 2$ die Ränge $R(z_{(1)}) = 1, R(z_{(2)}) = R(z_{(3)}) = 2,5, R(z_{(4)}) = 4$.

Für zweidimensionale ordinal skalierte Merkmale (X, Y) mit an den n Untersu-

chungseinheiten beobachteten Ausprägungen (x_i, y_i) für $i = 1, \dots, n$ kann man dann den **Spearmanschen Rangkorrelationskoeffizienten** r_s berechnen, der dem gewöhnlichen Korrelationskoeffizienten (nach Bravais-Pearson) r_{XY} der Rangzahlen entspricht.

Es gilt also:

$$r_s = \frac{\sum_{i=1}^n R(x_i)R(y_i) - n\overline{R(x)}\overline{R(y)}}{\sqrt{\left(\sum_{i=1}^n R(x_i)^2 - n\overline{R(x)}^2\right)\left(\sum_{i=1}^n R(y_i)^2 - n\overline{R(y)}^2\right)}}.$$

Wenn man r_s umformt, erhält man mit $d_i = R(x_i) - R(y_i)$ die einfache Formel

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \text{ falls alle } x_i \text{ und } y_i \text{ verschieden sind.}$$

Beispiel: Ein Professor hat in einem Semester Daten gesammelt und dabei festgestellt, daß die beste Klausur vom Studenten mit den meisten Übungsaufgabenpunkten und die schlechteste Klausur vom Studenten mit den wenigsten Übungsaufgabenpunkten geschrieben wurde, vgl. Tab. 25. Er vermutet deshalb einen Zusammenhang zwischen Punkten in der Klausur X und in den Übungsaufgaben Y.

Tab. 25: Klausur- und Übungspunkte von 10 Studenten

Student	A	B	C	D	E	F	G	H	I	J
Klausurpunkte	76	44	32	53	25	58	26	59	29	65
Rang der Klausur	1	6	7	5	10	4	9	3	8	2
Übungsaufgabenpunkte	122	67	68	101	42	59	118	79	83	89
Rang der Übungsaufgaben	1	8	7	3	10	9	2	6	5	4
d_i	0	-2	0	2	0	-5	7	-3	3	-2
d_i^2	0	4	0	4	0	25	49	9	9	4

Es ist

$$r_s = 1 - \frac{6 \cdot 104}{10(100 - 1)} = 0,37$$

dann ein Maß für den Zusammenhang zwischen Übungsaufgabenpunkten und Klausurergebnis.

Der Spearmansche Korrelationskoeffizient nimmt nicht nur bei streng linearem Zusammenhang zwischen X und Y den Wert 1 an, sondern auch dann, wenn die Beobachtungen monoton wachsend sind, d.h. wenn für alle $(x_i, y_i), (x_j, y_j)$ gilt:

mit $x_i < x_j$ ist auch $y_i < y_j$,

da er ja nur auf den Rangzahlen, nicht aber auf tatsächlich interpretierbaren Zahlenabständen beruht.

9.5. Der Kendallsche Rangkorrelationskoeffizient

Ein anderes Maß für die Korrelation, das man mit Hilfe der Rangzahlen ermittelt, ist **Kendalls τ** :

Wir ordnen zunächst den n Ausprägungen der Merkmale X und Y jeweils wie vorne beschrieben Rangzahlen zu. Dann ordnen wie die Paare $(x_1, y_1), \dots, (x_n, y_n)$ so, daß die Rangzahlen des Merkmals X der Größe nach geordnet sind und zwar von der kleinsten zur größten, wodurch auch eine eindeutige Reihenfolge der Rangzahlen von Y festgelegt ist. Dann bestimmen wir für jede Rangzahl $R(y_i)$ die Anzahl der Rangzahlen von Y , die kleiner oder gleich $R(y_i)$ sind und in der Anordnung rechts von $R(y_i)$ stehen. Für $R(y_i)$ bezeichnen wir diese Anzahl mit Q_i . Dann ergibt sich Kendall's τ zu

$$\tau = 1 - \frac{4 \sum_{i=1}^n Q_i}{n(n-1)};$$

τ liegt wie r_{XY} , r_F und r_s zwischen -1 und 1 .

Diese Vorgehensweise soll nun an einem Beispiel erläutert werden.

Beispiel: Es liegen die gleichen Daten wie beim letzten Beispiel vor, vgl. Tab. 25.

Wir ordnen zunächst die Ausprägungen von X (Klausurpunkte) der Größe nach und vergeben die Ränge. Das gleiche machen wir mit den Ausprägungen von Y (Übungspunkte) und ordnen dann die Merkmalspaare nach den Rangzahlen von X . Nun berechnet man die Q_i , vgl. Tab. 26.

Tab. 26: Arbeitstabelle zur Berechnung von Kendall's τ im Beispiel

Klausurpunkte (geordn.)	$x_{(1)} = 76$	$x_{(2)} = 65$	$x_{(3)} = 59$	$x_{(4)} = 58$	$x_{(5)} = 53$	$x_{(6)} = 44$	$x_{(7)} = 32$	$x_{(8)} = 29$	$x_{(9)} = 26$	$x_{(10)} = 25$
Rangzahl der Klau.- Punkte	1	2	3	4	5	6	7	8	9	10
zugehörige Übungspunkte	$= y_{(1)} = 122$	$= y_{(4)} = 89$	$= y_{(6)} = 79$	$= y_{(9)} = 59$	$= y_{(3)} = 101$	$= y_{(8)} = 67$	$= y_{(7)} = 68$	$= y_{(5)} = 83$	$= y_{(2)} = 118$	$= y_{(10)} = 42$
Rangzahl der Üb.-Punkte	1	4	6	9	3	8	7	5	2	10
Q_1, \dots, Q_{10}	0	2	3	5	1	3	2	1	0	0

In Tab. 26 ist z. B. $Q_4 = 5$, weil rechts von der Rangzahl 9 von Y fünf Rangzahlen von Y stehen, die kleiner als 9 sind, nämlich 3, 8, 7, 5 und 2.

Es ist also nach Tab. 26 gerade

$$\tau = 1 - \frac{4 \sum_{i=1}^{10} Q_i}{n(n-1)} = 1 - \frac{4 \cdot 17}{10 \cdot 9} = 1 - \frac{68}{90} = 0,244,$$

d. h. die Abhangigkeit zwischen Klausurpunkten und Ubungspunkten ist nicht sehr gro.

9.6. Der Yulesche Assoziationskoeffizient fur die Vierfeldertafel

Es fehlt nun noch der Fall, da (X, Y) nominal skaliert ist. Um ein Ma fur den Zusammenhang von X und Y zu finden, benotigt man die Haufigkeitsverteilung von (X, Y).

Wir wollen uns hier darauf beschranken, da X und Y jeweils genau 2 Auspragungen a_1, a_2 bzw. b_1, b_2 haben konnen. Die Kontingenztafel nennt man in diesem Fall **Vierfeldertafel**, sie hat die Form der **Tab. 27**.

Tab. 27: Vierfeldertafel

X \ Y	b_1	b_2	Σ
a_1	H_{11}	H_{12}	$H_{1 \cdot}$
a_2	H_{21}	H_{22}	$H_{2 \cdot}$
Σ	$H_{\cdot 1}$	$H_{\cdot 2}$	n

Als Ma fur die Abhangigkeit von X und Y kann man dann den **Yuleschen Assoziationskoeffizienten** verwenden, der durch

$$A_{XY} = \frac{H_{11}H_{22} - H_{12}H_{21}}{H_{11}H_{22} + H_{12}H_{21}} = \frac{h_{11}h_{22} - h_{12}h_{21}}{h_{11}h_{22} + h_{12}h_{21}}$$

definiert ist. Es ist $-1 \leq A_{XY} \leq 1$, und falls ein $H_{ij} = 0$ ist, so gilt: $|A_{XY}| = 1$.

Das Vorzeichen des Assoziationskoeffizienten ist nur in Verbindung mit der Vierfeldertafel interpretierbar.

Beispiel: 100 Personen (50 Frauen, 50 Manner) im Alter von 40–50 Jahren wurden befragt, ob sie Abitur haben oder nicht.

Dann kann man die Vierfeldertafeln aus **Tab. 28** und **29** verwenden:

Es ist

$$A_{XY} = \frac{7 \cdot 32 - 43 \cdot 18}{7 \cdot 32 + 43 \cdot 18} = -0,551 \quad \text{bei der Tafel aus Tab. 28 und}$$

$$A_{XY} = \frac{43 \cdot 18 - 7 \cdot 32}{43 \cdot 18 + 7 \cdot 32} = +0,551 \quad \text{bei der Tafel aus Tab. 29.}$$

Tab. 28: 1. Vierfeldertafel im Beispiel

	Frauen	Manner	Σ
Abitur	7	18	25
kein Abitur	43	32	75
Σ	50	50	100

Tab. 29: 2. Vierfeldertafel im Beispiel

	Männer	Frauen	\sum
Abitur	18	7	25
kein Abitur	32	43	75
\sum	50	50	100

Den ersten Assoziationskoeffizienten könnte man dann so interpretieren, daß es einen negativen Zusammenhang zwischen Frauen und Abitur gibt, und den zweiten so, daß es einen positiven Zusammenhang zwischen Männern und Abitur gibt.

Beispiel: Bei psychologischen Tests führt man oft denselben Test zweimal durch (**Retestmethode**), um den „Grad der Genauigkeit“ zu bestimmen, mit dem ein Persönlichkeits- oder Verhaltensmerkmal „gemessen“ wird. Wenn man die Versuchspersonen bzgl. eines nominal skalierten Merkmals mit Ausprägungen A und B untersucht, kann man mit Hilfe des Assoziationskoeffizienten ein Maß für diese Genauigkeit erhalten.

Im Idealfall, also wenn beide Tests die gleichen Ergebnisse liefern, ergibt sich die Vierfeldertafel aus **Tab. 30** und somit $A_{XY} = 1$.

Tab. 30: Vierfeldertafel der Retest-Methode im Idealfall

Test 1			
		A	B
Test 2	A	H_{11}	0
	B	0	H_{22}

Auch wenn nur im Feld (A, B) eine Null steht, ergibt sich $A_{XY} = 1$. Der Yulesche Assoziationskoeffizient ist deshalb in diesem Fall kein gutes Maß für den Zusammenhang.

Weitere Zusammenhangsmaße für nominal skalierte Merkmale findet man in Kap. VII.

10. Praktische Berechnung einiger Kenngrößen

10.1. Berechnung des arithmetischen Mittels und der Standardabweichung

Bei der Berechnung des arithmetischen Mittels und der Standardabweichung gibt es verschiedene Möglichkeiten, sich die Rechnung zu erleichtern.

Bestehen die einzelnen Werte aus wenigen Ziffern oder steht ein Rechner zur Verfügung, so wird man das arithmetische Mittel in üblicher Weise aus

$$\bar{x} = \frac{1}{n} (x_1 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

und die Standardabweichung aus

$$s = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right)}$$

berechnen. Hat man \bar{x} und s schon so aus n Elementen bestimmt und stellt dann fest, daß noch ein weiterer $(n+1)$ -ter Wert berücksichtigt werden muß, so sind folgende Regeln nützlich:

$$\begin{aligned}\bar{x}_{n+1} &= \frac{1}{n+1} (x_{n+1} + n \cdot \bar{x}), \\ s_{n+1} &= \sqrt{(n+1) \cdot (\bar{x}_{n+1} - \bar{x})^2 + \frac{n-1}{n} \cdot s^2} \\ &= \sqrt{\frac{1}{n+1} (x_{n+1} - \bar{x})^2 + \frac{n-1}{n} \cdot s^2}.\end{aligned}$$

Bei vielstelligen Einzelwerten kann man sich die Arbeit erleichtern, wenn man zunächst einen im Grunde genommen beliebigen Wert f von den Einzelwerten subtrahiert. Dieser Wert sollte möglichst so gewählt werden, daß alle Einzelwerte nach Subtraktion von f klein sind. Es ist dann

$$\begin{aligned}\bar{x} &= f + \frac{1}{n} \sum_{i=1}^n (x_i - f) \quad \text{und} \\ s &= \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n (x_i - f)^2 - n \cdot (\bar{x} - f)^2 \right)}.\end{aligned}$$

Beispiel: Wir wollen arithmetisches Mittel und Standardabweichung der in Tab. 31 zusammengestellten Einzelwerte berechnen. Da alle Zahlen in der Größenordnung 15000 liegen, subtrahieren wir zunächst $f = 15000$ von den Einzelwerten.

Tab. 31: Werte zum Beispiel: \bar{x} und s bei vielstelligen Einzelwerten

i	x_i	$x_i - f$
1	15146	146
2	15128	128
3	15033	33
4	15048	48
5	15117	117
6	15098	98
7	15057	57
8	15133	133
9	15029	29
10	15110	110

Es ist nun

$$\begin{aligned}\bar{x} &= f + \frac{1}{n} \sum_{i=1}^n (x_i - f) = 15000 + \frac{1}{10} \sum_{i=1}^{10} (x_i - f) = 15000 + \frac{899}{10} \\ &= 15089,9\end{aligned}$$

und

$$\begin{aligned}
 s &= \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 - n \cdot (\bar{x} - f)^2 \right)} = \sqrt{\frac{1}{9} (98265 - 80820,1)} \\
 &= \sqrt{1938,3222} \\
 &= 44,026.
 \end{aligned}$$

Sind die Einzelwerte so beschaffen, daß viele Stellen hinter dem Dezimalpunkt stehen, so kann man zunächst mit einer geeigneten Potenz von 10 (etwa $10^1 = 10$, $10^2 = 100$ oder $10^3 = 1000$) multiplizieren, denn ist

$$\begin{aligned}
 y_i &= x_i \cdot 10^k, \quad \text{so ist} \\
 \bar{x} &= \bar{y} \cdot 10^{-k} \quad \text{und} \quad s = s_y \cdot 10^{-k}.
 \end{aligned}$$

Möchte man gar von den Einzelwerten x_i zuerst eine Zahl subtrahieren und dann diese neue Zahl noch durch eine andere Zahl dividieren, möchte man sie also einer linearen Transformation

$$y_i = \frac{1}{k_1} (x_i - k_2)$$

unterziehen, um eine möglichst gute Größenordnung zum Rechnen zu erhalten, so kann man dies tun und dann zunächst \bar{y} und s_y in gewohnter Weise berechnen. Für die ursprünglichen Werte x_i gilt dann

$$\bar{x} = k_1 \cdot \bar{y} + k_2 \quad \text{und} \quad s = k_1 \cdot s_y.$$

Ist die Anzahl der zu berücksichtigenden Einzelwerte sehr groß (etwa $n > 25$), so ist es zweckmäßig, die Daten in Klassen einzuteilen. Diese Klassen sollten nun möglichst gleich groß sein. Sind sie es nicht, so können wir \bar{x} und s nach untenstehenden Formeln approximativ berechnen; vgl. Abschnitt 4.1 und 5.4.1:

$$\begin{aligned}
 \bar{x} &= \frac{1}{n} \sum_{j=1}^p m_j \cdot H_j, \\
 s &= \sqrt{\frac{1}{n-1} \left(\sum_{j=1}^p m_j^2 \cdot H_j - n \cdot \bar{x}^2 \right)}.
 \end{aligned}$$

Dabei bezeichnet p die Anzahl der Klassen, m_j die Klassenmitte der j -ten Klasse und H_j die Anzahl der Einzelwerte in der j -ten Klasse.

Beispiel: Wir wollen arithmetisches Mittel \bar{x} und Standardabweichung s der gruppierten Einzelwerte aus Tab. 32 bestimmen.

Tab. 32: Gruppierte Einzelwerte zur Berechnung von \bar{x} und s

j	Klassengrenzen	m_j	H_j
1	[0; 1)	0,5	6
2	[1; 3)	2,0	11
3	[3; 4)	3,5	13
4	[4; 7)	5,5	8
5	[7; 8)	7,5	7
6	[8; 10)	9,0	10

Es ergibt sich

$$\bar{x} = \frac{1}{55} \sum_{j=1}^6 m_j \cdot H_j = \frac{1}{55} \cdot 257 = 4,673 \quad \text{und}$$

$$s = \sqrt{\frac{1}{54} \left(\sum_{j=1}^6 m_j^2 \cdot H_j - 55 \cdot \bar{x}^2 \right)} = \sqrt{\frac{1}{54} (1650,5 - 1200,891)} \\ = 2,886.$$

Natürlich könnte man auch hier wieder von den Klassenmitteln, falls diese sehr viele Stellen haben, einen festen Betrag d subtrahieren:

$$\bar{x} = d + \frac{1}{n} \sum_{j=1}^p (m_j - d) H_j,$$

$$s = \sqrt{\frac{1}{n-1} \left(\sum_{j=1}^6 (m_j - d)^2 H_j - n (\bar{x} - d)^2 \right)}.$$

Kann man erreichen, daß alle p Klassen gleich breit sind, so wählt man zunächst eine beliebige – möglichst jedoch in der Mitte gelegene und stark besetzte – Klasse aus und gibt dieser die Nummer 0. Die Klassen oberhalb von dieser Klasse erhalten in aufsteigender Folge die Nummern 1, 2, 3, ..., die Klassen unterhalb erhalten in absteigender Folge die Nummern $-1, -2, \dots$. Zur Berechnung von \bar{x} und s berechnet man dann die Hilfsgrößen

$$S_1 = \sum_i H_i \cdot i \quad \text{und} \quad S_2 = \sum_i H_i \cdot i^2$$

und erhält daraus

$$\bar{x} = \frac{b}{n} S_1 + c \quad \text{sowie}$$

$$s = \sqrt{\frac{b^2}{n-1} (S_2 - \frac{1}{n} S_1^2)}.$$

Dabei bezeichnet b die überall gleiche Klassenbreite und c die Klassenmitte der mit 0 bezifferten Klasse. Da dieses Verfahren die Varianz überschätzt, ist es hier sinnvoll, die **Sheppardsche Korrektur** zu verwenden:

$$s' = \sqrt{s^2 - \frac{b^2}{12}}.$$

Beispiel: Für die gruppierten Daten aus Tab. 33 wollen wir arithmetisches Mittel und Standardabweichung bestimmen.

Die Klassenbreite b ist hier 2 und die Klassenmitte c der Klasse 0 ist 7. Es ergibt sich mit

$$S_1 = \sum_i H_i \cdot i = 3 \cdot (-3) + 6 \cdot (-2) + 5 \cdot (-1) + 12 \cdot 0 + 9 \cdot 1 + 3 \cdot 2 \\ + 7 \cdot 3 + 5 \cdot 4 \\ = -9 - 12 - 5 + 9 + 6 + 21 + 20 \\ = 30$$

und

Tab. 33: Beispiel zur Berechnung von \bar{x} und s bei gruppierten Daten mit gleicher Klassenbreite $b = 2$

Klassennr. i	Klassengrenzen	H_i
-3	[0; 2)	3
-2	[2; 4)	6
-1	[4; 6)	5
0	[6; 8)	12
1	[8; 10)	9
2	[10; 12)	3
3	[12; 14)	7
4	[14; 16)	5

$$\begin{aligned} S_2 &= \sum_i H_i \cdot i^2 = 3 \cdot 9 + 6 \cdot 4 + 5 \cdot 1 + 12 \cdot 0 + 9 \cdot 1 + 3 \cdot 4 + 7 \cdot 9 + 5 \cdot 16 \\ &= 27 + 24 + 5 + 9 + 12 + 63 + 80 \\ &= 220 \end{aligned}$$

hier nun

$$\bar{x} = \frac{b}{n} \cdot S_1 + c = \frac{2}{50} \cdot 30 + 7 = 8,2 \quad \text{sowie}$$

$$s = \sqrt{\frac{b^2}{n-1} \cdot (S_2 - \frac{1}{n} S_1^2)} = \sqrt{\frac{4}{49} \cdot \left(220 - \frac{900}{50}\right)} = 4,061.$$

Verwenden wir die Sheppardsche Korrektur, so ergibt sich

$$s' = \sqrt{s^2 - \frac{b^2}{12}} = \sqrt{16,49 - \frac{4}{12}} = 4,020$$

als korrigierte Standardabweichung.

10.2. Berechnung der mittleren absoluten Abweichung vom Median

Wollen wir die durchschnittliche Abweichung d von n Einzelwerten vom Median $\tilde{x}_{0,5}$ berechnen, so können wir uns diese Arbeit sehr einfach machen. Wir bestimmen zunächst

$$n^+ = \text{Anzahl der Werte } > \tilde{x}_{0,5},$$

$$S^+ = \text{Summe der Werte } > \tilde{x}_{0,5}$$

$$n^- = \text{Anzahl der Werte } < \tilde{x}_{0,5} \quad \text{und}$$

$$S^- = \text{Summe der Werte } < \tilde{x}_{0,5}.$$

Dann ist die durchschnittliche Abweichung d vom Median gerade

$$d = \frac{1}{n} [(S^+ - S^-) - (n^+ - n^-) \tilde{x}_{0,5}].$$

Beispiel: Wir wollen für die 17 Einzelwerte

$$\begin{aligned} x_1 &= 37, x_2 = 25, x_3 = 28, x_4 = 33, x_5 = 42, x_6 = 51, \\ x_7 &= 39, x_8 = 17, x_9 = 23, x_{10} = 39, x_{11} = 56, x_{12} = 46, \\ x_{13} &= 57, x_{14} = 21, x_{15} = 35, x_{16} = 12 \text{ und } x_{17} = 34 \end{aligned}$$

die mittlere Abweichung vom Median als Maß der Streuung bestimmen. Der Median ist hier

$$\tilde{x}_{0,5} = x_{15} = 35.$$

Weiter ist

$$n^+ = 8, \quad n^- = 8,$$

$$S^+ = x_1 + x_5 + x_6 + x_7 + x_{10} + x_{11} + x_{12} + x_{13} \\ = 37 + 42 + 51 + 39 + 39 + 56 + 46 + 57 = 367 \quad \text{und}$$

$$S^- = x_2 + x_3 + x_4 + x_8 + x_9 + x_{14} + x_{16} + x_{17} \\ = 25 + 28 + 33 + 17 + 23 + 21 + 12 + 34 = 193.$$

Damit ergibt sich

$$d = \frac{1}{n} [(S^+ - S^-) - (n^+ - n^-) \tilde{x}_{0,5}] = \frac{1}{17} [(367 - 193) - (8 - 8) 35] \\ = \frac{1}{17} \cdot 174 = 10,235$$

als durchschnittliche Abweichung vom Median.

Will man die durchschnittliche Abweichung vom Median aus gruppierten Daten bestimmen, so muß man zunächst die Klasse j ermitteln, in der der Median $\tilde{x}_{0,5}$ liegt und dann die Stelle i des Medians in der Gruppe bestimmen. Der Median berechnet sich dann zu

$$\tilde{x}_{0,5} = \text{untere Gruppengrenze der Mediengruppe} \\ + \frac{\text{Stelle des Medians in der Gruppe}}{\text{Anzahl der Elemente in der Gruppe}} \cdot \text{Gruppenbreite} \\ = u_j + \frac{i}{H_j} \cdot b_j.$$

Um nun d zu berechnen, unterteilt man die Mediengruppe j in zwei Klassen j_1 und j_2 und zwar so, daß die obere Klassengrenze von j_1 und die untere Klassengrenze von j_2 gerade der Median ist. Ist die Gesamtzahl von Werten gerade, so gehört der Median zur Gruppe j_1 , ist sie ungerade, so wird von da an mit $n - 1$ Werten weitergerechnet; der Median selbst entfällt bei der Berechnung. Die durchschnittliche Abweichung vom Median kann man dann aus

$$d = \frac{1}{n} (S^+ - S^-) \quad \text{bzw.} \quad d = \frac{1}{n-1} (S^+ - S^-)$$

bestimmen, wobei die Hilfsgrößen

$$S^- = \sum_{j=1}^{j_1} m_j \cdot H_j \quad \text{und} \quad S^+ = \sum_{j=j_2}^p m_j \cdot H_j$$

sind. Dabei bezeichnet m_j die Klassenmitte der j -ten Gruppe.

Beispiel: Wir wollen nun die Berechnung von d aus gruppierten Daten an einem Beispiel vorführen, vgl. Tab. 34.

Nun müssen wir zunächst den Median bestimmen. Die Anzahl der Beobachtungen

Tab. 34: Daten für die Berechnung von d

Klassennr. j	Klassengrenzen	H_j
1	(3; 6]	16
2	(6; 10]	19
3	(10; 12]	12
4	(12; 16]	15
5	(16; 18]	14
6	(18; 19]	17
7	(19; 23]	12

ist $n = 105$, d. h. der Median ist der 53. Wert. Er liegt also in der Klasse 4 und zwar an 6. Stelle. Die Klasse 4 hat die Breite

$$b_4 = 16 - 12 = 4$$

und somit ergibt sich der Median zu

$$\tilde{x}_{0,5} = u_j + \frac{i}{H_j} \cdot b_j = 12 + \frac{6}{15} \cdot 4 = 13,6.$$

Nun erstellen wir die Arbeitstabelle, s. **Tab. 35**.

Tab. 35: Arbeitstabelle zur Berechnung von d

Klassennr. j	Klassengrenzen	Klassenmitten m_j	H_j
1	(3; 6]	4,5	16
2	(6; 10]	8,0	19
3	(10; 12]	11,0	12
4 ₁	(12; 13,6]	12,8	5
4 ₂	(13,6; 16]	14,8	9
5	(16; 18]	17,0	14
6	(18; 19]	18,5	17
7	(19; 23]	21,0	12

Wir haben hier die 4. Klasse (Mediangruppe) gespalten in die Klassen 4₁ und 4₂. Da die Zahl $n = 105$ der Beobachtungen ungerade ist, haben wir den Median selbst weggelassen. Hier rechnen wir also mit 104 Werten weiter. Die Hilfsgrößen ergeben sich zu

$$S^- = \sum_{j=1}^{4_1} m_j \cdot H_j = 4,5 \cdot 16 + 8 \cdot 19 + 11 \cdot 12 + 12,8 \cdot 5 = 420 \quad \text{und}$$

$$S^+ = \sum_{j=4_2}^7 m_j \cdot H_j = 14,8 \cdot 9 + 17 \cdot 14 + 18,5 \cdot 17 + 21 \cdot 12 = 937,7.$$

Damit erhalten wir

$$d = \frac{1}{n-1} (S^+ - S^-) = \frac{1}{104} (937,7 - 420) = 4,978.$$

Kapitel II: Grundbegriffe der Wahrscheinlichkeitsrechnung

1. Ereignisse und Zufallsexperimente

Die Wahrscheinlichkeitsrechnung beschäftigt sich mit zufälligen Ereignissen und Erscheinungen. Die zugrundeliegenden Vorgänge sind dabei selbst von zufälliger Natur oder aber so komplex, daß wir nicht willens oder in der Lage sind, sie quantitativ zu erfassen. So läuft ein Münzwurf zwar nach wohlbekannten physikalischen Gesetzmäßigkeiten ab, doch sehen wir uns außerstande, das Ergebnis vorherzusagen. Dagegen ist das Ergebnis der in einem Kaufhaus am Jahresende stattfindenden Inventur prinzipiell durch eine detaillierte Erfassung sämtlicher Ein- und Verkäufe möglich. Der damit verbundene Aufwand läßt eine solche Vorgehensweise jedoch im allgemeinen nicht zu. Bei klinischen Untersuchungen kommt der Zufall oft bei der Auswahl der Versuchspersonen ins Spiel. Das ist z. B. auch so bei Meinungsumfragen und entsprechend bei der stichprobenhaften Kontrolle von Werkstücken in einem Industriebetrieb. Alles, was wir hier erwähnt haben, sind Vorgänge, die, zumindest im Prinzip, beliebig oft wiederholt werden können und jedesmal nach einer bestimmten Vorschrift ausgeführt werden, wobei die jeweiligen Ergebnisse vom Zufall abhängen, d. h. nicht im voraus bestimbar sind. Solch ein Vorgang wird **Zufallsexperiment** genannt. Die Menge aller möglichen Ergebnisse eines Zufallsexperimentes nennt man **Grundraum** oder **Ereignisraum**. Es ist üblich, dafür das Symbol Ω zu benutzen.

(**Zufällige**) **Ereignisse** sind Teilmengen von Ω und werden mit großen lateinischen Buchstaben bezeichnet. Die einelementigen Ereignisse nennt man auch **Elementarereignisse**. Man sagt, „das Ereignis A ist eingetreten“, wenn das Ergebnis des Zufallsexperimentes ein Element von A ist.

Beispiel:

- (a) Eine Münze wird geworfen. Die möglichen Ergebnisse dieses Zufallsexperimentes sind „Kopf“ und „Zahl“. Der Ereignisraum ist $\Omega = \{\text{Kopf}, \text{Zahl}\}$. Die einzigen Elementarereignisse sind $\{\text{Kopf}\}$ und $\{\text{Zahl}\}$.
- (b) Ein Würfel wird solange geworfen, bis zum ersten Mal drei Sechsen hintereinander kommen. Wir interessieren uns für die Anzahl der hierfür notwendigen Würfe; damit sind die Ergebnisse dieses Zufallsexperimentes die Zahlen 3, 4, 5, 6, 7, ... usw. Der Grundraum ist $\Omega = \{n : n \text{ natürliche Zahl } \geq 3\}$. Im Unterschied zu (a) ist der Grundraum hier unendlich und zwar – wie man sagt – abzählbar unendlich.
- (c) Durch die zufällige Auswahl einer Glühbirne und Bestimmung ihrer Lebensdauer ist ein Zufallsexperiment gegeben. Da als Lebensdauer jede reelle positive Zahl in Frage kommt, gilt $\Omega = \{x | x \geq 0\}$. Das Ereignis A, daß die Glühbirne höchstens 200 Stunden brennt, ist das Intervall $[0; 200]$.

Bei Zufallsexperimenten kann man nun an einem einzigen Ereignis interessiert sein, wie z. B. an 6 Richtigen im Lotto, aber auch an einer Vielzahl von Ereignissen, die zudem miteinander in Beziehung stehen können. So interessiert etwa, ob bei einem Fahrrad eine Reifenpanne auftritt, ob das Vorderrad dabei einen Platten hat oder das Hinterrad, oder gar beides der Fall ist. Alle diese Ereignisse lassen sich natürlich durch Teilmengen eines geeigneten Grundraumes Ω beschreiben. Beziehungen der

Ereignisse untereinander werden deutlich, wenn man sich der Mengenoperationen wie Vereinigungs-, Durchschnittsbildung usw. bedient. So ist z. B.

- $A \cup B$ das Ereignis, daß mindestens eines der beiden Ereignisse A und B eintritt;
- $A \cap B$ ist das Ereignis, daß sowohl A als auch B eintritt und
- $A - B$ ist das Ereignis, daß A aber nicht B eintritt.
- $\bar{A} = \Omega - A$ bedeutet, daß A nicht eintritt und heißt das zu A **komplementäre Ereignis**.

Mengendiagramme wie in Abb. 1 veranschaulichen uns dies.

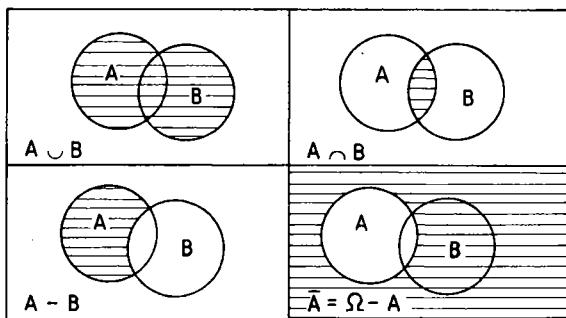


Abb. 1: Mengendiagramme

Bezeichnet hier A den Platten am Vorderrad und B den Platten am Hinterrad, dann beschreibt also $A \cup B$ gerade das Ereignis, daß überhaupt eine Panne auftritt, und $A \cap B$ bedeutet, daß beide Reifen platt sind. Wie man sieht, können auf diese Art und Weise eine Vielzahl von Ereignissen und deren Beziehungen untereinander beschrieben werden, und das sogar ohne sich der oft großen Mühe zu unterziehen, einen Grundraum Ω explizit anzugeben.

Beispiel: Bei der Produktion von Schrauben soll geprüft werden, ob die geforderte Länge, die zwischen 1,9 und 2,1 cm liegen soll, eingehalten wird. Das Zufallsexperiment besteht darin, eine Schraube zufällig auszusuchen und deren Länge zu messen. Sei A das Ereignis, daß die Schraube kürzer als 1,9 cm ist, und B das Ereignis daß die Schraube länger als 2,1 cm ist. Dann ist $A \cup B$ das Ereignis, daß die Schraube nicht die geforderte Länge hat und $A \cap B$ das Ereignis, daß die geforderte Länge eingehalten wird. Bezeichnet C das Ereignis, daß die Länge der Schraube mindestens 2 cm ist, so gibt uns $C \cap A \cup B$ das Ereignis an, daß die Länge der Schraube mindestens 2 cm und höchstens 2,1 cm ist.

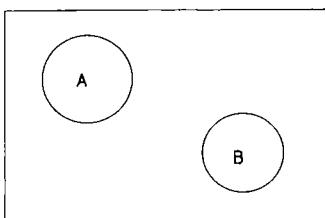


Abb. 2: Die Mengen A und B sind disjunkt

Den Grundraum Ω , der ja alle überhaupt möglichen Ereignisse enthält, nennt man auch das **sichere Ereignis** und dessen Komplement die leere Menge $\emptyset = \bar{\Omega}$ das **unmögliche Ereignis**. Zwei Ereignisse A und B, wie sie in Abb. 2 dargestellt sind, also mit $A \cap B = \emptyset$, heißen **unvereinbar** oder **disjunkt**. Für die Vereinigung zweier disjunkter Ereignisse schreibt man oft $A + B$ statt $A \cup B$. Die Menge aller Teilmengen von Ω heißt **Potenzmenge von Ω** ; sie wird mit $P(\Omega)$ bezeichnet.

Beispiel:

$$\Omega = \{1, 2, 3\}$$

$$P(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

Wie hier zu sehen ist, sind unter anderem die Elementarereignisse $\{1\}, \{2\}, \{3\}$ und Ω in der Potenzmenge enthalten.

Im folgenden sind noch einige Rechenregeln für Mengenoperationen zusammengestellt:

$$\begin{aligned} A \cup B &= B \cup A \\ A \cap B &= B \cap A \\ (A \cup B) \cup C &= A \cup (B \cup C) \\ (A \cap B) \cap C &= A \cap (B \cap C) \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \\ \underline{(A \cap B)} &= \bar{A} \cup \bar{B} \\ \underline{(A \cup B)} &= \bar{A} \cap \bar{B} \\ A \cap A &= A; A \cup A = A; A \cap \emptyset = \emptyset; A \cup \emptyset = A; \\ A \cap \Omega &= A; A \cup \Omega = \Omega; \\ A - B &= A \cap \bar{B}. \end{aligned}$$

2. Wahrscheinlichkeiten

Bei einem Zufallsexperiment kann man zwar nicht voraussagen, welches Ereignis eintritt, man hält jedoch oft das Eintreten einiger Ereignisse für mehr, anderer für weniger wahrscheinlich. So wird z. B. ein Lottospieler nicht überrascht sein, wenn er am Wochenende feststellt, daß er keine 6 Richtige hat. Der Spieler sagt: „Die Wahrscheinlichkeit, sechs Richtige im Lotto zu haben, ist sehr gering.“ Wie kann man diese „Wahrscheinlichkeit“ nun quantitativ erfassen? Interessieren wir uns einmal beim Würfeln für das Ereignis, daß eine Sechs auftritt. Der Grundraum ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $C = \{6\}$ beschreibt das uns interessierende Ereignis. Führen wir nun das folgende Experiment durch. Der Würfel wird immer wieder geworfen und nach jedem Wurf die Anzahl der bis dahin aufgetretenen Sechsen, also die absolute Häufigkeit $H_n(C)$ des Ereignisses C registriert. Dann kann man praktisch immer feststellen, daß sich die relative Häufigkeit $h_n(C) = H_n(C)/n$ des Ereignisses C bis zum n-ten Wurf, d. h. der Anteil der Sechsen bis dahin, mit wachsendem n bei etwa 1/6 stabilisiert. Das legt uns nahe, diesen Wert als Wahrscheinlichkeit von C zu interpretieren. Wir schreiben kurz

$$P(C) = \frac{1}{6} \quad (P = \text{probability}).$$

Interessiert man sich nur dafür, ob eine gerade Zahl auftritt, also für das Ereignis $G = \{2, 4, 6\}$, so wird man feststellen, daß sich die relative Häufigkeit $h_n(G)$ mit

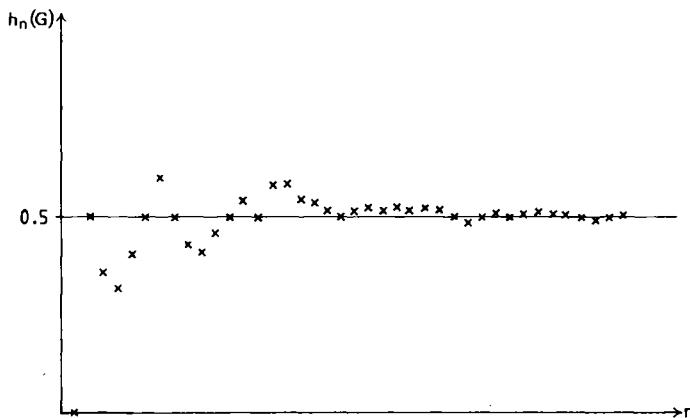


Abb. 3: Relative Häufigkeit eines geradzahligen Ergebnisses beim Würfelwurf

wachsendem n bei 0,5 einpendelt, vgl. auch **Abb. 3**. Damit haben wir für die Wahrscheinlichkeit dieses Ereignisses

$$P(G) = 0,5.$$

Auf diese Art und Weise lässt sich jedem Ereignis eines Zufallsexperimentes, und das nicht nur beim Würfeln, eine Wahrscheinlichkeit zuordnen (R. v. Mises 1931). Solch eine Wahrscheinlichkeit kann man durch folgende drei Eigenschaften, die man entsprechend auch bei den relativen Häufigkeiten findet, und die als **Kolmogoroffsche Axiome** bekannt sind, vollständig charakterisieren:

- (1) $P(A) \geq 0$ für jedes Ereignis $A \subset \Omega$ (**Positivität**)
- (2) $P(\Omega) = 1$ (**Normiertheit**)
- (3) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ für jede Folge (**σ -Additivität**)
paarweise disjunkter Ereignisse $A_i \subset \Omega$.

Die dritte Eigenschaft beinhaltet natürlich auch die **endliche Additivität**, d. h. für disjunkte Ereignisse A_1, \dots, A_n gilt

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (\text{Man setze nur } \emptyset = A_{n+1} = A_{n+2} = \dots \text{ in (3)}).$$

Speziell für nur zwei disjunkte Ereignisse A und B hat man

$$P(A \cup B) = P(A) + P(B).$$

Aus diesen drei charakteristischen Eigenschaften einer Wahrscheinlichkeit lassen sich sämtliche Rechenregeln für Wahrscheinlichkeiten folgern, wie z. B.

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \quad \text{für beliebiges } A \text{ und } B, \\ P(\bar{A}) &= 1 - P(A), \\ P(A - B) &= P(A) - P(B), \quad \text{falls } B \subset A \text{ usw.} \end{aligned}$$

Damit kann man, sofern die Wahrscheinlichkeit einiger Ereignisse schon bekannt ist, die Wahrscheinlichkeit vieler weiterer **rechnerisch** bestimmen.

Beispiel:

- (a) Es soll die Wahrscheinlichkeit bestimmt werden, daß beim Würfeln das Ergebnis eine Zwei oder eine Vier ist. Das interessierende Ereignis wird beschrieben durch die Menge

$$\{2, 4\} = \{2, 4, 6\} - \{6\} = G - C,$$

so daß

$$P(\{2, 4\}) = P(G) - P(C) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}.$$

- (b) In langjährigen Beobachtungen hat sich gezeigt, daß 1/85 der Geburten Zwillingssgeburen sind. Man wird daher für die Wahrscheinlichkeit des Ereignisses A einer Zwillingssgeburt 1/85 annehmen. Die Wahrscheinlichkeit, keine Zwillingssgeburt zu haben, ist dann

$$P(\bar{A}) = 1 - P(A) = 1 - \frac{1}{85} = \frac{84}{85}.$$

Es ist klar, daß bei der Bestimmung der Wahrscheinlichkeit eines Ereignisses mittels wiederholter Durchführung eines Zufallsexperimentes Ungenauigkeiten und Fehler auftreten können (vgl. Kap. III). Das ist besonders der Fall, wenn nur eine relativ kleine Anzahl von Versuchen durchgeführt werden kann, etwa aus ökonomischen, ethischen oder sonstigen Gründen. Ist jedoch überhaupt keine Wiederholbarkeit gegeben, wie es z. B. oft bei wirtschaftlichen Prozessen vorkommt, so läßt sich nur eine **subjektive Wahrscheinlichkeit** angeben, die sich etwa auf Meinungen und Einschätzungen von Experten gründen kann (**Delphi-Methode**). Näheres hierzu findet sich z. B. in Bamberg/Baur (1980) und der dort zitierten Literatur. Bei einigen, meist recht einfachen Zufallsexperimenten, kommt man schon allein durch eine Plausibilitätsbetrachtung zu der Auffassung, daß jedes Ergebnis die gleiche Wahrscheinlichkeit haben muß. So wird man aufgrund der Symmetrie eines Würfels jeder Zahl die gleiche Chance einräumen, als Ergebnis eines Wurfes zu erscheinen, und ihr somit die Wahrscheinlichkeit 1/6 zuordnen. Für solche Zufallsexperimente, bei denen also nur endlich viele, gleichwahrscheinliche Ergebnisse möglich sind, ergibt sich für ein beliebiges Ereignis A die Wahrscheinlichkeit

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ „günstigen“ Fälle}}{\text{Anzahl der möglichen Fälle}},$$

wobei $|A|$ die Anzahl der Elemente von A bezeichnet und entsprechend $|\Omega|$ die der Elemente von Ω . Auf diese Weise hat Laplace die Gewinnchancen von Glücksspielen berechnet. Deshalb spricht man hier auch von einer **Laplace-Wahrscheinlichkeit**.

Beispiel:

- (a) Zwei unterscheidbare Würfel werden einmal geworfen. Es soll die Wahrscheinlichkeit bestimmt werden, daß mit beiden Würfeln eine Sechs geworfen wird (Ereignis A).

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (6, 6)\}$$

$$A = \{(6, 6)\}$$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{36}$$

- (b) Wie oben werden zwei unterscheidbare Würfel einmal geworfen. Wie groß ist die Wahrscheinlichkeit, daß mindestens ein Würfel eine Sechs aufzeigt (Ereignis B)?

$$B = \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5)\}$$

$$P(B) = \frac{|B|}{|\Omega|} = \frac{11}{36}.$$

Zur Bestimmung der Laplaceschen Wahrscheinlichkeit bedarf es der Abzählung von unter Umständen sehr großen Mengen. Hierzu bedient man sich häufig der Kombinatorik, die man als die Kunst des Zählens bezeichnen kann.

3. Kombinatorik und Beispiele für die Berechnung von Laplace-Wahrscheinlichkeiten

In diesem Abschnitt werden die folgenden beiden Fragen beantwortet und als Hilfsmittel zur Berechnung von Laplace-Wahrscheinlichkeiten eingesetzt.

1. Wieviele Möglichkeiten gibt es, n Elemente **anzuordnen**?
2. Wieviele Möglichkeiten gibt es, von n Elementen k **auszuwählen**?

Im ersten Fall spricht man von **Permutationen** und im zweiten von **Kombinationen**.

3.1. Permutationen

Bei n Elementen gibt es

$$\frac{n!}{n_1! \cdot \dots \cdot n_r!} = \begin{cases} \text{Anordnungen, wenn alle Elemente verschieden sind und} \\ \text{Anordnungen, wenn nur r Typen existieren, wobei je-} \\ \text{weils } n_i \text{ Elemente vom Typ i sind.} \end{cases}$$

Beispiel: Für die Fertigungsreihenfolge von zehn unterschiedlichen Werkstücken gibt es $10! = 3628800$ verschiedene Möglichkeiten. Haben wir dagegen zehn Werkstücke, von denen drei von der Sorte A, fünf von der Sorte B und zwei von der Sorte C sind, so ist die Anzahl der möglichen Fertigungsreihenfolgen

$$\frac{10!}{3! \cdot 5! \cdot 2!} = \frac{3628800}{6 \cdot 120 \cdot 2} = 2520.$$

Sollen Stücke der gleichen Sorte jeweils hintereinander produziert werden, dann hat man nur noch $3! = 6$ Möglichkeiten.

3.2. Kombinationen

Aus einer Menge von n verschiedenen Elementen sollen k Stücke ausgewählt werden. Handelt es sich dabei um k verschiedene Stücke, man spricht dann von **Kombinationen ohne Wiederholung**, so gibt es dafür

$$\frac{n!}{(n - k)!} \quad \text{Möglichkeiten, wenn die Reihenfolge berücksichtigt wird und}$$

$$\binom{n}{k} = \frac{n!}{k!(n - k)!} \quad \text{Möglichkeiten, wenn die Reihenfolge unberücksichtigt bleibt.}$$

Läßt man dagegen zu, daß Elemente auch mehrfach ausgewählt werden, man spricht hier von **Kombinationen mit Wiederholung**, so sind es

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

n^k Möglichkeiten, wenn die **Reihenfolge berücksichtigt** wird und
Möglichkeiten, wenn die **Reihenfolge unberücksichtigt** bleibt.

Beispiel (Kombinationen ohne Wiederholung):

- (a) Beim Pferderennen gibt es, wenn acht Pferde am Start sind, für die Belegung der ersten drei Plätze $8!/(8-3)! = 336$ Möglichkeiten (Dreiereinlaufwette).
- (b) Beim Lotto hat man $\binom{49}{6} = 13983816$ Möglichkeiten, aus den insgesamt 49 Zahlen sechs Stück zu wählen.
- (c) Würfelt man dreimal hintereinander, so gibt es $\binom{6}{3} = 20$ Möglichkeiten, daß die Augenzahl bei jedem Wurf größer ist als beim Vorhergehenden.

Beispiel (Kombinationen mit Wiederholung):

- (a) Würfelt man viermal hintereinander, so sind insgesamt $6^4 = 1296$ Abläufe möglich.
- (b) Beim Knobeln mit drei Würfeln lassen sich insgesamt $\binom{6+3-1}{3} = 56$ unterschiedliche Figuren erzeugen.
- (c) Hat man vier verschiedene Sorten Bonbons, so gibt es $\binom{4+10-1}{10} = 286$ Möglichkeiten, eine Tüte mit zehn Bonbons zu füllen.

3.3. Beispiele zur Berechnung von Laplace-Wahrscheinlichkeiten

- (a) In einem Raum sind 30 Personen versammelt. Gefragt ist nach der Wahrscheinlichkeit, daß mindestens zwei am selben Tag Geburtstag haben. Der Grundraum Ω besteht aus den Elementarereignissen (Geburtstag der 1. Person), . . . , (Geburtstag der 30. Person).
 $|\Omega| = 365^{30}$ (Kombinationen mit Wiederholung mit Berücksichtigung der Reihenfolge)

Die Anzahl der Möglichkeiten, daß alle an verschiedenen Tagen Geburtstag haben ist

$$|A| = \frac{365!}{(365-30)!} \quad (\text{Kombinationen ohne Wiederholung mit Berücksichtigung der Reihenfolge}).$$

Damit ergibt sich die gesuchte Wahrscheinlichkeit zu

$$1 - \frac{|A|}{|\Omega|} = \frac{365^{30} - (365-30)!}{365^{30}} = 0,706.$$

- (b) Beim Lotto ist die Wahrscheinlichkeit sechs Richtige zu haben

$$\frac{1}{\binom{49}{6}} = \frac{1}{13983816} = 0,000000715.$$

- (c) Es soll die Wahrscheinlichkeit bestimmt werden, daß beim fünfmaligen Würfeln die größte Augenzahl eine Vier ist. Bei den fünf Würfen kommen mit einer Wahrscheinlichkeit $4^5/6^5$ Zahlen kleiner oder gleich vier vor. Die Wahrscheinlichkeit, daß nur Zahlen kleiner als vier vorkommen, beträgt $3^5/6^5$. Wir erhalten somit für die gesuchte Wahrscheinlichkeit

$$\frac{4^5 - 3^5}{6^5} = \frac{4^5 - 3^5}{6^5} = \frac{1024 - 243}{7776} = \frac{781}{7776} = 0,1.$$

- (d) Beim dreimaligen Würfeln gibt es $6!/(6-3)!$ Möglichkeiten, daß alle Würfe verschieden sind. Die Wahrscheinlichkeit dafür ist demnach

$$\frac{6!/(6-3)!}{6^3} = \frac{120}{216} = 0,5556.$$

- (e) Eine Urne enthalte drei weiße, drei schwarze und drei rote Kugeln. Entnehmen wir blind vier Kugeln, so ist die Wahrscheinlichkeit dafür, daß wir von jeder Farbe mindestens eine erhalten, gegeben durch

$$\frac{\binom{3}{2}\binom{3}{1}\binom{3}{1}}{\binom{9}{4}} + \frac{\binom{3}{1}\binom{3}{2}\binom{3}{1}}{\binom{9}{4}} + \frac{\binom{3}{1}\binom{3}{1}\binom{3}{2}}{\binom{9}{4}} = \frac{3 \cdot \binom{3}{2}\binom{3}{1}\binom{3}{1}}{\binom{9}{4}} = \frac{9}{14}.$$

4. Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Oft interessiert man sich für die Wahrscheinlichkeit eines Ereignisses A unter der Annahme, daß ein bestimmtes Ereignis B schon eingetreten ist. Man möchte z. B. wissen, mit welcher Wahrscheinlichkeit der Motor eines Pkws noch 20000 km hält, wenn man schon weiß, daß er bereits 100000 km ohne Störung gelaufen ist, oder wie groß die Wahrscheinlichkeit ist, an einem Herzinfarkt zu sterben, wenn es sich um eine männliche Person handelt, deren Alter mehr als 40 Jahre beträgt, oder aber wie wahrscheinlich es ist, daß das zweite Los ein Gewinn ist, wenn das erste eine Niete war. In solchen Situationen spricht man von **bedingten Wahrscheinlichkeiten**. Wie so eine bedingte Wahrscheinlichkeit nun konkret aussieht, wollen wir an einem **Beispiel** erläutern.

In der Oberstufe eines Gymnasiums wird durch eine Befragung festgestellt, wieviele weibliche bzw. männliche Schüler rauchen. Es ergaben sich dabei die Daten der **Tab. 1**.

Aus diesen Daten können wir z. B. entnehmen, daß die relative Häufigkeit der weiblichen Schüler $h(W) = 55/100$ beträgt, die der Raucher $h(R) = 62/100$, die der weiblichen Raucher $h(W \cap R) = 32/100$ usw. Diese Häufigkeiten sind alle auf die gesamte Schülerzahl von $H(W \cup M) = 100$ Personen bezogen. Interessieren wir uns aber speziell für die weiblichen Schüler, so werden wir uns auf deren Anzahl

Tab. 1: Rauchgewohnheiten nach Geschlecht

Geschlecht		W weiblich	M männlich	
Raucher				
R	ja	32	30	62
N	nein	23	15	38
		55	45	100

$H(W) = 55$ beziehen und erhalten damit die sogenannten **bedingten relativen Häufigkeiten unter der Bedingung W**

$$h(R|W) = \frac{H(R \cap W)}{H(W)} = \frac{32}{55} = \frac{32/100}{55/100} = \frac{h(R \cap W)}{h(W)} \quad \text{und}$$

$$h(N|W) = \frac{H(N \cap W)}{H(W)} = \frac{23}{55} = \frac{23/100}{55/100} = \frac{h(N \cap W)}{h(W)}.$$

Dadurch motiviert, setzen wir die **bedingte Wahrscheinlichkeit eines Ereignisses A unter der Bedingung B** fest als

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{für } P(B) > 0.$$

Beispiel: Ein Relais kann auf zwei verschiedene Arten ausfallen: durch Kurzschluß und durch Leerlauf. Beträgt der Anteil der defekten Relais in einem System 5% und der Anteil der durch Kurzschluß ausgefallenen Relais 3%, so beträgt die Wahrscheinlichkeit, daß ein defektes Relais durch Kurzschluß ausgefallen ist

$$P(\text{Kurzschluß}| \text{defekt}) = \frac{P(\text{defekt durch Kurzschluß})}{P(\text{defekt})} = \frac{0,03}{0,05} = 0,6.$$

Bis jetzt sind wir davon ausgegangen, daß die auf den Grundraum bezogenen Wahrscheinlichkeiten bekannt sind, um damit uns interessierende bedingte Wahrscheinlichkeiten zu berechnen. Oft befinden wir uns aber in der umgekehrten Situation. Mit relativ leicht zu bestimmenden bzw. bekannten bedingten Wahrscheinlichkeiten sollen eigentlich interessierende nicht bedingte Wahrscheinlichkeiten berechnet werden. Dazu werden wir zwei nützliche Hilfsmittel vorstellen, nämlich den **Satz von der totalen Wahrscheinlichkeit** und den **Multiplikationssatz für bedingte Wahrscheinlichkeiten**.

Beginnen wir mit einem **Beispiel**. Ein Produkt wird auf zwei unterschiedlichen Maschinen gefertigt. Die Maschine I fertigt 5000 Stück pro Tag mit einem Ausschuß von 10% und die Maschine II 6000 Stück mit einem Ausschuß von 8%. Wir interessieren uns für die nicht bedingte Wahrscheinlichkeit des Ereignisses D, daß ein zufällig beim Warenausgang ausgewähltes Produkt defekt ist. Bezeichnen A und B die Ereignisse, daß das Produkt von Maschine I bzw. II gefertigt wurde, so sind die bedingten Wahrscheinlichkeiten $P(D|A) = 0,1$ und $P(D|B) = 0,08$ bekannt. Außerdem kennt man

$$P(A) = \frac{5000}{5000 + 6000} = \frac{5}{11} \quad \text{und} \quad P(B) = 1 - P(A) = \frac{6}{11}.$$

Mit A und B sind auch die Ereignisse $D \cap A$ und $D \cap B$ disjunkt. Damit erhält man

$$\begin{aligned} P(D) &= P((D \cap A) \cup (D \cap B)) = P(D \cap A) + P(D \cap B) \\ &= \frac{P(D \cap A)}{P(A)} P(A) + \frac{P(D \cap B)}{P(B)} P(B) \\ &= P(D|A) P(A) + P(D|B) P(B) \\ &= 0,1 \cdot \frac{5}{11} + 0,08 \cdot \frac{6}{11} = 0,089. \end{aligned}$$

Die hier auftauchende Beziehung

$$P(D) = P(D|A) P(A) + P(D|B) P(B)$$

ist ein Spezialfall vom **Satz von der totalen Wahrscheinlichkeit**:

Sind A_1, \dots, A_k paarweise disjunkte Ereignisse mit $\bigcup_{i=1}^k A_i = \Omega$, so gilt für ein beliebiges Ereignis B

$$P(B) = \sum_{i=1}^k P(B|A_i) P(A_i).$$

Beispiel: Die Kundendienstniederlassung eines Herstellers für Haushaltsgeräte beschäftigt drei Mitarbeiter im Außendienst, von denen der erste nur halbtags arbeitet. Außerdem weiß man, daß dieser erste Mitarbeiter in 90% aller Fälle mit einem einzigen Kundenbesuch auskommt, während das bei den anderen beiden in 80% und 75% der Fälle so ist. Welche Wahrscheinlichkeit hat nun das Ereignis B, daß von dieser Firma eine Reparatur mit einer einzigen Fahrt zum Kunden erledigt wird? Bezeichnet A_i das Ereignis, daß der i-te Mitarbeiter kommt, so ist

$$P(B|A_1) = 0,9, \quad P(B|A_2) = 0,8 \quad \text{und} \quad P(B|A_3) = 0,75.$$

Aufgrund der unterschiedlich langen Arbeitszeiten kann man annehmen, daß $P(A_1) = 1/5$ und $P(A_2) = P(A_3) = 2/5$ gilt, so daß

$$\begin{aligned} P(B) &= P(B|A_1) P(A_1) + P(B|A_2) P(A_2) + P(B|A_3) P(A_3) \\ &= 0,9 \cdot \frac{1}{5} + 0,8 \cdot \frac{2}{5} + 0,75 \cdot \frac{2}{5} = 0,8. \end{aligned}$$

Wir haben hier gesehen, wie man durch die getrennte Untersuchung aller alternativen Möglichkeiten unter denen ein Ereignis B auftreten kann, die Wahrscheinlichkeit dieses Ereignisses mit dem Satz von der totalen Wahrscheinlichkeit bestimmt. Eine in gewissem Sinne konträre Vorgehensweise ist, wenn zu diesem Zweck das interessierende Ereignis B durch das gleichzeitige Eintreffen mehrerer Einzelereignisse A_1, \dots, A_k beschrieben wird, um dann die Wahrscheinlichkeit von B mit Hilfe des **Multiplikationssatzes für bedingte Wahrscheinlichkeiten** zu berechnen, der lautet

$$\begin{aligned} P(A_1 \cap \dots \cap A_k) &= P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \cdot \dots \\ &\quad \cdot P(A_k|A_1 \cap \dots \cap A_{k-1}) \end{aligned}$$

Beispiel: Eine Firma produziert Elektrogeräte für den Export mit zwei unterschiedlichen Anschlußsteckern D und E. An einem Tag werden 80 Geräte in ein Land ausgeliefert, für das die Steckverbindung D vorgeschrieben ist. Versehentlich werden fünf Geräte mit der Steckverbindung E ausgeliefert. Wie groß ist die Wahrscheinlichkeit des Ereignisses B, daß ein Händler, der drei Geräte abnimmt nur Geräte mit Steckverbindung D erhält? A_i bezeichnet das Ereignis, daß das i-te Gerät mit Steckverbindung D ausgerüstet ist. Dann ist $B = A_1 \cap A_2 \cap A_3$, und wir berechnen die gesuchte Wahrscheinlichkeit zu

$$\begin{aligned} P(B) &= P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\ &= \frac{75}{80} \cdot \frac{74}{79} \cdot \frac{73}{78} = 0,822. \end{aligned}$$

Die Berechnung solcher Wahrscheinlichkeiten vereinfacht sich bei „unabhängigen“ Ereignissen, wie es z. B. die einzelnen Ergebnisse beim mehrmaligen Würfeln, Fabrikationsfehler an verschiedenen Einzelteilen eines industriellen Produkts, Heilerfolge eines Medikamentes an unterschiedlichen Personen und wiederholte physikalische Messungen oft sind. Wir sagen, zwei Ereignisse A und B sind **(stochastisch) unabhängig**, wenn gilt

$$P(A \cap B) = P(A)P(B),$$

was gleichbedeutend ist mit

$$P(B|A) = P(B) \text{ bzw. } P(A|B) = P(A).$$

Allgemein heißen A_1, \dots, A_k stochastisch unabhängig, falls

$$P(A_{i_1} \cap \dots \cap A_{i_l}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_l})$$

für alle möglichen $1 \leq i_1 < i_2 < \dots < i_l \leq k$ und alle l.

Es sei bemerkt, daß aus der paarweisen Unabhängigkeit i. a. nicht die Unabhängigkeit der Ereignisse A_1, \dots, A_k folgt.

Beispiel:

- (a) Ein Werkstück muß von zwei unabhängig voneinander arbeitenden Maschinen bearbeitet werden. Es ist bekannt, daß Maschine A 40% und Maschine B 20% Ausschuß liefert. Man möchte nun wissen, mit welcher Wahrscheinlichkeit ein einwandfreies Werkstück hergestellt wird. Ist A das Ereignis, daß Maschine A korrekt arbeitet und B das Ereignis, daß Maschine B fehlerfrei arbeitet, so sucht man

$$\begin{aligned} P(A \cap B) &= P(A)P(B) = (1 - P(\bar{A}))(1 - P(\bar{B})) \\ &= (1 - 0,4)(1 - 0,2) = 0,48. \end{aligned}$$

Wir haben hier ein sogenanntes **Seriensystem** vorliegen, das man sich wie in **Abb. 4** veranschaulichen kann.

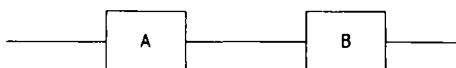


Abb. 4: Seriensystem aus zwei Maschinen A und B

- (b) Das Rotsignal einer Ampel wird mit zwei Glühbirnen A und B , die gleichzeitig brennen, ausgerüstet und ist daher intakt, wenn wenigstens eine Glühbirne noch brennt. Da die Birnen alle 90 Tage ausgewechselt werden, interessiert uns die Wahrscheinlichkeit dafür, daß das Rotsignal mindestens 90 Tage intakt ist. Sind A und B die Ereignisse, daß die Birnen A und B mindestens 90 Tage halten und ist 0,97 die Wahrscheinlichkeit dafür, so erhält man für die Wahrscheinlichkeit des uns interessierenden Ereignisses $A \cup B$ den Wert

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A)P(B) \\ &= 0,97 + 0,97 - 0,97 \cdot 0,97 \\ &= 0,9991. \end{aligned}$$

Hier hat man es mit einem sogenannten **Parallelsystem** zu tun, wie es in Abb. 5 veranschaulicht wird.

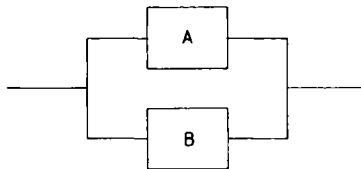


Abb. 5: Parallelsystem aus zwei Glühbirnen A und B

5. Die Bayessche Formel

Durch einen zu spät erkannten Fabrikationsfehler sind in einer Automobilproduktion genau 20 defekte Lenkgetriebe eingebaut worden. In einer Rückrufaktion werden alle 200000 Wagen dieser Serie überprüft und ein als fehlerhaft eingestuftes Lenkgetriebe durch ein neues ersetzt. Dabei wird mit 99prozentiger Sicherheit die Überprüfung zu einem korrekten Ergebnis führen. Wie groß ist die Wahrscheinlichkeit, daß ein ausgewechseltes Lenkgetriebe auch defekt ist? Mit der Bezeichnung D für das Ereignis eines defekten Lenkgetriebes und A für das eines ausgewechselten, schreiben sich die vorliegenden Informationen kurz

$$P(D) = 0,0001, \quad P(A|D) = 0,99 \quad \text{und} \quad P(A|\bar{D}) = 0,01.$$

Gesucht ist $P(D|A)$. Mit der Definition der bedingten Wahrscheinlichkeit und dem Satz von der totalen Wahrscheinlichkeit erhält man

$$\begin{aligned} P(D|A) &= \frac{P(D \cap A)}{P(A)} = \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|\bar{D})P(\bar{D})} \\ &= \frac{0,99 \cdot 0,0001}{0,99 \cdot 0,0001 + 0,01 \cdot 0,9999} \\ &= 0,01. \end{aligned}$$

Fast alle ausgewechselten Lenkgetriebe sind demnach in Ordnung. Die hier benutzte Formel ist unter dem Namen **Bayessche Formel** bekannt und lautet für paar-

weise disjunkte Ereignisse A_1, \dots, A_k mit $\bigcup_{i=1}^k A_i = \Omega$ und ein beliebiges Ereignis B

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^k P(B|A_j)P(A_j)}.$$

Beispiel: Die Produktion eines Bauteils laufe über drei unabhängig voneinander arbeitende Fließbänder und falle jeweils in zwei Qualitätsstufen 1 und 2 an. Die fertigen Teile werden zum Sortieren zu einer gemeinsamen Sammelstelle gebracht. Wir interessieren uns für die Wahrscheinlichkeit, daß ein der Qualitätsstufe 1 zugeordnetes Bauteil von Band I stammt. A_i sei das Ereignis, daß das Bauteil von Band i stammt und B beschreibe die Zuordnung zur Qualitätsstufe 1. Produziert Band I 750 Bauteile/Std., von denen 80% in die Qualitätsstufe 1 fallen, sind es bei Band II 800 Bauteile/Std. mit 85% aus Qualitätsstufe 1 und bei Band III 1000 Bauteile/Std. mit 65% der Stufe 1, so berechnet sich die gesuchte Wahrscheinlichkeit mit der Formel von Bayes zu

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &= \frac{0,8 \cdot 750 / (750 + 800 + 1000)}{0,8 \cdot 750 / 2550 + 0,85 \cdot 800 / 2550 + 0,65 \cdot 1000 / 2550} \\ &= \frac{0,8 \cdot 750}{0,8 \cdot 750 + 0,85 \cdot 800 + 0,65 \cdot 1000} \\ &= 0,31. \end{aligned}$$

6. Zufallsvariable und Verteilungen

Es werde zweimal hintereinander gewürfelt. Alle dabei nur denkbaren Ereignisse sind Teilmengen des Grundraumes $\Omega = \{(i, j) : i, j = 1, \dots, 6\}$. So wird etwa das Ereignis, daß die Summe der Augenzahlen gleich 4 ist, durch die Menge $A = \{(1, 3)(2, 2)(3, 1)\}$ beschrieben. Bei ausschließlichem Interesse für die Würfelsummen bietet sich jedoch eine weit elegantere Beschreibungsweise an. Wir führen dazu auf Ω die Funktion X ein mit $X((i, j)) = i + j$ und schreiben $X(i, j) = 4$ bzw. $X = 4$, wenn die Augensumme gleich 4 ist; $10 \leq X \leq 12$ bedeutet dann, daß die Summe zwischen 10 und 12 liegt; $X \geq 7$ sagt uns, daß ihr Wert mindestens 7 beträgt usw. Solch eine Funktion X , die Ergebnissen eines Zufallsexperimentes sinnvoll reelle Zahlen zuordnet, so daß für die mit Hilfe dieser Funktion beschriebenen Ereignisse Wahrscheinlichkeiten angebbar sind, nennt man Zufallsvariable. Dabei können, wie hier, die Ergebnisse des Zufallsexperimentes selbst schon Zahlen sein, aber es kann sich bei den Ergebnissen auch um Ausprägungen qualitativer Merkmale handeln. Beim n-maligen Werfen einer Münze ist z. B. die Anzahl des Auftretens von Kopf eine Zufallsvariable. Für den einzelnen Münzwurf ist allein schon durch 1 ~ Kopf und 0 ~ Wappen eine Zufallsvariable erklärt. Man spricht sogar von einer Zufallsvariablen, wenn die interessierende reelle Größe direkt als Ergebnis eines Zufallsexperimentes auftaucht, wie z. B. die Lebensdauer einer Glühbirne, der Ernteertrag eines Weizenfeldes, die tägliche Anzahl der Besucher eines Museums, die Augenzahl eines Würfels usw.

Allgemein sagt man, X sei eine **Zufallsvariable**, wenn die Werte dieser Größe reelle Zahlen sind, die durch ein Zufallsexperiment bestimmt werden und wenn für die Ereignisse, die man damit beschreiben kann, Wahrscheinlichkeiten angebar sind. Den Wert x , den die Zufallsvariable X bei der Durchführung des Zufallsexperiments annimmt, nennt man **Realisation von X** .

Oft interessiert man sich nicht nur für eine Zufallsvariable, sondern gleich für mehrere Zufallsvariablen X_1, \dots, X_n . Werden diese in einem Vektor $\mathbf{X} = (X_1, \dots, X_n)$ zusammengefaßt, so spricht man von einem **Zufallsvektor**. Das hat man z. B., wenn bei einer stichprobenhaften Kontrolle von Schrauben sowohl die Länge als auch der Durchmesser wichtig ist, oder wenn beim 4maligen Würfeln außer der Summe auch noch die größte und die kleinste Augenzahl interessiert usw. Man sieht, daß eigentlich jedes Zufallsexperiment durch eine Zufallsvariable X bzw. einen Zufallsvektor \mathbf{X} beschrieben werden kann. Ereignisse können dann allgemein in der Form $X \in B$ angegeben werden (mit B als Teilmenge im Wertebereich von X).

In diesem Zusammenhang spricht man nicht einfach von einer Wahrscheinlichkeit, sondern von der **Wahrscheinlichkeitsverteilung von X** . Damit wollen wir uns nun beschäftigen. Zur Festlegung einer Wahrscheinlichkeitsverteilung reicht es, die Wahrscheinlichkeiten für eine geeignete Auswahl von Ereignissen anzugeben. Die Wahrscheinlichkeiten aller übrigen Erzeugnisse lassen sich dann nämlich daraus berechnen. Bei einer einzelnen Zufallsvariablen X können dabei die Wahrscheinlichkeiten der Ereignisse $X \leq t$ als Grundlage dienen. So erhält man die Funktion

$$F_X(t) = P(X \leq t),$$

die man als **Verteilungsfunktion der Zufallsvariablen X** bezeichnet.

Beispiel: (a) Ist X die Lebensdauer einer Glühbirne, dann gibt $F_X(t)$ gerade die Wahrscheinlichkeit an, daß die Birne den Zeitpunkt t nicht überlebt. Will man nun wissen, mit welcher Wahrscheinlichkeit die Lebensdauer zwar größer ist als 800 Std., aber 1000 Std. nicht übersteigt, so berechnet man

$$\begin{aligned} P(800 < X \leq 1000) &= P(X \leq 1000) - P(X \leq 800) \\ &= F_X(1000) - F_X(800). \end{aligned}$$

Die Wahrscheinlichkeit, daß die Lebensdauer mehr als 1000 Std. beträgt, erhält man durch

$$P(X > 1000) = 1 - P(X \leq 1000) = 1 - F_X(1000).$$

Zur konkreten Berechnung fehlt uns nur noch die Verteilungsfunktion. Nehmen wir dazu einmal an, daß die Wahrscheinlichkeit für den Ausfall einer Glühbirne in einem Intervall zwischen t und $t + \Delta t$ ungefähr proportional der Länge Δt des Zeitintervalls ist (und nicht vom Zeitpunkt t selbst abhängt), d. h.

$$P(X \leq t + \Delta t | X > t) \approx \lambda \cdot \Delta t.$$

und somit nach der Formel für bedingte Wahrscheinlichkeiten

$$\frac{P(t < X \leq t + \Delta t)}{P(X > t)} \approx \lambda \cdot \Delta t \quad \text{bzw.} \quad \frac{P(t < X \leq t + \Delta t)}{\Delta t} \approx \lambda \cdot P(X > t)$$

so erhalten wir im Grenzübergang

$$\lim_{\Delta t \rightarrow 0} \frac{P(t < X \leq t + \Delta t)}{\Delta t} = \lambda \cdot P(X > t) = \lambda \cdot (1 - F_X(t)).$$

Den Proportionalitätsfaktor $\lambda > 0$ bezeichnet man als **Ausfallrate**. Die einzige Verteilungsfunktion, die dieses erfüllt, ist nun

$$F_X(t) = \begin{cases} 1 - e^{-\lambda t}, & \text{falls } t \geq 0 \\ 0, & \text{sonst} \end{cases}, \text{ vgl. Abb. 6;}$$

sie ist unter dem Namen **Exponentialverteilung mit Parameter λ** bekannt ($E(x(\lambda))$).

(b) X sei die Augenzahl beim Würfeln. Dann ist $P(X = 1) = P(X = 2) = \dots = P(X = 6) = 1/6$. Man berechnet z. B.

$$F_X(3) = P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{3}{6} = \frac{1}{2},$$

$$F_X(1,5) = P(X \leq 1,5) = P(X = 1) = \frac{1}{6} \text{ usw.}$$

So erhält man schließlich die Verteilungsfunktion

$$F_X(t) = \begin{cases} 0 & t < 1 \\ 1/6 & 1 \leqq t < 2 \\ 2/6 & 2 \leqq t < 3 \\ 3/6 & 3 \leqq t < 4 \\ 4/6 & 4 \leqq t < 5 \\ 5/6 & 5 \leqq t < 6 \\ 1 & 6 \leqq t \end{cases}$$

Diese Funktion springt an den Stellen 1, 2, ..., 6 jeweils um $1/6$ und ist sonst konstant, s. Abb. 6.

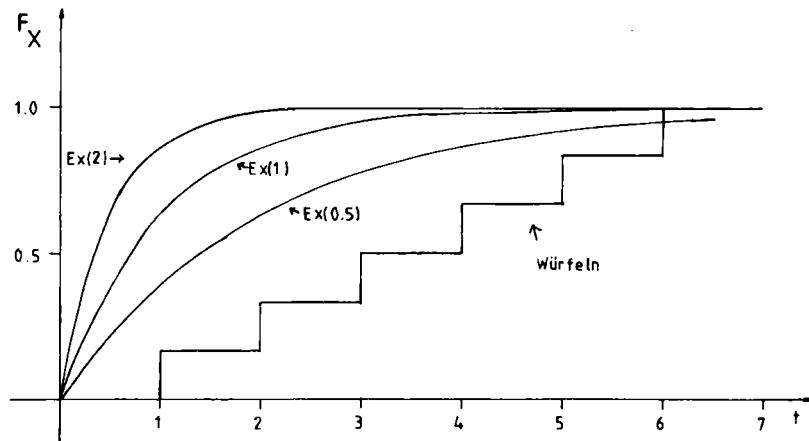


Abb. 6: Verteilungsfunktion der Zufallsvariablen $X = \text{Augenzahl beim Würfelwurf sowie der } E(x(0,5))- \text{, } E(x(1))- \text{ und } E(x(2))- \text{Verteilung}$

Die Verteilungsfunktion einer Zufallsvariablen X wird durch folgende Eigenschaften charakterisiert:

1. F_X ist monoton nicht fallend, d. h. für $t_1 < t_2$ ist $F_X(t_1) \leq F_X(t_2)$.
2. F_X ist rechtsseitig stetig, d. h. der rechtsseitige Grenzwert

$$\lim_{\substack{h \rightarrow 0 \\ h > 0}} F_X(t + h) = F_X(t).$$

3. $\lim_{t \rightarrow \infty} F_X(t) = 1$ und $\lim_{t \rightarrow -\infty} F_X(t) = 0$.

Wir haben im obigen Beispiel zwei ganz verschiedene Typen von Verteilungsfunktionen kennengelernt; die eine war stetig und die andere eine Sprungfunktion. Dabei fällt beim Würfelexperiment auf, daß die Verteilung der Zufallsvariablen X noch einfacher durch die Angabe der Wahrscheinlichkeiten $P(X = 1), \dots, P(X = 6)$ beschrieben wird. In so einer Situation, in der eine Zufallsvariable X nur endlich viele oder abzählbar unendlich viele Werte annehmen kann, sagt man, X ist **diskret verteilt**. Sind x_1, x_2, \dots die möglichen Realisationen von X , man bezeichnet die Gesamtheit dieser Werte auch als **Trägermenge von X** , so ist die Wahrscheinlichkeitsverteilung durch die Angabe der Werte $P(X = x_i)$ eindeutig bestimmt. Die zugehörige Verteilungsfunktion ist eine Sprungfunktion und berechnet sich zu

$$F_X(t) = \sum_{i: x_i \leq t} P(X = x_i).$$

Haben wir dagegen eine stetige Verteilungsfunktion, wie z. B. bei der Exponentialverteilung, so gilt

$$\begin{aligned} P(X = x) &= P(X \leq x) - P(X < x) = F_X(x) - \lim_{\substack{h \rightarrow 0 \\ h < 0}} P(X \leq x + h) \\ &= F_X(x) - \lim_{\substack{h \rightarrow 0 \\ h < 0}} F_X(x + h) = 0, \end{aligned}$$

d. h. jede Realisation von X hat die Wahrscheinlichkeit Null. Aber auch hier ist trotzdem oft eine ähnliche Darstellung der Verteilungsfunktion wie im diskreten Fall möglich, nämlich

$$F_X(t) = \int_{-\infty}^t f_X(\xi) d\xi.$$

Dabei wird die nichtnegative Funktion f_X **Dichtefunktion** oder kurz **Dichte** genannt. Die Verteilungsfunktion an der Stelle x , also der Wert $F_X(x)$, gibt somit gerade die Größe der Fläche oberhalb der x -Achse an, die unterhalb des Graphen der Funktion f_X zwischen $-\infty$ und x liegt; die Größe der Gesamtfläche unterhalb des Graphen von f_X (und oberhalb der x -Achse) ist wegen $F_X(+\infty) = 1$ natürlich gerade Eins.

Jede Wahrscheinlichkeitsverteilung einer Zufallsvariablen X , für die eine Dichte existiert, heißt **stetige Verteilung (Verteilung mit Dichte)**. Durch die Angabe einer Dichte, die in vielen Fällen als Ableitung der Verteilungsfunktion bestimmt werden kann, ist eine Wahrscheinlichkeitsverteilung eindeutig bestimmt.

Beispiel: Aus der Verteilungsfunktion der Exponentialverteilung erhalten wir durch Differentiation die zugehörige Dichte, s. Abb. 7,

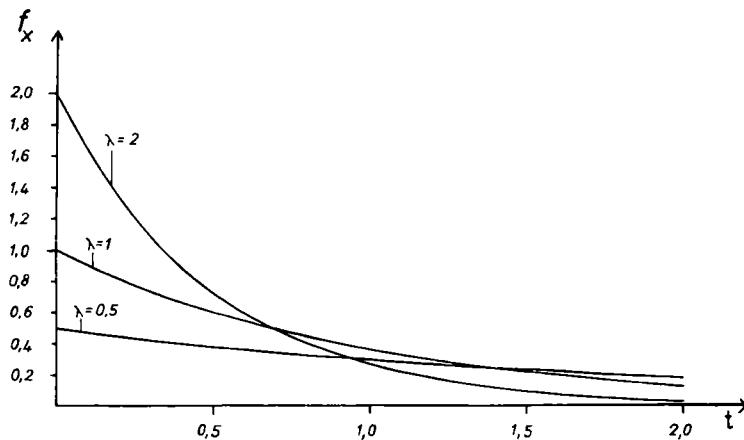


Abb. 7: Dichtefunktion der Ex(1)-, Ex(2)- und Ex(0,5)-Verteilung

$$f_X(t) = F'_X(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

Damit ergibt sich z. B. für die Wahrscheinlichkeit

$$P(a < X \leq b) = \int_a^b \lambda e^{-\lambda t} dt.$$

Man sieht hier, daß an der Wahrscheinlichkeitsdichte leichter als an der Verteilungsfunktion zu erkennen ist, welche Realisationen von X mehr und welche weniger häufig auftreten.

Wir wollen nun noch zeigen, wie man die Wahrscheinlichkeitsverteilung eines Zufallsvektors $X = (X_1, \dots, X_n)$ beschreibt. Dabei reicht es im allgemeinen nicht aus, die Verteilungen der einzelnen Komponenten X_i anzugeben. So läßt sich z. B. beim zweimaligen Würfeln allein aus der Kenntnis der Wahrscheinlichkeit dafür, daß die Würfelsumme größer als sieben ist, und dafür, daß die größte Augenzahl kleiner als sechs ist, nicht schließen, mit welcher Wahrscheinlichkeit beides der Fall ist. Solche Fragen lassen sich jedoch beantworten, wenn die (**multivariate, mehrdimensionale**) **Verteilungsfunktion** von X

$$F_X(t_1, \dots, t_n) = P(X_1 \leq t_1, \dots, X_n \leq t_n)$$

bekannt ist. Die Verteilung einer einzelnen Zufallsvariablen X_i bezeichnet man als **i-te Randverteilung** des Zufallsvektors X und deren Verteilungsfunktion berechnet sich aus der obigen wie folgt

$$\begin{aligned} F_{X_i}(t_i) &= P(X_i \leq t_i) = P(X_1 \leq \infty, \dots, X_i \leq t_i, \dots, X_n \leq \infty) \\ &= \lim_{\substack{t_j \rightarrow \infty \\ j \neq i}} F_X(t_1, \dots, t_n). \end{aligned}$$

Die Verteilung des Zufallsvektors X nennt man auch die **gemeinsame Verteilung** der X_i und F_X die **gemeinsame Verteilungsfunktion**. Analog zum eindimensionalen Fall spricht man hier von einem **diskret verteilten Zufallsvektor**, wenn die Anzahl der möglichen Realisationen abzählbar ist und somit die Wahrscheinlichkeitsver-

teilung durch die Angabe der Wahrscheinlichkeiten für diese Realisationen eindeutig festgelegt werden kann. Bezeichnet z. B. beim zweimaligen Würfeln X_1 die Summe und X_2 die größte der beiden Augenzahlen, so ist deren gemeinsame Verteilung durch die Wahrscheinlichkeiten $P(X_1 = i, X_2 = j)$ festgelegt. Von einem **stetig verteilten Zufallsvektor** spricht man, wenn sich die Verteilungsfunktion darstellen lässt in der Form

$$F_X(t_1, \dots, t_n) = \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_n} f_X(\xi_1, \dots, \xi_n) d\xi_1 \dots d\xi_n.$$

Die Funktion f_X nennt man **gemeinsame Dichte** der X_i .

7. Unabhängigkeit und Funktionen von Zufallsvariablen

7.1. Unabhängigkeit von Zufallsvariablen

Beim zweimaligen Würfeln bezeichne X_1 die Augenzahl beim ersten und X_2 die beim zweiten Wurf. Dann ist z. B. das Ereignis $X_2 = 6$, daß beim zweiten Mal eine Sechs gewürfelt wird, unabhängig von $X_1 < 3$, daß beim ersten Mal eine Zahl kleiner als drei gewürfelt wurde; das Erscheinen einer geraden Zahl beim zweiten Wurf, d. h. $X_2 \in \{2, 4, 6\}$, ist unabhängig davon, ob im ersten Wurf eine ungerade Zahl erschienen ist, d. h. unabhängig von $X_1 \in \{1, 3, 5\}$ usw. Zwei Ereignisse, von denen das eine mit Hilfe von X_1 darstellbar ist und das andere mit Hilfe von X_2 , erweisen sich hier immer als unabhängig voneinander. Man sagt die Zufallsvariablen X_1 und X_2 sind **(stochastisch) unabhängig**. Allgemein heißen Zufallsvariablen X_1, \dots, X_n (stochastisch) unabhängig, wenn für jede Auswahl von höchstens n Ereignissen, die jeweils getrennt durch verschiedene von diesen Zufallsvariablen beschrieben werden können, Unabhängigkeit vorliegt. Zum Nachweis der stochastischen Unabhängigkeit reicht es allerdings, Ereignisse der Form $X_i \leq x_i$ zu betrachten. Ist nämlich die gemeinsame Verteilungsfunktion der X_i gleich dem Produkt der einzelnen Verteilungsfunktionen, d. h.

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \dots P(X_n \leq x_n),$$

so folgt daraus schon die stochastische Unabhängigkeit der X_i . Besitzen X_1, \dots, X_n eine gemeinsame Dichte f , so ist Unabhängigkeit auch gleichbedeutend damit, daß f als Produkt der Einzeldichten darstellbar ist, d. h.

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n).$$

Sind dagegen X_1, \dots, X_n diskret verteilt, dann lässt sich auch aus der Gültigkeit von

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n)$$

auf die Unabhängigkeit der X_i schließen.

7.2. Funktionen von Zufallsvariablen

Wir wollen uns nun mit den Verteilungen einiger Funktionen von unabhängigen Zufallsvariablen beschäftigen, nämlich der **Summe**, des **Maximums** und des **Minimums**. Außerdem soll noch die **lineare Transformation** einer einzelnen Zufallsvariablen behandelt werden, die es uns ermöglicht, Zufallsvariable zu standardisieren.

A. Lineare Transformation; Normalverteilung

Hat eine Zufallsvariable X die Verteilungsfunktion F_X , so ist die Verteilungsfunktion der linearen Transformation $Y = \alpha \cdot X + \beta$ mit $\alpha > 0$ gegeben durch

$$F_Y(y) = P(Y \leq y) = P(\alpha \cdot X + \beta \leq y) = P(X \leq \frac{y - \beta}{\alpha}) = F_X\left(\frac{y - \beta}{\alpha}\right).$$

Besitzt X eine Dichte f_X , so hat Y die Dichte

$$f_Y(y) = \frac{1}{\alpha} f_X\left(\frac{y - \beta}{\alpha}\right).$$

Beispiel: Gehen wir davon aus, daß die Länge X von Schrauben, die auf einer bestimmten Maschine gefertigt werden, verteilt ist gemäß einer Verteilung mit der Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}.$$

Die Länge schwankt also um einen mittleren Wert μ , wobei die Schwankungen umso größer sind, je größer der Parameter σ ist. Diese Verteilung heißt **Normalverteilung (Gaußverteilung)** mit den Parametern μ und $\sigma^2 > 0$, in Zeichen $N(\mu, \sigma^2)$. Durch die lineare Transformation $Y = (X - \mu)/\sigma$ erhält man die sogenannte **Standardnormalverteilung** $N(0, 1)$ mit der Dichte

$$\varphi(y) = \sigma \cdot f_X\left((y + \frac{\mu}{\sigma})\sigma\right) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot y^2}.$$

Deren Verteilungsfunktion wird mit Φ bezeichnet und ist im Anhang vertafelt. Damit läßt sich nun z. B. die Wahrscheinlichkeit bestimmen, daß die Länge einer Schraube zwischen x_1 und x_2 liegt. Es ist nämlich

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= P\left(\frac{x_1 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right). \end{aligned}$$

Ausführlich wird in Kap. IV, Abschnitt 1 auf die Normalverteilung eingegangen.

B. Summe; Faltung; Binomialverteilung

a) Sind X_1, \dots, X_n unabhängig und diskret verteilt, so ist die Verteilung der Summe dieser Zufallsvariablen gegeben durch

$$P\left(\sum_{i=1}^n X_i = x\right) = \sum_{\substack{\sum x_i = x \\ i=1}} P(X_1 = x_1) \dots P(X_n = x_n)$$

bzw. durch die Verteilungsfunktion

$$P\left(\sum_{i=1}^n X_i \leq x\right) = \sum_{\substack{\sum x_i \leq x \\ i=1}} P(X_1 = x_1) \dots P(X_n = x_n).$$

Beispiel: Der Anteil unehelicher Geburten im District of Columbia im Jahre 1957 betrug 5,88% (Vital Statistics-Special Reports 47(8) (1960): "Illegitimate Births in the United States 1938–1957"). Wie groß ist die Wahrscheinlichkeit, daß bei fünf Geburten höchstens eine uneheliche ist? Es bezeichne X_i den Status für die i -te Geburt:

$$X_i = \begin{cases} 1, & \text{Geburt unehelich} \\ 0, & \text{Geburt ehelich} \end{cases}$$

Die Zufallsvariable $Y = \sum_{i=1}^5 X_i$ gibt dann die Anzahl der unehelichen unter fünf Geburten an. Nehmen wir an, daß die X_i unabhängig sind (was bei der Größe der Population gerechtfertigt scheint), so ergibt sich

$$\begin{aligned} P(Y \leq 1) &= P(Y = 0) + P(Y = 1) \\ &= P(X_1 = \dots = X_5 = 0) + P(X_1 = 1, X_2 = \dots = X_5 = 0) \\ &\quad + P(X_2 = 1, X_1 = X_3 = X_4 = X_5 = 0) + \dots \\ &\quad + P(X_5 = 1, X_1 = X_2 = X_3 = X_4 = 0) \\ &= (0,9412)^5 + (0,9412)^4 \cdot 0,0588 + \dots + (0,9412)^4 \cdot 0,0588 \\ &= (0,9412)^5 + (0,9412)^4 \cdot 0,0588 \cdot 5 \\ &\approx 0,96932. \end{aligned}$$

Es gilt nämlich wegen der Unabhängigkeit $P(X_1 = 0, \dots, X_5 = 0) = P(X_1 = 0) \cdot \dots \cdot P(X_5 = 0)$ usw. Allgemein erhält man für die Verteilung der Zufallsvariablen Y :

$$\begin{aligned} P(Y = i) &= \binom{5}{i} \cdot 0,0588^i \cdot (1 - 0,0588)^{5-i}, \quad i = 0, 1, \dots, 5, \\ P(Y \leq j) &= \sum_{k=0}^j \binom{5}{k} \cdot 0,0588^k \cdot (1 - 0,0588)^{5-k}, \quad j = 0, 1, \dots, 5. \end{aligned}$$

Man nennt solch eine Verteilung **Binomialverteilung** mit Parametern 5 und 0,0588. Allgemein heißt eine Zufallsvariable X **binomialverteilt mit den Parametern n und p** ($B(n, p)$), falls

$$P(X = i) = \binom{n}{i} \cdot p^i \cdot (1 - p)^{n-i}$$

gilt. Die Zufallsvariablen X_i sind dann offensichtlich $B(1, 0,0588)$ verteilt.

Wie hier im Beispiel gilt immer, daß die Summe von n unabhängigen $B(1, p)$ -verteilten Zufallsvariablen $B(n, p)$ verteilt ist. Ausführlich wird auf die Binomialverteilung in Abschnitt 3 von Kap. IV eingegangen.

- b) Bei stetig verteilten unabhängigen Zufallsvariablen X_1, \dots, X_n ist die Summe verteilt mit der Dichte

$$f_{\sum X_i}(x) = \int \cdots \int f_{X_1}(x_1) \cdots f_{X_n}(x_n) dx_1 \cdots dx_n.$$

$$\sum_{i=1}^n x_i = x$$

Speziell für die Summe zweier Zufallsvariablen erhält man

$$f_{X_1+X_2}(x) = \int_{-\infty}^{\infty} f_{X_1}(x-y) f_{X_2}(y) dy \quad (\text{Faltung}).$$

Beispiel: Zwei verschiedene Bleche werden aufeinander geschweißt. Bezeichnen X_1 und X_2 die Dickenabweichungen der beiden Bleche vom Sollwert, so weicht die Gesamtdicke um $X_1 + X_2$ vom Sollwert ab. Sind X_1 und X_2 unabhängig und $N(0, 1)$ verteilt, so ist $X_1 + X_2$ verteilt mit der Dichte

$$\begin{aligned} f_{X_1+X_2}(x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x-y)^2} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}y^2} dy \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \cdot e^{-\frac{1}{2}\frac{x^2}{2}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{1/2}} \cdot e^{-\frac{1}{2}\frac{(y-\frac{x}{2})^2}{1/2}} dy \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \cdot e^{-\frac{1}{2}\frac{x^2}{2}}, \end{aligned}$$

also $N(0, 2)$ verteilt. Allgemein erweist sich die Summe einer $N(\mu_1, \sigma_1^2)$ - und einer davon unabhängigen $N(\mu_2, \sigma_2^2)$ -verteilten Zufallsgröße als $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ -verteilt.

C. Maximum

Sind X_1, \dots, X_n unabhängig voneinander, so ergibt sich für die Verteilungsfunktion des Maximums

$$\begin{aligned} P(\max_{1 \leq i \leq n} X_i \leq x) &= P(X_i \leq x \text{ für alle } i) = P(X_1 \leq x) \cdot \dots \cdot P(X_n \leq x) \\ &= F_{X_1}(x) \dots F_{X_n}(x). \end{aligned}$$

Beispiel: Fünf Glühbirnen, deren Lebensdauer X_i jeweils exponentialverteilt ist mit Ausfallrate $\lambda = 0,00125$ (1|Std.), werden parallel geschaltet. Das Maximum der X_i gibt uns dann die Gesamtlebensdauer des Systems an. Die Verteilungsfunktion der Gesamtlebensdauer ist also gegeben durch

$$P(\max_{1 \leq i \leq 5} X_i \leq x) = \begin{cases} (1 - e^{-\lambda x})^5, & x \geq 0 \\ 0, & \text{sonst} \end{cases}$$

Damit erhält man für die Wahrscheinlichkeit, daß das Parallelsystem höchstens 800 Std. intakt ist, den Wert $(1 - e^{-0,00125 \cdot 800})^5 = 0,1$. Der entsprechende Wert für eine einzelne Glühbirne ist dagegen $1 - e^{-0,00125 \cdot 800} = 0,63$.

D. Minimum

Sind X_1, \dots, X_n unabhängig voneinander, so ergibt sich für die Verteilungsfunktion des Minimums

$$\begin{aligned} P(\min_{1 \leq i \leq n} X_i \leq x) &= 1 - P(\min_{1 \leq i \leq n} X_i > x) = 1 - P(X_i > x \text{ für alle } i) \\ &= 1 - (1 - F_{X_1}(x)) \cdot \dots \cdot (1 - F_{X_n}(x)). \end{aligned}$$

Beispiel: Werden fünf Glühbirnen in Serie geschaltet, so fällt das Gesamtsystem aus, wenn eine der Birnen ausfällt. Sind die Lebensdauern X_i der einzelnen Glühbirnen exponentialverteilt mit Ausfallrate λ , so ergibt sich für die Verteilungsfunktion der Lebensdauer des Gesamtsystems

$$P\left(\min_{1 \leq i \leq 5} X_i \leq x\right) = \begin{cases} 1 - (1 - e^{-\lambda x})^5 = 1 - e^{-5\lambda x}, & x \geq 0 \\ 0, & \text{sonst,} \end{cases}$$

also eine Exponentialverteilung mit der Ausfallrate 5λ .

8. Kenngrößen von Zufallsvariablen

Ähnlich wie bei der deskriptiven Statistik die Häufigkeitsverteilungen durch Kenngrößen beschrieben werden, so werden es hier die Wahrscheinlichkeitsverteilungen von Zufallsgrößen. Dabei werden uns ebenso wie dort Lage- und Streuungsparameter begegnen. Sogar die Bezeichnungen Mittelwert, Median oder Varianz und Standardabweichung usw. tauchen wieder auf. Wiederholt man nämlich ein durch eine Zufallsvariable X beschriebenes Zufallsexperiment genügend oft, so stimmen meist die Kenngrößen der beobachteten Häufigkeitsverteilung ungefähr mit den analogen Größen der Wahrscheinlichkeitsverteilung von X überein.

8.1. Lageparameter

A. Der Erwartungswert

Beschäftigen wir uns zunächst mit den **Lageparametern**. Der wichtigste ist dabei der **Erwartungswert** oder **Mittelwert** einer Verteilung. Ist X eine stetige Zufallsvariable mit der Dichte f , so erhält man ihn, falls er existiert, durch

$$E X = \int_{-\infty}^{\infty} x \cdot f(x) dx.$$

Bei einer diskreten Zufallsvariablen mit den möglichen Realisierungen x_1, x_2, \dots ist entsprechend

$$E X = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i).$$

Beispiel:

(a) Für die Standardnormalverteilung $N(0, 1)$ ergibt sich für den Erwartungswert

$$\sqrt{2\pi} \cdot E X = \int_{-\infty}^{\infty} x \cdot e^{-\frac{1}{2}x^2} dx = \int_0^{\infty} |x| \cdot e^{-\frac{1}{2}x^2} dx - \int_{-\infty}^0 |x| \cdot e^{-\frac{1}{2}x^2} dx = 0$$

(b) Ist X exponentialverteilt mit Parameter $\lambda > 0$, so ist

$$\begin{aligned} E X &= \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda x} dx = \left[x \cdot \lambda \left(-\frac{1}{\lambda}\right) e^{-\lambda x} \right]_0^{\infty} - \int_0^{\infty} \lambda \left(-\frac{1}{\lambda}\right) e^{-\lambda x} dx \\ &= 0 + \frac{1}{\lambda} \int_0^{\infty} \lambda \cdot e^{-\lambda x} dx = \frac{1}{\lambda}. \end{aligned}$$

Beobachtet man eine genügend große Anzahl von Glühbirnen, bei denen die Lebensdauer einer einzelnen Birne durch X beschrieben werden kann, so ergibt sich praktisch immer als durchschnittliche Lebensdauer dieser Wert $1/\lambda$, also die besagte Übereinstimmung mit dem Mittelwert der beobachteten Häufigkeitsverteilung.

- (c) Eine Zufallsvariable X , die zählt, wie oft ein relativ seltenes Ereignis auftritt, kann oft als **Poissonverteilung** angenommen werden, d. h.

$$P(X = i) = \frac{\lambda^i}{i!} \cdot e^{-\lambda}.$$

Deren Erwartungswert berechnet sich zu

$$EX = \sum_{i=0}^{\infty} i \cdot \frac{\lambda^i}{i!} \cdot e^{-\lambda} = \sum_{i=1}^{\infty} \lambda \cdot \frac{\lambda^{i-1}}{(i-1)!} e^{-\lambda} = \lambda.$$

Daß Erwartungswerte durchaus nicht immer existieren müssen, zeigt das folgende

Beispiel: Hat X die Verteilung $N(\mu, \lambda^2)$ und ist die davon unabhängige Zufallsvariable Y standardnormalverteilt, dann hat X/Y eine **Cauchyverteilung** mit der Dichte

$$f(t) = \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (t - \mu)^2}.$$

Da $\int_0^{\infty} t \cdot f(t) dt = \infty = - \int_{-\infty}^0 t \cdot f(t) dt,$

existiert deren Erwartungswert nicht.

Die Berechnung von Erwartungswerten wird oft einfacher, wenn man folgende Regeln benutzt:

1. Ist $Y = g(X)$, dann erhält man den Erwartungswert, sofern er existiert, durch

$$EY = \int_{-\infty}^{\infty} g(x) f_X(x) dx, \text{ falls } X \text{ stetig bzw.}$$

$$EY = \sum_{i=1}^{\infty} g(x_i) P(X = x_i), \text{ falls } X \text{ diskret ist.}$$

Hat man speziell $Y = \alpha \cdot X + \beta$, so ergibt sich $EY = \alpha \cdot EX + \beta$. [Ist g konkav (konvex), so ist $Eg(X) \leq (\geq) g(EX)$; **Jensensche Ungleichung**.]

2. Für beliebige Zufallsvariablen X_1, \dots, X_n gilt

$$E(X_1 + X_2 + \dots + X_n) = EX_1 + EX_2 + \dots + EX_n.$$

3. Sind X_1, \dots, X_n unabhängige Zufallsvariablen, so ist

$$E(X_1 \cdot X_2 \cdot \dots \cdot X_n) = EX_1 \cdot EX_2 \cdot \dots \cdot EX_n.$$

Beispiel:

- (a) Hat X die Verteilung $N(0, 1)$, dann ist $Y = \sigma \cdot X + \mu$ normalverteilt mit den Parametern μ und σ^2 . Mit der ersten Rechenregel ergibt sich

$$EY = \sigma \cdot EX + \mu = \sigma \cdot 0 + \mu = \mu.$$

Der Erwartungswert einer $N(\mu, \sigma^2)$ -Verteilung ist also gleich μ .

- (b) X_1, \dots, X_n seien unabhängig und $B(1, p)$ -verteilt. Die Summe dieser Zufalls-

variablen ist $B(n, p)$ -verteilt, so daß sich der folgende Erwartungswert berechnet:

$$\begin{aligned} E(X_1 + \dots + X_n) &= EX_1 + \dots + EX_n = n \cdot EX_1 \\ &= n(0 \cdot P(X_1 = 0) + 1 \cdot P(X_1 = 1)) = n \cdot p. \end{aligned}$$

B. Der Median und andere Quantile

Als weiteres Lagemaß ist noch der **Median** $\xi_{0,5}$ gebräuchlich. Das ist gerade diejenige Zahl, für die gilt

$$P(X \leq \xi_{0,5}) \geq \frac{1}{2} \text{ und } P(X \geq \xi_{0,5}) \geq \frac{1}{2}.$$

Mit Hilfe der sogenannten **verallgemeinerten inversen Verteilungsfunktion**

$$F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\},$$

die mit der normalen Inversen übereinstimmt, falls es sie gibt, läßt sich schreiben: $\xi_{0,5} = F_X^{-1}(0,5)$. Man bezeichnet $\xi_\alpha = F_X^{-1}(\alpha)$ auch als α **Quantil** einer Verteilung, so daß hier der Spezialfall eines 0,5 Quantils vorliegt. Außerdem haben noch das 0,25 und das 0,75 Quantil spezielle Namen, nämlich **unteres** und **oberes Quartil**. Zu erwähnen ist noch, daß das $1 - \alpha$ Quantil gelegentlich als α **Fraktil** bezeichnet wird.

Besitzt eine Zufallsvariable X eine stetige Verteilung, so gilt für das α Quantil ξ_α gerade

$$F_X(\xi_\alpha) = \int_{-\infty}^{\xi_\alpha} f_X(x) dx = \alpha;$$

d. h. das α Quantil einer stetigen Verteilung ist derjenige Wert ξ_α , für den gilt, daß die Größe der Fläche oberhalb der x-Achse und unterhalb des Graphen der Dichtefunktion f_X von $-\infty$ bis ξ_α gerade α ist, vgl. auch Abb. 8.

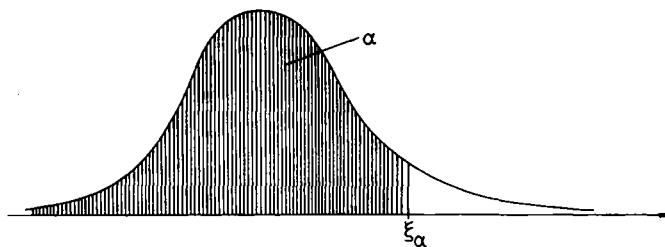


Abb. 8: α -Quantil ξ_α einer stetigen Verteilung mit obiger Dichtefunktion

Ist die Verteilung von X dagegen diskret, so gilt für das α Quantil ξ_α

$$F_X(\xi_\alpha) \geq \alpha$$

und für jedes $x < \xi_\alpha$ ist

$$F_X(x) < \alpha,$$

vgl. auch Abb. 9.

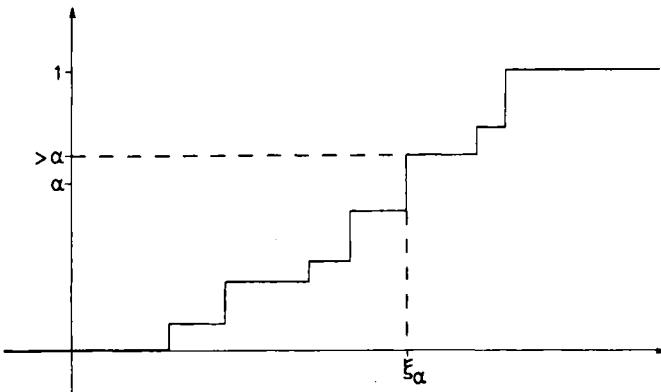


Abb. 9: α -Quantil ξ_α einer diskreten Verteilung mit obiger Verteilungsfunktion

Beispiel:

- (a) Bei der Normalverteilung $N(\mu, \sigma^2)$ ist wegen

$$\int_{-\infty}^{\mu} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{-\infty}^{\mu} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0,5$$

der Median identisch mit dem Erwartungswert μ . Das ist allgemein so bei jeder symmetrischen Verteilung.

Die **Quantile der Standardnormalverteilung** werden mit u_α bezeichnet; es ergibt sich aus der Tabelle im Anhang z. B.: $u_{0,99} = 2,3264$ und wegen der Symmetrie der Dichte der $N(0, 1)$ -Verteilung ergibt sich $u_\alpha = -u_{1-\alpha}$, d. h. $u_{0,01} = -2,3264$.

- (b) Ist X exponentialverteilt mit Parameter λ , dann berechnet sich das α Quantil ξ_α wegen $F_X(\xi_\alpha) = \alpha$ aus

$$\alpha = 1 - e^{-\lambda \cdot F_X^{-1}(\alpha)} \text{ zu } \xi_\alpha = F_X^{-1}(\alpha) = -\frac{\ln(1-\alpha)}{\lambda}.$$

So hat z. B. der Median den Wert $(\ln 2)/\lambda$.

- (c) Die Zufallsvariable X sei $B(5, 0,6)$ -verteilt. Es ist

$$\begin{aligned} F_X(1) &= \binom{5}{0} \cdot 0,6^0 \cdot (1-0,6)^{5-0} + \binom{5}{1} \cdot 0,6^1 \cdot (1-0,6)^{5-1} \\ &= 0,0102 + 0,0768 = 0,0870, \end{aligned}$$

$$\begin{aligned} F_X(2) &= F_X(1) + \binom{5}{2} \cdot 0,6^2 \cdot (1-0,6)^{5-2} \\ &= 0,0870 + 0,2304 = 0,3174 \text{ und} \end{aligned}$$

$$F_X(3) = 0,6630.$$

So ergibt sich z. B. für das 0,1 Quantil der Wert 2 und für den Median der Wert 3.

C. Der Modalwert

Als letztes Lagemaß soll noch der **Modus (Modalwert)** erwähnt werden. Bei diskreten Verteilungen ist das der Wert mit der größten Wahrscheinlichkeit und bei stetigen der Wert, an dem die Dichtefunktion maximal wird.

8.2. Streuungsparameter

A. Die Varianz und die Standardabweichung; standardisierte Zufallsvariable; Tschebyscheffsche Ungleichung

Kommen wir nun zu den **Streuungs- oder Dispersionsmaßen**, von denen Varianz und Standardabweichung die wichtigsten sind. Die **Varianz** einer Zufallsgröße X ist

$$\text{Var } X = E(X - EX)^2.$$

Dafür benutzt man auch häufig das Symbol σ^2 . Die Quadratwurzel aus der Varianz, die dann folglich mit σ abgekürzt wird, nennt man **Standardabweichung**. Die Bedeutung dieser Größen wird durch die **Ungleichung von Tschebyscheff** unterstrichen. Für eine beliebige positive Zahl c gilt nämlich

$$P(|X - EX| \geq c) \leq \frac{\text{Var } X}{c^2}.$$

Wählt man nun speziell c als ein Vielfaches der Standardabweichung so hat man

$$P(|X - EX| \geq k \cdot \sigma) \leq \frac{\sigma^2}{(k \cdot \sigma)^2} = \frac{1}{k^2}.$$

Damit ergibt sich z. B., daß eine Zufallsvariable X nur mit der Wahrscheinlichkeit 0,111 Werte annehmen kann, die $3 \cdot \sigma$ oder mehr vom Erwartungswert abweichen. Dies ist dennoch meist eine recht grobe Abschätzung, wie etwa der Vergleich mit dem entsprechenden Wert 0,003 für die Normalverteilung zeigt.

Beispiel: Eine Zufallsgröße X habe eine **Rechteckverteilung** auf dem Intervall $[0; a]$, d. h. eine Verteilung mit der Dichte

$$f_X(x) = \begin{cases} \frac{1}{a}, & \text{falls } 0 \leq x \leq a \\ 0, & \text{sonst} \end{cases}$$

Man sagt auch, X ist **gleichverteilt** auf $[0; a]$. Dann ist

$$EX = \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_0^a x \cdot \frac{1}{a} dx = \frac{1}{2} \cdot a^2 \cdot \frac{1}{a} = \frac{a}{2}$$

und damit

$$\begin{aligned} \text{Var } X &= E(X - EX)^2 = \int_{-\infty}^{\infty} \left(x - \frac{1}{2} a \right)^2 \cdot f_X(x) dx \\ &= \int_0^a \left(x^2 - x \cdot a + \frac{a^2}{4} \right) \cdot \frac{1}{a} dx \\ &= \left(\frac{1}{3} \cdot a^3 - \frac{1}{2} \cdot a^2 \cdot a + a \cdot \frac{a^2}{4} \right) \cdot \frac{1}{a} = \frac{1}{12} \cdot a^2. \end{aligned}$$

Für die Standardabweichung erhält man $\sigma = \sqrt{\text{Var } X} = 0,29 \cdot a$. Hier ist sogar schon $P(|X - EX| \geq 2 \cdot \sigma) = 0$.

Zur Berechnung von Varianzen sind folgende Regeln nützlich:

1. $\text{Var } X = EX^2 - (EX)^2$ (**Verschiebungssatz, Satz von Steiner**)
2. $\text{Var}(\alpha \cdot X + \beta) = \alpha^2 \text{Var } X$
3. Sind X_1, \dots, X_n unabhängige Zufallsvariablen, dann gilt die **Gleichung von Bienaymé**:

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var } X_1 + \text{Var } X_2 + \dots + \text{Var } X_n.$$

(Allgemein ist für beliebige Zufallsvariablen X_1, \dots, X_n

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var } X_i + 2 \cdot \sum_{i < j} \text{Cov}(X_i, X_j),$$

wobei $\text{Cov}(X_i, X_j) = E(X_i - EX_i)(X_j - EX_j)$ als **Kovarianz** von X_i und X_j bezeichnet wird.)

Beispiel:

- (a) Ist X eine exponentialverteilte Zufallsvariable mit Parameter λ , so hat man wegen $EX = 1/\lambda$ und

$$\begin{aligned} EX^2 &= \int_0^\infty x^2 \cdot \lambda \cdot e^{-\lambda x} dx = [-x^2 \cdot e^{-\lambda x}]_0^\infty + \int_0^\infty 2 \cdot x \cdot e^{-\lambda x} dx \\ &= 0 + \frac{2}{\lambda} \cdot EX = \frac{2}{\lambda^2} \end{aligned}$$

nach dem Satz von Steiner:

$$\text{Var } X = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

- (b) X sei $N(0, 1)$ -verteilt. Dann ist $EX = 0$ und $\text{Var } X = 1$. Folglich hat die Zufallsvariable $Y = \sigma \cdot X + \mu$, die $N(\mu, \sigma^2)$ -verteilt ist, als Varianz den Wert σ^2 .
- (c) Es seien X_1, \dots, X_n unabhängig und $B(1, p)$ -verteilt. Da X_i nur die Werte 0 und 1 annimmt, ist $X_i^2 = X_i$ und somit $EX_i^2 = EX_i = p$. Demnach gilt

$$\text{Var } X_i = EX_i^2 - (EX_i)^2 = p - p^2 = p(1-p).$$

$X_1 + \dots + X_n$ ist $B(n, p)$ -verteilt. Daher ergibt sich nun

$$\text{Var}(X_1 + \dots + X_n) = \text{Var } X_1 + \dots + \text{Var } X_n = n \cdot p(1-p).$$

- (d) Ist X eine Zufallsvariable mit Erwartungswert μ und Varianz σ^2 , dann hat man für die sogenannte **standardisierte Zufallsvariable** $Y = (X - \mu)/\sigma$:

$$\text{Var } Y = \frac{1}{\sigma^2} \text{Var } X = 1 \text{ und } EY = \frac{1}{\sigma} (EX - \mu) = 0.$$

B. Der Variationskoeffizient

Die Varianz (bzw. die Standardabweichung) ist ein Maß für die **absolute Größe** der

Streuungen um den Erwartungswert. Dagegen mißt der **Variationskoeffizient** $\sqrt{\text{Var } X} / E X$ die Abweichungen relativ zur Größe des Erwartungswertes.

C. Der Quartilsabstand

Als letztes Streuungsmaß sei hier noch der **Quartilsabstand** erwähnt, der nun wieder einen Eindruck von der **absoluten** Größe der Streuungen liefert. Er berechnet sich als Differenz des oberen und unteren Quartils.

Beispiel: Bei der Exponentialverteilung mit Parameter λ hat das α Quantil den Wert $-\ln(1 - \alpha)/\lambda$. Das obere Quartil, d. h. das 0,75 Quantil, ist somit gleich $-(\ln 0,25)/\lambda$ und das untere Quartil, d. h. das 0,25 Quantil, ist gleich $-(\ln 0,75)/\lambda$. Es ergibt sich hier nun der **Quartilsabstand**

$$-\frac{\ln 0,25}{\lambda} + \frac{\ln 0,75}{\lambda} = \frac{\ln(0,75/0,25)}{\lambda} = \frac{\ln 3}{\lambda}.$$

8.3. Momente von Zufallsvariablen; Schiefe; Exzeß

Eine weitergehende Charakterisierung der Verteilung einer Zufallsvariablen X als durch Lage- und Streuungsmaße läßt sich durch die **Momente** $E X^k$ bzw. die **zentralen Momente** $E(X - E X)^k$ vornehmen. Dabei haben wir das erste Moment $E X$ schon als Erwartungswert und das zweite zentrale Moment $E(X - E X)^2$ als Varianz kennengelernt. Aus dem dritten zentralen Moment leitet man die Größe $E(X - E X)^3/\sqrt{\text{Var } X}$ ab, die bei symmetrischen Verteilungen immer den Wert Null hat, und als **Schiefe** bezeichnet wird. Mit dem vierten zentralen Moment wird die mit **Exzeß** oder **Wölbung** bezeichnete Größe $E(X - E X)^4/(\text{Var } X)^2 - 3$ erklärt. Sie kann als Maß dafür angesehen werden, wie stark sich die „Wölbung“ der Dichte von der Normalverteilung unterscheidet.

Beispiel:

- (a) Ist $N(\mu, \sigma^2)$ die Verteilung von X , so zeigt man durch Rückführung des k -ten zentralen Moments auf das $(k - 2)$ -te, $(k - 4)$ -te usw. mittels partieller Integration, daß

$$E(X - E X)^k = \begin{cases} 1 \cdot 3 \cdots (k - 1) \cdot \sigma^k, & \text{falls } k \text{ gerade} \\ 0, & \text{falls } k \text{ ungerade} \end{cases}$$

Damit erhält man für die Schiefe und den Exzeß den Wert Null.

- (b) Die Schiefe einer Exponentialverteilung mit Parameter λ berechnet man wie folgt. Es ist

$$E X = 1/\lambda, E X^2 = 2/\lambda^2, E X^3 = 6/\lambda^3$$

und damit

$$\begin{aligned} E(X - E X)^3 &= E\left(X^3 - \frac{3}{\lambda}X^2 + \frac{3}{\lambda^2}X - \frac{1}{\lambda^3}\right) \\ &= \frac{6}{\lambda^3} - \frac{3}{\lambda} \frac{2}{\lambda^2} + \frac{3}{\lambda^2} \frac{1}{\lambda} - \frac{1}{\lambda^3} = \frac{2}{\lambda^3}, \end{aligned}$$

so daß

$$\frac{E(X - EX)^3}{\sqrt{\text{Var } X}^3} = \frac{2/\lambda^3}{\sqrt{1/\lambda^2}^3} = 2.$$

Die Schiefe ist also unabhängig von der Ausfallsrate λ .

8.4. Kovarianz und Korrelation von Zufallsvariablen

Wir haben bisher nur Kenngrößen für einzelne Zufallsvariablen vorgestellt. Bei der **Kovarianz** handelt es sich dagegen um eine Größe, die den Zusammenhang zwischen zwei Zufallsvariablen X und Y mißt. Sie ist erklärt durch:

$$\text{Cov}(X, Y) = E(X - EX)(Y - EY)$$

und wurde schon benutzt zur Berechnung der Varianz einer Summe von Zufallsvariablen. Hat die Kovarianz den Wert Null, so sagt man X und Y sind unkorreliert. Das ist immer so bei stochastisch unabhängigen Zufallsvariablen. Die Umkehrung ist aber i. a. falsch. Bei normalverteilten Zufallsvariablen allerdings beinhaltet die Unkorreliertheit auch die stochastische Unabhängigkeit.

Bei linearen Transformationen der Zufallsvariablen ändert sich der Wert der Kovarianz in der folgenden Weise

$$\text{Cov}(\alpha \cdot X + \beta, \gamma \cdot Y + \delta) = \alpha \cdot \gamma \cdot \text{Cov}(X, Y).$$

Dagegen tritt keine Änderung auf, wenn statt der Kovarianz die sogenannte **Korrelation**

$$\varrho = \varrho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var } X} \cdot \sqrt{\text{Var } Y}} = \text{Corr}(X, Y)$$

benutzt wird. Es handelt sich hier um eine normierte Größe, deren Wert immer zwischen -1 und $+1$ liegt. Zur Berechnung von Kovarianzen ist es noch nützlich zu wissen, daß folgendes gilt:

$$\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y);$$

$$\text{Cov}(X, Y_1 + Y_2) = \text{Cov}(X, Y_1) + \text{Cov}(X, Y_2).$$

Man überlegt sich leicht, daß dies aus den Rechenregeln für den Erwartungswert folgt.

Beispiel: X_1 und X_2 seien unabhängige $N(0, 1)$ -verteilte Zufallsvariable. Wir setzen:

$$Y_1 = X_1 \quad \text{und} \quad Y_2 = \varrho \cdot X_1 + \sqrt{1 - \varrho^2} \cdot X_2 \quad \text{für} \quad \varrho \neq 0.$$

Für diese, dann abhängigen Zufallsvariablen Y_1 und Y_2 berechnet sich die Kovarianz zu

$$\begin{aligned} \text{Cov}(Y_1, Y_2) &= \text{Cov}(X_1, \varrho \cdot X_1 + \sqrt{1 - \varrho^2} \cdot X_2) \\ &= \text{Cov}(X_1, \varrho \cdot X_1) + \text{Cov}(X_1, \sqrt{1 - \varrho^2} \cdot X_2) \\ &= \varrho \cdot \text{Var } X_1 + \sqrt{1 - \varrho^2} \cdot \text{Cov}(X_1, X_2) \\ &= \varrho \cdot 1 + \sqrt{1 - \varrho^2} \cdot 0 = \varrho. \end{aligned}$$

Da $\text{Var } Y_1 = \text{Var } X_1 = 1$ und

$$\begin{aligned}\text{Var } Y_2 &= \text{Var}(\varrho \cdot X_1 + \sqrt{1 - \varrho^2} \cdot X_2) \\ &= \varrho^2 \cdot \text{Var } X_1 + (1 - \varrho^2) \text{Var } X_2 = 1,\end{aligned}$$

ergibt sich für die Korrelation $\varrho(Y_1, Y_2)$ ebenfalls der Wert ϱ . Betrachten wir nun statt Y_1 und Y_2 die Zufallsvariablen

$$Z_1 = \sigma_1 \cdot Y_1 + \mu_1 \quad \text{und} \quad Z_2 = \sigma_2 \cdot Y_2 + \mu_2,$$

so erhalten wir

$$\text{Cov}(Z_1, Z_2) = \sigma_1 \cdot \sigma_2 \cdot \text{Cov}(Y_1, Y_2) = \sigma_1 \cdot \sigma_2 \cdot \varrho \quad \text{und}$$

$$\varrho(Z_1, Z_2) = \varrho(Y_1, Y_2) = \varrho.$$

Die gemeinsame Verteilung der beiden Zufallsvariablen Z_1 und Z_2 aus dem obigen Beispiel ist eine sogenannte **bivariate Normalverteilung** mit der Dichte

$$f_{Z_1, Z_2}(z_1, z_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \cdot \exp \left\{ -\frac{1}{2(1-\varrho^2)} \left(\frac{(z_1 - \mu_1)^2}{\sigma_1^2} + \frac{(z_2 - \mu_2)^2}{\sigma_2^2} - 2\varrho \cdot \frac{(z_1 - \mu_1)(z_2 - \mu_2)}{\sigma_1\sigma_2} \right) \right\}.$$

Als **Randverteilungen** haben wir $N(\mu_1, \sigma_1^2)$ als Verteilung von Z_1 und $N(\mu_2, \sigma_2^2)$ als Verteilung von Z_2 . Die Dichte der gemeinsamen Verteilung von Z_1 und Z_2 lässt sich auch in einer etwas anderen Form schreiben. Dazu führen wir die Vektoren

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad \text{und} \quad \mu = \begin{pmatrix} E Z_1 \\ E Z_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

sowie die **Kovarianzmatrix**

$$\Sigma = \text{Cov } Z = \begin{pmatrix} \text{Var } Z_1 & \text{Cov}(Z_1, Z_2) \\ \text{Cov}(Z_2, Z_1) & \text{Var } Z_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho \\ \sigma_1\sigma_2\varrho & \sigma_2^2 \end{pmatrix}$$

ein. Wir schreiben nun

$$f_Z(z) = \left(\frac{1}{\sqrt{2\pi}} \right)^2 \cdot \frac{1}{\sqrt{|\Sigma|}} \cdot \exp \left(-\frac{1}{2} (z - \mu)' \Sigma^{-1} (z - \mu) \right).$$

Ist $\varrho = 0$, so sind Z_1 und Z_2 unkorreliert und in diesem Fall auch stochastisch unabhängig, denn dann ist

$$f_Z(z) = \frac{1}{2\pi\sigma_1\sigma_2} \cdot \exp \left(-\frac{1}{2} (z - \mu)' \begin{pmatrix} \sigma_1^{-2} & 0 \\ 0 & \sigma_2^{-2} \end{pmatrix} (z - \mu) \right) = f_{Z_1}(z_1) \cdot f_{Z_2}(z_2).$$

Hierbei bezeichnet $|\Sigma|$ die Determinante und Σ^{-1} die inverse Matrix von Σ , vgl. z.B. Hartung/Elpelt (1984, Kap. I).

9. Grenzwertsätze

Wir wollen uns in diesem Abschnitt damit beschäftigen, wie sich der Mittelwert \bar{X}_n von unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n verhält, wenn deren Anzahl n laufend erhöht wird. Die X_i können z. B. Ergebnisse von wiederholten Durchführungen eines Zufallsexperimentes sein.

Das **Gesetz der großen Zahlen** beinhaltet die Aussage, daß sich der Mittelwert \bar{X}_n mit wachsendem n immer mehr um den gemeinsamen Erwartungswert μ der X_i konzentriert. Genau gilt (falls Erwartungswert μ und Varianz σ^2 der X_i existieren):

$$P\left(\left|\frac{1}{n} \cdot \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für alle Zahlen } \varepsilon > 0.$$

Aus der Tschebyscheff-Ungleichung ergibt sich nämlich:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var} \bar{X}_n}{\varepsilon^2} = \frac{\frac{1}{n} \cdot \sigma^2}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0;$$

man sagt: \bar{X}_n konvergiert stochastisch gegen μ .

Beispiel: Ein Zufallsexperiment werde immer wieder durchgeführt und jedesmal interessiere, ob ein Ereignis A eingetreten ist oder nicht. Führen wir die Zufallsvariablen

$$X_i = \begin{cases} 1, & \text{falls beim i-ten Mal A eintritt} \\ 0, & \text{sonst} \end{cases}$$

ein, so ist deren Mittelwert gerade die relative Häufigkeit des Ereignisses A und deren Erwartungswert die Wahrscheinlichkeit p von A. Das Gesetz der großen Zahlen macht hier nun die Aussage, daß sich die relative Häufigkeit des Ereignisses A mit wachsender Anzahl von Wiederholungen des Zufallsexperimentes bei p stabilisiert.

Wird das Zufallsexperiment im obigen Beispiel durch eine Zufallsvariable X beschrieben und wählen wir speziell $A = \{X \leq t\}$, so liefert uns das Gesetz der großen Zahlen die Aussage, daß der Wert der empirischen Verteilungsfunktion an der Stelle t mit wachsendem n stochastisch gegen den entsprechenden Wert der Verteilungsfunktion von X konvergiert, d. h.

$$P(|S_n(t) - F_X(t)| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für alle Zahlen } \varepsilon > 0.$$

Der **Satz von Glivenko-Cantelli** verschärft diese Aussage zu

$$P\left(\sup_t |S_n(t) - F_X(t)| \xrightarrow{n \rightarrow \infty} 0\right) = 1.$$

Im Gegensatz zu den bisher behandelten Sätzen macht der nun folgende Grenzwertsatz eine Aussage über die Form der Verteilung des Mittelwertes. Dazu werden die Zufallsvariablen \bar{X}_n standardisiert, d. h. man bildet

$$Z_n = \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma},$$

so daß unabhängig von n gilt: $E Z_n = 0$ und $\text{Var} Z_n = 1$.

Der **zentrale Grenzwertsatz** sagt nun aus, daß die Verteilungsfunktion von Z_n gegen die Standardnormalverteilung konvergiert, d. h.

$$P(Z_n \leq z) \xrightarrow[n \rightarrow \infty]{} \Phi(z).$$

Man sagt auch: Z_n ist approximativ $N(0, 1)$ -verteilt. Der obige Satz wird oft als Rechtfertigung dafür angeführt, daß die beobachteten Zufallsvariablen als normalverteilt angenommen werden können, insbesondere dann, wenn sie aus dem (additiven) Zusammenwirken vieler Einzeleinflüsse resultieren.

Beispiel: Eine Münze wird tausendmal geworfen. Wir interessieren uns für die Wahrscheinlichkeit, daß höchstens 475 mal die Zahl kommt. Bezeichnet X_i das Ergebnis des i-ten Wurfes, wobei $X_i = 1$ bedeuten soll, daß es sich um die Zahl, und $X_i = 0$, daß es sich um den Kopf handelt, so fragen wir gerade nach

$$P\left(\sum_{i=1}^{1000} X_i \leq 475\right).$$

Da $E X_i = 0,5$ und $\text{Var } X_i = 0,5^2$, erhalten wir dafür mit dem zentralen Grenzwertsatz näherungsweise

$$\begin{aligned} P\left(\sqrt{1000} \cdot \frac{\sum_{i=1}^{1000} X_i - 0,5}{0,5} \leq \sqrt{1000} \cdot \frac{0,475 - 0,5}{0,5}\right) \\ \simeq \Phi\left(\sqrt{1000} \cdot \frac{0,475 - 0,5}{0,5}\right) = \Phi(-1,58) = 0,057. \end{aligned}$$

Einen Eindruck von der Güte der Approximation durch die Normalverteilung liefert uns der Satz von **Berry-Esséen**, der, falls die Größe $E|X_i - \mu|^3$ bekannt ist, uns die folgende Abschätzung ermöglicht

$$\sup_z \left| P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z\right) - \Phi(z) \right| \leq C \cdot \frac{E|X_i - \mu|^3}{\sqrt{n} \cdot \sigma^3}.$$

Dabei ist für die Konstante C bekannt, daß $0,4097 \leq C \leq 0,7975$ gilt. Oft wird man sogar feststellen, daß die Güte der Anpassung noch besser ist als hier angegeben.

Wir wollen nun noch speziell auf die Approximation der $B(n, p)$ -Verteilung eingehen. Wie wir wissen kann man sie als Verteilung der Summe von n unabhängigen $B(1, p)$ -verteilten Zufallsvariablen auffassen und somit wegen der Gültigkeit des zentralen Grenzwertsatzes durch die Normalverteilung approximieren, wie das auch im obigen Beispiel gemacht wurde. Eine zufriedenstellende Annäherung ist in der Regel zu erwarten, wenn $n \cdot p \cdot (1 - p) \geq 9$. Andernfalls läßt sich, zumindest wenn n hinreichend groß ist, noch der **Grenzwertsatz von Poisson** anwenden:

Ist X_n $B(n, p_n)$ -verteilt und gilt $n \cdot p_n \xrightarrow{n \rightarrow \infty} \lambda > 0$, so hat man:

$$P(X_n = k) \xrightarrow[n \rightarrow \infty]{} \frac{\lambda^k}{k!} \cdot e^{-\lambda},$$

also Konvergenz gegen die Poissonverteilung.

Kapitel III: Statistische Schlußweisen

In Kapitel II haben wir uns eingehend damit beschäftigt, wie man einen Zufallsvorgang mit Hilfe einer Zufallsvariablen bzw. eines Zufallsvektors und einer zugehörigen Wahrscheinlichkeitsverteilung beschreiben kann. Dabei haben wir uns praktisch nicht darum gekümmert, wie wir die Informationen zur Festlegung dieser Größen erhalten. Ergeben sich die Zufallsvariablen selbst meist in kanonischer Weise aus einer einfachen Beschreibung des Zufallsexperiments und der Angabe der dabei interessierenden Ergebnisse, so ist die Bestimmung der Wahrscheinlichkeitsverteilung mit oft erheblichen Schwierigkeiten verbunden und zudem mit Fehlern und Ungenauigkeiten behaftet. Das wird schon an einem einfachen Beispiel deutlich. Interessieren wir uns bei einem Zufallsexperiment einmal nur dafür, ob ein bestimmtes Ereignis eintritt oder nicht. Dieses Ereignis kann etwa die erfolgreiche Behandlung mit einem Medikament, das Versagen eines technischen Gerätes, der Sieg beim Kartenspiel oder ähnliches sein. Eine angemessene Beschreibung liefert dann die Zufallsvariable, die den Wert 1 annimmt, wenn das betreffende Ereignis eintritt und sonst 0 ist. Wir können dann zwar annehmen, daß die Zufallsvariable $B(1, p)$ -verteilt ist, vgl. 7.2 in Kap. II, kennen den Wert p aber nicht. Informationen über diese Größe können wir nur durch die möglichst wiederholte Durchführung des Zufallsexperiments erhalten. Dabei wird sich bei einer großen Anzahl von Wiederholungen die relative Häufigkeit für das Auftreten des interessierenden Ereignisses praktisch immer bei p stabilisieren. Es liegt jedoch in der Natur der Sache, daß exakte Aussagen über die Größe von p nicht möglich sind. Das gilt um so mehr, je weniger Beobachtungen uns zur Verfügung stehen. Wir können mit den beobachteten Werten x_1, \dots, x_n nur eine mehr oder weniger genaue Schätzung von p angeben. Dabei lassen sich allerdings oft noch Angaben über die Genauigkeit von solchen **statistischen Schätzungen** machen. Im Prinzip ähnliche Überlegungen hat man, wenn zur Beschreibung des Zufallsexperiments eine andere als die $B(1, p)$ -Verteilung, wie z. B. die Normalverteilung, die Exponentialverteilung, die Poissonverteilung usw., zugrundegelegt werden kann, deren Parameter ebenfalls unbekannt sind.

Mit den hier angesprochenen statistischen Schätzungen läßt sich die Richtigkeit irgendwelcher Aussagen über die Parameter einer Verteilung nicht überprüfen. So kann zwar z. B. die Erfolgswahrscheinlichkeit eines neuen Medikaments aufgrund von Ergebnissen mehrmaliger Anwendungen geschätzt werden. Die Frage, ob dieses Medikament denn nun besser sei als ein bekanntes, dessen Erfolgswahrscheinlichkeit man kennt, läßt sich aber nicht mit genügender Sicherheit durch den einfachen Vergleich der beiden Werte beantworten. Dazu bedarf es eines **statistischen Tests**. Mit solchen Tests läßt sich z. B. auch überprüfen, ob der Parameter einer Verteilung in einem vorgegebenen Intervall liegt, ob er positiv ist, ob er größer ist als der entsprechende, auch unbekannte Parameter einer anderen Verteilung, es lassen sich Hypothesen über die Form einer Verteilung überprüfen usw.

Zur weiteren Veranschaulichung greifen wir noch einmal auf die aus Kap. II bereits bekannte Exponentialverteilung mit Parameter λ zurück. Wir waren dort davon ausgegangen, daß der Parameter λ eine bekannte Größe ist. Diese Größe können wir interpretieren als die Ausfallrate von Glühbirnen, mithin als Parameter eines realen Problems. Die **statistische Modellbildung**, die wir vorgenommen haben, bestand nun darin, λ als einzigen Parameter des Modells zu wählen und als Vertei-

lungstyp die Exponentialverteilung festzulegen. Bei praktischen Fragestellungen sind der oder die Parameter i. d. R. aber unbekannt. Man möchte dann

- den Wert des oder der Parameter bestimmen, z. B. die Ausfallrate von Glühbirnen,
- Aussagen über den oder die Parameter überprüfen, z. B. möchte man wissen, ob die Ausfallrate von Glühbirnen kleiner oder gleich 1/800 [1/Std.] ist.

Hierzu beobachtet man z. B. in einem Experiment n Glühbirnen und stellt deren Lebensdauern x_1, \dots, x_n fest. Diese Lebensdauern werden dann angesehen als **Realisationen** von lebensdauerbeschreibenden Zufallsvariablen X_1, \dots, X_n , die unabhängig und identisch exponentialverteilt sind mit gleichem unbekannten Parameter λ . Man sagt dann auch: x_1, \dots, x_n ist eine **einfache Zufallsstichprobe** bzw. kurz **Stichprobe** vom Umfang n aus einer Exponentialverteilung mit Parameter λ . Anhand dieser Stichprobe x_1, \dots, x_n werden nun mit Hilfe statistischer Verfahren obige Fragestellungen beantwortet.

In diesem Kapitel wollen wir uns einige grundsätzliche Gedanken zu den statistischen Schätzungen und den statistischen Tests machen. Man spricht hier, im Unterschied zur deskriptiven Statistik, in der man primär an einer Beschreibung des Datenmaterials interessiert ist, von der **induktiven** (oder **schließenden**) **Statistik**.

1. Schätzen von Parametern

Aufgrund der Beobachtungen x_1, \dots, x_n von Zufallsgrößen X_1, \dots, X_n , die z. B. die Wiederholungen eines Zufallsexperiments beschreiben können, soll eine „möglichst gute“ Schätzung eines unbekannten Parameters θ der Verteilungen der X_i angegeben werden. Wir suchen also eine Funktion $\hat{\theta}(x_1, \dots, x_n)$, die der Zufallsstichprobe x_1, \dots, x_n einen Wert zuordnet, der möglichst nahe oder gleich dem wahren Parameter θ ist, für den wir uns interessieren.

Beispiel: Beim Sportunterricht wird die Zeit des Siegers beim 100 m-Lauf von 10 Schülern gleichzeitig gestoppt. Dabei ergaben sich die Ergebnisse der **Tab. 1**.

Tab. 1: Gestoppte Zeiten des Siegers beim 100 m-Lauf

Nr. der Messung	1	2	3	4	5
Gestoppte Zeit in sec	11,0	11,1	10,7	11,5	11,2
Nr. der Messung	6	7	8	9	10
Gestoppte Zeit in sec	10,9	10,6	11,2	11,4	10,8

Wie man sieht, sind die obigen Messungen nicht identisch. Wir können aber annehmen, daß die Meßwerte x_i Realisationen von unabhängigen $N(\mu, \sigma^2)$ -verteilten Zufallsgrößen mit unbekanntem Mittelwert μ und unbekannter Varianz σ^2 sind. Das bedeutet, daß die Meßwerte zufällig um die wahre Zeit μ schwanken. Es liegt deshalb nahe, diesen Wert durch den Mittelwert der Beobachtungen zu schätzen, d. h. wir verwenden die Schätzfunktion

$$\hat{\mu}(x_1, \dots, x_{10}) = \frac{1}{10} \sum_{i=1}^{10} x_i = \bar{x}.$$

Für dieses Beispiel ergibt sich dann $\hat{\mu} = 11,04$.

Da eine statistische Schätzung auf den Ergebnissen eines Zufallsexperiments beruht, unterliegt auch sie dem Zufall. Eine vernünftig erscheinende, oft auch realisierbare Forderung an solch eine Schätzfunktion ist, daß sie wenigstens im Mittel den richtigen Wert liefert, d.h. es sollte gelten

$$E\hat{\theta}(X_1, \dots, X_n) = \theta.$$

Eine Schätzfunktion mit dieser Eigenschaft nennt man **erwartungstreu**.

Beispiel: Verwenden wir zur Schätzung des Parameters μ einer $N(\mu, \sigma^2)$ -Verteilung den Mittelwert der Beobachtungen, also die Schätzfunktion $\hat{\theta}(x_1, \dots, x_n) = \bar{x}$, so ist diese Schätzung erwartungstreu, denn

$$E\hat{\theta}(X_1, \dots, X_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n EX_i = \frac{1}{n} n \cdot \mu = \mu.$$

Eine weitere, wünschenswerte Eigenschaft einer Schätzmethode ist, daß sie immer genauere Ergebnisse liefert, je mehr Beobachtungen zur Verfügung stehen. Bezeichnen wir die Schätzfunktion für θ , die auf n Beobachtungen beruht, mit $\hat{\theta}_n$, so sollte gelten

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{für jedes } \varepsilon > 0.$$

Der Parameter θ wird dann asymptotisch richtig geschätzt. Man sagt, der Schätzer $\hat{\theta}_n$ sei **konsistent** für θ .

Beispiel: Nach dem Gesetz der großen Zahlen, vgl. Abschnitt 9 in Kap. II, konvergiert der Mittelwert der Beobachtungen von identisch verteilten, unabhängigen Zufallsvariablen X_1, \dots, X_n stochastisch gegen den Erwartungswert der X_i . Sind die X_i $N(\mu, \sigma^2)$ -erteilt, so ist aber gerade $EX_i = \mu$. Der Mittelwert der Beobachtungen ist also ein konsistenter Schätzer für den Parameter μ einer Normalverteilung.

Die hier vorgestellten Konzepte **Erwartungstreue** und **Konsistenz** sagen noch relativ wenig über die Güte eines Schätzers aus. So bleiben diese Eigenschaften z.B. oft dann noch erhalten, wenn etwa nur jede zweite Beobachtung berücksichtigt wird. Als Beurteilungskriterium für die Güte eines Schätzers eignet sich dagegen der **mittlere quadratische Fehler** (**mean squared error** = **MSE**):

$$MSE(\theta, \hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

Man kann auch schreiben

$$MSE = \text{Varianz} + (\text{Bias})^2.$$

Die Größe $E\hat{\theta} - \theta$ nennt man nämlich **Bias** oder **Verzerrung** des Schätzers $\hat{\theta}$ und es gilt

$$\begin{aligned} E(\hat{\theta} - \theta)^2 &= E\hat{\theta}^2 - 2(E\hat{\theta})\theta + \theta^2 \\ &= E\hat{\theta}^2 - (E\hat{\theta})^2 + (E\hat{\theta})^2 - 2(E\hat{\theta})\theta + \theta^2 \\ &= \text{Var}\hat{\theta} + (E\hat{\theta} - \theta)^2. \end{aligned}$$

Ist $\hat{\theta}$ erwartungstreu für θ , so ist der Bias natürlich Null und somit der mittlere quadratische Fehler gerade die Varianz des Schätzers.

Beispiel: Der Schätzer \bar{x} für den Mittelwert μ der $N(\mu, \sigma^2)$ -Verteilung hat den mittleren quadratischen Fehler

$$\text{MSE} = \text{Var } \bar{X} = \left(\frac{1}{n} \right)^2 \sum_{i=1}^n \text{Var } X_i = \left(\frac{1}{n} \right)^2 n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

Bei einem erwartungstreuen Schätzer ist eine umso stärkere Konzentration der Schätzungen um den wahren Wert zu erwarten, je kleiner dessen Varianz ist. Man ist deshalb bestrebt, einen **erwartungstreuen Schätzer mit gleichmäßig minimaler Varianz** zu erhalten, d.h. einen Schätzer $\hat{\theta}$, so daß für eine beliebige andere erwartungstreue Schätzfunktion $\tilde{\theta}$ gilt: $\text{Var } \hat{\theta} \leq \text{Var } \tilde{\theta}$. Diese Eigenschaft hat z.B. der Mittelwert zur Schätzung des Parameters μ der $N(\mu, \sigma^2)$ -Verteilung.

Nach den bisherigen Bemerkungen über die Eigenschaften statistischer Schätzer sollen nun Prinzipien zur Herleitung solcher Schätzer dargestellt werden. Dabei kann man durchaus bei der Anwendung verschiedener Prinzipien zu den gleichen Schätzern gelangen.

A. Momentenmethode

Man erhält eine Schätzung nach der Momentenmethode, wenn man den zu schätzenden Parameter durch die Momente der Verteilung ausdrückt und diese in der so erhaltenen Formel durch die empirischen Momente ersetzt. Dabei ist das **k-te empirische Moment** der Stichprobe x_1, \dots, x_n gegeben durch

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k.$$

Beispiel: Die Lebensdauer von Glühbirnen kann als exponentialverteilt angenommen werden, mit einem als Ausfallrate zu interpretierenden Parameter λ . Aufgrund der beobachteten Lebensdauern x_1, \dots, x_n von n Glühbirnen soll nun dieser Parameter geschätzt werden. Wir wissen, daß der Erwartungswert der Exponentialverteilung gerade $1/\lambda$ beträgt. Der interessierende Parameter λ kann also durch das 1. Moment ausgedrückt werden in der Form

$$\lambda = \frac{1}{E X_i}.$$

Als Schätzung nach der Momentenmethode erhält man so

$$\hat{\lambda} = \frac{1}{m_1} = \frac{1}{\bar{x}} = \frac{n}{\sum_{i=1}^n x_i}.$$

B. Maximum-Likelihood-Methode

Wir wollen uns hier zunächst mit der Schätzung für den Parameter θ einer **diskreten** Verteilung beschäftigen. Dazu bestimmen wir die Wahrscheinlichkeit für das Auftreten der beobachteten Stichprobe x_1, \dots, x_n unter der Annahme einer jeden der überhaupt in Frage kommenden Verteilungen. Der **Maximum-Likelihood-Schätzer (ML-Schätzer)** ist gerade der Parameter derjenigen Verteilung, die dabei den größten Wert liefert, bei deren Vorliegen die beobachtete Stichprobe also am ehesten zu

erwarten ist. Wir wollen nun im folgenden Beispiel einen derartigen Schätzer bestimmen.

Beispiel: Um den Anteil der übergewichtigen Jugendlichen zu bestimmen, wird eine Stichprobe von 1000 Jugendlichen untersucht. Setzt man $x_i = 1$, wenn bei der i-ten Person Übergewicht festgestellt wurde und sonst $x_i = 0$, so kann man annehmen, daß die x_i Realisationen von unabhängigen $B(1, p)$ -verteilten Zufallsvariablen sind. Es interessiert der Parameter p , der den Anteil der übergewichtigen Jugendlichen angibt. Die Wahrscheinlichkeit, daß die beobachtete Stichprobe auftritt, wenn der Parameter p vorliegt, ist

$$P(X_1 = x_1, \dots, X_{1000} = x_{1000}) = \binom{1000}{k} p^k (1-p)^{1000-k}$$

mit $k = \sum_{i=1}^{1000} x_i$.

Der ML-Schätzer ist nun gerade der Wert p , der diese Funktion, die man auch **Likelihood** L der Beobachtung nennt, maximiert. Man kann ihn natürlich auch genauso gut durch die Maximierung des Logarithmus dieser Funktion, der sogenannten **log-Likelihood**, erhalten. Das bedeutet hier, daß

$$\begin{aligned} \ln L_{1000}(p) &= \ln \left[\binom{1000}{k} p^k (1-p)^{1000-k} \right] \\ &= \ln \left(\frac{1000}{k} \right) + k \ln p + (1000 - k) \ln (1-p) \end{aligned}$$

in Abhängigkeit von p zu maximieren ist. Da diese Funktion in p beliebig oft differenzierbar ist, bestimmen wir dazu die Nullstellen der 1. Ableitung.

Es ergibt sich aus

$$\frac{k}{\hat{p}} - \frac{1000 - k}{1 - \hat{p}} = 0$$

die einzige Nullstelle

$$\hat{p} = \frac{k}{1000} = \frac{1}{1000} \sum_{i=1}^{1000} x_i.$$

Für ein relatives Extremum in $(0, 1)$ kommt also nur dieser Wert in Betracht. Eine Untersuchung der 2. Ableitung und eine Randwertbetrachtung, d.h. eine Betrachtung der Funktion an den Randpunkten $p = 0$ und $p = 1$, gibt uns die Gewißheit, daß hier tatsächlich ein absolutes Maximum vorliegt. Der **Maximum-Likelihood-Schätzer** ist also gerade der relative Anteil übergewichtiger Personen in der Stichprobe. Da p der Erwartungswert der X_i ist, hätten wir nach der **Momentenmethode** den gleichen Schätzer erhalten.

Wir kommen nun zur Schätzung für den Parameter θ einer **stetigen** Verteilung. Analog zum diskreten Fall heißt hier der Wert der gemeinsamen Dichtefunktion der X_i für die vorliegende Stichprobe $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ **Likelihood** der Beobachtung x_1, \dots, x_n . Auch hier ist der Schätzer $\hat{\theta}$ ein **Maximum-Likelihood-Schätzer**, falls er die Likelihood der Stichprobe maximiert.

Beispiel: X_1, \dots, X_n seien stochastisch unabhängige, auf $[0; \theta]$ gleichverteilte Zufallsvariablen, d.h. mit der Dichte

$$f_{X_i}(x_i) = \begin{cases} 1/\theta, & 0 \leq x_i \leq \theta \\ 0, & \text{sonst} \end{cases}$$

Wir wollen nun den ML-Schätzer für θ bestimmen. Als Likelihood der Stichprobe x_1, \dots, x_n ergibt sich für

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) = \begin{cases} 1/\theta^n, & \text{falls für alle } x_i \text{ gilt } 0 \leq x_i \leq \theta \\ 0, & \text{sonst} \end{cases}$$

Diese Funktion wird maximal, wenn $\theta \geq x_1, \dots, x_n$ und θ dabei möglichst klein ist. Als ML-Schätzer ergibt sich also

$$\hat{\theta} = \max_{1 \leq i \leq n} x_i.$$

Dieses Beispiel zeigt, daß man nicht jeden ML-Schätzer mittels der Differentialrechnung bestimmen kann.

C. Methode der kleinsten Quadrate

Wir gehen davon aus, daß die Erwartungswerte der Zufallsvariablen Y_1, \dots, Y_n , deren Realisierungen wir beobachten, in einer uns bekannten Weise von dem interessierenden Parameter(vektor) abhängen, also daß gilt

$$EY_i = g_i(\theta)$$

mit bekannten Funktionen g_i .

Im einfachsten Fall, so z.B. wenn die Y_i unabhängig und identisch $N(\theta, \sigma^2)$ -verteilt sind, ist $g_i(\theta) = \theta$ für alle i . Eine Beziehung der Form

$$EY_i = \theta_1 x_i + \theta_2$$

erhält man, wenn z. B. Y_i das Gewicht einer männlichen Person der Größe x_i angibt. Die bekannte Faustregel: {Gewicht in kg = Körpergröße in cm - 100} besagt dann, daß $\theta_1 = 1$ und $\theta_2 = -100$ ist. Auch alle anderen Regressions- und Varianzanalysemmodelle, vgl. Kap. X und XI, passen in den hier abgesteckten Rahmen.

Meist werden Realisationen einer Zufallsvariablen am häufigsten in der Nähe ihres Erwartungswertes zu beobachten sein, dies gilt umso mehr, je kleiner die Varianz dieser Zufallsgröße ist. Hier bedeutet das, daß die Beobachtungen y_i in der Regel nahe bei $g_i(\theta)$ zu erwarten sind. Es ist deshalb naheliegend, als Schätzung $\hat{\theta}$ für den unbekannten Parameter θ einen Wert zu wählen, der den Abstand zwischen $g_i(\hat{\theta})$ und y_i möglichst klein macht, d.h. $\hat{\theta}$ so zu bestimmen, daß gilt

$$\sum_{i=1}^n (y_i - g_i(\hat{\theta}))^2 \leq \sum_{i=1}^n (y_i - g_i(\theta))^2$$

für alle möglichen Parameterwerte θ .

Man nennt dann diesen Wert $\hat{\theta}$ einen Schätzer nach der **Methode der kleinsten Quadrate**.

Beispiel: Man hat eine Stichprobe von n_1 Widerständen mit dem unbekannten Mittel θ [Ohm]. Außerdem ist eine zweite Stichprobe des Umfangs n_2 von Zwi-

schensteckern vorhanden, in denen jeweils $a = 3$ solcher Widerstände hintereinandergeschaltet sind und die deshalb den Mittelwert $a \cdot \theta$ [Ohm] besitzen. Wir bezeichnen mit y_1, \dots, y_{n_1} die Werte der ersten und mit $y_{n_1+1}, \dots, y_{n_1+n_2}$ die Werte der zweiten Stichprobe. Wegen

$$EY_i = g_i(\theta) = \begin{cases} \theta & , \text{ falls } 1 \leq i \leq n_1 \\ a \cdot \theta & , \text{ falls } n_1 + 1 \leq i \leq n_1 + n_2 \end{cases}$$

ist der Schätzer $\hat{\theta}$ nach der Methode der kleinsten Quadrate eine Lösung des Minimierungsproblems

$$\sum_{i=1}^{n_1} (y_i - \theta)^2 + \sum_{i=n_1+1}^{n_1+n_2} (y_i - a\theta)^2 \rightarrow \min.$$

Die erste Ableitung der zu minimierenden Funktion hat genau eine Nullstelle, die sich aus

$$-2 \sum_{i=1}^{n_1} (y_i - \hat{\theta}) - 2a \sum_{i=n_1+1}^{n_1+n_2} (y_i - a\hat{\theta}) = 0$$

ergibt zu

$$\hat{\theta} = \frac{\sum_{i=1}^{n_1} y_i + a \sum_{i=n_1+1}^{n_1+n_2} y_i}{n_1 + a^2 n_2}.$$

Man zeigt, daß dies tatsächlich das Minimum ist und somit der gesuchte Schätzer nach der Methode der kleinsten Quadrate. Es gilt übrigens, daß $\hat{\theta}$ erwartungstreu ist, da nämlich

$$E\hat{\theta} = \frac{n_1 \cdot \theta + a^2 n_2 \theta}{n_1 + a^2 n_2} = \theta.$$

Als Schätzung erhält man z. B. mit den Werten $n_1 = 10, n_2 = 20$,

$$\frac{1}{n_1} \sum_{i=1}^{n_1} y_i = 100 \quad \text{und} \quad \frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} y_i = 290:$$

$$\hat{\theta} = \frac{10 \cdot 100 + 3 \cdot 20 \cdot 290}{10 + 9 \cdot 20} = \frac{18400}{190} = 96,8.$$

2. Konfidenzintervalle

Anstatt einen unbekannten Parameter θ durch einen einzigen Wert $\hat{\theta}$ zu schätzen, von dem man höchstens weiß, daß er meist in der Nähe von θ liegt, möchte man oft lieber einen möglichst kleinen Bereich angeben, in dem der gesuchte Parameter zu finden ist. Da man die Informationen über θ durch ein Zufallsexperiment erhält, ist es natürlich i. a. unmöglich, ein Verfahren anzugeben, das immer so einen Bereich liefert. Es gibt jedoch Methoden, die nur mit einer kleinen Wahrscheinlichkeit Bereiche liefern, die den unbekannten Parameter θ nicht enthalten. Ist diese **Irrtumswahrscheinlichkeit** höchstens gleich α , erhalten wir also mit einer Wahrscheinlichkeit von mindestens $1 - \alpha$ einen Bereich, in dem der unbekannte Wert liegt, so nennt man einen mit dieser Methode gewonnenen Bereich **Konfidenzbereich zum Niveau** $1 - \alpha$. Ist dieser Bereich ein Intervall, so spricht man von einem **Konfidenzintervall**.

zum Niveau $1 - \alpha$, kurz auch $1 - \alpha$ Konfidenzintervall (auch **Vertrauensbereich** bzw. **Vertrauensintervall** genannt).

Eine häufig erfolgreiche Vorgehensweise zur Konstruktion von Konfidenzintervallen soll nun erklärt und anschließend an einem Beispiel demonstriert werden.

Zunächst verschafft man sich einen Punktschätzer $\hat{\theta}$ für den interessierenden unbekannten Parameter θ . Die Verteilung der Schätzfunktion $\hat{\theta}$ wird normalerweise von dem unbekannten Parameter θ abhängen, ebenso wie deren Erwartungswert $\mu(\theta)$ und deren Varianz $\sigma^2(\theta)$. Bildet man nun die standardisierte Zufallsvariable

$$\frac{\hat{\theta} - \mu(\theta)}{\sigma(\theta)},$$

so hat diese den Erwartungswert 0 und die Varianz 1 und, wenn man Glück hat, ist sogar deren Verteilung unabhängig von θ . Dann lassen sich Zahlen u_1 und u_2 finden, so daß

$$P\left(u_1 \leq \frac{\hat{\theta} - \mu(\theta)}{\sigma(\theta)} \leq u_2\right) \geq 1 - \alpha \quad \text{für alle } \theta.$$

Diejenigen Werte θ für die gilt

$$u_1 \leq \frac{\hat{\theta} - \mu(\theta)}{\sigma(\theta)} \leq u_2$$

bilden nun einen Konfidenzbereich zum Niveau $1 - \alpha$. Damit dieser Bereich nicht zu groß wird, sollten u_1 und u_2 so gewählt werden, daß die obige Wahrscheinlichkeit möglichst nahe bei $1 - \alpha$ liegt, bzw. gleich $1 - \alpha$ ist. Ist der so erhaltene Bereich ein Intervall, dann hat man ein $1 - \alpha$ Konfidenzintervall. Man erhält oft sogenannte **einseitige Konfidenzintervalle**, wenn man $u_1 = -\infty$ oder $u_2 = \infty$ setzt. Es soll noch angemerkt werden, daß man sich bei der Transformation der Zufallsvariablen $\hat{\theta}$ nicht auf die hier angegebene beschränken muß. Es ist allein ausschlaggebend, daß man schließlich eine Zufallsvariable erhält, deren Verteilung unabhängig von θ ist.

Beispiel: Betrachten wir noch einmal das Beispiel aus Abschnitt 1, in dem die Zeit eines 100 m-Läufers von 10 Schülern gleichzeitig gestoppt wurde. Die Meßwerte x_1, \dots, x_n können als Realisationen von $n = 10$ unabhängigen $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen aufgefaßt werden. Wir setzen voraus, daß die Varianz $\sigma^2 = 0,09$ bekannt ist. Die Schätzfunktion \bar{X} für den Parameter μ hat den Erwartungswert μ und die Varianz $\frac{1}{n} \sigma^2$, so daß

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

nach $N(0, 1)$ verteilt ist und damit

$$P\left(-u_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq u_{1-\alpha/2}\right) = 1 - \alpha,$$

wobei $u_{1-\alpha/2}$ das $1 - \alpha/2$ Quantil der Standardnormalverteilung ist. Das $1 - \alpha$ Konfidenzintervall für den Parameter μ ergibt sich nun zu

$$\left[\bar{x} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}; \bar{x} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right].$$

Wir sehen, daß die Länge dieses Intervalls mit wachsendem Stichprobenumfang immer kleiner wird. Soll eine vorgegebene Länge d nicht überschritten werden, so muß gelten

$$n \geq \left(\frac{2\sigma u_{1-\alpha/2}}{d} \right)^2,$$

vgl. auch Kap. IV, Abschnitt 1.

Die einseitigen Konfidenzintervalle ergeben sich wegen

$$P\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \geq -u_{1-\alpha}\right) = P\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq u_{1-\alpha}\right) = 1 - \alpha$$

zu

$$\left(-\infty; \bar{x} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\right] \quad \text{und} \quad \left[\bar{x} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}; \infty\right).$$

Aus der Tabelle im Tafelanhang entnimmt man für $\alpha = 0,05$ den Wert $u_{1-\alpha/2} = 1,96$, so daß sich mit den Werten aus Tab. 1 das 0,95 Konfidenzintervall $[10,85; 11,23]$ ergibt. Für die einseitigen Konfidenzintervalle erhält man wegen $u_{0,95} = 1,645$ die Bereiche $(-\infty; 11,20)$ und $[10,88; \infty)$.

Die oben beschriebene Prozedur zur Konstruktion eines Konfidenzintervalls ist jedoch nicht immer anwendbar. So kann es passieren, daß die Schätzfunktion $\hat{\theta}_n$ sich nicht in eine Größe transformieren läßt, deren Verteilung unabhängig ist von den unbekannten Parametern. Man kann aber oft noch, zumindest für große Stichprobenumfänge, die Verteilung der Zufallsvariablen

$$\sqrt{n} \frac{\hat{\theta}_n - \mu(\theta)}{\sigma} \quad \text{mit} \quad \sigma(\theta) = \frac{\sigma}{\sqrt{n}}$$

durch die Standardnormalverteilung approximieren. Dazu muß gelten

$$P\left(\sqrt{n} \frac{\hat{\theta}_n - \mu(\theta)}{\sigma} \leq z\right) \xrightarrow{n \rightarrow \infty} \Phi(z).$$

Verwenden wir nun das sogenannte **asymptotische Konfidenzintervall** zum Niveau $1 - \alpha$

$$\left[\hat{\theta}_n - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}; \hat{\theta}_n + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} \right],$$

so konvergiert die Wahrscheinlichkeit dafür, daß man auf diese Weise ein Intervall erhält, in dem der wahre Parameter liegt, mit wachsendem Stichprobenumfang n gegen $1 - \alpha$. Das gleiche gilt auch, wenn der in der Regel unbekannte Parameter σ durch einen konsistenten Schätzer $\hat{\sigma}_n$ ersetzt wird. In so einem Fall erreicht man mit kleineren Stichprobenumfängen unter Umständen eine Verbesserung, wenn statt der Quantile der $N(0, 1)$ -Verteilung passende Quantile der t-Verteilung benutzt werden, s. Kap. IV, Abschnitt 1.3. Asymptotische Konfidenzintervalle sollte man jedoch nur dann benutzen, wenn man sonst keine andere Möglichkeit sieht, sich ein Konfidenzintervall zu beschaffen. Deren tatsächliches Niveau ist nämlich nicht bekannt. Man weiß nur, daß es bei „genügend großen“ Stichprobenumfängen „in der Nähe“ von $1 - \alpha$ liegt.

3. Prognose- und Toleranzintervalle

Im Abschnitt 2 haben wir uns mit Konfidenzintervallen beschäftigt, also Bereichen, bei denen wir mit hoher vorgegebener Wahrscheinlichkeit vermuten können, daß sie den uns unbekannten Parameter einer Verteilung enthalten. Diese Bereiche wurden aufgrund der Realisierungen x_1, \dots, x_n einer Zufallsvariablen mit eben dieser Verteilung gewonnen. Oft interessiert aber eigentlich gar nicht der Parameter dieser Verteilung, sondern man möchte wissen, wie die Ergebnisse von zukünftigen Zufallsexperimenten sind, man möchte voraussagen, welche Realisierung als nächstes beobachtet wird, wie der Mittelwert oder irgend eine andere Funktion einer zukünftigen Stichprobe y_1, \dots, y_m sein wird o. ä. Kennen wir die Verteilung der Zufallsvariablen, dann bereitet es meist keine Schwierigkeiten, ein Intervall anzugeben, das solche zukünftigen Ergebnisse mit hoher Wahrscheinlichkeit enthält. So liegt z. B. eine einzelne Beobachtung mit der Wahrscheinlichkeit p in dem Intervall, das als untere Grenze das $\frac{1-p}{2}$ Quantil und als obere das $\frac{1+p}{2}$ Quantil der Verteilung besitzt.

So ein Intervall nennt man übrigens **Toleranzintervall mit minimaler Trefferwahrscheinlichkeit p** (und Sicherheitswahrscheinlichkeit 1). In der Regel sind jedoch die Parameter der Verteilung unbekannt und müssen somit aufgrund einer Stichprobe x_1, \dots, x_n geschätzt werden. Man wird dann nicht immer, auch wenn die Intervalle wegen der Unsicherheit der Parameterschätzung etwas größer gewählt werden, als wenn die Verteilung bekannt wäre, ein solches mit der Trefferwahrscheinlichkeit p erhalten. Es ist jedoch oft möglich ein Verfahren anzugeben, das mit einer vorgegebenen Wahrscheinlichkeit $1 - \alpha$ solche Intervalle liefert. Man spricht dann von einem **Toleranzintervall mit minimaler Trefferwahrscheinlichkeit p und Sicherheitswahrscheinlichkeit $1 - \alpha$** . Die Wahrscheinlichkeit auf diese Art das nächste Ergebnis richtig zu **prognostizieren** ist mindestens $p \cdot (1 - \alpha)$. Allgemein spricht man von einem **Prognoseintervall zum Niveau $1 - \alpha$** (mit Trefferwahrscheinlichkeit $1 - \alpha$), wenn die Wahrscheinlichkeit, das nächste Ergebnis richtig zu prognostizieren, mindestens $1 - \alpha$ beträgt. So ein Intervall soll im folgenden Beispiel angegeben werden.

Beispiel: Es soll der Mittelwert \bar{y} einer zukünftigen Stichprobe y_1, \dots, y_m von Realisationen unabhängiger $N(\mu, \sigma^2)$ -verteilter Zufallsvariablen prognostiziert werden. Dabei sei σ^2 bekannt und μ werde aus einer Stichprobe x_1, \dots, x_n durch deren

Mittelwert \bar{x} geschätzt. Dann gilt, da $\text{Var } \bar{X} = \frac{1}{n} \sigma^2$ und $\text{Var } \bar{Y} = \frac{1}{m} \sigma^2$ und somit

$$\text{Var}(\bar{X} - \bar{Y}) = \frac{1}{n} \sigma^2 + \frac{1}{m} \sigma^2 \text{ gilt,}$$

$$\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

ist nach $N(0, 1)$ verteilt. Also erhält man

$$1 - \alpha = P\left(-u_{1-\alpha/2} \leq \frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \leq u_{1-\alpha/2}\right)$$

$$= P\left(\bar{X} - \sqrt{\frac{1}{n} + \frac{1}{m}} \sigma u_{1-\alpha/2} \leq \bar{Y} \leq \bar{X} + \sqrt{\frac{1}{n} + \frac{1}{m}} \sigma u_{1-\alpha/2}\right).$$

Demnach ist

$$\left[\bar{x} - \sqrt{\frac{1}{n} + \frac{1}{m}} \sigma u_{1-\alpha/2}; \bar{x} + \sqrt{\frac{1}{n} + \frac{1}{m}} \sigma u_{1-\alpha/2} \right]$$

ein Prognoseintervall für den Mittelwert \bar{y} zum Niveau $1 - \alpha$. Für $m = 1$ ist das ein Prognoseintervall für eine einzelne künftige Beobachtung.

4. Statistische Tests

In diesem Abschnitt geht es um Verfahren, die die Richtigkeit irgendwelcher Aussagen über die Verteilung einer Zufallsvariablen überprüfen sollen, etwa, ob ein Parameter größer oder kleiner als ein bestimmter Wert ist oder in einem bestimmten Intervall liegt, ob die Zufallsvariable normalverteilt ist usw. Solche Verfahren können natürlich nicht immer die richtige Entscheidung liefern, da die Ergebnisse eines zufälligen Vorgangs als Entscheidungsgrundlage dienen. Beim Vorliegen einiger der möglichen Verteilungen ist die Wahrscheinlichkeit falsch zu entscheiden jedoch durch einen kleinen vorgegebenen Wert α beschränkt. Nehmen wir einmal an, es sollte überprüft werden, ob ein Parameter der tatsächlich vorliegenden Verteilung in einem bestimmten Bereich liegt oder nicht. Diese beiden verschiedenen Möglichkeiten wollen wir **Null-** und **Alternativhypothese** nennen und mit H_0 und H_1 bezeichnen. Interessieren wir uns z. B. dafür, ob der Parameter μ einer $N(\mu, \sigma^2)$ -Verteilung größer oder kleiner als ein Wert μ_0 ist, so könnte etwa $\mu \leq \mu_0$ die Nullhypothese und $\mu > \mu_0$ die Alternativhypothese sein.

Man beobachtet nun eine Stichprobe x_1, \dots, x_n und soll sich aufgrund dieser Werte für eine der beiden Hypothesen entscheiden. Dabei kann man einerseits den Fehler machen, daß, obwohl H_0 vorliegt, man sich für H_1 entscheidet, also einen sogenannten **Fehler 1. Art** machen. Andererseits kann man sich auch fälschlicherweise für H_0 entscheiden. Man spricht dann von einem **Fehler 2. Art**, vgl. auch **Tab. 2**. Man weiß natürlich nicht, ob man in einer konkreten Situation einen Fehler macht, sondern lediglich, welcher Art dieser ist. Ist aber bekannt, daß das verwandte Entscheidungsverfahren nur mit einer Wahrscheinlichkeit von **höchstens α** den Fehler 1. Art (α -Fehler) macht, so spricht man von einem **Test zum Niveau α** (bzw. **Niveau- α -Test**), egal mit welcher Wahrscheinlichkeit β der Fehler 2. Art (β -Fehler) vorkommt. Man ist jedoch darum bemüht, solche Tests zu finden, die neben

Tab. 2: Fehlerarten beim Testen

		H_0	H_1
Entscheidung für	Es liegt vor		
		–	Fehler 2. Art (β -Fehler)
H_0			–
H_1	Fehler 1. Art (α -Fehler)		–

der Einhaltung des vorgegebenen **Signifikanzniveaus** α die Wahrscheinlichkeit für den Fehler 2. Art möglichst klein machen. Über die Größe dieser Wahrscheinlichkeit werden oft keine Angaben gemacht.

Wie sieht so ein Test nun konkret aus? Nehmen wir an, es soll getestet werden

$$H_0: \theta \leq \theta_0 \quad \text{gegen} \quad H_1: \theta > \theta_0,$$

d.h. man möchte wissen, ob der Parameter θ der vorliegenden Verteilung größer oder kleiner als ein Wert θ_0 ist. In so einem Fall ordnet man oft mit Hilfe einer sogenannten **Teststatistik (Prüfgröße)** T den Beobachtungen x_1, \dots, x_n eine Zahl $T(x_1, \dots, x_n)$ zu, bei der ein großer Wert eher für das Vorliegen von H_1 und ein kleiner eher für das von H_0 spricht. Außerdem sucht man eine Zahl $c_{1-\alpha}$, für die gilt

$$P_\theta(T > c_{1-\alpha}) \leq \alpha \quad \text{für alle } \theta \in H_0,$$

wobei der Index θ andeuten soll, daß die Wahrscheinlichkeit bei Vorliegen von θ gemeint ist.

Trifft man nun die Entscheidung für H_1 , falls $T > c_{1-\alpha}$, und die Entscheidung für H_0 , falls $T \leq c_{1-\alpha}$, so handelt es sich bei dieser **Entscheidungsregel** um einen Test zum Niveau α . Der Wert $c_{1-\alpha}$ heißt **kritischer Wert** des Tests, der Bereich $T > c_{1-\alpha}$ **Ablehnbereich** und der Bereich $T \leq c_{1-\alpha}$ **Annahmebereich** des Tests. Entscheidet man sich für H_0 , so sagt man auch, die „Nullhypothese wird angenommen“ bzw. die „Nullhypothese kann nicht verworfen werden“ und bei einer Entscheidung für H_1 , die „Nullhypothese wird verworfen“ bzw. die „Alternative wird angenommen“ bzw. die „Alternativhypothese H_1 ist signifikant (zum Niveau α)“ bzw. die „Alternative ist zum Niveau α statistisch gesichert“.

Beispiel: Eine Maschine verpackt Zucker in Tüten zu je 500 g. Man kann dabei annehmen, daß die tatsächliche Füllmenge $N(\mu, \sigma^2)$ -verteilt ist, also um einen mittleren Wert μ schwankt. Der Parameter σ sei vom Hersteller der Maschine mit 2,7 g angegeben. Man möchte nun wissen, ob die Maschine im Mittel mehr oder weniger als $\mu_0 = 500$ g abfüllt. Dazu wird die Füllmenge von 9 Tüten ermittelt, die von dieser Maschine verpackt wurden. Die Werte sind in Tab. 3 wiedergegeben.

Tab. 3: Füllmengen von Zuckertüten

Laufende Nr. der Messung	1	2	3	4	5	6	7	8	9
Füllmenge x_i in g	506	502	500	505	499	505	501	504	498

Wir formulieren das Testproblem

$$H_0: \mu \leq \mu_0 \quad \text{gegen} \quad H_1: \mu > \mu_0.$$

Ist der Mittelwert \bar{x} der beobachteten Stichprobe x_1, \dots, x_n relativ groß, so spricht das sicherlich für das Vorliegen von H_1 . Ist er dagegen relativ klein, so würde man eher annehmen, daß es sich um H_0 handelt. Wir benutzen deshalb als Teststatistik die Größe $\sqrt{n}(\bar{X} - \mu_0)/\sigma$, die $N(0, 1)$ -verteilt ist, wenn der wahre Parameter μ_0 ist. Es gilt also, falls $\mu = \mu_0$ ist,

$$P_\mu \left(\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} > u_{1-\alpha} \right) = \alpha.$$

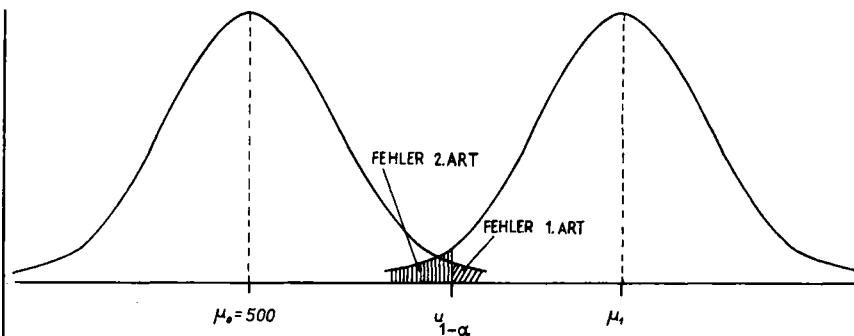


Abb. 1: Einseitiger Einstichprobengaußtest

Ist dagegen der wahre Parameter $\mu < \mu_0$, so ist

$$P_\mu \left(\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} > u_{1-\alpha} \right) < \alpha.$$

Man hat also einen Test zum Niveau α , wenn man

die Entscheidung für H_1 trifft, falls $\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} > u_{1-\alpha}$ und

die Entscheidung für H_0 trifft, falls $\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} \leq u_{1-\alpha}$.

Diesen Test nennt man (einseitigen) **Einstichprobengaußtest**, vgl. Abb. 1. Es ergibt sich hier $\bar{x} = 502,22$. Wählt man $\alpha = 0,025$, so daß $u_{1-\alpha} = 1,96$, dann gilt

$$\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} = \sqrt{9} \frac{502,22 - 500}{2,7} = 2,47 > 1,96 = u_{1-\alpha}.$$

Man entscheidet sich also für H_1 , d. h. dafür, daß das mittlere Füllgewicht mehr als 500 g beträgt. Was bedeutet nun, daß wir diese Entscheidung mit einem Test zum Niveau $\alpha = 0,025$ getroffen haben? Wir können natürlich nicht sicher sein, daß wir uns hier richtig entschieden haben. Das mittlere Füllgewicht kann durchaus auch kleiner oder gleich 500 g sein. Wir wissen allerdings, daß wir in so einem Fall die obige Entscheidung nur mit einer Wahrscheinlichkeit von höchstens $\alpha = 0,025$ getroffen hätten. Wir können uns also ziemlich sicher sein, daß wir uns hier richtig entschieden haben. Anders wäre es gewesen, wenn die Stichprobe eine Entscheidung für H_0 geliefert hätte. Dann wäre ein Irrtum nämlich gleichbedeutend damit, daß tatsächlich H_1 vorgelegen hätte, wir also einen Fehler 2. Art gemacht hätten. Die Wahrscheinlichkeit für den Fehler 2. Art ist hier jedoch nicht bekannt.

Wir haben im obigen Beispiel gesehen, daß durch die Angabe des Testniveaus nur die Entscheidung für H_1 abgesichert wird. Wir kennen dadurch nämlich nicht die Wahrscheinlichkeit für den Fehler 2. Art. Eine vollständige Charakterisierung eines Tests erhält man nur durch die Angabe einer **Gütfunktion** $g(\theta)$, dabei ist $g(\theta)$ die Wahrscheinlichkeit sich für H_1 zu entscheiden, falls θ der wahre Parameter ist. Einen Test zum Niveau α hat man also, wenn

$$g(\theta) \leq \alpha \quad \text{für alle } \theta \in H_0.$$

Ist $\theta \in H_1$, so gibt $1 - g(\theta)$ die Wahrscheinlichkeit für den Fehler 2. Art an. Die

Funktion $1 - g(\mu)$ bezeichnet man auch mit **OC (Operationscharakteristik)**. Sie gibt ebenso wie die Gütfunktion $g(\mu)$ eine vollständige Charakterisierung des zugehörigen Tests.

Beispiel: Wir wollen die Gütfunktion für den Einstichprobengaußtest aus dem vorigen Beispiel berechnen. Es ist

$$g(\mu) = P_\mu \left(\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} > u_{1-\alpha} \right).$$

Da bei Vorliegen von μ

$$\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} \text{ nach } N \left(\sqrt{n} \frac{\mu - \mu_0}{\sigma}, 1 \right)$$

verteilt ist, ergibt sich

$$g(\mu) = 1 - \Phi \left(u_{1-\alpha} - \sqrt{n} \frac{\mu - \mu_0}{\sigma} \right).$$

Der Verlauf dieser Gütfunktion ist in Abb. 2 dargestellt.

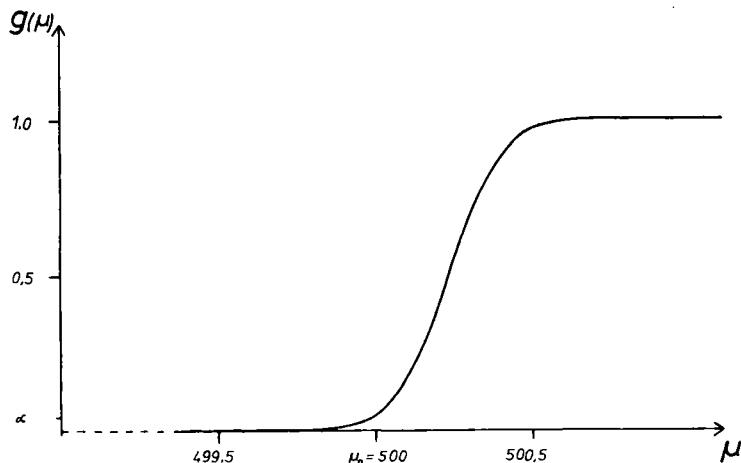


Abb. 2: Gütfunktion des einseitigen Einstichprobengaußtests mit $\mu_0 = 500$, $\sigma_0 = 1$, $n = 9$ und $u_{1-\alpha} = 1,645$

Möchte man nun sicherstellen, daß bei unserem Beispiel ein Überschreiten des Sollwertes von 500 g um 2 g oder mehr mit einer Wahrscheinlichkeit von mindestens $\gamma = 0,6$ auch wirklich entdeckt wird, soll die Wahrscheinlichkeit für den Fehler 2. Art bei einem tatsächlich vorliegenden Wert $\mu \geq \mu_1 = 502$ also höchstens 0,4 betragen, so muß gelten

$$g(\mu) \geq \gamma \quad \text{für alle } \mu \geq \mu_1.$$

Das bedeutet hier

$$\Phi \left(u_{1-\alpha} - \sqrt{n} \frac{\mu - \mu_0}{\sigma} \right) \leq 1 - \gamma \quad \text{für } \mu \geq \mu_1$$

und somit

$$n \geq \frac{\sigma^2}{(\mu_1 - \mu_0)^2} (u_{1-\alpha} - u_{1-\beta})^2.$$

Für unser konkretes Beispiel heißt das

$$n \geq \frac{2,7^2}{(502 - 500)^2} (1,96 + 0,26)^2 = 8,98.$$

Man braucht also mindestens 9 Testbeobachtungen, damit bei dem hier verwandten Gaußtest zum Niveau $\alpha = 0,025$ schon ein um 2 g zu hohes durchschnittliches Füllgewicht mit einer Wahrscheinlichkeit von mindestens 0,6 entdeckt wird.

Wir wollen noch kurz auf eine Vorgehensweise eingehen, die häufig von Praktikern und statistischen Programmpaketen gewählt wird. Man kann sich natürlich fragen, bei welchem Niveau der verwandte Test gerade noch abgelehnt hätte. Ist diese Größe, die man **Level attained**, **p-Wert** (**p-Value**) oder **descriptive Level** nennt, kleiner oder gleich dem **vorher festgesetzten Niveau** (!), so lehnt man die Nullhypothese ab und andernfalls wird sie angenommen. Auf diese Weise kommt man zu der gleichen Entscheidung wie bei der oben beschriebenen Vorgehensweise. Dabei ist die Forderung, das Niveau vorher festzulegen, eine Forderung an die Ehrlichkeit des Statistikers. Man könnte nämlich jede Nullhypothese verwerfen, wenn das Niveau nachträglich etwas größer als der Level attained festgesetzt würde.

Beispiel: Bei dem in diesem Abschnitt als Beispiel verwandten Gaußtest ist der Level attained gerade derjenige Wert $\tilde{\alpha}$, für den gilt:

$$\sqrt{n} \frac{\bar{x} - \mu_0}{\sigma} = u_{1-\tilde{\alpha}}, \quad \text{also} \quad \tilde{\alpha} = 1 - \Phi\left(\sqrt{n} \frac{\bar{x} - \mu_0}{\sigma}\right).$$

Es ergibt sich konkret $\tilde{\alpha} = 1 - \Phi(2,47) = 0,007$.

Dieser Wert ist kleiner als das vorher festgelegte Niveau $\alpha = 0,025$, so daß auch bei dieser Vorgehensweise wie erwartet die Nullhypothese verworfen wird.

5. Beurteilungskriterien für statistische Tests

Die Eigenschaften eines Tests werden durch die Gütfunktion vollständig beschrieben, wie wir schon einmal im vorigen Abschnitt erwähnt haben. Wie sollte diese Funktion nun aussehen? Ideal wäre eine Gütfunktion mit

$$g(\theta) = 0 \quad \text{für } \theta \in H_0 \quad \text{und} \quad g(\theta) = 1 \quad \text{für } \theta \in H_1,$$

so daß der zugehörige Test mit der Wahrscheinlichkeit 1 richtig entscheidet. Das ist natürlich nicht möglich. Durch die Festsetzung des **Niveaus** α wird allerdings die Wahrscheinlichkeit für den Fehler 1. Art durch α begrenzt, d.h.

$$g(\theta) \leq \alpha \quad \text{für alle } \theta \in H_0.$$

Eine oftmals vernünftige Forderung an einen Niveau- α -Test, die auch von vielen solchen Tests erfüllt wird, wie z. B. von dem Gaußtest, ist die **Unverfälschtheit**, d. h.

$$g(\theta) \geq \alpha \quad \text{für alle } \theta \in H_1.$$

Das bedeutet, daß die Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn diese tatsächlich nicht vorliegt, mindestens so groß sein sollte, als wenn sie vorgele-

gen hätte. Außerdem wird man i. d. R. erwarten können, daß der Test mit wachsender Beobachtungszahl n immer besser wird, d.h.

$$g(\theta) \rightarrow 1, \text{ falls } n \rightarrow \infty, \text{ für } \theta \in H_1.$$

Diese Eigenschaft nennt man **Konsistenz**, genauso wie bei den Parameterschätzungen. Ebenso wie dort sagen auch hier die beiden Eigenschaften Unverfälschtheit und Konsistenz noch sehr wenig über die Güte aus. Dagegen liegt ein Optimum vor, wenn man sagen kann, es handele sich um einen **gleichmäßig besten** Niveau- α -Test, d.h., wenn dessen Gütfunktion g auf H_1 **gleichmäßig besser** ist als die Gütfunktion \tilde{g} eines beliebigen anderen Niveau- α -Tests für das gleiche Problem:

$$g(\theta) \geq \tilde{g}(\theta) \text{ für alle } \theta \in H_1.$$

Diese Eigenschaft hat z. B. der einseitige Gaußtest.

Für das damit im vorigen Abschnitt behandelte Testproblem gibt es also keinen Niveau- α -Test mit geringerer Wahrscheinlichkeit für den Fehler 2. Art.

Wir sind bisher immer davon ausgegangen, daß wir das Niveau eines Tests exakt angeben können. Das ist aber nicht immer der Fall. Dann ist es aber manchmal noch möglich zu zeigen, daß der gewünschte Wert α zumindest nicht überschritten wird. Man kann nämlich gelegentlich zeigen, daß die Gütfunktion auf einer größeren, aber eventuell einfacher strukturierten Menge \tilde{H}_0 als es die Nullhypothese H_0 ist, durch α beschränkt wird. Benutzt man den Test nun als Niveau- α -Test für das obige Problem, so kann man mit einem sogenannten **konservativen** Verhalten rechnen, d.h. einem längeren Festhalten an der Nullhypothese als es eigentlich notwendig wäre. Dienen asymptotische Überlegungen zur Konstruktion eines Tests, so läßt sich über das Niveau oft nur sagen, daß es bei genügend großem Stichprobenumfang in der Nähe von α liegt. Ist z. B. nur bekannt, daß X_1, \dots, X_n unabhängig identisch verteilt sind (nicht notwendig normalverteilt) mit bekannter Varianz σ^2 und unbekanntem Erwartungswert μ , und benutzt man für das Testproblem

$$H_0: \mu \leq \mu_0 \text{ gegen } H_1: \mu > \mu_0$$

die Teststatistik

$$T_n(x_1, \dots, x_n) = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma}$$

mit dem kritischen Wert $u_{1-\alpha}$, so hat das Niveau dieses Tests ungefähr den Wert α , wenn der Stichprobenumfang genügend groß ist. Nach dem zentralen Grenzwertsatz gilt nämlich

$$P(T_n > u_{1-\alpha}) \rightarrow \alpha \text{ für } n \rightarrow \infty.$$

6. Arten von Hypothesen und allgemeine Bemerkungen

Wir hatten bislang Hypothesen der Form

$$H_0: \theta \leq \theta_0 \text{ gegen } H_1: \theta > \theta_0$$

kennengelernt. Diese und die Hypothesen

$$H_0: \theta \geq \theta_0 \text{ gegen } H_1: \theta < \theta_0$$

bezeichnet man als **einseitige Hypothesen**.

Diese Art von Hypothesen reicht jedoch nicht zur Beschreibung aller statistischen Testprobleme aus. Die Gründe hierfür sind:

- (1) Manchmal möchte man auch Hypothesen über einen Parametervektor prüfen.
- (2) Beschreibt der Parameter θ einen Unterschied, z. B. einen Behandlungsunterschied zwischen zwei Medikamenten, so möchte man prüfen, ob überhaupt eine Wirkung vorhanden ist, d.h. es interessiert

$$H_0: \theta = 0 \text{ gegen } H_1: \theta \neq 0.$$

Allgemein heißt $H_0: \theta = \theta_0$ gegen $H_1: \theta \neq \theta_0$ eine **zweiseitige Hypothese**.

Die Testprobleme zu den obigen Hypothesen heißen dann **einseitige**, bzw. **zweiseitige Testprobleme**.

Testet man nicht Hypothesen über den Parameter einer Verteilung, sondern über den Typ einer Verteilung, etwa darüber, ob es sich bei den beobachteten Werten um Realisationen normalverteilter Größen handelt, so spricht man von einem **Anpassungstest**. Im Gegensatz zu den übrigen Testproblemen, den sogenannten **Signifikanztests**, liegt hier das Interesse darin, die Nullhypothese nicht zu verwerfen; daher sollte hier das Niveau α mitunter auch hoch (z. B: $\alpha = 0,9$) gewählt werden.

Bedeutet z. B. die Gültigkeit der Nullhypothese, daß eine Normalverteilung vorliegt, so möchte man diese Hypothese natürlich annehmen. Man kann dann aber aus der Nichtablehnung der Nullhypothese **nicht schließen**, daß eine Normalverteilung auch tatsächlich vorliegt. Die irrtümliche Annahme der Nullhypothese ist nämlich ein Fehler 2. Art, über dessen Wahrscheinlichkeit man hier keine Aussagen machen kann. Vor einer unkritischen Übernahme der gelegentlich anzutreffenden Vorgehensweise, bei der man zunächst einen Anpassungstest rechnet, der die Normalverteilungsannahme nicht verwirft, und anschließend einen Normalverteilungstest über die Parameter durchführt, kann nur gewarnt werden.

Falls eine Nullhypothese H_0 nur aus einem Element besteht, z. B. beim Test auf Standardnormalverteilung, so spricht man von einer **einelementigen Nullhypothese**. Bestünde H_0 z. B. aus allen Normalverteilungen, so hieße H_0 **zusammengesetzte Nullhypothese**. Tests für obige Probleme werden wir z. B. in Kap. IV kennenlernen.

7. Nichtparametrische (verteilungsfreie) Verfahren

In manchen Anwendungen scheint es nicht immer gerechtfertigt anzunehmen, daß die beobachtete Stichprobe einer relativ kleinen parametrischen Klasse von Verteilungen, z. B. $N(\mu, \sigma^2)$, $Ex(\lambda)$, entstammt. Möchte man auch in diesen Fällen statistische Aussagen (**Inferenz**) treffen, so bedient man sich der sogenannten **nichtparametrischen** oder auch **verteilungsfreien** Verfahren, die uns in verschiedenen Kapiteln begegnen werden.

Entstammt die betrachtete Stichprobe x_1, \dots, x_n aus einer Grundgesamtheit mit stetiger Verteilungsfunktion, so werden bei diesen nichtparametrischen Verfahren die jeweils vorgegebenen Niveaus exakt eingehalten. Bei diskreten zugrundeliegenden Verteilungen können diese Verfahren allerdings auch angewandt werden, wenn die Zahl der **Bindungsgruppen** in der Stichprobe nicht „zu klein“ ist und die Größen der Bindungsgruppen nicht zu groß sind. Unter einer **Bindung** versteht man hierbei den Sachverhalt, daß zwei Werte in der Stichprobe gleich sind, und demzufolge

besteht eine Bindungsgruppe aus allen Elementen in der Stichprobe, die den selben Wert besitzen.

Beispiel:

Haben wir etwa

$$x_1 = 7, x_2 = 3, x_3 = 7, x_4 = 5, x_5 = 7 \text{ und } x_6 = 3$$

beobachtet, so ist die Anzahl der Bindungsgruppen drei (mit den Größen 3, 2 und 1), nämlich

$$x_1 = x_3 = x_5 = 7, x_2 = x_6 = 3 \text{ und } x_4 = 5.$$

Ein Teil der verteilungsfreien Verfahren benutzt sogenannte **Ordnungsstatistiken**

$$X_{(1)}, X_{(2)}, \dots, X_{(n)},$$

die bedeuten, daß ihre Realisationen

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

der Größe nach geordnet sind, d. h. also

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Beobachtet man eine Stichprobe x_1, \dots, x_n , so bewirken also die Ordnungsstatistiken, daß die Stichprobenelemente der Größe nach angeordnet und entsprechend umbenannt werden. Treten Bindungen auf, so wird über die Anordnung an den betreffenden Stellen zufällig entschieden.

Beispiel:

Im obigen Beispiel erhält man etwa

$$x_{(1)} = x_6 = 3 = x_2 = x_{(2)} < x_{(3)} = x_4 = 5 < \text{usw.}$$

Andere nichtparametrische Verfahren benutzen sogenannte **Rangstatistiken**

$$R(X_1), \dots, R(X_n),$$

deren Realisationen

$$R(x_1) = r_1, \dots, R(x_n) = r_n$$

die Rangzahlen zur Stichprobe x_1, \dots, x_n angeben. Unter **Rangzahl eines Stichprobenelementes** versteht man hierbei zumeist die Platznummer, die ein Stichprobenelement einnimmt, wenn man die gesamte Stichprobe der Größe nach ordnet. Ist etwa x_5 der zweitkleinste Wert in einer Stichprobe, so erhält x_5 die Rangzahl 2, d. h. $x_5 = x_{(2)}$ und $R(x_5) = 2$.

Treten Bindungen auf, so wird in der Regel über die für eine Bindungsgruppe in Frage kommenden Ränge gemittelt und diese **mittlere Rangzahl (midrank)** allen Elementen dieser Gruppe zugeordnet.

Beispiel:

Betrachten wir wieder die Stichprobe aus obigem Beispiel, so ergibt sich

$$R(x_2) = R(x_6) = (1 + 2)/2 = 1,5,$$

$$R(x_4) = 3 \text{ und}$$

$$R(x_1) = R(x_3) = R(x_5) = (4 + 5 + 6)/3 = 5.$$

Es sei darauf hingewiesen, daß einige Verfahren spezielle Vorschriften zur Vergabe von Rangzahlen benutzen.

8. Zufällige Auswahl, Randomisation

Statistische Aussagen basieren auf Stichprobenwerten, die zum Teil erst in einem Experiment an Objekten aus einer interessierenden Grundgesamtheit beobachtet werden müssen. Um bei der Auswahl der Objekte systematische Fehler zu vermeiden, muß diese **zufällig** erfolgen. Hierzu bedient man sich sogenannter (gleichverteilter) **Zufallszahlen**, vgl. auch Abschnitt 2.1 in Kap. IV, aus Computern oder Tabellen; im Anhang ist eine solche Tabelle zu finden. Man geht dann so vor, daß man zunächst die in Betracht kommenden Objekte durchnumeriert und dann gemäß der ermittelten Zufallszahlen diejenigen Objekte auswählt, die in das betreffende Experiment einzubeziehen sind. Diese Vorgehensweise nennt man auch **Randomisation**. Möchte man etwa eine Stichprobe von $n = 6$ Ratten aus einem Käfig mit 100 Ratten einem Lernexperiment unterziehen, so wäre es ungünstig, 6 mal in den Käfig zu greifen und jeweils die erstbeste Ratte, die man greifen kann, herauszunehmen. Dann würde man nämlich lediglich die langsamsten und eventuell schwachen bzw. sogar kranken Tiere in die Stichprobe einbeziehen, da die vitalen Tiere so schnell sind, daß man sie gar nicht so leicht greifen kann; die Ergebnisse des Lernexperiments würden hierdurch u. U. stark verfälscht. Zur Vermeidung dieses systematischen Fehlers numeriert man nun alle 100 Ratten von 00 bis 99 durch, bestimmt dann 6 zweistellige (auf dem Intervall [0;99] gleichverteilte ganzzahlige) Zufallszahlen und wählt für das Experiment diejenige Ratten aus, die eine Nummer tragen, welche unter den Zufallszahlen ist.

Sind in dem Käfig z. B. nur 82 Ratten, so werden diese entsprechend von 00 bis 81 durchnumeriert und von den zweistelligen Zufallszahlen, die man z. B. aus einer Tabelle gewinnt, nur die ersten sechs berücksichtigt, die kleiner als 82 sind.

Beispiel: In Tabellen sind häufig fünfstellige Zufallszahlen angegeben. Wir greifen eine beliebige Zeile und Ziffernspalte heraus (dies könnte man natürlich auch wieder etwa durch Würfeln von eventuellen systematischen Fehlern befreien) und notieren von der zugehörigen Ziffer ausgehend zeilenweise die nachstehenden Zahlenkombinationen. Dies könnte zum Beispiel folgendermaßen aussehen :

..... 738 19524 57826 80561 2

Nun fassen wir jeweils zwei Ziffern zu einer zweistelligen Zufallszahl zusammen :

73 81 95 24 57 82 68 05 61 ...

und nehmen für unser Experiment nun die 6 Tiere aus dem Käfig mit 82 Ratten, die die Nummern

73, 81, 24, 57, 68 und 05
tragen.

Die eben geschilderte Vorgehensweise ist jedoch nicht zu empfehlen, wenn die Anzahl der zulässigen Zufallszahlen zu gering ist, etwa wenn man 6 Ratten aus einem Käfig mit insgesamt nur 17 Tieren zufällig auswählen möchte. Sind a Objekte aus einer Grundgesamtheit von n Objekten auszuwählen und ist $n - 1$ eine k -stellige Zahl, so nehmen wir k -stellige Zufallszahlen Z und berechnen $X = 0, Z \cdot n$.

Die gewünschten, auf dem Intervall $[0; n - 1]$ gleichverteilten ganzzahligen Zu-

fallszahlen Y ergeben sich dann als die jeweils größte ganze Zahl, die kleiner als das berechnete X oder gleich X ist; man schreibt dafür auch

$$Y = [X], \text{ (Gaußklammer } X\text{).}$$

Beispiel: Sollen 6 Ratten aus 17 Ratten ausgewählt werden, so müssen wir von zweistelligen Zufallszahlen Z zwischen 00 ... 99 ausgehen. Liegt etwa die Zufallszahlenfolge, vgl. obiges Beispiel,

73 81 95 24 57 82 68 05 61 ...

vor, so ergeben sich zunächst die Hilfsgrößen X zu

$$0,73 \cdot 17 = 12,41,$$

$$0,81 \cdot 17 = 13,77, 16,15, 4,08, 9,69, 13,94, 11,56, 0,85, 10,37$$

und die entsprechenden auf dem Intervall [0;16] gleichverteilten Zufallszahlen Y sind dann

12, 13, 16, 04, 09, 13, 11, 00, 10.

Die 6 auszuwählenden – unter den von 00 bis 16 durchnumerierte – Ratten sind dann diejenigen, die die Nummern

12, 13, 16, 04, 09 und 11

tragen (da die 13 doppelt vorkommt).

Entsprechend geht man vor, wenn Objekte innerhalb eines Experiments auf verschiedene Plätze zu verteilen sind. Will man etwa mehrere Maschinen miteinander vergleichen, so muß man die Bediener den Maschinen **zufällig zuordnen**, um Bedienereinflüsse auszugleichen. Hierzu wählt man entsprechend oben geschilderter Vorgehensweise zunächst die benötigten n Bediener zufällig aus einer Gesamtheit von möglichen Bedienern aus. Diese numeriert man dann von 0 bis n – 1 beliebig durch und wählt nun mittels passender Zufallszahlen, wie oben beschrieben, zunächst diejenigen aus, die Maschine 1 bedienen sollen, dann die Bediener, die an Maschine 2 arbeiten sollen, usw.

Ausführlich wird auf die hier angeschnittene Problematik der Gewinnung von „unverfälschten“, zufälligen Stichproben und der Versuchsplanung in Kap. V. eingegangen.

9. Notation von Zufallsvariablen

Aus Notationsgründen werden wir in der Bezeichnungsweise nicht immer streng zwischen Zufallsvariablen und ihren Realisationen unterscheiden. Ist von Wahrscheinlichkeiten die Rede, so ist klar, daß sich dies auf die den jeweiligen Realisationen zugrundeliegenden Zufallsvariablen bezieht; ebenso bei Verteilungen, Erwartungswert, Varianz etc.

Ebenfalls zur Schreiberleichterung sei die Existenz von Momenten etc. stets vorausgesetzt, wenn wir sie verwenden.

Kapitel IV:

Spezielle Verteilungen und statistische Schlüsse über Kenngrößen von Verteilungen mittels einer Meßreihe (Stichprobe)

In den bisherigen Kapiteln haben wir bereits einige statistische Verteilungen kennengelernt. Darüber hinaus haben wir im Kap. III gesehen, wie man aufgrund einer Stichprobe vom Umfang n aus einer parametrischen Klasse von Verteilungen (man spricht auch vom Einstichprobenproblem) Aussagen über den unbekannten Parameter macht. In diesem Kapitel werden wir einige der wichtigsten statistischen Verteilungen kennenlernen, gleichzeitig die zugehörigen parametrischen Tests und Schätzer. Die uns bereits bekannte Normalverteilung sowie die aus ihr abgeleiteten Verteilungen spielen hierbei eine entscheidende Rolle.

Aus Vollständigkeitsgründen und zur Bequemlichkeit des Lesers werden wir auch die Eigenschaften der uns bereits bekannten Verteilungen, die in einem der vorigen Kapitel erwähnt wurden, wieder aufführen.

Neben den schon erwähnten parametrischen (Einstichproben-)Tests für die einzelnen Verteilungsklassen werden wir die zugehörigen Anpassungstests und im Abschnitt 5 auch noch verteilungsfreie (nichtparametrische) Verfahren für statistische Schlüsse anhand einer einzelnen Meßreihe vorstellen. Die Problematik **sequentieller Tests** wird im Abschnitt 6 erörtert.

1. Die Normalverteilung und daraus abgeleitete Verteilungen

1.1 Die Normalverteilung und ihre Bedeutung

Die Häufigkeitsverteilung von n Merkmalsausprägungen weist häufig eine glockenförmige Gestalt auf, die annähernd durch die Funktion

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

beschrieben werden kann; diese Funktion ist gerade die Dichte einer Normalverteilung $N(\mu, \sigma^2)$ bzw. einer normalverteilten Zufallsvariablen X mit Erwartungswert μ und Varianz σ^2 . Sie ist symmetrisch um μ , d.h. sie nimmt für

$$\mu - \Delta \quad \text{und} \quad \mu + \Delta$$

den gleichen Wert an, und hat Wendepunkte an den Stellen

$$\mu - \sigma \quad \text{und} \quad \mu + \sigma,$$

vgl. auch **Abb. 1**.

Die große Bedeutung der Normalverteilung für die Statistik folgt aus dem uns bereits bekannten Zentralen Grenzwertsatz. Er erlaubt es nämlich, eine Zufallsvariable, die sich additiv aus unabhängigen Einzelwirkungen zusammensetzt, als angenähert normalverteilt anzunehmen, siehe Abschnitt 9 in Kap. II.

Hinzu kommt nun noch, daß die statistische Theorie normalverteilter Beobachtungen am weitesten entwickelt ist. Die sich daraus ergebenden Schätzer und Tests haben dann noch oft die Eigenschaft, daß sie gegenüber Abweichungen von der

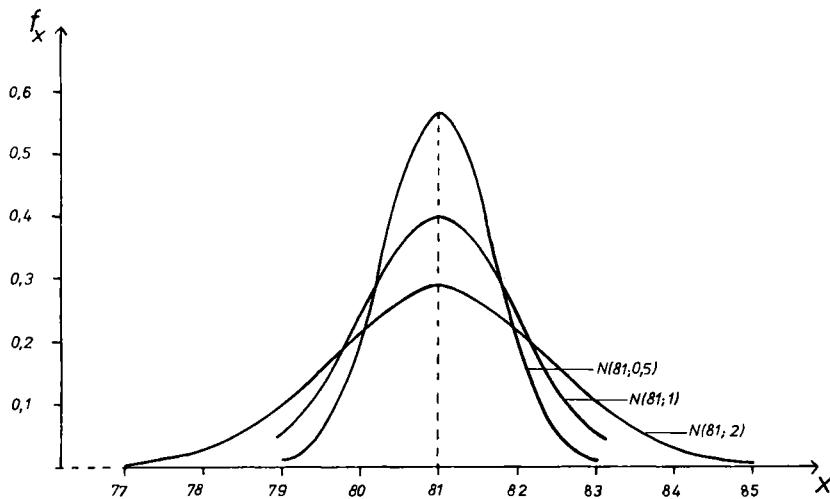


Abb. 1: Dichtefunktion der $N(81, 0,5)$ -, $N(81, 1)$ - und $N(81, 2)$ -Verteilung

Normalverteilungsannahme relativ unempfindlich (**robust**) sind (zumindest für große Stichprobenumfänge), wie wir bereits am Beispiel des Einstichproben-Gaußtests gesehen hatten, vgl. auch Kap. III, Abschnitt 5. Zudem lassen sich Beobachtungen, die keiner normalverteilten Grundgesamtheit entstammen, in vielen Fällen durch geeignete Transformationen in Daten aus einer Normalverteilung überführen, vgl. Kap. V, Abschnitt 4.5.

Während wir hier nur eine einzelne normalverteilte Zufallsvariable X betrachten werden, gehen wir in Kap. VIII und Kap. XI ausführlich auf den Vergleich zweier bzw. mehrerer Normalverteilungen ein.

Wie schon oben erwähnt wurde, sind die beiden Kenngrößen Erwartungswert und Varianz bei der Normalverteilung identisch mit den Parametern μ und σ^2 . Der Erwartungswert μ ist dabei irgendeine reelle Zahl und die Varianz σ^2 eine positive reelle Zahl. Für die k -ten zentralen Momente, vgl. Abschnitt 8.3 in Kapitel II, der $N(\mu, \sigma^2)$ -Verteilung gilt

$$E(X - EX)^k = \begin{cases} 0 & , \text{ falls } k \text{ ungerade} \\ 1 \cdot 3 \cdot 5 \cdot \dots \cdot (k-1) \sigma^k, & \text{falls } k \text{ gerade.} \end{cases}$$

Damit erhält man sowohl für die Schiefe als auch für den Exzeß den Wert Null. Aus der Symmetrie der Normalverteilung ergibt sich direkt, daß ihr Median mit dem Erwartungswert μ identisch ist. Die Verteilungsfunktion der $N(\mu, \sigma^2)$ -Verteilung ist

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}} d\xi, \quad \text{vgl. Abb. 2.}$$

Da dieses Integral nicht elementar berechenbar ist, wird die Verteilungsfunktion der Normalverteilung tabelliert. Dies wird dadurch erleichtert, daß sich die Normalverteilung stets in eine Standardnormalverteilung $N(0, 1)$ transformieren läßt:

Ist nämlich X nach $N(\mu, \sigma^2)$ -verteilt (in Kurzschreibweise: $X \sim N(\mu, \sigma^2)$), so ist

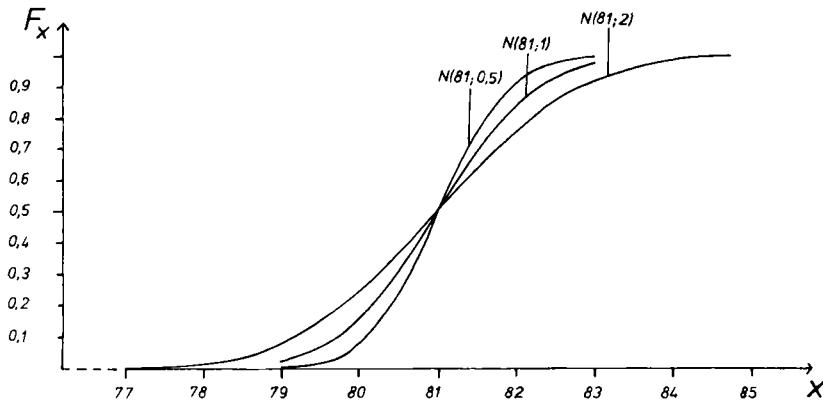


Abb. 2: Verteilungsfunktion der $N(81, 0,5)$ -, $N(81, 1)$ - und $N(81, 2)$ -Verteilung

$Y = (X - \mu)/\sigma$ standardnormalverteilt. Es reicht also aus, die Verteilungsfunktion

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot e^{-\xi^2/2} d\xi$$

der Standardnormalverteilung zu tabellieren, vgl. Tabelle 1 im Anhang. Die Dichte der $N(0, 1)$ -Verteilung ist gerade

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}.$$

Die Verteilungsfunktion und die Dichte der $N(\mu, \sigma^2)$ -Verteilung erhält man damit zu

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad \text{und} \quad f_X(x) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x - \mu}{\sigma}\right).$$

Allgemein gilt, daß die Zufallsvariable

$$Y = \alpha \cdot X + \beta \quad \text{mit} \quad \alpha \neq 0$$

$N(\alpha \cdot \mu + \beta, \alpha^2 \cdot \sigma^2)$ -verteilt ist, wenn X nach $N(\mu, \sigma^2)$ verteilt ist.

Beispiel: Die Länge X eines Werkstücks wird in m gemessen. Es sei bekannt, daß $X \sim N(0,8, 0,0001)$. Wird das Werkstück nun konstant um 1 cm verlängert, so gibt

$$Y = 100X + 1$$

die Länge des Werkstücks in cm an; es ist dann $Y \sim N(81, 1)$ und man berechnet

$$\begin{aligned} P(80 \leq Y \leq 82) &= F_Y(82) - F_Y(80) \\ &= \Phi\left(\frac{82 - 81}{1}\right) - \Phi\left(\frac{80 - 81}{1}\right) \\ &= \Phi(1) - \Phi(-1) \\ &= 2\Phi(1) - 1 = 0,6826. \end{aligned}$$

Im Beispiel haben wir von der Symmetrie der Normalverteilung Gebrauch ge-

macht; es ist nämlich

$$\Phi(-x) = 1 - \Phi(x) \quad \text{wegen} \quad \varphi(x) = \varphi(-x).$$

Dies bedeutet für $X \sim N(\mu, \sigma^2)$:

$$P(X \leq \mu + x) = 1 - P(X \leq \mu - x) = P(X \geq \mu - x), \quad \text{s. Abb. 3.}$$

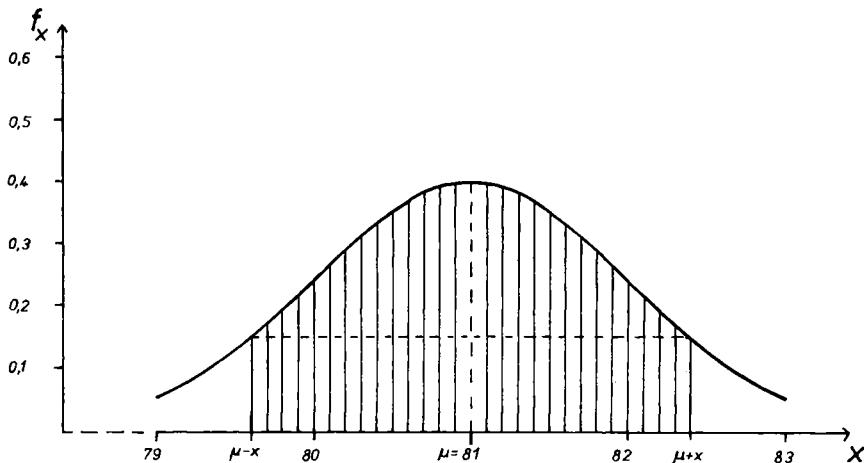


Abb. 3: Dichtefunktion der $N(81, 1)$ -Verteilung mit eingetragenem x -Bereich um μ

Aus diesen Überlegungen folgt, daß es reicht, $\Phi(x)$ für Werte $x \geq 0$ zu tabellieren.

Im Kapitel III haben wir bereits folgende wichtige Eigenschaft der Normalverteilung benutzt:

Sind X_1, \dots, X_n **stochastisch unabhängige** $N(\mu_i, \sigma_i^2)$ -verteilte Zufallsvariable, $i = 1, \dots, n$, dann ist

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Damit ist das arithmetische Mittel $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ aus n $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen gemäß einer $N\left(\mu, \frac{\sigma^2}{n}\right)$ -Verteilung verteilt.

Beispiel: Liegen etwa neun Werkstücke aus einer $N(81, 1)$ -Verteilung vor, so ist die Wahrscheinlichkeit, daß das arithmetische Mittel der Längen in $[80; 82]$ fällt, gerade gleich

$$P(80 \leq \bar{X} \leq 82) = \Phi\left(\frac{82 - 81}{\sqrt{\frac{1}{9}}}\right) - \Phi\left(\frac{80 - 81}{\sqrt{\frac{1}{9}}}\right) = 2\Phi(3) - 1 = 0,9974.$$

Sind die Parameter μ und σ^2 einer normalverteilten Grundgesamtheit **bekannt**, so berechnet sich die Wahrscheinlichkeit, daß ein zufällig entnommenes Element in die Intervalle

$$\begin{aligned}\mu \pm \sigma, \mu \pm 2\sigma, \mu \pm 3\sigma &\text{ fällt, zu} \\ P(\mu - \sigma \leq X \leq \mu + \sigma) &= 0,6827, \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 0,9545 \quad \text{bzw.} \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 0,9973.\end{aligned}$$

Beispiel: Die Wahrscheinlichkeit, daß ein zufällig gezogenes Werkstück eine Länge außerhalb des Intervalls [78; 84] hat, beträgt also nur 0,0027.

Die oben angegebenen Toleranzintervalle heißen **1-, 2-, 3- σ Intervalle**. Ist von einer Verteilung nur die Varianz bekannt, so läßt sich mit Hilfe der Tschebyscheffschen Ungleichung, s. Kap. II, Abschnitt 8.2.A die Wahrscheinlichkeit für eine Beobachtung außerhalb des 3σ Intervalls um den Erwartungswert nur durch $\frac{1}{9} = 0,111$ abschätzen.

Die α -Quantile $\Phi^{-1}(\alpha)$ der $N(0, 1)$ -Verteilung bezeichnen wir mit u_α . Es gilt

$$u_\alpha = -u_{1-\alpha}.$$

Diese Quantile sind für einige Werte $0 < \alpha < 1$ im Anhang gesondert vertafelt. Das α -Quantil einer $N(\mu, \sigma^2)$ -Verteilung $w_\alpha(\mu, \sigma^2)$ bestimmt sich zu

$$w_\alpha(\mu, \sigma^2) = \sigma \cdot u_\alpha + \mu.$$

Für **Simulationsstudien** ist es noch wichtig zu wissen, wie man auf einem Rechner Realisationen normalverteilter Zufallsvariablen erzeugen kann. Man geht dabei aus von unabhängigen Realisationen x_i einer auf dem Intervall $[0; 1]$ gleichverteilten Zufallsgröße, vgl. Abschnitt 2.1.1. Bildet man nun

$$y = \frac{\sum_{i=1}^n x_i - \frac{n}{2}}{\sqrt{n/12}},$$

so hat man es – wegen der Gültigkeit des zentralen Grenzwertsatzes – für genügend großes n näherungsweise mit der Realisation einer $N(0, 1)$ -verteilten Zufallsgröße zu tun. Oft reicht dabei schon $n = 12$ aus, so daß

$$y = \sum_{i=1}^{12} x_i - 6$$

ist. Dann können Extremwerte jedoch nicht auftauchen, denn es gilt $-6 \leq y \leq 6$. Eine andere Möglichkeit bietet die Berechnung der Werte

$$y_1 = \sqrt{-2 \cdot \ln x_1} \cdot \cos(2\pi x_2) \quad \text{und} \quad y_2 = \sqrt{-2 \cdot \ln x_1} \cdot \sin(2\pi x_2).$$

Diese können als Realisationen zweier **unabhängiger** $N(0, 1)$ -verteilter Zufallsgrößen aufgefaßt werden.

Zur $N(\mu, \sigma^2)$ -Verteilung kommt man schließlich durch die Multiplikation dieser Werte mit σ und die anschließende Addition von μ .

1.2. Einige in enger Beziehung zur Normalverteilung stehende Verteilungen

1.2.1 Aus der Normalverteilung abgeleitete Verteilungen

A. Die gestutzte Normalverteilung

Wie wir gesehen haben, ist die Dichte der Normalverteilung für alle x strikt positiv. Damit ist auch (theoretisch) die Wahrscheinlichkeit, daß die Länge X des Werkstücks im Beispiel aus Abschnitt 1.1 kleiner oder gleich 0 ist, von Null verschieden:

$$P(X \leq 0) = \Phi(-81) = \int_{-\infty}^{-81} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} d\xi > 0.$$

Es gilt jedoch

$$P(X \leq 0) \leq \frac{1}{81} \cdot \varphi(-81) = 10^{-1427}.$$

Vom praktischen Standpunkt aus ist dieser Defekt, daß eine Größe (hier: Länge des Werkstückes), die nur positive Werte annehmen kann, mit einer positiven (theoretischen) Wahrscheinlichkeit negative Werte annimmt, in der Regel vernachlässigbar.

Es gibt jedoch auch Beispiele, bei denen das nicht so ist. In solchen Fällen bedient man sich häufig der gestutzten Normalverteilung.

Beispiel: Aus einem Sortiment Schrauben, deren Längen einer $N(10, 0,09)$ -Verteilung folgen (Angaben in mm), werden alle Schrauben entfernt, deren Länge nicht zwischen 9,5 und 10,5 mm liegt. Aus dem reduzierten Sortiment wird zufällig eine Schraube gezogen, deren Länge durch die Zufallsvariable Z wiedergegeben wird. Wie groß ist die Wahrscheinlichkeit, daß die Schraube eine Länge von höchstens 10,4 mm hat, d. h. wie groß ist

$$P(Z \leq 10,4)?$$

Im Grunde handelt es sich bei der oben gesuchten Wahrscheinlichkeit um eine bedingte Wahrscheinlichkeit: Gesucht ist nämlich die Wahrscheinlichkeit, daß die Länge einer Schraube kleiner als 10,4 mm ist, gegeben, daß ihre Länge zwischen 9,5 und 10,5 mm liegt. Damit erhalten wir, vgl. Abschnitt 4 in Kap. II,

$$P(Z \leq 10,4) = \frac{P(9,5 \leq X \leq 10,4)}{P(9,5 \leq X \leq 10,5)},$$

wobei X die Länge einer Schraube im Ausgangskollektiv bezeichnet. Es ist also

$$\begin{aligned} P(Z \leq 10,4) &= \frac{\Phi\left(\frac{10,4 - 10}{\sqrt{0,09}}\right) - \Phi\left(\frac{9,5 - 10}{\sqrt{0,09}}\right)}{\Phi\left(\frac{10,5 - 10}{\sqrt{0,09}}\right) - \Phi\left(\frac{9,5 - 10}{\sqrt{0,09}}\right)} \\ &= \frac{\Phi\left(\frac{0,4}{0,3}\right) - \Phi\left(\frac{-0,5}{0,3}\right)}{\Phi\left(\frac{0,5}{0,3}\right) - \Phi\left(\frac{-0,5}{0,3}\right)} = \frac{0,9082 - 0,0485}{0,9031} = 0,952. \end{aligned}$$

Im Vergleich hierzu ist die Wahrscheinlichkeit, daß eine Schraube aus dem Ausgangssortiment eine Länge von höchstens 10,4 mm hat, gerade 0,9082.

Die Zufallsvariable Z hat also die Verteilungsfunktion

$$P(Z \leq z) = \begin{cases} 0 & , \text{ falls } z < 9,5 \\ \frac{\Phi\left(\frac{z-10}{\sqrt{0,09}}\right) - \Phi\left(\frac{9,5-10}{\sqrt{0,09}}\right)}{\Phi\left(\frac{10,5-10}{\sqrt{0,09}}\right) - \Phi\left(\frac{9,5-10}{\sqrt{0,09}}\right)} & , \text{ falls } 9,5 \leq z \leq 10,5 \\ 1 & , \text{ sonst.} \end{cases}$$

Die Normalverteilung $N(10, 0,09)$ wird sozusagen an den beiden Stellen 9,5 und 10,5 „abgeschnitten“ bzw. „gestutzt“. Man spricht von einer (**zweiseitig gestutzten Normalverteilung**).

Die Verteilungsfunktion einer zweiseitig gestutzten $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen Z ist allgemein, wenn z_u der untere und z_o der obere Abschneidepunkt ist, vgl. Abb. 4,

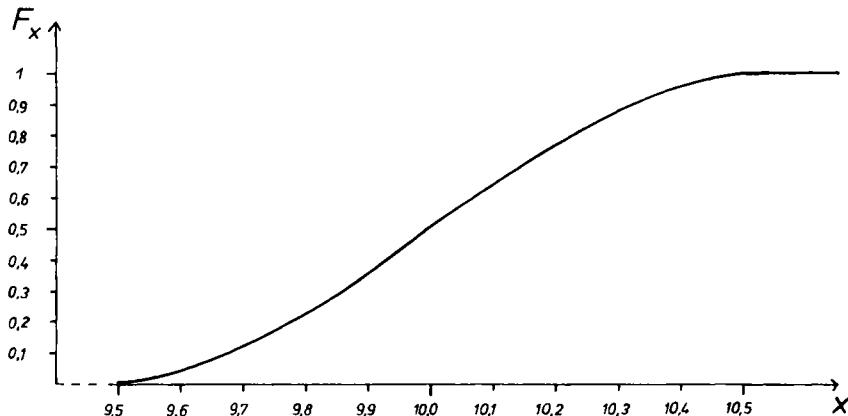


Abb. 4: Verteilungsfunktion der gestutzten $N(10, 0,09)$ -Verteilung

$$F_Z(z) = \frac{F_Z(z) - F_Z(z_u)}{F_Z(z_o) - F_Z(z_u)} = \frac{\Phi(y) - \Phi(y_u)}{\Phi(y_o) - \Phi(y_u)} \quad \text{für } z_u \leq z \leq z_o.$$

Dabei ist $y = \frac{z-\mu}{\sigma}$ und $y_u = \frac{z_u-\mu}{\sigma}$, $y_o = \frac{z_o-\mu}{\sigma}$ sind die standardisierten Abschneidepunkte. Erwartungswert und Varianz der zweiseitig gestutzten Normalverteilung sind mit obigen Größen durch

$$EZ = \mu + \frac{\varphi(y_u) - \varphi(y_o)}{\Phi(y_o) - \Phi(y_u)} \cdot \sigma \quad \text{und}$$

$$\text{Var } Z = \sigma^2 \left\{ 1 + \frac{y_u \varphi(y_u) - y_o \varphi(y_o)}{\Phi(y_o) - \Phi(y_u)} - \left(\frac{\varphi(y_u) - \varphi(y_o)}{\Phi(y_o) - \Phi(y_u)} \right)^2 \right\}$$

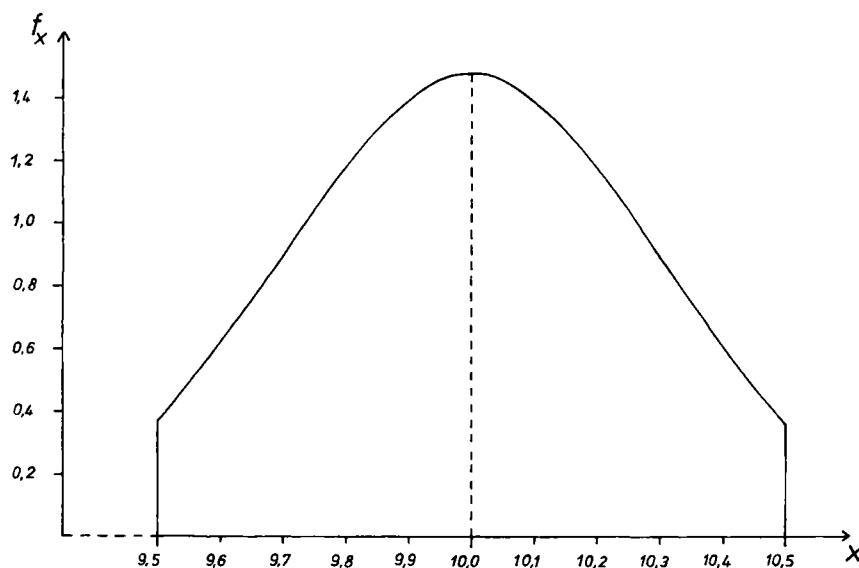


Abb. 5: Dichtefunktion der gestutzten $N(10, 0,09)$ -Verteilung

gegeben und die Dichte der Verteilung von Z ist, vgl. **Abb. 5**,

$$f_Z(z) = \frac{1}{\sigma} \frac{\varphi(y)}{\Phi(y_0) - \Phi(y_u)} \quad \text{für } z_u \leq z \leq z_0.$$

Ist $z_u = -\infty$, so heißt Z rechtssseitig gestutzt, ist $z_0 = \infty$ linksseitig gestutzt.

Auch z. B. bei der Analyse normalverteilter Lebensdauern von irgendwelchen Objekten spielt die gestutzte Normalverteilung in spezieller Form eine Rolle. Da Lebensdauern natürlich nie negative Werte annehmen können, ist die Dichte normalverteilter Lebensdauern auf dem Intervall $[0; \infty)$ konzentriert. Die Dichte muß also einseitig an der Stelle $z_u = 0$ gestutzt sein.

Die Dichtefunktion ist dann durch

$$f_Z(z) = \begin{cases} 0 & , \text{ falls } z < 0 \\ \frac{1}{a\sigma\sqrt{2\pi}} e^{-(z-\mu)^2/(2\sigma^2)}, & \text{falls } z \geq 0 \end{cases}$$

gegeben; dabei ist

$$a = \int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-(\xi-\mu)^2/(2\sigma^2)} d\xi.$$

Ist μ größer als 3σ , was in praktischen Fällen häufig der Fall ist, so liegt a nahe bei Eins und kann weggelassen werden, so daß wieder mit der üblichen Dichte der $N(\mu, \sigma^2)$ -Verteilung gearbeitet werden kann.

Gestutzte Verteilungen (truncated distributions) lassen sich auch allgemein definieren. Man erhält sie, wenn etwa aus einer Grundgesamtheit mit Hilfe eines soge-

nannten **Zensierungsmechanismus** einige Stichprobenelemente entfernt werden, wie es z. B. bei klinischen Langzeitstudien durch Wohnortwechsel der Patienten o. ä. vorkommt.

B. Die Lognormalverteilung

In vielen Fällen kann man nicht davon ausgehen, daß eine Meßreihe x_1, \dots, x_n einer normalverteilten Grundgesamtheit entstammt, da die Verteilung des Merkmals X nicht symmetrisch um einen Wert ist. Besteht der Verdacht auf eine rechtschiefe Verteilung, d. h. $\gamma_1 = E(X - EX)^3 / \sqrt{\text{Var } X^3} > 0$, so kann oft durch Logarithmieren erreicht werden, daß die Verteilung (der logarithmierten Werte) nahezu symmetrisch und sogar einer Normalverteilung gut angepaßt ist, vgl. Abschnitt 4.5 in Kapitel V. Beispiele für solche Fälle sind z. B. oftmals die Umsatzziffern von Unternehmen, Konzentrationsmessungen, Empfindsamkeitsmessungen bei Verabreichung von Medikamenten und Wachstumsuntersuchungen, vgl. auch Abschnitt 3.3 in Kap. XIII. Ist nun $\ln X$ eine $N(\mu, \sigma^2)$ -verteilte Zufallsvariable, so heißt X selbst logarithmisch normalverteilt. Die Verteilungsfunktion von X ist dann

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(\ln X \leq \ln x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma\xi} e^{-(\ln \xi - \mu)^2/(2\sigma^2)} d\xi, \end{aligned}$$

und die Dichte der Verteilung von X ist, vgl. Abb. 6,

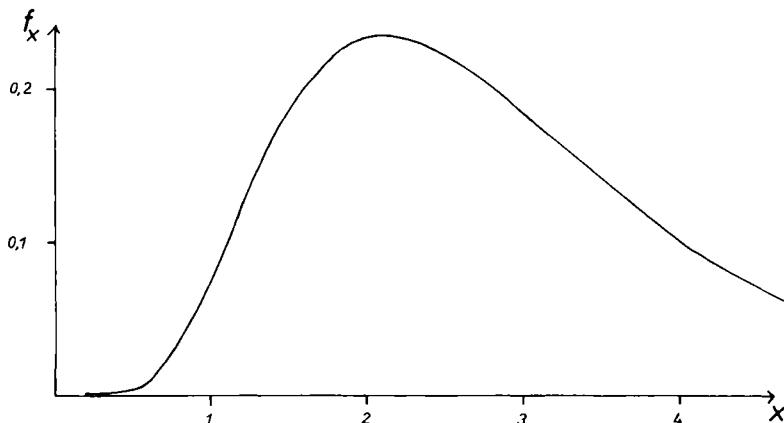


Abb. 6: Dichtefunktion der Lognormalverteilung mit $\mu = 1$ und $\sigma^2 = 0,5$

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-(\ln x - \mu)^2/(2\sigma^2)}, & \text{falls } x > 0 \\ 0 & \text{sonst.} \end{cases}$$

Der Erwartungswert der lognormalverteilten Zufallsvariablen X ist

$$EX = e^{\mu + \sigma^2/2}$$

ihrer Varianz ist

$$\text{Var } X = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1).$$

Weiter berechnet sich der Median zu e^μ und die Schiefe zu

$$\sqrt{e^{\sigma^2} - 1} (e^{\sigma^2} + 2) (> 0, \text{ d.h. die Verteilung ist rechtsschief}).$$

1.2.2. Prüfverteilungen

Bei der Ableitung des Einstichproben-Gauß-Tests im Abschnitt 4 des Kap. III war uns die Prüfgröße

$$\sqrt{n} \frac{\bar{x} - \mu}{\sigma_0},$$

begegnet. Wir hatten dort gesehen, daß die Kenntnis der Verteilung dieser Größe für die Durchführung des Tests entscheidend war. Als (Prüf-)Verteilung hatten wir die $N(0, 1)$ -Verteilung ermittelt, die uns die Festlegung des kritischen Wertes gestattete.

Bei der Durchführung statistischer Tests sind die Verteilungen von Größen wie

$$\sum_{i=1}^n (x_i - \bar{x})^2, \quad \frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}} \quad \text{usw. , mit } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

für normalverteilte x_i (d.h. für normalverteilte Zufallsvariablen X_i mit Realisationen x_i) ungeheuer wichtig (wie wir noch sehen werden). Deshalb bezeichnet man solche Verteilungen, die im folgenden aufgeführt werden, auch als **Prüfverteilungen**.

A. Die χ^2 -Verteilung

Sind X_1, \dots, X_n (stochastisch) unabhängige $N(0, 1)$ -verteilte Zufallsvariablen, so heißt die Verteilung von $\sum_{i=1}^n X_i^2$ eine zentrale χ^2 -Verteilung (**Chi-Quadrat-Verteilung**) mit n Freiheitsgraden, in Zeichen

$$\sum_{i=1}^n X_i^2 \sim \chi_n^2.$$

Die Größe $Z = \sum_{i=1}^n (X_i + \mu_i)^2$ heißt χ^2 -verteilt mit n Freiheitsgraden und Nichtzentralitätsparameter $\delta^2 = \sum_{i=1}^n \mu_i^2$ ($Z \sim \chi_n^2(\delta^2)$).

Ist $\delta^2 = 0$, d.h. $\mu_1 = \dots = \mu_n = 0$, so ergibt sich hieraus gerade als Spezialfall die (zentrale) χ^2 -Verteilung mit n Freiheitsgraden.

Hier wollen wir lediglich auf die zentrale χ^2 -Verteilung eingehen, die die Dichte

$$f_{\chi_n^2}(x) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{n/2-1} e^{-x/2}, \quad \text{für } x > 0,$$

besitzt. Eine solche Dichte ist in Abb. 7 graphisch dargestellt. Die Funktion Γ bezeichnet hier die **Gammafunktion**

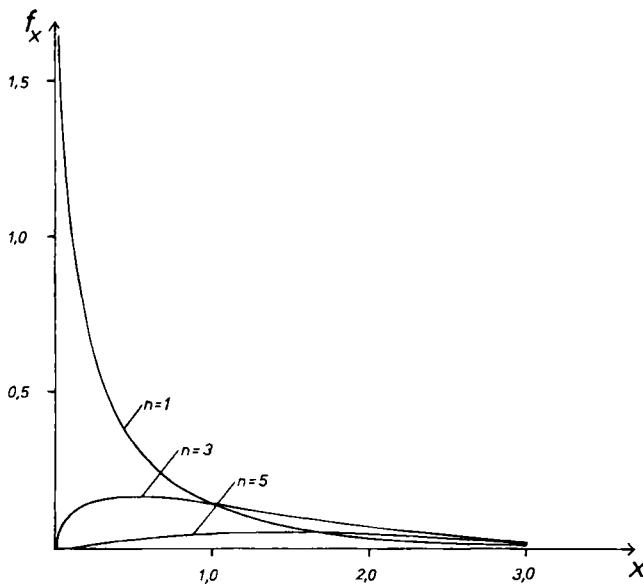


Abb. 7: Dichtefunktion der χ^2_1 -, χ^2_3 - und χ^2_5 -Verteilung

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt,$$

für die $\Gamma(x+1) = x \cdot \Gamma(x)$ gilt.

Für ganzzahlige Werte $x = n$ ergibt sich gerade $\Gamma(n) = (n-1)!$. Die Gammafunktion ist im Anhang vertafelt.

Der Erwartungswert und die Varianz einer χ_n^2 -verteilten Zufallsvariablen Z sind

$$EZ = n \quad \text{und} \quad \text{Var}Z = 2n.$$

Die Verteilung von Z besitzt die Schiefe $2\sqrt{2}/\sqrt{n}$ und den Exzeß $12/n$.

Für einige n und einige α sind die Quantile

$$\chi_{n; \alpha}^2 = F_{\chi_n^2}^{-1}(\alpha)$$

im Anhang vertafelt. Für große n lässt sich die Verteilungsfunktion $F_{\chi_n^2}(x)$ durch die Normalverteilung $N(n, 2n)$ approximieren:

$$F_{\chi_n^2}(x) = P(Z \leq x) \simeq \Phi\left(\frac{x - n}{\sqrt{2n}}\right).$$

Sind x_1, \dots, x_n Realisationen von stochastisch unabhängig $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen, dann lässt sich bei **bekanntem** μ die Varianz σ^2 schätzen durch

$$s_\mu^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \mu)^2.$$

Für die Verteilung dieser Größe (bzw. der zugehörigen Zufallsvariablen) gilt

$$n \cdot \frac{s_\mu^2}{\sigma^2} \sim \chi_n^2.$$

Ist dagegen μ unbekannt, so benutzt man die empirische Varianz

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

als Schätzwert für σ^2 . Für deren Verteilung (bzw. für die der zugehörigen Zufallsvariablen) gilt dann

$$(n-1) \frac{s^2}{\sigma^2} \sim \chi_{n-1}^2.$$

B. Die t-Verteilung

Eine Zufallsvariable

$$Z = \frac{X_0}{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n X_i^2}}$$

heißt (zentral) **t-verteilt mit n Freiheitsgraden** ($Z \sim t_n$), falls X_0, X_1, \dots, X_n (stochastisch) unabhängige standardnormalverteilte Zufallsvariablen sind. Die t-Verteilung geht auf Student zurück und wird daher manchmal auch **Studentsche t-Verteilung** genannt. Ihre Dichte ist

$$f_{t_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Diese Dichte ist symmetrisch um Null, vgl. Abb. 8. Ihre Verteilungsfunktion ist

$$F_{t_n}(x) = \int_{-\infty}^x f_{t_n}(\xi) d\xi.$$

Der Erwartungswert einer t_n -verteilten Zufallsvariablen existiert für $n \geq 2$ und ist – wie der Median – aufgrund der Symmetrie der Dichte gerade Null. Ist $n \geq 3$, so existiert auch die Varianz; diese ist $\frac{n}{n-2}$.

Die α -Quantile der t_n -Verteilung $t_{n; \alpha}$ sind für einige n und α im Anhang vertafelt; dabei haben wir uns die Tatsache zunutze gemacht, daß gilt

$$t_{n; \alpha} = -t_{n; 1-\alpha}.$$

Für große n nähert sich die Verteilungsfunktion einer t_n -verteilten Zufallsvariablen Z immer mehr der Standardnormalverteilung, so daß dann statt der Quantile $t_{n; \alpha}$ die Quantile u_α der Standardnormalverteilung verwandt werden können.

In der Statistik spielt die t-Verteilung eine bedeutende Rolle. Der Grund hierfür wird im folgenden deutlich. Sind $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ -verteilte (stochastisch) unabhängige Zufallsvariablen, so sind die Zufallsvariablen

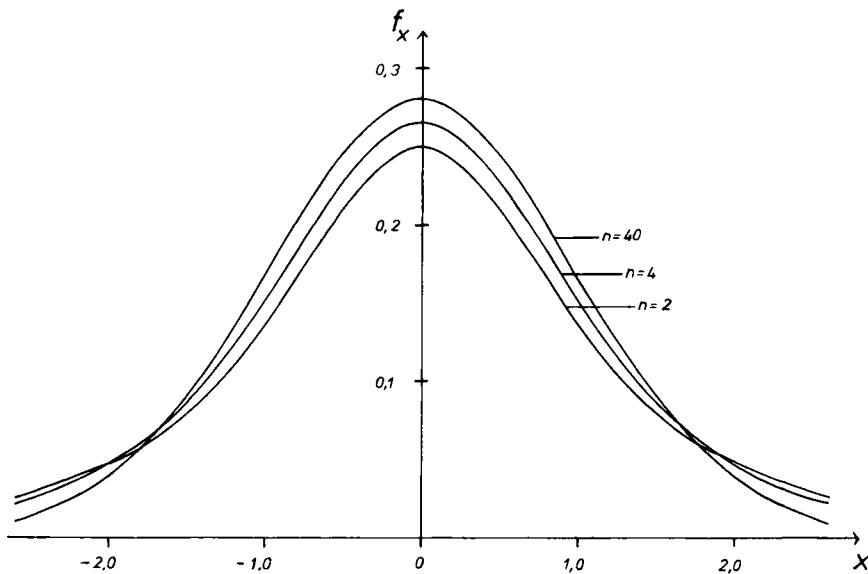


Abb. 8: Dichtefunktion der t_2 -, t_4 - und t_{40} -Verteilung

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{und} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

(stochastisch) unabhängig und es gilt

$$\sqrt{n} \frac{\bar{X} - \mu}{\sqrt{S^2}} \sim t_{n-1}.$$

Beispiel: Entstammen x_1, \dots, x_n einer normalverteilten Grundgesamtheit, so ist die Wahrscheinlichkeit, daß der unbekannte Parameter μ in einem $2s/\sqrt{n}$ Intervall um \bar{x} liegt gerade

$$\begin{aligned} P(\mu \in [\bar{X} - 2S/\sqrt{n}, \bar{X} + 2S/\sqrt{n}]) &= P(|\sqrt{n} \frac{\bar{X} - \mu}{S}| \leq 2) \\ &= 2P(0 \leq \sqrt{n} \frac{\bar{X} - \mu}{S} \leq 2) \\ &= 2(F_{t_{n-1}}(2) - F_{t_{n-1}}(0)) \\ &= 2F_{t_{n-1}}(2) - 1 \end{aligned}$$

und damit aus der t-Verteilung ablesbar.

Die t_1 -Verteilung ist uns bereits unter dem Namen Cauchy-Verteilung (mit Parametern $\mu = 0, \lambda = 1$), vgl. Abschnitt 8.1 in Kap. II, bekannt.

Unter einer nichtzentralen t_n -Verteilung – mit Nichtzentralitätsparameter δ – verstehen wir die Verteilung von $\sqrt{n} V/W$, wobei $V \sim N(\delta, 1)$, $W^2 \sim \chi_n^2$ und V, W unabhängig sind. Ihre Dichte ist i. a. nicht mehr geschlossen darstellbar; wir verweisen auf Kühlmeyer (1970).

C. Die F-Verteilung

Eine Zufallsvariable

$$Z = \left(\frac{1}{m} \sum_{i=1}^m Y_i^2 \right) / \left(\frac{1}{n} \sum_{i=1}^n V_i^2 \right)$$

heißt (zentral) **F-verteilt mit m und n Freiheitsgraden** ($Z \sim F_{m,n}$), wenn Y_1, \dots, Y_m und V_1, \dots, V_n (stochastisch) unabhängige standardnormalverteilte Zufallsvariablen sind. Eine $F_{m,n}$ -verteilte Zufallsvariable ergibt sich somit gerade als Quotient einer χ_m^2 - und einer χ_n^2 -verteilten Zufallsvariablen, multipliziert mit $\frac{n}{m}$.

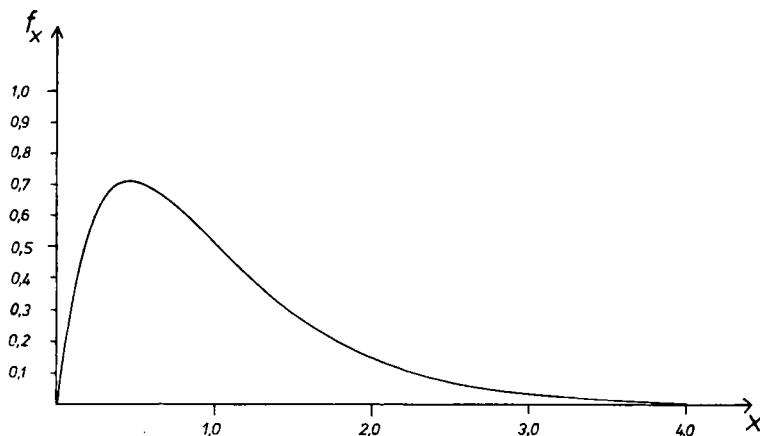


Abb. 9: Dichtefunktion der $F_{4,40}$ -Verteilung

Die Dichte der $F_{m,n}$ -Verteilung, vgl. auch Abb. 9, ist

$$f_{F_{m,n}}(x) = \frac{\left(\frac{m}{n}\right)^{m/2} \cdot x^{m/2 - 1}}{B\left(\frac{m}{2}, \frac{n}{2}\right)} (1 + \frac{m}{n}x)^{-\frac{m+n}{2}} \quad \text{für } x > 0;$$

dabei bezeichnet

$$B(p, q) = \frac{\Gamma(p) \cdot \Gamma(q)}{\Gamma(p+q)}$$

die **Eulersche Betafunktion**, und $\Gamma(\cdot)$ die Gammafunktion.

Ist $n > 2$, so existiert der Erwartungswert

$$EZ = \frac{n}{n-2},$$

und ist $n > 4$, so existiert die Varianz

$$\text{Var } Z = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

einer $F_{m,n}$ -verteilten Zufallsvariablen Z . Ist n groß, so läßt sich die Verteilung von $m \cdot Z$ auch durch eine χ_m^2 -Verteilung approximieren. Einige Quantile $F_{n,m;\alpha}$ der $F_{n,m}$ -Verteilung sind im Anhang vertafelt. Beim Nachschlagen von Tafelwerten macht man sich zunutze, daß gilt

$$F_{m,n;1-\alpha} = \frac{1}{F_{n,m;\alpha}} \quad \text{und} \quad t_{n;(1+\alpha)/2} = \sqrt{F_{1,n;\alpha}}.$$

Wie bei der χ^2 - und der t -Verteilung lassen sich auch bei der F -Verteilung Nichtzentralitätsparameter erklären (**nichtzentrale F-Verteilung**). Hierauf soll jedoch nicht eingegangen werden.

1.3. Punktschätzungen und Konfidenz-, Prognose- und Toleranz-Intervalle bei normalverteilter Grundgesamtheit

In diesem Abschnitt wollen wir Punkt- und Intervallschätzungen für die Parameter einer $N(\mu, \sigma^2)$ -verteilten Gesamtheit angeben. Neben den allgemein üblichen Schätzern berücksichtigen wir dabei insbesondere auch Schätzverfahren, die vornehmlich in experimentellen und technischen Bereichen zur Anwendung kommen.

1.3.1. Schätzen der Parameter μ und σ^2

Ein erwartungstreuer Schätzer für den Erwartungswert μ ist das arithmetische Mittel von n beobachteten Werten x_1, \dots, x_n (aus dieser $N(\mu, \sigma^2)$ -Verteilung)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Die Varianz σ^2 innerhalb der Grundgesamtheit ist durch

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

erwartungstreu schätzbar. Der – allerdings nicht erwartungstreue – Maximum-Likelihood Schätzer für σ^2 ist gegeben durch

$$s_{ML}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Ist der Erwartungswert μ bekannt, so wird jedoch σ^2 erwartungstreu geschätzt durch

$$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$$

Die Standardabweichung σ ist schätzbar durch

$$s_{ML} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2};$$

meist wird die Standardabweichung jedoch durch

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

oder durch den erwartungstreuen Schätzer

$$s' = \gamma_n \cdot s$$

geschätzt wobei gilt:

$$\gamma_n = \sqrt{\frac{n-1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

Für „große“ n ist γ_n nahezu Eins, für kleine n ergeben sich die Werte aus **Tab. 1**.

Tab. 1: Werte γ_n für $n < 20$

n	γ_n	n	γ_n
2	1,253	11	1,025
3	1,127	12	1,024
4	1,085	13	1,020
5	1,064	14	1,019
6	1,052	15	1,017
7	1,042	16	1,017
8	1,036	17	1,015
9	1,032	18	1,015
10	1,028	19	1,013

Vornehmlich im technischen Bereich werden die Parameter auch wie folgt geschätzt, und zwar der Mittelwert μ durch

$$m' = \frac{x_{(k)} + x_{(n-k+1)}}{2}.$$

Dieser Schätzer ist bei beliebiger Wahl von k erwartungstreu. (Dabei ist $x_{(j)}$ der j -größte Wert in der Beobachtungsreihe). Für kleine n ($n \leq 10$) wird die Standardabweichung σ durch

$$s'' = \frac{1}{\alpha_n} (x_{(n)} - x_{(1)}) = \frac{1}{\alpha_n} R_n \quad (= \frac{1}{\alpha_n} \cdot \text{Spannweite})$$

geschätzt, wobei die α_n in **Tabelle 2** angegeben sind.

Tab. 2: α_n für $2 \leq n \leq 10$

n	α_n	n	α_n
2	1,129	7	2,703
3	1,692	8	2,849
4	2,058	9	2,967
5	2,326	10	3,077
6	2,532		

Weitere in diesem Bereich übliche Angaben von Schätzern erhält man bei Aufteilung einer größeren Stichprobe vom Umfang n in p kleinere Stichproben $j = 1, \dots, p$ mit jeweiligem Umfang q :

$$\begin{array}{llll} x_1 \dots x_q & x_{q+1} \dots x_{2q} & \dots & x_{q(p-1)+1} \dots x_{qp} \\ (j=1) & (j=2) & & (j=p) \end{array} = x_n.$$

Die Schätzung für μ wird dann angegeben als

$$\bar{x} = \frac{1}{p} \sum_{j=1}^p \bar{x}_j, \quad \bar{x}_j = \text{„Mittelwert der } j\text{-ten Stichprobe“},$$

(natürlich ist \bar{x} gleich dem Mittelwert m der gesamten Stichprobe), und die Schätzung für σ wird angegeben als

$$\bar{s} = \frac{1}{p} \sum_{j=1}^p s_j, \quad s_j = \text{„Standardabweichung der } j\text{-ten Stichprobe“}.$$

Den Schätzer \bar{s} kann man natürlich noch mit γ_q aus Tabelle 1 korrigieren,

$$\bar{s}' = \gamma_q \bar{s}.$$

Eine andere gebräuchliche Schätzung für σ ist noch

$$\bar{s}'' = \frac{1}{\alpha_q p} \sum_{j=1}^p R_j = \frac{1}{\alpha_q} \bar{R}, \quad R_j = \text{„Spannweite der } j\text{-ten Stichprobe“}.$$

Diese Schätzmethoden sind i. a. jedoch nicht verwendbar, es sei denn, die Beobachtungen werden in der Reihenfolge verarbeitet, in der sie erhoben wurden.

Beispiel: (a) Bei einer Sportveranstaltung wurden die Laufzeiten auf der 100 m Strecke von $n = 16$ zufällig ausgewählten Teilnehmern gestoppt. Die gelaufenen Zeiten [in sec] finden sich in Tab. 3.

Tab. 3: Zeiten der 100 m-Läufer

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Zeit	13,2	12,7	11,4	12,9	14,2	11,7	11,3	12,4	12,6	13,9	11,2	13,1	14,3	12,5	12,2	12,3

Es ergibt sich hier

$$m = \bar{x} = 12,619$$

$$s_{ML} = 0,937, \quad s' = 0,951 \quad \text{und}$$

$$s^2 = 0,936 \quad (s = 0,9675).$$

(b) Von einer Maschine wurden Wollknäuel gewickelt. Nacheinander wurden $n = 10$ zufällig ausgewählte Knäuel gewogen. Die Gewichte [in g] sind in Tab. 4 angegeben.

Tab. 4: Gewichte der Wollknäuel

i	1	2	3	4	5	6	7	8	9	10
Gewicht	50,74	50,35	49,61	50,39	49,47	50,25	49,58	50,10	50,67	49,86

Wir wollen nun für dieses Beispiel alle vorgestellten Schätzer bestimmen. Es ergibt sich

$$\begin{aligned} m &= 50,102, \\ m' &= \frac{49,61 + 50,39}{2} = 50 \quad \text{für } k = 3, \\ m' &= \frac{49,47 + 50,74}{2} = 50,105 \quad \text{für } k = 1, \\ s_{ML} &= 0,4323, \\ s' &= 1,028 \cdot 0,4556 = 0,4684, \\ s^2 &= 0,2076 \quad (s = 0,4556), \\ s'' &= \frac{1}{3,077} (50,74 - 49,47) = 0,4128, \end{aligned}$$

und für eine Aufteilung in $p = 2$ Stichproben vom Umfang $q = 5$:

$$\begin{aligned} \bar{x} &= \frac{1}{2} (\bar{x}_1 + \bar{x}_2) = \frac{1}{2} (50,112 + 50,092) = 50,102, \\ \bar{s} &= \frac{1}{2} (\bar{s}_1 + \bar{s}_2) = \frac{1}{2} (0,546 + 0,411) = 0,4785, \\ \bar{s}' &= \gamma_5 \cdot \bar{s} = 1,064 \cdot 0,4785 = 0,5091, \\ \bar{s}'' &= \frac{1}{\alpha_5} \bar{R} = \frac{1}{2,326} \cdot \frac{1}{2} ([50,74 - 49,47] + [50,67 - 49,58]) = \frac{1}{2,326} 1,18 \\ &= 0,5073. \end{aligned}$$

1.3.2. Konfidenzintervalle für μ , σ^2 und σ

Wir kommen nun zur Bestimmung von Konfidenzintervallen für den Mittelwert μ , die Varianz σ^2 und die Standardabweichung σ bei normalverteilten Grundgesamtheiten.

Ein **Konfidenzintervall für μ (bei bekanntem σ^2)** haben wir schon in Kapitel III kennengelernt:

$$\left[\bar{x} - u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} ; \bar{x} + u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

ist ein $(1 - \alpha)$ Konfidenzintervall für μ .

Beispiel: (b) Ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ für den Mittelwert der von der Maschine gewickelten Wollknäuel ist (Die Standardabweichung σ sei vom Hersteller der Maschine mit $\sigma = 0,4$ angegeben.):

$$\left[50,102 - 1,960 \cdot \frac{0,4}{\sqrt{10}} ; 50,102 + 1,960 \cdot \frac{0,4}{\sqrt{10}} \right] = [49,85; 50,35].$$

Wie wir in Abschnitt 1.2.2.B gesehen haben, ist

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1}.$$

Daher gilt

$$P(t_{n-1; \alpha/2} \leq \frac{\bar{X} - \mu}{S} \sqrt{n} \leq t_{n-1; 1-\alpha/2}) = 1 - \alpha,$$

woraus man das folgende $1 - \alpha$ Konfidenzintervall für μ bei unbekanntem σ^2 erhält:

$$\left[\bar{X} - t_{n-1; 1-\alpha/2} \frac{s}{\sqrt{n}} ; \bar{X} + t_{n-1; 1-\alpha/2} \frac{s}{\sqrt{n}} \right].$$

Beispiel: (a) In unserem Beispiel mit den 100 m Läufern ergibt sich mit $s = \sqrt{s^2} = \sqrt{0,936} = 0,9675$ ein Konfidenzintervall zum Niveau $1 - \alpha = 0,98$ zu

$$\begin{aligned} & \left[12,619 - 2,602 \cdot \frac{0,9675}{\sqrt{16}} ; 12,619 + 2,602 \cdot \frac{0,9675}{\sqrt{16}} \right] \\ &= [12,619 - 0,629 ; 12,619 + 0,629] \\ &= [11,990 ; 13,248], \end{aligned}$$

d.h., „mit Wahrscheinlichkeit 0,98 liegt die mittlere Laufzeit μ in diesem Intervall“.

Aus Abschnitt 1.2.2 wissen wir, daß die Größe

$$\frac{(n-1)S^2}{\sigma^2}$$

χ^2 -verteilt ist mit $n-1$ Freiheitsgraden; also gilt

$$P(\chi^2_{n-1; \alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1; 1-\alpha/2}) = 1 - \alpha.$$

Damit läßt sich ein $1 - \alpha$ Konfidenzintervall für die Varianz σ^2 einer $N(\mu, \sigma^2)$ -verteilten Grundgesamtheit angeben als

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1; 1-\alpha/2}} ; \frac{(n-1)s^2}{\chi^2_{n-1; \alpha/2}} \right].$$

Ist der Mittelwert μ der zugrundeliegenden Verteilung bekannt, so können wir auch

$$\left[\frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi^2_{n; 1-\alpha/2}} ; \frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi^2_{n; \alpha/2}} \right]$$

als $1 - \alpha$ Konfidenzintervall für σ^2 verwenden, denn es gilt

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2.$$

Aus dem Vorhergehenden ergeben sich direkt folgende Konfidenzintervalle für die Standardabweichung σ :

$$\left[s \sqrt{\frac{n-1}{\chi^2_{n-1; 1-\alpha/2}}} ; s \sqrt{\frac{n-1}{\chi^2_{n-1; \alpha/2}}} \right]$$

und bei bekanntem Mittelwert μ auch noch

$$\left[\sqrt{\frac{\sum_i (x_i - \mu)^2}{\chi^2_{n; 1-\alpha/2}}}; \sqrt{\frac{\sum_i (x_i - \mu)^2}{\chi^2_{n; \alpha/2}}} \right].$$

Beispiel: (a) Wir wollen nun im Beispiel der 100 m Läufer Konfidenzintervalle zum Niveau $1 - \alpha = 0,9$ für die Varianz σ^2 sowie die Standardabweichung σ berechnen. Wir kennen natürlich den Mittelwert μ nicht, so daß die hier eingehende χ^2 -Verteilung 15 Freiheitsgrade besitzt. Es ist

$$\left[\frac{15 \cdot 0,936}{24,996}; \frac{15 \cdot 0,936}{7,261} \right] = [0,562; 1,934]$$

ein Konfidenzintervall für σ^2 und

$$[\sqrt{0,562}; \sqrt{1,934}] = [0,7497; 1,3907]$$

ein Konfidenzintervall für die Standardabweichung σ .

Wie wir bei den Punktschätzungen bemerkt haben, wird im technischen Bereich oft die Standardabweichung durch

$$s''' = \frac{1}{\alpha_q p} \sum_{j=1}^p R_j = \frac{1}{\alpha_q} \bar{R}$$

mit α_q aus Tab. 2 geschätzt. Das diesem Punktschätzer entsprechende Konfidenzintervall zum Niveau $1 - \alpha$ ergibt sich, da

$$\frac{\frac{1}{\alpha_q} \bar{R} - \sigma}{\delta_q \sigma} \sqrt{p}$$

approximativ standardnormalverteilt ist, aus der Gleichung

$$P \left[u_{\alpha/2} \leq \frac{\frac{1}{\alpha_q} \bar{R} - \sigma}{\delta_q \sigma} \sqrt{p} \leq u_{1-\alpha/2} \right] = 1 - \alpha$$

zu

$$\left[\frac{\bar{R}}{\alpha_q + u_{1-\alpha/2} \frac{\alpha_q \delta_q}{\sqrt{p}}}; \frac{\bar{R}}{\alpha_q - u_{1-\alpha/2} \frac{\alpha_q \delta_q}{\sqrt{p}}} \right],$$

wobei δ_q der Tab. 5 entnommen werden kann.

Beispiel: (b) In unserem Beispiel aus der Wollknäuelwicklung wählen wir, wie schon bei der Punktschätzung s''' für σ , $q = 5$ und $p = 2$. Es ergab sich dort

$$\bar{R} = 1,18.$$

Hiermit erhalten wir nun ein Konfidenzintervall für σ zum Niveau $1 - \alpha = 0,9$:

$$\left[\frac{1,18}{2,326 + 1,645 \frac{2,326 \cdot 0,371}{\sqrt{2}}}; \frac{1,18}{2,326 - 1,645 \frac{2,326 \cdot 0,371}{\sqrt{2}}} \right] = [0,354; 0,892].$$

Tab. 5: δ_q für $2 \leq q \leq 10$

q	δ_q
2	0,756
3	0,525
4	0,427
5	0,371
6	0,335
7	0,308
8	0,288
9	0,272
10	0,259

1.3.3. Prognose- und Toleranzintervalle

Schließlich wollen wir noch Prognose- und Toleranzintervalle bei normalverteilten Grundgesamtheiten bestimmen (vgl. auch Abschnitt 3 in Kap. III). Ein **Prognoseintervall für den Mittelwert \bar{y} einer künftigen Beobachtungsreihe y_1, \dots, y_m mit Trefferwahrscheinlichkeit $1 - p$** aus einer Meßreihe x_1, \dots, x_n lässt sich wegen

$$\frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n-1} \quad (\bar{y} \text{ Realisation von } \bar{Y})$$

und damit

$$P\left(t_{n-1; p/2} \leq \frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \leq t_{n-1; 1-p/2}\right) = 1 - p$$

bestimmen zu

$$\left[\bar{x} - t_{n-1; 1-p/2} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{x} + t_{n-1; 1-p/2} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{1}{m}} \right].$$

Ein solches Prognoseintervall enthält dann mit Wahrscheinlichkeit $1 - p$ das arithmetische Mittel von m zukünftigen Beobachtungswerten aus der Grundgesamtheit. Wählt man speziell $m = 1$, so liefert dieses Intervall natürlich einen Bereich, in dem ein zukünftiger Beobachtungswert mit $100(1 - p)$ -prozentiger Sicherheit liegt.

Kommen wir nun zur Bestimmung von Toleranzintervallen. Ein **Toleranzintervall für eine zukünftige Stichprobe y_1, \dots, y_m** lässt sich aufgrund von n vorliegenden Beobachtungen x_1, \dots, x_n bestimmen. In einem solchen Intervall liegen dann mindestens $p \cdot 100\%$ der zukünftigen Beobachtungen (**minimale Trefferwahrscheinlichkeit p**) und das mit einer statistischen Sicherheit von $(1 - \alpha) 100\%$ (**Sicherheitswahrscheinlichkeit $1 - \alpha$**).

Ist die Varianz σ^2 der normalverteilten Grundgesamtheit, aus der die Beobachtungswerte x_1, \dots, x_n stammen, bekannt, so lässt sich ein Toleranzintervall mit minimaler Trefferwahrscheinlichkeit p und Sicherheitswahrscheinlichkeit $1 - \alpha$ für

m zukünftige Beobachtungen aus dieser Grundgesamtheit in der Form

$$[\bar{x} - l\sigma; \bar{x} + l\sigma]$$

angeben. Der sogenannte **Toleranzfaktor l** ist dabei aus der Gleichung

$$\Phi\left(\frac{u_{1-\alpha/2}}{\sqrt{n}} + 1\right) - \Phi\left(\frac{u_{1-\alpha/2}}{\sqrt{n}} - 1\right) = p$$

zu bestimmen. Dies ist mit Hilfe der Tabelle der Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung, die im Anhang zu finden ist, möglich.

Ist die Varianz σ^2 der Grundgesamtheit unbekannt und muß aus der Stichprobe x_1, \dots, x_n geschätzt werden, so verwendet man

$$[\bar{x} - \lambda s; \bar{x} + \lambda s]$$

als Toleranzintervall mit minimaler Trefferwahrscheinlichkeit p und Sicherheitswahrscheinlichkeit $1 - \alpha$. Dabei kann man den **Toleranzfaktor λ** aus folgender Näherung bestimmen:

$$\lambda \approx r \sqrt{\frac{n-1}{\chi^2_{n-1; \alpha}}},$$

wobei r die Lösung folgender Gleichung ist:

$$p = \frac{1}{\sqrt{2\pi}} \int_{-r/\sqrt{n}}^{r/\sqrt{n}} e^{-x^2/2} dx.$$

Der Toleranzfaktor λ hängt also von n , p sowie $1 - \alpha$ ab. Für einige Wertekombinationen dieser Größen kann der Toleranzfaktor λ der Tab. 6 entnommen werden.

Beispiel: (a) Wir wollen ein Prognoseintervall mit Trefferwahrscheinlichkeit $1 - p = 0,9$ für die mittlere gestoppte Zeit von $m = 10$ Läufern einer weiteren Stichprobe y_1, \dots, y_{10} aus dieser Grundgesamtheit (Teilnehmer an gleichartigen Sportveranstaltungen) berechnen.

Mit

$$\bar{x} = 12,619,$$

$$s = \sqrt{s^2} = 0,9675,$$

$$n = 16,$$

$$m = 10 \quad \text{und} \quad t_{n-1; 1-p/2} = t_{15; 0,95} = 1,735$$

ergibt sich:

$$\begin{aligned} & \left[\bar{x} - t_{15; 0,95} s \sqrt{\frac{1}{n} + \frac{1}{m}}; \bar{x} + t_{15; 0,95} s \sqrt{\frac{1}{n} + \frac{1}{m}} \right] \\ &= \left[12,619 - 1,735 \cdot 0,9675 \cdot \sqrt{\frac{1}{16} + \frac{1}{10}}; 12,619 + 1,735 \cdot 0,9675 \cdot \sqrt{\frac{1}{16} + \frac{1}{10}} \right] \\ &= [12,619 - 0,684; 12,619 + 0,684] \\ &= [11,935; 13,303], \end{aligned}$$

Tab. 6: Toleranzfaktoren λ

n \ p	1 - $\alpha = 0,90$		1 - $\alpha = 0,95$		1 - $\alpha = 0,99$	
	0,90	0,95	0,90	0,95	0,90	0,95
2	15,98	18,80	32,02	37,67	160,19	188,49
3	5,85	6,92	8,38	9,92	18,93	22,40
4	4,17	4,94	5,37	6,37	9,40	11,15
5	3,50	4,15	4,28	5,08	6,61	7,86
6	3,13	3,72	3,71	4,41	5,34	6,35
7	2,90	3,45	3,37	4,01	4,61	5,49
8	2,74	3,26	3,14	3,73	4,15	4,94
9	2,63	3,13	2,97	3,53	3,82	4,55
10	2,54	3,02	2,84	3,38	3,58	4,27
12	2,40	2,86	2,66	3,16	3,25	3,87
14	2,31	2,76	2,53	3,01	3,03	3,61
16	2,25	2,68	2,44	2,90	2,87	3,42
18	2,19	2,61	2,37	2,82	2,75	3,28
20	2,15	2,56	2,31	2,75	2,66	3,17
25	2,08	2,47	2,21	2,63	2,49	2,97
30	2,03	2,41	2,14	2,55	2,39	2,84
50	1,92	2,28	2,00	2,38	2,16	2,58
100	1,82	2,17	1,87	2,23	1,98	2,36
500	1,72	2,05	1,74	2,07	1,78	2,12
∞	1,65	1,96	1,65	1,96	1,65	1,96

d.h. die mittlere gestoppte Zeit einer künftigen Stichprobe von $m = 10$ Läufern wird mit der Wahrscheinlichkeit 0,90 in diesem Intervall liegen. Wir bestimmen nun ein Toleranzintervall zur Sicherheitswahrscheinlichkeit $1 - \alpha = 0,9$ mit der minimalen Trefferwahrscheinlichkeit $p = 0,95$ für die 100 m-Zeiten.

Mit

$$\begin{aligned}\bar{x} &= 12,619, \\ s &= 0,9675 \quad \text{und} \quad n = 16, \\ \lambda &= \lambda(16; 0,9; 0,95) = 2,68\end{aligned}$$

ergibt sich:

$$\begin{aligned}[\bar{x} - \lambda s; \bar{x} + \lambda s] &= [12,619 - 2,68 \cdot 0,9675; 12,619 + 2,68 \cdot 0,9675] \\ &= [10,026; 15,212],\end{aligned}$$

d.h. mindestens 95% aller Läufer dieser Grundgesamtheit werden mit Wahrscheinlichkeit 0,9 eine Zeit zwischen 10,026 und 15,212 sec laufen.

Tab. 7: Gegenüberstellung von Konfidenz-, Prognose- und Toleranzintervallen

KI für μ $1 - \alpha = 0,98$ σ^2 unbekannt	PI für \bar{y} $1 - p = 0,90$ $m = 10$	TI für y_1, \dots, y_m $p = 0,95$ $1 - \alpha = 0,9$
[11,990; 13,248]	[11,935; 13,303]	[10,026; 15,212]

Eine Gegenüberstellung von Konfidenz-, Prognose- und Toleranzintervall bei den 100 m-Zeiten ist in **Tab. 7** gegeben.

Diese Gegenüberstellung zeigt (trotz verschiedener Niveaus $1 - \alpha$) schon, daß ein Konfidenzintervall kürzer ist als ein Prognoseintervall. Die größte Länge hat das Toleranzintervall, was aber natürlich noch von der speziellen Wahl der Trefferwahrscheinlichkeit p abhängt.

1.4. Bestimmung von benötigten Stichprobenumfängen bei Intervallschätzungen

In diesem Abschnitt wollen wir uns damit beschäftigen, wie man vorgegebene Genauigkeiten von Konfidenz-, Prognose- und Toleranzintervallen einhalten kann.

1.4.1. Einhaltung absoluter Genauigkeiten

Beschäftigen wir uns zunächst mit der **Absicherung einer vorgegebenen Breite d von Konfidenzintervallen für den Parameter μ** einer normalverteilten Grundgesamtheit.

Ist die **Varianz σ^2** der Grundgesamtheit **bekannt**, so können wir aus dem Konfidenzintervall für μ , vgl. Abschnitt 1.3, ersehen, daß die Breite des Konfidenzintervalls gerade

$$d = \frac{2 \cdot u_{1-\alpha/2} \cdot \sigma}{\sqrt{n}}$$

ist. Wir geben nun die Breite d vor und bestimmen dann die kleinste ganze Zahl n für die

$$n \geq \left(\frac{2 \cdot u_{1-\alpha/2} \cdot \sigma}{d} \right)^2$$

ist. Eine Stichprobe von diesem Umfang liefert uns dann ein Konfidenzintervall der gewünschten Breite.

Beispiel: Eine Maschine stellt Schrauben her, deren Länge – laut Herstellerangabe – normalverteilt ist mit Varianz $\sigma^2 = 0,04$. Wir wollen nun ein Konfidenzintervall für die erwartete Länge μ der Schrauben zum Niveau $1 - \alpha = 0,95$ bestimmen, das nicht länger als $d = 0,2$ ist. Wir berechnen den hierzu benötigten Stichprobenumfang n von Schrauben aus der Gesamtheit. Es ergibt sich wegen

$$\left(\frac{2 \cdot u_{0,975} \cdot 0,2}{0,2} \right)^2 = (2 \cdot 1,96)^2 = 15,37,$$

daß wir $n = 16$ wählen müssen.

Ist die **Varianz σ^2** der Grundgesamtheit **unbekannt**, so ist es schon etwas schwieriger, den benötigten Stichprobenumfang n zur Absicherung einer Breite d des Konfidenzintervalls für den Erwartungswert μ zu bestimmen, denn wir müssen die Varianz σ^2 aus einer kleinen Vorstichprobe abschätzen.

Stein (1945) schlägt vor, zunächst n_0 Beobachtungen zu machen, daraus die Varianz σ^2 abzuschätzen und dann das kleinste n mit

$$n \geq \left[\frac{2 \cdot t_{n_0-1; 1-\alpha/2} \cdot s}{d} \right]^2$$

zu bestimmen. Es ist hierbei die Abschätzung s^2 der Varianz σ^2 gerade

$$s^2 = \frac{1}{n_0 - 1} \sum_{i=1}^{n_0} (x_i - \bar{x})^2.$$

Dieses zweistufige Verfahren schätzt den Stichprobenumfang oft zu groß ab, vor allem bei ungeschickter Wahl von n_0 (zur Bestimmung von n_0 , bei gewisser Vorinformation über die Lage von σ vgl. man Seelbinder (1953)). Deshalb wird in der Praxis oft das folgende **mehrstufige Verfahren** angewandt.

Man zieht eine kleine Stichprobe $x_1, \dots, x_{n'}$ vom Umfang n' und bestimmt deren empirische Standardabweichung

$$s = \sqrt{\frac{1}{n' - 1} \sum_{i=1}^{n'} (x_i - \bar{x})^2}.$$

Da sich wegen

$$d = \frac{2 \cdot t_{n'-1; 1-\alpha/2} \cdot s}{\sqrt{n'}}$$

ergibt, daß

$$n'' \approx \left(\frac{2 \cdot t_{n''-1; 1-\alpha/2} \cdot s}{d} \right)^2$$

ist, suchen wir dann unter Zuhilfenahme der Tabelle der t-Verteilung aus dem Anhang die kleinste ganze Zahl n'' , für die

$$n'' \geq \left(\frac{2 \cdot t_{n''-1; 1-\alpha/2} \cdot s}{d} \right)^2$$

gilt.

Dann ziehen wir eine Stichprobe vom Umfang n'' , bestimmen s neu als deren empirische Standardabweichung und suchen die kleinste ganze Zahl n''' mit

$$n''' \geq \left(\frac{2 \cdot t_{n'''-1; 1-\alpha/2} \cdot s}{d} \right)^2.$$

Dieses Verfahren wiederholen wir solange bis das Konfidenzintervall die Breite d nicht überschreitet. Dabei werden im allgemeinen die bereits gemachten Beobachtungen im jeweils nächsten Schritt wieder mitgezählt, d. h. man macht dann jeweils nur noch $n'' - n'$, $n''' - n''$ usw. neue Beobachtungen. Hierbei muß auf jeden Fall sichergestellt sein, daß die Rahmenbedingungen unverändert bleiben.

Beispiel: Es sei bekannt, daß die Verlängerung der Schlafdauer durch Einnahme eines Schlafmittels normalverteilt ist. Wir wollen nun ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ mit maximaler Breite $d = 2$ für die erwartete Schlafverlängerung μ bestimmen. Dazu müssen wir zunächst die benötigte Anzahl n von Versuchspersonen berechnen.

Da wir die Varianz der Grundgesamtheit nicht kennen, wählen wir zunächst einmal $n' = 5$ Personen aus. Die Schlafverlängerungen durch Einnahme des Schlafmittels sind in Tab. 8 zusammengestellt.

Tab. 8: Schlafverlängerung x_i bei $n' = 5$ Versuchspersonen (in Std.)

i	1	2	3	4	5
x_i	1,7	-0,5	3,4	2,8	3,1

Aus diesen 5 Ergebnissen berechnen wir

$$s = \sqrt{\frac{1}{n' - 1} \sum_{i=1}^{n'} (x_i - \bar{x})^2} = \sqrt{\frac{1}{4} \sum_{i=1}^5 (x_i - \bar{x})^2} = 1,59.$$

Damit ergibt sich mit Hilfe der Tafel für die t-Verteilung wegen

$$13 = n'' \geq \left(\frac{2 \cdot t_{12; 0,975} \cdot 1,59}{2} \right)^2 = 12,015 \quad \text{und}$$

$$12 = n'' \leq \left(\frac{2 \cdot t_{11; 0,975} \cdot 1,59}{2} \right)^2 = 12,236$$

gerade $n'' = 13$. Anhand der mit 13 Versuchspersonen erhaltenen Daten berechnen wir nun zunächst das neue s und, falls das sich jetzt ergebende Konfidenzintervall die gewünschte Höchstbreite d überschreitet, bestimmen wir ein n''' . Dieses Verfahren führen wir solange durch bis wir ein Konfidenzintervall der Breite d gefunden haben.

Kommen wir nun zu einem rein **sequentiellen Verfahren** zur Bestimmung des Stichprobenumfangs. Wir ziehen zunächst eine Stichprobe x_1, x_2 vom Umfang $n = 2$ und bestimmen deren empirische Varianz

$$s_2^2 = \sum_{i=1}^2 (x_i - \bar{x})^2.$$

Ist

$$2 < \left(\frac{2 s_2 t_{1; 1-\alpha/2}}{d} \right)^2,$$

so machen wir eine weitere Beobachtung x_3 und berechnen

$$s_3^2 = \frac{1}{2} \sum_{i=1}^3 (x_i - \bar{x})^2;$$

ist dann

$$3 < \left(\frac{2 s_3 t_{2; 1-\alpha/2}}{d} \right)^2,$$

so machen wir eine Beobachtung x_4 usw.

Allgemein heißt das:

Ist die n -te Beobachtung x_n gemacht, so berechnet man

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n-2}{n-1} s_{n-1}^2 + \frac{1}{n-1} x_n^2 + \bar{x}_{n-1}^2 - \frac{n}{n-1} \bar{x}_n^2$$

mit

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{n-1}{n} \bar{x}_{n-1} + \frac{1}{n} x_n$$

und überprüft, ob gilt

$$n < \left(\frac{2 s_n t_{n-1; 1-\alpha/2}}{d} \right)^2.$$

Ist dies der Fall, so macht man noch eine $(n+1)$ -te Beobachtung. Ist dagegen

$$n \geq \left(\frac{2 s_n t_{n-1; 1-\alpha/2}}{d} \right)^2,$$

so gibt man mit Hilfe der gemachten Beobachtungen ein Konfidenzintervall an. Dieses Konfidenzintervall hält allerdings in der Regel das vorgegebene Niveau $1 - \alpha$ nicht genau ein. Poethke/Schmitz (1980) zeigen diesen Defekt anhand von Simulationsstudien auf, vgl. Tab. 9.

Tab. 9: Vorgegebene α -Werte und aufgrund von 20000 Simulationsläufen geschätzte $\hat{\alpha}$ -Werte $\hat{\alpha}$ (mit 99% Sicherheit Schwankungen von weniger als 0,5%)

α	$\hat{\alpha}$
10%	17%
5%	9,49%
2%	4,4%

Deshalb wird folgende **Modifikation des sequentiellen Verfahrens** vorgeschlagen:

Man verwendet das sequentielle Verfahren, beginnend mit einem Stichprobenumfang n_1 . Hat man dann das gesuchte n gefunden, so macht man nachträglich noch k Beobachtungen und erhält mit

$$[\bar{x} - d/2; \bar{x} + d/2],$$

wobei

$$\bar{x} = \frac{1}{n+k} \sum_{i=1}^{n+k} x_i$$

ist, ein Konfidenzintervall zum Niveau $1 - \alpha$ der vorgegebenen Breite. Die Anzahl k der zusätzlichen Beobachtungen muß dabei im vorhinein geeignet gewählt werden. Numerische Ergebnisse von Osborne/Simons (1970) und Rosenthal (1977) legen z. B. für $\alpha = 0,05$ und $n_1 = 3$ eine Wahl von $k = 5$ nahe.

Beispiel: Wir wollen ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ mit vorgegebener Breite $d = 3$ für den Erwartungswert μ einer normalverteilten Grundgesamtheit bestimmen. Wir wählen $n_1 = 3$ und $k = 5$. Die ersten 3 Beobachtungen sind:

$$x_1 = 3,17, x_2 = 5,73, x_3 = 4,95.$$

Somit ist

$$s_3^2 = \frac{1}{2} \sum_{i=1}^3 (x_i - \bar{x})^2 = 1,722 \quad \text{und}$$

$$3 < \left(\frac{2 \cdot t_{2; 0,975}}{3} \right)^2 \cdot 1,722 = \left(\frac{2 \cdot 4,303}{3} \right)^2 \cdot 1,722 = 14,17.$$

Wir müssen also noch eine vierte Beobachtung machen:

$$x_4 = 3,38.$$

Es ergibt sich

$$s_4^2 = 1,530 \quad \text{und}$$

$$4 < \left(\frac{2 \cdot t_{3; 0,975}}{3} \right)^2 \cdot 1,530 = \left(\frac{2 \cdot 3,182}{3} \right)^2 \cdot 1,530 = 6,89,$$

so daß noch eine neue Beobachtung gemacht werden muß:

$$x_5 = 4,21.$$

Nun ist

$$s_5^2 = 1,149 \quad \text{und}$$

$$5 \geq \left(\frac{2 \cdot t_{4; 0,975}}{3} \right)^2 \cdot 1,149 = \left(\frac{2 \cdot 2,776}{3} \right)^2 \cdot 1,149 = 3,94.$$

Wir machen nun also noch die $k = 5$ zusätzlichen Beobachtungen:

$$x_6 = 4,39, x_7 = 5,23, x_8 = 3,10, x_9 = 4,21 \quad \text{und} \quad x_{10} = 3,76.$$

Damit ist nun

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 4,213 \quad \text{und}$$

$$[\bar{x} - d/2; \bar{x} + d/2] = [4,213 - 1,5; 4,213 + 1,5] \\ = [2,713; 5,713]$$

ist ein Konfidenzintervall der gewünschten Breite zum Niveau 0,95 für den Erwartungswert μ der Grundgesamtheit.

Verfahren zur Bestimmung von Stichprobenumfängen zur Einhaltung von Genauigkeiten in prozentualer Abhängigkeit vom Mittelwert der Grundgesamtheit findet man in Abschnitt 1.4.2.

Kommen wir nun zur Bestimmung des Stichprobenumfangs n zur **Einhaltung einer vorgegebenen Breite d eines Prognoseintervalls für den Mittelwert \bar{y} einer zukünftigen Stichprobe y_1, \dots, y_m mit fest gewähltem Umfang m und gewählter Trefferwahrscheinlichkeit $1 - p$.** Kennen wir die Varianz σ^2 der zugrundeliegenden normalverteilten Grundgesamtheit nicht, so ist – vgl. Abschnitt 1.3 – die Breite d eines Prognoseintervalls für den Mittelwert \bar{y} einer zukünftigen Stichprobe vom Umfang m mit der Trefferwahrscheinlichkeit $1 - p$ gerade

$$d = 2 \cdot t_{n-1; 1-p/2} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{1}{m}},$$

d. h. der Umfang n der Stichprobe, aus der das Prognoseintervall bestimmt werden soll, ist die kleinste ganze Zahl n , für die gilt:

$$\frac{n \cdot m}{n + m} \geq \left(\frac{2 \cdot t_{n-1; 1-p/2} \cdot s}{d} \right)^2.$$

Da wir die Varianz σ^2 der Grundgesamtheit nicht kennen, haben wir sie durch die empirische Varianz s^2 einer kleinen Vorstichprobe vom Umfang n' abgeschätzt.

Mit Hilfe dieser Abschätzung bestimmen wir dann die kleinste ganze Zahl n'' mit:

$$\frac{n'' \cdot m}{n'' + m} \geq \left(\frac{2 \cdot t_{n''-1; 1-p/2} \cdot s}{d} \right)^2.$$

Wir machen nun n'' Beobachtungen, bestimmen das dazugehörige s und überprüfen, ob das jetzt erhaltene Prognoseintervall eine Breite von höchstens d besitzt. Ist dies nicht der Fall, so wiederholen wir dieses Vorgehen, erhalten entsprechend eine Zahl n''' , machen n''' (bzw. $n''' - n''$ weitere) Beobachtungen, überprüfen die Breite des sich ergebenden Prognoseintervalls usw.

Beispiel: Wir suchen den benötigten Stichprobenumfang n , um ein Prognoseintervall mit Trefferwahrscheinlichkeit $1 - p = 0,95$ und maximaler Breite $d = 1$ für das mittlere Gewicht \bar{y} von Wollknäueln einer zukünftigen Stichprobe vom Umfang $m = 20$ zu bestimmen. Wir wissen, daß die Grundgesamtheit der von einer Maschine produzierten Wollknäuel normalverteiltes Gewicht hat. Die Varianz σ^2 des Gewichts kennen wir nicht.

Zur Abschätzung von σ durch s wählen wir zunächst $n' = 5$ Knäuel zufällig aus. Die Gewichte der Knäuel sind in **Tab. 10** zusammengefaßt.

Tab. 10: Gewichte x_i von $n' = 5$ Knäuel (in g)

i	1	2	3	4	5
x_i	49,6	50,6	48,4	49,7	50,8

Es ergibt sich als erste Abschätzung nun:

$$s = \sqrt{\frac{1}{4} \sum_{i=1}^5 (x_i - \bar{x})^2} = 0,955.$$

Mit Hilfe der Tafel für die t-Verteilung ergibt sich für

$$n'' = 55: 14,667 = \frac{n'' \cdot m}{n'' + m} \geq \left(\frac{2 \cdot t_{55-1; 0,975} \cdot s}{d} \right)^2 = 14,665,$$

$$n'' = 54: 14,595 = \frac{n'' \cdot m}{n'' + m} \geq \left(\frac{2 \cdot t_{54-1; 0,975} \cdot s}{d} \right)^2 = 14,680.$$

Damit ist die gesuchte Zahl $n'' = 55$.

Wir wollen nun **Stichprobenumfänge** bestimmen, die nötig sind, damit ein Toleranzintervall für eine zukünftige Stichprobe aus einer normalverteilten Grundgesamtheit eine vorgegebene Breite d nicht überschreitet.

Ist die **Varianz σ^2** der Grundgesamtheit **bekannt**, so ist die Breite des Toleranzintervalls mit minimaler Trefferwahrscheinlichkeit p und Sicherheitswahrscheinlichkeit $1 - \alpha$ aus einer Stichprobe vom Umfang n gegeben durch

$$d = 2 \cdot \sigma \cdot l,$$

wobei l aus der Gleichung

$$\Phi\left(\frac{u_{1-\alpha/2}}{\sqrt{n}} + 1\right) - \Phi\left(\frac{u_{1-\alpha/2}}{\sqrt{n}} - 1\right) = p$$

bestimmt wird. Geben wir nun d vor und wollen den nötigen Stichprobenumfang zur Einhaltung der Breite d bestimmen, so berechnen wir zunächst

$$l = \frac{d}{2\sigma}$$

und suchen dann die kleinste ganze Zahl n , für die gilt:

$$\Phi\left(\frac{u_{1-\alpha/2}}{\sqrt{n}} + 1\right) - \Phi\left(\frac{u_{1-\alpha/2}}{\sqrt{n}} - 1\right) \geq p.$$

Bei der Bestimmung dieses Stichprobenumfangs n nehmen wir natürlich die Tabelle der Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung zu Hilfe, die im Anhang zu finden ist.

Beispiel: Wir wollen nun ein Intervall der Breite $d = 0,3$ bestimmen, daß mit Wahrscheinlichkeit $1 - \alpha = 0,99$ mindestens $100 p\% = 80\%$ einer Grundgesamtheit von Schrauben mit normalverteiltem Durchmesser enthält. Die Varianz σ^2 der Durchmesser ist laut Herstellerangabe $\sigma^2 = 0,01$. Um den hierzu benötigten Stichprobenumfang n zu bestimmen, berechnen wir zunächst

$$l = \frac{d}{2 \cdot \sigma} = \frac{0,3}{0,2} = 1,5.$$

Da für $n = 18$ gilt:

$$\Phi\left(\frac{u_{0,995}}{\sqrt{18}} + 1,5\right) - \Phi\left(\frac{u_{0,995}}{\sqrt{18}} - 1,5\right) = 0,80 \geq 0,80 = p$$

und für $n = 17$ gilt:

$$\Phi\left(\frac{u_{0,995}}{\sqrt{17}} + 1,5\right) - \Phi\left(\frac{u_{0,995}}{\sqrt{17}} - 1,5\right) = 0,79 \nleq 0,80 = p,$$

ist der gesuchte Stichprobenumfang aus der Grundgesamtheit der interessierenden Schrauben natürlich $n = 18$.

Ist die Varianz σ^2 unbekannt und wollen wir absichern, daß ein Toleranzintervall eine vorgegebene Breite d nicht überschreitet, so müssen wir zunächst aus einer kleinen Vorstichprobe vom Umfang n' die Standardabweichung σ durch s abschätzen und dann

$$\lambda' = \frac{d}{2s}$$

bestimmen. Dann suchen wir in der Tabelle 6 der Toleranzfaktoren λ aus Abschnitt 1.3 das kleinste n'' für das

$$\lambda \leq \lambda'$$

gilt. Dann ziehen wir eine Stichprobe vom Umfang n'' , schätzen s neu ab und überprüfen, ob die Breite $2 \cdot \lambda \cdot s$ des jetzt erhaltenen Toleranzintervalls den gewünschten Wert d überschreitet. Ist dies der Fall, so wiederholen wir das Vorgehen analog zu oben.

Beispiel: Nehmen wir einmal an, daß wir die Varianz σ^2 des normalverteilten Durchmessers einer Grundgesamtheit von Schrauben nicht kennen. Wir suchen

nun den nötigen Stichprobenumfang n , um die Breite $d = 0,7$ des Toleranzintervalls für den Durchmesser der Schrauben mit minimaler Trefferwahrscheinlichkeit $p = 0,9$ und Sicherheitswahrscheinlichkeit $1 - \alpha = 0,95$ zu garantieren. Wir ziehen zunächst eine Stichprobe vom Umfang $n' = 5$. Die Durchmesser der 5 Schrauben können Tab. 11 entnommen werden.

Tab. 11: Durchmesser x_i von $n' = 5$ Schrauben (in mm)

i	1	2	3	4	5
x_i	5,76	5,84	5,65	5,97	5,63

Hier ergibt sich nun wegen

$$s = \sqrt{\frac{1}{4} \sum_{i=1}^5 (x_i - \bar{x})^2} = 0,14$$

als Toleranzfaktor

$$\lambda' = \frac{d}{2 \cdot s} = \frac{0,7}{0,28} = 2,5.$$

Aus der Tab. 6 der Toleranzfaktoren λ aus Abschnitt 1.3. ergibt sich gerade $n'' = 15$,

da das zugehörige λ kleiner als 2,5 ist und für $n = 14$ das λ größer als 2,5 ist.

1.4.2. Einhaltung prozentualer Genauigkeiten

Bei **Konfidenzintervallen** zum Niveau $1 - \alpha$ ist es für den Praktiker oftmals nicht von Interesse, eine bestimmte Breite d des Konfidenzintervalls abzusichern, sondern vielmehr, eine **prozentuale Genauigkeit $\varepsilon \cdot 100\%$, bezogen auf das arithmetische Mittel**, einzuhalten. Die erforderliche Genauigkeit hängt dabei vom Verwendungszweck ab, insbesondere, wenn mit den errechneten Mittelwerten noch weiter gearbeitet werden soll. Etwa bei Arbeitszeitaufnahmen in der Produktherstellung zur Kostenanalyse eines Gesamtprodukts oder zur Festlegung von Akkordlöhnen wird die erforderliche Genauigkeit der benötigten Mittelwerte häufig durch Vorschriften oder Abmachungen von Vertragspartnern festgelegt; vgl. REFA (1971).

Bei der Festlegung einer prozentualen Genauigkeit $\varepsilon \cdot 100\%$ liegt der Gedanke zugrunde, daß kleine (Mittel-)Werte absolut gesehen exakter bestimmt werden sollen als große. Zum Beispiel sind bei Zeitaufnahmen zur Festsetzung von Akkordlöhnen die (durchschnittlichen) Zeitdauern kurzer Arbeitsvorgänge präziser zu erfassen als die längerer Arbeiten, da letztere, über einen festen Zeitraum gesehen, weniger häufig in die Gesamtkalkulation eingehen (Fehlerfortpflanzungsgesetz, vgl. Abschnitt 4.2 in Kap. V.).

A. Das Variationszahlverfahren

Wir können ein Konfidenzintervall für den Parameter μ einer normalverteilten Grundgesamtheit natürlich schreiben als

$$[\bar{x} - \varepsilon \cdot \bar{s}; \bar{x} + \varepsilon \cdot \bar{s}],$$

wenn wir

$$\varepsilon = \frac{t_{n-1; 1-\alpha/2}}{\sqrt{n}} \cdot \frac{s}{\bar{x}} = \frac{t_{n-1; 1-\alpha/2}}{\sqrt{n}} \cdot v$$

setzen. Dabei ist

$$v = \frac{s}{\bar{x}} = \frac{1}{\bar{x}} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

natürlich der empirischen Variationskoeffizient; vgl. Abschnitt 5 in Kap. I. Daher wird in einigen Bereichen der Praxis auch vom **Variationszahlverfahren** gesprochen.

Wir geben nun die gewünschte Genauigkeit ε vor und schätzen den Variationskoeffizienten v aus einer kleinen Stichprobe vom Umfang n' ab. Dann bestimmen wir die kleinste ganze Zahl n'' , für die gilt:

$$n'' \geq \left(\frac{t_{n''-1; 1-\alpha/2} v}{\varepsilon} \right)^2,$$

ziehen eine Stichprobe vom Umfang n'' und überprüfen, ob die gewünschte Genauigkeit erreicht wird; wenn nicht, so wiederholen wir das Verfahren. Dabei geht man in der Praxis allerdings so vor, daß man im zweiten Schritt nur noch $n'' - n'$ neue Beobachtungen macht und die ersten n' Beobachtungen weiterhin verwertet, falls sich die Rahmenbedingungen nicht verändern. (In grober Näherung ergibt sich stets $n'' \approx \left(\frac{2 \cdot v}{\varepsilon} \right)^2$ für $\alpha = 0,05$ und $n'' \geq 20$.) Zeichnerisch läßt sich das Problem mit Hilfe sogenannter **Leitertafeln** lösen, vgl. Abb. 10.

Beispiel: Für den normalverteilten Durchmesser einer Grundgesamtheit von Schrauben wollen wir ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ mit Genauigkeit $\varepsilon = 0,005$ angeben. Wir suchen den dazu benötigten Stichprobenumfang n nach dem Variationszahl-Verfahren.

Wir ziehen zunächst eine Stichprobe vom Umfang $n' = 7$ (die Ergebnisse sind in Tab. 12 zusammengestellt) und berechnen daraus den Variationskoeffizienten

Tab. 12: Durchmesser x_i der $n' = 7$ Schrauben (in mm)

i	1	2	3	4	5	6	7
x_i	5,37	5,46	5,52	5,39	5,41	5,49	5,45

$$v = \frac{s}{\bar{x}} = \frac{\sqrt{\frac{1}{6} \sum_{i=1}^7 (x_i - \bar{x})^2}}{\frac{1}{7} \sum_{i=1}^7 x_i} = \frac{0,054}{5,441} = 0,01.$$

Wegen

$$18 = n'' \geq \left(\frac{t_{17; 0,975} \cdot 0,01}{0,005} \right)^2 = 17,81 \quad \text{und}$$

$$17 = n'' \geq \left(\frac{t_{16; 0,975} \cdot 0,01}{0,005} \right)^2 = 17,98$$

ergibt sich $n'' = 18$, usw.

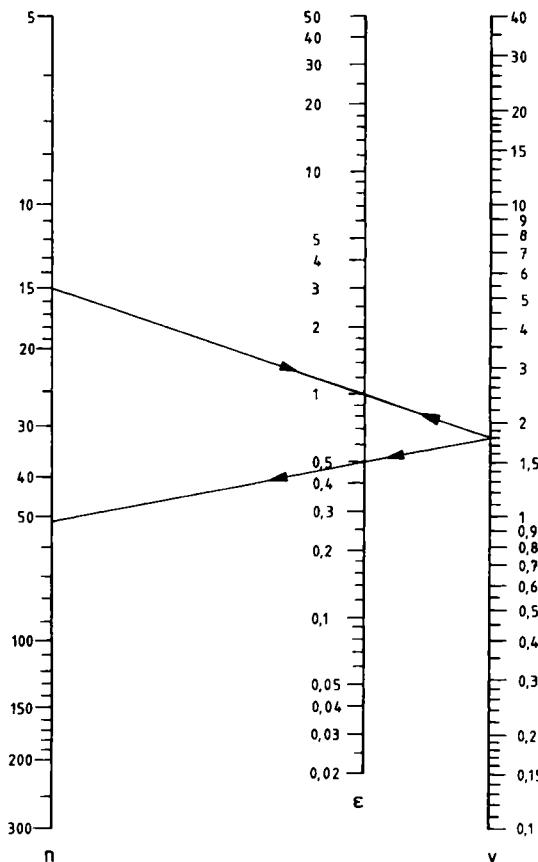


Abb. 10: Leitertafel zur Ermittlung des Stichprobenumfangs n bei vorgegebener Genauigkeit ε mit Hilfe des Variationskoeffizienten v (Angabe von ε und v in Prozent), $1 - \alpha = 0,95$, **Ablesebeispiele:** 1. $n = 15$, $v = 1,8\% \Rightarrow \varepsilon = 1\%$ 2. $v = 1,8\%$, $\varepsilon = 0,5\% \Rightarrow n = 52$

B. Das Streuzahlverfahren

Setzen wir nun

$$\varepsilon = \frac{t_{n-1; 1-\alpha/2}}{\sqrt{n}} z$$

mit

$$z = \begin{cases} \frac{R}{\alpha_n \cdot \bar{x}}, & \text{für } n \leq 10 \\ \frac{\bar{R}}{\alpha_5 \cdot \bar{x}}, & \text{für } n > 10 \end{cases}$$

wobei R die Spannweite einer Stichprobe, \bar{R} die mittlere Spannweite mehrerer kleiner Stichproben vom Umfang 5 bezeichnet und α_n der Tabelle 2 aus Abschnitt 1.3 entnommen werden kann, so ergibt sich ein anderes Verfahren – oft **Streuzahl-**

verfahren ($z \hat{=} \text{Streuzahl}$) genannt – zur Bestimmung des Stichprobenumfangs n zur Absicherung der Genauigkeit ε eines $1 - \alpha$ Konfidenzintervalls für den Parameter μ einer normalverteilten Grundgesamtheit.

Geben wir die gewünschte Genauigkeit ε vor und bestimmen die Streuzahl z aus einer Vorstichprobe vom Umfang n' , so ergibt sich der zunächst benötigte Stichprobenumfang als kleinste ganze Zahl n'' mit

$$n'' \geq \left(\frac{t_{n'-1; 1-\alpha/2} \cdot R}{\varepsilon \alpha_{n'} \bar{x}} \right)^2, \quad \text{falls } n' \leq 10,$$

und als kleinste ganze Zahl n'' mit

$$n'' \geq \left(\frac{t_{n'-1; 1-\alpha/2} \cdot \bar{R}}{\varepsilon \alpha_5 \bar{x}} \right)^2, \quad \text{falls } n' > 10, \text{ wobei } \alpha_5 = 2,326 \text{ ist.}$$

Ist mit diesen n'' Beobachtungen die gewünschte Genauigkeit nicht erreicht, so wiederholen wir dieses Vorgehen analog zu oben. Zeichnerisch lässt sich dieses Problem mit Hilfe der Leitertafel aus **Abb. 11** lösen.

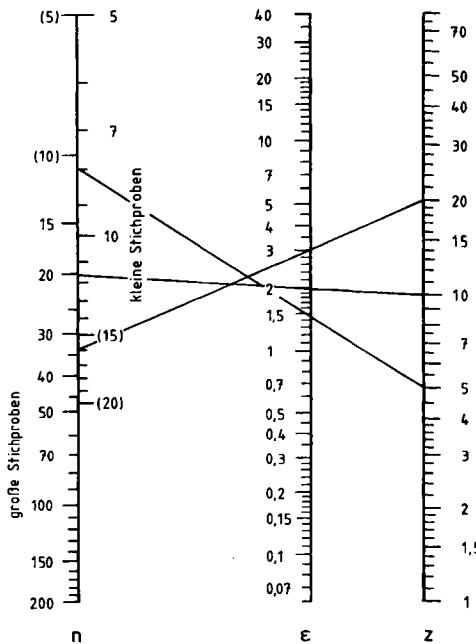


Abb. 11: Leitertafel zur Ermittlung des Stichprobenumfangs n bei vorgegebener Genauigkeit ε mit Hilfe der Streuzahl z (Angabe von ε und z in Prozent), $1 - \alpha = 0,95$,

Ablesebeispiele: 1. Abschätzungen für z werden aus kleiner Vorstichprobe (Umfang $n' \leq 15$) gewonnen:

a) $n' = 5, \varepsilon = 1,45\%, z = 5\% \Rightarrow n = 8$

b) $n' = 6, \varepsilon = 3\%, z = 20\% \Rightarrow n = 16 > 15$, also von vorne beginnen mit Vorstichprobe vom Umfang 16

2. Abschätzungen für z werden aus großer Vorstichprobe (Umfang $n' > 15$) gewonnen:

$n' = 17, \varepsilon = 2\%, z = 10\% \Rightarrow n = 20$

Beispiel: Für den Durchmesser von Schrauben (s. voriges Beispiel, Tab. 12) wollen wir aufgrund der dort vorliegenden Vorstichprobe den benötigten Stichprobenumfang aufgrund des Streuzahlverfahrens bestimmen. Es ist

$$z = \frac{R}{\alpha_7 \cdot \bar{x}} = \frac{5,52 - 5,37}{2,703 \cdot 5,441} = 0,0102,$$

und somit ergibt sich für $\varepsilon = 0,005$ und $1 - \alpha = 0,95$ wegen

$$19 = n'' \geq \left(\frac{t_{18; 0,975} \cdot 0,0102}{0,005} \right)^2 = 18,37 \quad \text{und}$$

$$18 = n'' \nleq \left(\frac{t_{17; 0,975} \cdot 0,0102}{0,005} \right)^2 = 18,53$$

mit Hilfe des Streuzahlverfahrens $n'' = 19$. Das Streuzahlverfahren liefert hier also nicht das gleiche Ergebnis wie das Variationszahlverfahren.

Wir wollen hier noch den nächsten Schritt vorführen. Zunächst ziehen wir jetzt eine Stichprobe vom Umfang 20 (kleinste durch 5 teilbare Zahl, die nicht kleiner als 19 ist). Die Meßergebnisse sind in **Tab. 13** zusammengefaßt.

Tab. 13: Durchmesser x_i von $n'' = 20$ Schrauben

i	1	2	3	4	5	6	7	8	9	10
x_i	5,46	5,37	5,55	5,42	5,39	5,51	5,36	5,42	5,49	5,45
i	11	12	13	14	15	16	17	18	19	20
x_i	5,53	5,48	5,46	5,39	5,42	5,48	5,50	5,41	5,39	5,48

Hiermit ergibt sich

$$R_1 = 5,55 - 5,37 = 0,18,$$

$$R_2 = 5,51 - 5,36 = 0,15,$$

$$R_3 = 5,53 - 5,39 = 0,14 \quad \text{und}$$

$$R_4 = 5,50 - 5,39 = 0,11.$$

Somit ist

$$z = \frac{\bar{R}}{\alpha_5 \bar{x}} = \frac{0,145}{2,326 \cdot 5,448} = 0,01144$$

und wegen

$$23 = n''' \geq \left(\frac{t_{22; 0,975} \cdot 0,01144}{0,005} \right)^2 = 22,52,$$

$$22 = n''' \nleq \left(\frac{t_{21; 0,975} \cdot 0,01144}{0,005} \right)^2 = 22,65$$

ergibt sich $n''' = 23$ usw.

1.5. Testen von Parameter-Hypothesen und Bestimmung des benötigten Stichprobenumfangs

1.5.1. Testen von Hypothesen über die Parameter einer normalverteilten Grundgesamtheit

Wir wollen hier Hypothesen über die Parameter μ und σ^2 einer normalverteilten Grundgesamtheit testen. Wir werden stets auf ein- und zweiseitige Fragestellungen eingehen. An dieser Stelle verzichten wir auf die Vorstellung sequentieller Testverfahren, da hierauf speziell in Abschnitt 6 eingegangen wird. In den konkreten Beispielen greifen wir wieder auf die Zahlen der beiden Beispiele aus Abschnitt 1.3 zurück. In den nun folgenden Ausführungen werden wir im Beispiel verschiedene Tests auf die Daten der Beispiele aus Abschnitt 1.3 anwenden, obwohl dies nicht unproblematisch ist. Gewöhnlich sollte man sich stets im vorhinein auf einen Test festlegen. Sei weiterhin x_1, \dots, x_n eine Stichprobe dieser $N(\mu, \sigma^2)$ -Verteilung.

A. Hypothesen über den Mittelwert μ

Zunächst wollen wir Hypothesen über den Mittelwert μ testen. Die einseitigen Hypothesen und Alternativen sind dabei

$$\begin{aligned} H_{01}: \mu \leq \mu_0 &\text{ gegen } H_{11}: \mu > \mu_0 \quad \text{und} \\ H_{02}: \mu \geq \mu_0 &\text{ gegen } H_{12}: \mu < \mu_0. \end{aligned}$$

Im zweiseitigen Testproblem wird

$$H_{03}: \mu = \mu_0 \quad \text{gegen} \quad H_{13}: \mu \neq \mu_0$$

getestet. Alle diese Tests werden zum Niveau α durchgeführt.

Wir müssen nun verschiedene Fälle unterscheiden:

(a) Die Varianz σ^2 ist bekannt

In diesem Falle können wir natürlich den **Gauß-Test**, den wir schon in Kap. III, Abschnitt 4 kennengelernt haben, anwenden.

Testen wir H_{01} gegen H_{11} , so verwerfen wir die Nullhypothese, falls

$$z > u_{1-\alpha}, \quad \text{mit} \quad z = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n},$$

ist. Beim Test von H_{02} gegen H_{12} verwerfen wir dagegen die Hypothese H_{02} im Falle

$$z < u_\alpha.$$

Im zweiseitigen Testproblem schließlich wird H_{03} verworfen, falls

$$|z| > u_{1-\alpha/2}$$

ist. Dies ergibt sich aufgrund der Symmetrie der $N(0, 1)$ -Verteilung.

Beispiel: (b) Nehmen wir an, daß die Wickelmaschine vom Hersteller mit einer Standardabweichung von $\sigma = 0,4$ angegeben ist.

Wir wollen nun zum Niveau $\alpha = 0,05$ die zweiseitige Hypothese

$$H_{03}: \mu = 50$$

gegen die Alternative

$$H_{13}: \mu \neq 50$$

testen, vgl. Tab. 4.

Es ist $u_{1-\alpha/2} = u_{0,975} = 1,96$ und

$$|z| = \frac{50,102 - 50}{0,4} \sqrt{10} = 0,806.$$

Somit können wir unsere Nullhypothese nicht verwerfen.

(b) Die Varianz σ^2 ist unbekannt

In diesem Fall benutzen wir den **Studentschen t-Test** mit der Prüfgröße

$$t = \frac{\bar{x} - \mu_0}{s} \sqrt{n};$$

Wir müssen hier die einseitige Nullhypothese H_{01} verwerfen, wenn

$$t > t_{n-1; 1-\alpha}$$

ist.

Die Hypothese H_{02} wird dann verworfen, wenn gilt

$$t < t_{n-1; \alpha}$$

und die zweiseitige Hypothese müssen wir verwerfen, wenn

$$|t| > t_{n-1; 1-\alpha/2}$$

(wegen der Symmetrie der zentralen t-Verteilung).

Beispiel: (a) Bei den 100 m Zeiten unserer Läufer aus Abschnitt 1.3, vgl. Tab. 3, wollen wir die Hypothese H_{02} : die Läufer entstammen einer normalverteilten Grundgesamtheit mit Erwartungswert $\mu \geq 13,3$; d.h.

$$H_{02}: \mu \geq 13,3$$

gegen die Alternative

$$H_{12}: \mu < 13,3$$

zum Niveau $\alpha = 0,05$ testen.

Es ergibt sich wegen $t_{n-1; \alpha} = t_{15; 0,05} = -1,753$ und

$$t = \frac{12,619 - 13,3}{0,9675} \sqrt{16} = -2,816,$$

daß die Nullhypothese H_{02} zu diesem Niveau verworfen werden muß.

B. Hypothesen über die Varianz σ^2

Wir wollen nun **Hypothesen über die Varianz σ^2** einer normalverteilten Grundgesamtheit testen. Dabei ergeben sich keine grundlegenden Unterschiede daraus, ob der Mittelwert μ **bekannt** oder **unbekannt** ist, so daß wir beide Fälle zusammen abhandeln können.

Die interessierenden Nullhypotesen (sowie Alternativen) für solche Tests über σ^2 sind im einseitigen Falle

$$\begin{aligned} H_{01}: \sigma^2 &\leq \sigma_0^2 \quad \text{gegen} \quad H_{11}: \sigma^2 > \sigma_0^2, \\ H_{02}: \sigma^2 &\geq \sigma_0^2 \quad \text{gegen} \quad H_{12}: \sigma^2 < \sigma_0^2, \end{aligned}$$

und im zweiseitigen Testproblem

$$H_{03}: \sigma^2 = \sigma_0^2 \quad \text{gegen} \quad H_{13}: \sigma^2 \neq \sigma_0^2.$$

Wie wir schon in Abschnitt 1.2 gesehen haben, ist die Größe, deren Realisation

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (\text{für unbekanntes } \mu) \text{ bzw.}$$

$$\chi^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} \quad (\text{für bekanntes } \mu)$$

ist, für $\sigma^2 = \sigma_0^2$ χ^2 -verteilt mit $n-1$ Freiheitsgraden bzw. n Freiheitsgraden.

Die Entscheidungsregeln müssen also für unser Testproblem so lauten:

Wir verwerfen die Nullhypothese H_{01} , falls gilt

$$\chi^2 > \chi_{n-1; 1-\alpha}^2 \quad (\text{bzw. } \chi^2 > \chi_{n; 1-\alpha}^2)$$

und die Nullhypothese H_{02} , falls gilt

$$\chi^2 < \chi_{n-1; \alpha}^2 \quad (\text{bzw. } \chi^2 < \chi_{n; \alpha}^2).$$

Das zweiseitige Problem verlangt im Falle

$$\begin{aligned} \chi^2 &> \chi_{n-1; 1-\alpha/2}^2 \quad \text{oder} \quad \chi^2 < \chi_{n-1; \alpha/2}^2 \\ (\text{bzw. } \chi^2 &> \chi_{n; 1-\alpha/2}^2 \quad \text{oder} \quad \chi^2 < \chi_{n; \alpha/2}^2) \end{aligned}$$

ein Verwerfen der Hypothese H_{03} .

Beispiel: (b) Wir wollen überprüfen, ob die Herstellerangabe $\sigma_0 = 0,4$ für die Standardabweichung der Maschinen zur Wollknäuelaufwicklung eventuell auf $\sigma = 0,3$ zu verbessern ist. Wir testen dazu die Hypothese

$$H_{03}: \sigma^2 = 0,09$$

gegen die Alternative

$$H_{13}: \sigma^2 \neq 0,09$$

zum Niveau $\alpha = 0,1$.

Da uns der exakte Mittelwert μ nicht bekannt ist, müssen wir dabei

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

als Teststatistik benutzen. Die kritische Werte ergeben sich aus einer χ^2 -Verteilung mit $n-1 = 9$ Freiheitsgraden. Man erhält

$$\chi^2 = \frac{9 \cdot 0,2076}{0,09} = 20,76 \quad \text{und}$$

$$\chi_{9; 0,95}^2 = 16,919, \chi_{9; 0,05}^2 = 3,325.$$

Damit muß wegen

$$20,76 > 16,919$$

die Nullhypothese $H_0: \sigma^2 = 0,09$ verworfen werden.

1.5.2. Bestimmung des Stichprobenumfangs n beim Testen von Hypothesen über den Erwartungswert μ einer normalverteilten Grundgesamtheit bei vorgegebenem Fehler 1. Art α und Fehler 2. Art β

Testen wir die Hypothese

$$H_0: \mu \leq \mu_0 \quad \text{gegen} \quad H_1: \mu > \mu_0$$

zum Niveau α und geben an einer Stelle $\mu_1 > \mu_0$ einen Fehler 2. Art β vor, so muß, um beide Fehler einhalten zu können,

$$n \geq \left(\frac{(u_{1-\alpha} + u_{1-\beta})\sigma}{\mu_1 - \mu_0} \right)^2$$

sein. Dabei bezeichnet σ die bekannte Standardabweichung in der Grundgesamtheit. Aus eben dieser Formel bestimmt sich n auch, wenn wir

$$H_0: \mu \geq \mu_0 \quad \text{gegen} \quad H_1: \mu < \mu_0$$

zum Niveau α testen bei vorgegebenem Fehler 2. Art β an einer Stelle $\mu_1 < \mu_0$. Im zweiseitigen Testproblem zum Niveau α

$$H_0: \mu = \mu_0 \quad \text{gegen} \quad H_1: \mu \neq \mu_0$$

ergibt sich n bei vorgegebenem β -Fehler an einer Stelle $\mu_1 \neq \mu_0$ zu

$$n \geq \left(\frac{(u_{1-\alpha/2} + u_{1-\beta})\sigma}{\mu_1 - \mu_0} \right)^2.$$

In den exakten Formeln ist hier σ^2 , die Varianz in der Grundgesamtheit, als bekannt vorausgesetzt. Näherungsweise lassen sich die Formeln natürlich auf Situationen übertragen, wo σ zwar unbekannt ist, aber eine gute Abschätzung s vorhanden ist.

Beispiel: Wir wollen den nötigen Stichprobenumfang n bestimmen, um beim Test zum Niveau $\alpha = 0,025$

$$H_0: \mu \geq 15 \quad \text{gegen} \quad H_1: \mu < 15$$

an der Stelle $\mu_1 = 12$ einen Fehler 2. Art von $\beta = 0,05$ abzusichern. Die Varianz σ^2 der Grundgesamtheit nehmen wir einmal als $\sigma^2 = 9$ an. Es ergibt sich

$$n \geq \left(\frac{(u_{1-\alpha} + u_{1-\beta})\sigma}{\mu_0 - \mu_1} \right)^2 = \left(\frac{(1,96 + 1,645)3}{15 - 12} \right)^2 = \left(\frac{10,815}{3} \right)^2 = 12,996,$$

d.h. wir müssen mindestens eine Stichprobe vom Umfang $n = 13$ ziehen. Damit haben wir dann sichergestellt, daß – falls H_0 vorliegt – man sich höchstens mit Wahrscheinlichkeit $\alpha = 0,025$ (falschlicherweise) für die Alternative H_1 entscheidet und daß – falls in Wirklichkeit der wahre Parameter μ kleiner als $\mu_1 = 12$ ist – man sich höchstens mit Wahrscheinlichkeit $\beta = 0,05$ (falschlicherweise) für die Hypothese H_0 entscheidet.

Liegt der wahre Parameter zwischen 12 und 15, so kennt man den Fehler 2. Art nicht. Daher ist μ_1 (hier $\mu_1 = 12$) stets so zu wählen, daß die Abweichung zwischen μ_0 und μ_1 nicht von praktischem Interesse ist.

1.6. Anpassungstests an die Normalverteilung

In diesem Abschnitt wollen wir den χ^2 -Anpassungstest sowie den Kolmogoroff-Smirnov-Test für Einstichprobenprobleme kennenlernen. Liegen n unabhängige Beobachtungen x_1, \dots, x_n vor, so überprüfen solche Anpassungstests die Hypothese, daß die Beobachtungen aus einer Grundgesamtheit stammen, welche bezüglich des betrachteten Merkmals – in diesem Falle – normalverteilt ist.

A. Der χ^2 -Anpassungstest

Mit Hilfe des χ^2 -Anpassungstests auf Normalverteilung können wir die Hypothese

$$H_0: \text{die Grundgesamtheit ist } N(\mu_0, \sigma_0^2)\text{-verteilt}$$

gegen die Alternative

$$H_1: \text{die Grundgesamtheit ist nicht } N(\mu_0, \sigma_0^2)\text{-verteilt}$$

zum Signifikanzniveau α testen (mit spezifizierten Werten μ_0 und σ_0^2).

Wir gehen dabei wie folgt vor:

1. Schritt: Wir unterteilen das reelle Intervall $(-\infty, \infty)$ in k Klassen.
2. Schritt: Nun müssen wir feststellen, wie viele unserer beobachteten Meßwerte x_1, \dots, x_n in jeder Klasse liegen. Diese Anzahlen bezeichnen wir mit O_i , $i = 1, \dots, k$ (dabei steht O für „observed“).
3. Schritt: Wir berechnen die Wahrscheinlichkeit $p_i = P(X \in I_i)$, $i = 1, \dots, k$, mit denen eine Beobachtung unter der Hypothese H_0 in der i -ten Klasse liegt. Die Zahl $E_i = n p_i$, $i = 1, \dots, k$, (E steht für „expected“) ist dann die Zahl der unter H_0 erwarteten Beobachtungen in der i -ten Klasse.
4. Schritt: Wir berechnen für unsere konkret vorliegende Beobachtungsreihe die Größe

$$T = \sum_{i=1}^k \frac{1}{E_i} (O_i - E_i)^2,$$

die unter H_0 asymptotisch χ^2 -verteilt ist mit $k - 1$ Freiheitsgraden. Diese Approximation ist hinreichend genau, falls nicht mehr als 20% der E_i kleiner als 5 sind und kein Wert E_i kleiner als 1 ist.

Wir müssen die Nullhypothese H_0 nun zum Signifikanzniveau α verwerfen, falls gilt:

$$T > \chi_{k-1; 1-\alpha}^2.$$

Möchte man die Nullhypothese

$$H_0: \text{die Stichprobe ist } N(\mu, \sigma^2)\text{-verteilt gegen}$$

$$H_1: \text{die Stichprobe ist nicht normalverteilt}$$

testen, wenn man μ und σ^2 nicht auf bestimmte Werte festlegt, so ist es auch noch möglich, den χ^2 -Test zu verwenden.

Schätzt man die Parameter μ und σ^2 nach der Maximum-Likelihood-Methode aus den **Klassenhäufigkeiten**, so reduziert sich die Anzahl der Freiheitsgrade um 2.

Verwendet man hingegen die Stichprobenschätzer $m = \bar{x}$ und $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$, so ist dieses Vorgehen nur näherungsweise richtig. Allerdings ist es in solchen Fällen korrekt, H_0 zu verwerfen, falls

$$T > \chi^2_{k-1; 1-\alpha},$$

bzw. nicht zu verwerfen, falls

$$T \leq \chi^2_{k-1-2; 1-\alpha},$$

gilt, vgl. Chernoff/Lehmann (1954). Hierbei ist man beim Verwerfen von H_0 allerdings zu konservativ.

Völlig analog läßt sich der χ^2 -Anpassungstest für beliebige Verteilungen durchführen. Werden dabei aus der Stichprobe geschätzte Parameter benutzt, so sind entsprechend die Freiheitsgrade von χ^2 um die Anzahl der geschätzten Parameter zu erniedrigen.

B. Der Kolmogoroff-Smirnov-Anpassungstest

Soll überprüft werden, ob die unbekannte Verteilungsfunktion $F(x)$ der betrachteten Grundgesamtheit mit einer hypothetischen Verteilungsfunktion $F_0(x)$ übereinstimmt, so kann man z. B. auch den Kolmogoroff-Smirnov-Test verwenden. Dabei muß vorausgesetzt werden, daß die hypothetische Verteilung $F_0(x)$ stetig ist. Erfüllt $F_0(x)$ diese Bedingung nicht, so ist der Test konservativ, hält also länger an der Hypothese der Gleichheit fest als geboten.

Für „kleine“ Stichprobenumfänge ist der Kolmogoroff-Smirnov-Anpassungstest besser geeignet als der χ^2 -Test, da letzterer nur approximativ arbeitet; s. auch Slakter (1965) für detaillierte Vergleiche.

Die Hypothese

$$H_0: F(x) = F_0(x) \text{ für alle } x$$

gegen die Alternative

$$H_1: F(x) \neq F_0(x) \text{ für wenigstens einen Wert von } x$$

wird getestet mittels der Prüfgröße

$$\sqrt{n} D_n, \text{ mit } D_n = \sup_x |F_0(x) - S_n(x)|,$$

wobei $S_n(x)$ die empirische Verteilungsfunktion der Beobachtungen x_1, \dots, x_n bezeichnet, vgl. Kap. I, d.h.

$$S_n(x) = \begin{cases} 0 & x < x_i \text{ für alle } i = 1, \dots, n \\ \frac{k}{n}, & \text{falls } x \geq x_i \text{ für genau } k \text{ Beobachtungen } x_i \text{ aus } x_1, \dots, x_n. \\ 1 & x \geq x_i \text{ für alle } i = 1, \dots, n \end{cases}$$

Die Größe D_n gibt den größten vertikalen Abstand zwischen hypothetischer und empirischer Verteilungsfunktion an, vgl. Abb. 12 und Abschnitt C.

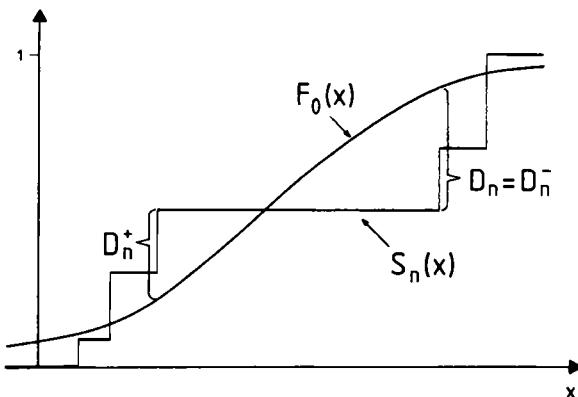


Abb. 12: Prüfgröße des Kolmogoroff-Smirnov-Anpassungstests

Die Hypothese H_0 wird nun zum Niveau α verworfen, wenn gilt

$$\sqrt{n} D_n \geq d_{n; 1-\alpha}$$

wobei die Quantile $d_{n; 1-\alpha}$ für $\alpha = 0,20, 0,10, 0,05, 0,02, 0,01$ aus der Tab. 14 entnommen werden können. Ist das gerade vorliegende n nicht in Tab. 14 zu finden, so nehme man das angegebene Quantil zum nächst größeren n .

Tab. 14: Kritische Werte $d_{n; 1-\alpha}$ für den Kolmogoroff-Smirnov-Anpassungstest

n	5	8	10	20	40	> 40
$d_{n; 0,80}$	1,00	1,01	1,02	1,04	1,05	1,08
$d_{n; 0,90}$	1,14	1,16	1,17	1,19	1,20	1,23
$d_{n; 0,95}$	1,26	1,28	1,29	1,31	1,33	1,36
$d_{n; 0,98}$	1,40	1,43	1,45	1,47	1,49	1,52
$d_{n; 0,99}$	1,50	1,53	1,55	1,57	1,59	1,63

Dieses Vorgehen gilt für alle (stetigen) Verteilungsfunktionen $F_0(x)$. Speziell im Falle einer $N(\mu_0, \sigma_0^2)$ -Normalverteilung hat man $F_0(x) = \Phi\left(\frac{x - \mu_0}{\sigma_0}\right)$, mit spezifizierten Werten μ_0 und σ_0^2 . Dabei bezeichnet $\Phi(z)$ die Verteilungsfunktion der Standardnormalverteilung.

Interessiert man sich nicht für die Nullhypothese, daß die zugrundeliegende Grundgesamtheit einer speziellen $N(\mu_0, \sigma_0^2)$ -Verteilung (mit spezifizierten Werten μ_0 und σ_0^2) entstammt, sondern für die Nullhypothese

H'_0 : die Verteilung der Grundgesamtheit ist eine Normalverteilung $N(\mu, \sigma^2)$ (mit nicht festgelegten μ und σ^2),

so ist der Kolmogoroff-Smirnov-Test nicht direkt anwendbar.

Hier werden dann μ durch \bar{x} und σ^2 durch s^2 aus der Stichprobe geschätzt. Nun liegt es nahe, als Teststatistik

$$\sqrt{n} L_n^{\text{norm}}, \quad \text{mit} \quad L_n^{\text{norm}} = \sup_x \left| S_n(x) - \Phi\left(\frac{x - \bar{x}}{s}\right) \right|$$

zu verwenden. Die Verteilung von L_n^{norm} hängt nicht mehr von den unbekannten Parametern μ und σ^2 ab, vgl. David/Johnson (1948). Andererseits ist die Verteilung aber eine andere als die von D_n . Jedoch ist bekannt, daß bei Verwendung der kritischen Werte von $\sqrt{n} D_n$ der unten stehende Test (extrem) konservativ ist, vgl. Lilliefors (1967).

Die Bestimmung der kritischen Werte selbst ist mit Schwierigkeiten verbunden. Aus diesem Grunde hat Lilliefors (1967) Simulationsstudien durchgeführt und Tabellen der (simulierten) Quantile bereitgestellt. Umfangreichere Simulationsstudien von Stephens ergaben die von Pearson/Hartley (1972) angegebenen kritischen Werte.

Die obige Hypothese H_0' auf Vorliegen einer Normalverteilung wird nun verworfen, wenn gilt

$$\sqrt{n} L_n^{\text{norm}} \geq l_{n; 1-\alpha}^{\text{norm}},$$

wobei $l_{n; 1-\alpha}^{\text{norm}}$ der **Tab. 15** für $\alpha = 0,10, 0,05$ und $0,01$ entnommen werden kann.

Tab. 15: Kritische Werte $l_{n; 1-\alpha}^{\text{norm}}$ zum Test auf nicht spezifizierte Normalverteilung

n	5	8	10	20	30	> 30
$l_{n; 0,90}^{\text{norm}}$	0,72	0,74	0,76	0,79	0,80	0,81
$l_{n; 0,95}^{\text{norm}}$	0,76	0,81	0,82	0,85	0,88	0,89
$l_{n; 0,99}^{\text{norm}}$	0,91	0,94	0,94	1,03	1,03	1,04

Häufig ist man auch nur daran interessiert, ob die unbekannte Verteilungsfunktion $F(x)$ der betrachteten Grundgesamtheit unterhalb oder oberhalb einer hypothetischen Verteilungsfunktion $F_0(x)$ liegt, d.h. ob $P(X \leq x) \leq F_0(x)$ bzw. $P(X \leq x) \geq F_0(x)$ für alle x gilt, wenn X die das interessierende Merkmal der Grundgesamtheit beschreibende Zufallsvariable ist; x_1, \dots, x_n sind Realisationen von X .

Es werden dann die Hypothesen

$$H_{01}: F(x) \geq F_0(x) \text{ für alle } x$$

gegen die Alternative

$$H_{11}: F(x) < F_0(x) \text{ für mindestens ein } x$$

mittels der Teststatistik

$$\sqrt{n} D_n^-, \quad \text{mit} \quad D_n^- = \sup_x \{F_0(x) - S_n(x)\}$$

und

$$H_{02}: F(x) \leq F_0(x) \text{ für alle } x$$

gegen die Alternative

$$H_{12}: F(x) > F_0(x) \text{ für mindestens ein } x,$$

mittels der Teststatistik

$$\sqrt{n} D_n^+, \quad \text{mit} \quad D_n^+ = \sup_x \{S_n(x) - F_0(x)\},$$

zum Niveau α geprüft, indem man

H_{01} verwirft, wenn $\sqrt{n} D_n^- > d_{n; 1 - 2\alpha}$ und

H_{02} verwirft, wenn $\sqrt{n} D_n^+ > d_{n; 1 - 2\alpha}$.

Dabei sind die Quantile $d_{n; 1 - 2\alpha}$ für $\alpha = 0,10, 0,05, 0,025, 0,01, 0,005$ aus der Tabelle 14 zu verwenden. Die anfangs eingeführte Größe D_n ist natürlich stets gleich dem größeren der Werte D_n^- und D_n^+ , vgl. Abb. 12.

Erstellt man eine Wertetabelle für die empirische Verteilungsfunktion, indem man die Funktion an allen Beobachtungspunkten, d.h. an allen Sprungstellen, berechnet, so berechnet man zweckmäßigerweise auch die Werte der hypothetischen Verteilungsfunktion an diesen Stellen und trägt diese beiden Funktionen dann in ein Schaubild ein, dem man dann entnehmen kann, an welchen Stellen die benötigten maximalen vertikalen Abstände auftreten und wie groß diese sind. Verzichtet man auf ein Schaubild, so kann man die benötigten Abstände auch der Wertetabelle entnehmen:

Ist $S_n(x_i) \geq F_0(x_i)$, so notiere man den Abstand $S_n(x_i) - F_0(x_i)$. Das Maximum dieser Differenzen ist D_n^+ .

Ist $S_n(x_j) \leq F_0(x_i)$, so berechne man die Differenz $F_0(x_i) - S_n(x_j)$, wobei x_j der nächstkleinere Wert zu x_i ist; ist x_i schon der kleinste Beobachtungswert, so ist natürlich hier $S_n(x_j) = 0$ zu setzen, d.h. man notiert noch den Wert von F_0 am kleinsten Beobachtungswert; das Maximum dieser Werte ist dann D_n^- . Wie bereits gesagt, ist dann

$$D_n = \max(D_n^-, D_n^+).$$

C. Ein Beispiel

Aus einer Menge von Schrauben werden zufällig $n = 20$ Stück ausgewählt und ihre Durchmesser gemessen (in mm), vgl. Tab. 16.

Tab. 16: Schraubendurchmesser

Nr.	1	2	3	4	5	6	7	8	9	10
\varnothing	0,79	0,68	0,75	0,73	0,69	0,77	0,76	0,74	0,73	0,68
<hr/>										
Nr.	11	12	13	14	15	16	17	18	19	20
\varnothing	0,72	0,75	0,71	0,76	0,69	0,72	0,70	0,77	0,71	0,74

Zum Niveau $\alpha = 0,05$ wollen wir mit Hilfe des χ^2 -Tests und des Kolmogoroff-Smirnov-Tests überprüfen, ob der Durchmesser der gemessenen Schrauben einer Normalverteilung mit Parametern $\mu = 0,75$ und $\sigma^2 = 0,001$ entstammt.

C1. χ^2 -Anpassungstest

1. Schritt: Wir bilden 3 Klassen aus dem Intervall $(-\infty, \infty)$. Nun berechnen wir die Klassengrenzen. Es sollen dabei in jedem der drei Intervalle theoretisch gleich viele Beobachtungen liegen, also $I_1 = (-\infty, w_{0,33}]$, $I_2 = (w_{0,33}, w_{0,66}]$ und $I_3 = (w_{0,66}, \infty)$, wobei w_α das α -Quantil der $N(0,75, 0,001)$ -Verteilung ist. Damit ist dann $p_i = 0,33$ für $i = 1, 2, 3$. Wir bestimmen die Quantile $u_{0,33}$ und $u_{0,66}$ der Standard-

normalverteilung und transformieren diese in die Quantile $w_{0,33}$ bzw. $w_{0,66}$ der $N(0,75, 0,001)$ -Verteilung (vgl. Abschnitt 1.1):

$$w_{0,33} = 0,75 + 0,032 \cdot (-0,43) = 0,736$$

$$w_{0,66} = 0,75 + 0,032 \cdot 0,43 = 0,764.$$

2., 3. Schritt: Wir bestimmen nun für die 3 Klassen die Größen E_i und O_i (vgl. **Tab. 17**).

Tab. 17: Arbeitstabelle zum χ^2 -Anpassungstest im Beispiel

i	1 $(-\infty; 0,736]$	2 $(0,736; 0,764]$	3 $(0,764; \infty)$
O_i	11	6	3
E_i	6,67	6,67	6,67

Somit ergibt sich

$$T = \sum_{i=1}^3 \frac{1}{E_i} (O_i - E_i)^2 = 4,898 \quad \text{und} \quad \chi^2_{3-1; 0,95} = 5,991.$$

Die Hypothese kann demnach nicht verworfen werden.

C2. Kolmogoroff-Smirnov-Anpassungstest

Hier müssen wir zunächst die empirische Verteilungsfunktion $S_n(x)$ unserer Beobachtungen bestimmen, und dann aus der Beziehung

$$F_0(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - 0,75}{0,032}\right), \quad \text{vgl. Abschnitt 1.1,}$$

die Werte der hypothetischen Verteilungsfunktion berechnen, vgl. **Tab. 18**.

Tab. 18: Arbeitstabelle zum Kolmogoroff-Smirnov-Test im Beispiel

Nr. i	x_i	$S_n(x_i)$	$F_0(x_i)$	$S_n(x_i) - F_0(x_i)$	$F_0(x_i) - S_n(x_j)$
2;10	0,68	0,10	0,014	0,086	0,014
5;15	0,69	0,20	0,030	0,170	
17	0,70	0,25	0,059	0,191	
13;19	0,71	0,35	0,105	0,245	
11;16	0,72	0,45	0,173	0,277	
4;9	0,73	0,55	0,267	0,283	< 0
8;20	0,74	0,65	0,378	0,272	
3;12	0,75	0,75	0,500	0,250	
7;14	0,76	0,85	0,622	0,228	
6;18	0,77	0,95	0,733	0,217	
1	0,79	1,00	0,894	0,106	

Damit ist, vgl. auch **Abb. 13**,

$$D_n = \sup_x |F_0(x) - S_n(x)| = 0,283.$$

Wegen

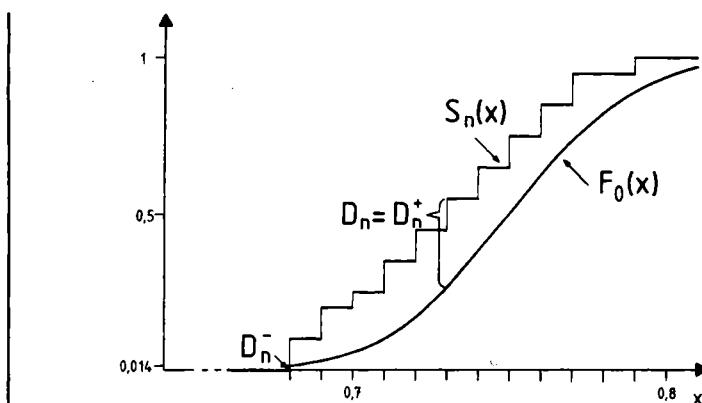


Abb. 13: Empirische Verteilungsfunktion zu den Daten des Beispiels und Verteilungsfunktion der $N(0,75, 0,001)$ -Verteilung

$$\sqrt{n} D_n = \sqrt{20} \cdot 0,283 = 1,266 < 1,31 = d_{20; 0,95}$$

kann die Nullhypothese nicht verworfen werden.

Für die möglichen einseitigen Tests der Hypothesen H_{01} und H_{02} ergeben sich die benötigten Hilfsgrößen zu $D_n^+ = D_n = 0,283$ und $D_n^- = 0,014$.

Wir wollen nun noch die Hypothese

H'_0 : Die Grundgesamtheit ist normalverteilt (μ, σ^2 unbekannt)

zum Niveau $\alpha = 0,05$ testen. Wir schätzen zunächst μ und σ^2 aus der Stichprobe:

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = 0,7295,$$

$$s^2 = \frac{1}{19} \sum_{i=1}^{20} (x_i - 0,7295)^2 = 0,00104$$

Tab. 19: Arbeitstabelle zum Kolmogoroff-Smirnov-Test im Beispiel (bei nichtspezifizierten μ und σ^2).

Nr. i	x_i	$S_n(x_i)$	$\Phi\left(\frac{x_i - \bar{x}}{s}\right)$	$S_n(x_i) - \Phi\left(\frac{x_i - \bar{x}}{s}\right)$	$\Phi\left(\frac{x_i - \bar{x}}{s}\right) - S_n(x_i)$
2;10	0,68	0,10	0,06	0,04	0,06
5;15	0,69	0,20	0,11	0,09	
17	0,70	0,25	0,18	0,07	
13;19	0,71	0,35	0,27	0,08	
11;16	0,72	0,45	0,38	0,07	
4;9	0,73	0,55	0,51	0,04	< 0
8;20	0,74	0,65	0,63	0,02	
3;12	0,75	0,75	0,74	0,01	
7;14	0,76	0,85	0,83	0,02	
6;18	0,77	0,95	0,90	0,05	
1	0,79	1,00	0,97	0,03	

und erstellen dann die Tabelle der empirischen und der theoretischen Verteilungsfunktion, vgl. **Tab. 19**.

Damit ist

$$L_n^{\text{norm}} = 0,09 \quad \text{und somit} \quad \sqrt{n} L_n^{\text{norm}} = \sqrt{20} \cdot 0,09 = 0,402.$$

Als kritischen Wert entnehmen wir der Tab. 15 gerade

$$l_{n; 1-\alpha}^{\text{norm}} = l_{20; 0,95}^{\text{norm}} = 0,85,$$

und somit kann die Normalverteilungshypothese nicht verworfen werden.

1.7 Weitere Verfahren zum Testen von Normalverteilungshypothesen

1.7.1. Test auf Schiefe und Exzeß

Testet man simultan die Hypothesen, daß die Beobachtungswerte einer Meßreihe einer Verteilung mit 1. Schiefe $\gamma_1 = 0$ und 2. Exzeß $\gamma_2 = 0$ entstammen und braucht keine der beiden Hypothesen zu verworfen, so bedeutet das, daß die Hypothese „die Beobachtungen entstammen einer Normalverteilung“ nicht verworfen wird.

Wir überprüfen simultan (d.h. die Gesamthypothese wird nur dann nicht verworfen, wenn die beiden Teilhypotesen nicht verworfen werden) die Hypothesen

$$\begin{aligned} H_{01}: \gamma_1 &= 0 \quad \text{gegen} \quad H_{11}: \gamma_1 \neq 0 \quad \text{und} \\ H_{02}: \gamma_2 &= 0 \quad \text{gegen} \quad H_{12}: \gamma_2 \neq 0. \end{aligned}$$

Ein Punktschätzer für die Schiefe γ_1 der den Beobachtungswerten x_1, \dots, x_n zugrundeliegenden Verteilung ist

$$c_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^3},$$

und für den Exzeß γ_2 ist

$$c_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} - 3$$

eine Punktschätzung.

Wir verwerfen die Nullhypothese $H_{01}: \gamma_1 = 0$ zum Niveau α , falls gilt:

$$|c_1| > g_{1; \alpha/2}$$

und die Nullhypothese $H_{02}: \gamma_2 = 0$ zum Niveau α , wenn

$$c_2 > g_{2; \alpha/2} \quad \text{oder} \quad c_2 < g_{2; 1-\alpha/2}$$

ist. Die Fraktile $g_{1; \alpha}$ und $g_{2; \alpha}$ sind z.B. bei Pearson/Hartley (1970) vertafelt.

1.7.2 Überprüfung der Normalverteilungsannahme mit Hilfe von Wahrscheinlichkeitspapier

Haben wir Werte x_1, \dots, x_n beobachtet und wollen überprüfen, ob diesen Werten eine Normalverteilung zugrundeliegt, so können wir dies mit Hilfe eines sehr einfachen Verfahrens tun. (Diese Überprüfung ist allerdings nur recht ungenau.)

Wir berechnen zunächst die empirische Verteilungsfunktion $S_n(x)$ unserer Beobachtungsreihe an jeder der Stellen x_1, \dots, x_n . Dann tragen wir die Punkte $(x_i, S_n(x_i))$ für $i = 1, \dots, n$ auf Wahrscheinlichkeitspapier ab (Die Ordinate ist nach dem Gaußschen Integral

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\xi^2/2} d\xi$$

aufgeteilt, vgl. auch das Beispiel); solches Papier ist im Fachhandel erhältlich. Ist dann die Grundgesamtheit, aus der unsere Beobachtungen (n muß „ziemlich groß“ sein, in praxi $n \geq 10$) stammen, normalverteilt, so wird die empirische Verteilungsfunktion der theoretischen Verteilungsfunktion einer Normalverteilung gut angenähert sein. Dies bedeutet gerade, daß die auf Wahrscheinlichkeitspapier aufgetragenen Punkte nahezu auf einer Geraden liegen. Ist dies der Fall, so kann man umgekehrt auch schließen, daß die zugrundeliegende Gesamtheit normalverteilt ist. Die oben erwähnte Ungenauigkeit dieser Methode röhrt daher, daß natürlich über die Güte der Annäherung an eine Gerade per Augenmaß entschieden werden muß.

Beispiel: Auf einem Feld, das in zwanzig gleichgroße Parzellen à $\frac{1}{10}$ ha Fläche aufgeteilt ist, wurde Weizen angebaut. Anhand von Wahrscheinlichkeitspapier soll überprüft werden, ob die Erträge pro $\frac{1}{10}$ ha normalverteilt sind. Die Erträge der zwanzig Parzellen sind in Tab. 20 angegeben.

Tab. 20: Erträge von Weizen in kg pro Parzelle

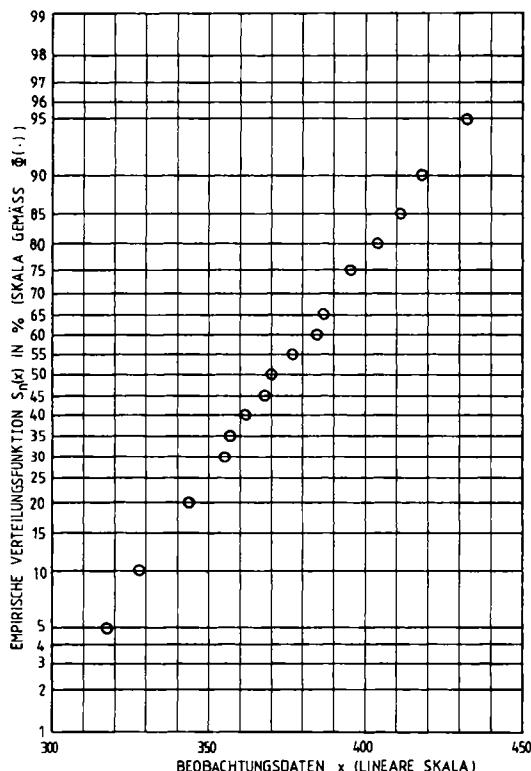
Parzelle i	1	2	3	4	5	6	7	8	9	10
Ertrag x_i	344	432	317	356	448	418	385	362	377	344
Parzelle i	11	12	13	14	15	16	17	18	19	20
Ertrag x_i	404	370	356	395	387	328	357	368	411	395

Zunächst muß nun die empirische Verteilungsfunktion $S_n(x)$ der Parzellenerträge bestimmt werden, vgl. Tab. 21. Die Punkte $(x_i, S_n(x_i))$ sind in Abb. 14 auf Wahrscheinlichkeitspapier eingetragen worden. Wie man sieht, liegen diese Punkte in etwa auf einer Geraden, was die Normalverteilungshypothese bestätigt.

In Kap. XIV werden alternativ zum Wahrscheinlichkeitspapier Wurzeldiagramme und Q-Q-Plots vorgestellt. Ein Vergleich verschiedener Normalitätstests, der auch die hier vorgestellten beinhaltet, findet sich bei Shapiro et al. (1968).

Tab. 21: Empirische Verteilungsfunktion der Weizenerträge pro Parzelle

i	x_i	$S_n(x_i)$
3	317	0,05
16	328	0,10
1;10	344	0,20
4;13	356	0,30
17	357	0,35
8	362	0,40
18	368	0,45
12	370	0,50
9	377	0,55
7	385	0,60
15	387	0,65
14;20	395	0,75
11	404	0,80
19	411	0,85
6	418	0,90
2	432	0,95
5	448	1,00

**Abb. 14:** Daten des Beispiels auf Wahrscheinlichkeitspapier

2. Die Gleichverteilung und die Dreiecksverteilung

2.1. Die stetige Gleichverteilung

2.1.1. Die eindimensionale Gleichverteilung und ihre Anwendung in der Computersimulation

Die wohl einfachste stetige Verteilung ist die der diskreten Laplace-Verteilung analoge stetige **Gleichverteilung** $U(a, b)$ auf dem Intervall $[a; b]$, die auch **Rechteckverteilung** genannt wird.

Eine so verteilte Zufallsgröße X beschreibt ein Zufallsexperiment, bei dem jede Zahl aus dem Intervall $[a; b]$ die gleiche Chance hat, Realisation von X zu sein. Die Wahrscheinlichkeit, daß eine Zahl außerhalb dieses Intervalls vorkommt, ist Null. Damit hat X die in **Abb. 15** dargestellte Dichte

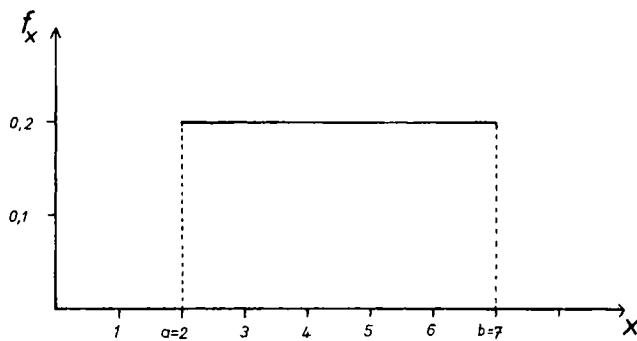


Abb. 15: Dichtefunktion der Rechteckverteilung $U(2, 7)$

$$f_x(x) = \begin{cases} \frac{1}{b-a}, & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases},$$

und die zugehörige in **Abb. 16** dargestellte Verteilungsfunktion ist

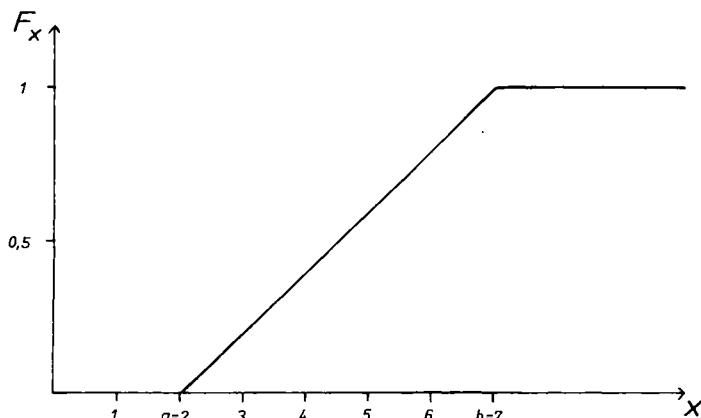


Abb. 16: Verteilungsfunktion der Rechteckverteilung $U(2, 7)$

$$F_X(x) = \begin{cases} 0 & , \quad x < a \\ \frac{x-a}{b-a}, & \text{falls } a \leq x \leq b \\ 1 & , \quad x > b. \end{cases}$$

Als Erwartungswert von X ergibt sich

$$EX = \frac{b+a}{2},$$

und für die Varianz erhält man

$$\text{Var } X = \frac{(b-a)^2}{12}.$$

Die zentralen Momente ungerader Ordnung sind alle Null und die gerader Ordnung berechnen sich zu

$$E(X - EX)^k = \frac{1}{k+1} \left(\frac{b-a}{2} \right)^k.$$

Eine bedeutsame Anwendung der Gleichverteilung findet man in der **Computersimulation**. Auf dem Rechner lassen sich nämlich recht einfach Zahlen erzeugen, die man als unabhängige Realisationen einer $U(0, 1)$ -verteilten Zufallsgröße X ansehen kann. Transformiert man nun diese Zahlen mit F_X^{-1} , so können die Ergebnisse als Realisationen einer Zufallsgröße Y mit der Verteilungsfunktion F aufgefaßt werden.

Beispiel: Es soll das Ausfallverhalten von Glühbirnen auf einem Rechner simuliert werden. Dabei kann angenommen werden, daß deren Lebensdauern exponentialverteilt sind mit dem bekannten Parameter λ . Wir müssen uns also Zahlen verschaffen, die als Realisationen einer so verteilten Zufallsgröße Y angesehen werden können. Dazu erzeugen wir uns zunächst im Rechner mit einem sogenannten **Zufallszahlengenerator** unabhängige Realisationen x_i einer $U(0, 1)$ -verteilten Zufallsgröße X. Die Inverse der Verteilungsfunktion

$$F_Y(y) = 1 - e^{-\lambda y}$$

der Exponentialverteilung ist gegeben durch

$$F_Y^{-1}(x) = -\frac{1}{\lambda} \ln(1-x),$$

so daß die Werte $y_i = -\frac{1}{\lambda} \ln(1-x_i)$ als Realisationen einer exponentialverteilten Zufallsgröße angesehen werden können. (Da mit X auch $1-X$ $U(0, 1)$ -verteilt ist, gilt das ebenso für die einfacher zu berechnenden Werte $-\frac{1}{\lambda} \ln x_i$.) Somit kann die Zahl y_i als Lebensdauer der (fiktiven) i-ten Glühbirne angesehen werden.

Obwohl das hier vorgestellte Verfahren bei beliebigen Verteilungen anwendbar ist, wird man es, da meist große Mengen von **Zufallszahlen** benötigt werden, nur dann einsetzen, wenn das Invertieren der Verteilungsfunktion nicht zu aufwendig bzw. zu zeitintensiv ist.

Nun wollen wir noch andeuten, wie gleichverteilte Zufallszahlen im Rechner erzeugt werden können. Bei dem Vorgang, der uns diese Zahlen liefert, handelt es sich in aller Regel nicht, wie man zunächst meinen könnte, um einen zufälligen, sondern um einen **deterministischen** Vorgang: Man berechnet eine neue Zufallszahl jeweils aus den Werten der vorhergehenden. Die so erhaltenen Zahlen werden deshalb auch **Pseudozufallszahlen** genannt. Eine einfache, häufig verwandte Methode, solche Pseudozufallszahlen zu erzeugen, ist die **multiplikative Kongruenzmethode**: Man gibt sich zunächst ganze Zahlen a und m mit $a < m$ vor. Dann setzt man einen ganzzahligen Wert x'_1 fest und berechnet daraus Werte x'_2, x'_3, \dots gemäß

$$x'_{i+1} = ax'_i \pmod{m},$$

d.h. es wird jeweils der ganzzahlige Rest der Division ax'_i/m bestimmt. Die Zahlen $x_i = x'_i/m$ werden dann (bei geeigneter Wahl von a und m) als unabhängige Realisationen einer $U(0, 1)$ -verteilten Zufallsgröße angesehen.

Beispiel: Wir setzen $a = 3$ und $m = 5$. Als Startwert benutzen wir $x'_1 = 1$. Dann ist

$$x'_2 = 3 \cdot x'_1 \pmod{5} = 3 \pmod{5} = 3,$$

$$x'_3 = 3 \cdot x'_2 \pmod{5} = 9 \pmod{5} = 4,$$

$$x'_4 = 3 \cdot x'_3 \pmod{5} = 12 \pmod{5} = 2,$$

$$x'_5 = 3 \cdot x'_4 \pmod{5} = 6 \pmod{5} = 1.$$

Man sieht, daß dieser **Zufallszahlengenerator** eine Periode der Länge 4 besitzt. Es ist also sinnlos, hiermit mehr als 4 Zahlen zu erzeugen.

Geeignete Werte für a und m sind z.B. $a = 65539$ und $m = 2^{31}$, bei denen sich dann eine Periodenlänge von 2^{29} ergibt. Die Periodenlänge ist jedoch nur eines der Kriterien zur Beurteilung eines Generators für Pseudozufallszahlen; vgl. hierzu Ahrens et al. (1970), Dieter (1971), Schmitz/Lehmann (1976). Dort werden auch einige weitere Möglichkeiten zur Erzeugung von Zufallszahlen behandelt.

2.1.2. Die zweidimensionale Gleichverteilung

Neben der bereits vorgestellten eindimensionalen Rechteckverteilung gibt es auch mehrdimensionale Rechteckverteilungen. So ist z.B. die **zweidimensionale Rechteckverteilung** eines Zufallsvektors (X, Y) gegeben durch die Dichte

$$f_{(X, Y)}(x, y) = \begin{cases} \frac{1}{b_1 - a_1} \cdot \frac{1}{b_2 - a_2}, & \text{für } a_1 \leq x \leq b_1 \text{ und } a_2 \leq y \leq b_2 \\ 0, & \text{sonst.} \end{cases}$$

Dabei sind X und Y unabhängig voneinander und einzeln auch wieder rechteckverteilt, nämlich gemäß einer $U(a_1, b_1)$ - bzw. $U(a_2, b_2)$ -Verteilung.

Beispiel: Ein Sportschützenverein benutzt gelegentlich als Ziel eine kleine rechteckige Scheibe. Da diese Scheibe in einer verhältnismäßig großen Entfernung aufgestellt wird, sind die Treffer gleichmäßig auf ihr verteilt. Es passiert nun das Mißgeschick, daß die letzte vorhandene Scheibe durchbricht und nur noch das in Abb. 17 unterhalb des Graphen der Funktion T liegende Stück zu gebrauchen ist. Trotzdem soll, um vergleichbare Ergebnisse zu erhalten, das gesamte Rechteck als Ziel aufgefaßt werden. Ein Treffer kann jedoch nur nachgewiesen werden, wenn er in der Fläche unterhalb des Graphen der Funktion T liegt. Wie groß ist nun die Wahrscheinlichkeit, daß wir einen Treffer nachweisen können? Wir erhalten hierfür

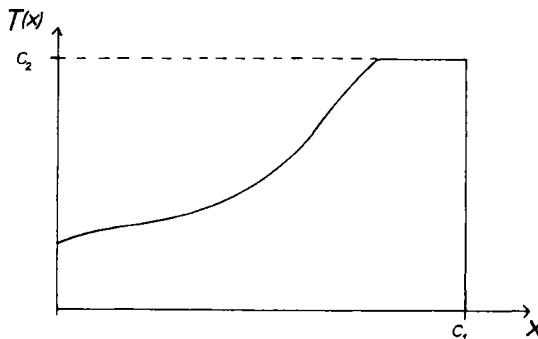


Abb. 17: Treffen einer Scheibe (geometrische Wahrscheinlichkeit)

$$\begin{aligned}
 P(Y < T(X)) &= \int_{\{(x, y) : y < T(x)\}} f_{(x, y)}(x, y) dx dy \\
 &= \int_0^{c_1} \left(\int_0^{T(x)} \frac{1}{c_1} \frac{1}{c_2} dy \right) dx \\
 &= \int_0^{c_1} \frac{1}{c_1} \cdot \frac{1}{c_2} \cdot T(x) dx = \frac{1}{c_1 \cdot c_2} \cdot \int_0^{c_1} T(x) dx.
 \end{aligned}$$

Die gesuchte Wahrscheinlichkeit ist also gleich dem Verhältnis der Fläche unterhalb von T zur Gesamtfläche des Rechtecks. Dieser Wert ist, zumindest bei einer großen Anzahl von Treffern, ungefähr gleich dem Anteil der nachweisbaren Treffer. Damit lässt sich nun die gesamte Trefferzahl näherungsweise bestimmen.

2.2. Die Dreiecksverteilung

Haben wir uns in Abschnitt 2.1.2 mit der **gemeinsamen Verteilung** zweier unabhängiger Zufallsgrößen X und Y beschäftigt, so wollen wir uns nun der Verteilung der **Summe** von X und Y zuwenden. Wir beschränken uns dabei auf den noch relativ einfach zu behandelnden Fall, daß X und Y die gleiche Rechteckverteilung $U(a, b)$ besitzen. Die Verteilung der Summe $Z = X + Y$ ist dann eine **gleichschenklige**

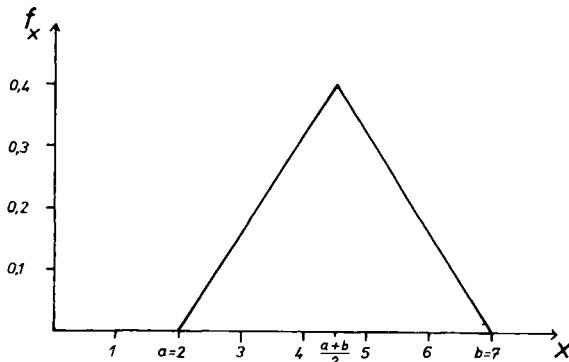


Abb. 18: Dichtefunktion der Simpson-Verteilung

Dreiecksverteilung mit den Parametern a und b . Die Dichte dieser Verteilung, die man auch **Simpson-Verteilung** nennt, ist in Abb. 18 dargestellt und berechnet sich zu

$$f_Z(z) = \begin{cases} \frac{2}{b-a} \left(\frac{z-a}{b-a} \middle| \frac{1}{2} \right), & \text{falls } a \leq z \leq \frac{a+b}{2} \\ \frac{2}{b-a} \left(\frac{b-z}{b-a} \middle| \frac{1}{2} \right), & \text{falls } \frac{a+b}{2} \leq z \leq b \\ 0 & , \text{ sonst} \end{cases}$$

$$= \begin{cases} \frac{2}{b-a} \left(1 - \frac{2}{b-a} \cdot \left| z - \frac{a+b}{2} \right| \right), & \text{falls } a \leq z \leq b \\ 0 & , \text{ sonst}. \end{cases}$$

Man hat hier

$$EZ = \frac{a+b}{2} \quad \text{und} \quad \text{Var } Z = \frac{1}{24} (b-a)^2;$$

der Modus (Modalwert) der Simpson-Verteilung ist wegen der Symmetrie der Verteilung identisch mit dem Erwartungswert von Z .

Der Vollständigkeit halber wollen wir hier auch noch die Dichte der **allgemeinen Dreiecksverteilung** angeben, die neben den Parametern a und b zusätzlich den Modalwert H als Parameter besitzt. Für eine so verteilte Zufallsgröße Z ist

$$f_Z(z) = \begin{cases} \frac{2}{b-a} \left(\frac{z-a}{b-a} \middle| \frac{H-a}{b-a} \right), & \text{für } a \leq z \leq H \\ \frac{2}{b-a} \left(\frac{b-z}{b-a} \middle| \frac{b-H}{b-a} \right), & \text{für } H \leq z \leq b \\ 0 & , \text{ sonst} \end{cases}$$

$$= \begin{cases} \frac{2}{b-a} \cdot \frac{z-a}{H-a}, & \text{für } a \leq z \leq H \\ \frac{2}{b-a} \cdot \frac{b-z}{b-H}, & \text{für } H \leq z \leq b \\ 0 & , \text{ sonst}, \end{cases}$$

$$EZ = \frac{1}{3} (a + b + H) \quad \text{und}$$

$$\text{Var } Z = \frac{1}{18} [(b-a)^2 - (b-a)(H-a) + (H-a)^2].$$

Speziell für $H = \frac{a+b}{2}$ ergibt sich wieder die oben vorgestellte gleichschenklige Dreiecksverteilung; ist $H = a$ oder $H = b$, so spricht man von **einseitigen (links- bzw. rechtsseitigen) Dreiecksverteilungen**.

2.3. Punkt- und Intervallschätzungen für die Gleichverteilung

Als **Punktschätzer für die Parameter a und b** können wir natürlich den kleinsten Wert $x_{(1)}$ bzw. den größten Wert $x_{(n)}$ einer Beobachtungsreihe x_1, \dots, x_n , die einer $U(a, b)$ -Verteilung entstammt, benutzen. Aufgrund dessen liegt es nahe auch das $1 - \alpha$ **Konfidenzintervall für $b - a$** – also für die Länge des Intervalls, auf dem die Beobachtungen gleichverteilt sind – mit Hilfe dieser Punktschätzer abzuleiten. Die Differenz $b - a$ kann nämlich durch die Spannweite

$$R = x_{(n)} - x_{(1)}$$

der Beobachtungsreihe geschätzt werden. Normieren wir noch die Spannweite mit gewissen Zahlen β und γ , so daß

$$P\left(\frac{R}{\beta} \leq b - a \leq \frac{R}{\gamma}\right) = 1 - \alpha$$

gilt, dann erhalten wir mit

$$\left[\frac{R}{\beta}; \frac{R}{\gamma}\right]$$

ein Konfidenzintervall zum Niveau $1 - \alpha$ für die Differenz $b - a$. Dabei lassen sich die Normierungskonstanten β und γ aus

$$n\gamma^{n-1} - (n-1)\gamma^n = \frac{\alpha}{2} \quad \text{bzw.} \quad n\beta^{n-1} - (n-1)\beta^n = 1 - \frac{\alpha}{2}$$

bestimmen.

Beispiel: Weiß man, daß die Länge von Abfallstücken einer Schneidemaschine gleichverteilt ist mit Parametern a und b , so möchte man nun aufgrund einer Stichprobe vom Umfang $n = 20$ Abfallstücke die Punktschätzer für a und b sowie ein Konfidenzintervall für die Differenz $b - a$ zum Niveau 0,90 berechnen.

Es wurden in der Meßreihe die in **Tab. 22** angegebenen Längen der 20 Abfallstücke festgestellt.

Tab. 22: Längen der 20 Abfallstücke (der Größe nach geordnet)

Nr.	1	2	3	4	5	6	7	8	9	10
Länge	11,1	12,4	12,9	13,1	14,6	15,0	15,3	15,9	16,2	16,6
Nr.	11	12	13	14	15	16	17	18	19	20
Länge	17,1	17,8	17,9	18,5	18,6	18,9	19,4	19,8	20,0	20,2

Als Punktschätzer für a bzw. b erhalten wir somit

$$x_{(1)} = 11,1 \quad \text{bzw.} \quad x_{(n)} = 20,2.$$

Es ist dann

$$R = x_{(n)} - x_{(1)} = 20,2 - 11,1 = 9,1.$$

Wir bestimmen nun die Werte β und γ :

$$n\beta^{n-1} - (n-1)\beta^n = 1 - \frac{\alpha}{2} \Leftrightarrow 20\beta^{19} - 19\beta^{20} = 0,95$$

$$\Leftrightarrow \beta^{19}(20 - 19\beta) = 0,95$$

$$\Rightarrow \beta \approx 0,982;$$

$$n\gamma^{n-1} - (n-1)\gamma^n = \frac{\alpha}{2} \Leftrightarrow 20\gamma^{19} - 19\gamma^{20} = 0,05$$

$$\Leftrightarrow \gamma^{19}(20 - 19\gamma) = 0,05$$

$$\Rightarrow \gamma \approx 0,784.$$

Es ergibt sich also das gesuchte Konfidenzintervall für $b - a$ zu

$$\left[\frac{R}{\beta}; \frac{R}{\gamma} \right] = \left[\frac{9,1}{0,982}; \frac{9,1}{0,784} \right] = [9,267; 11,607],$$

d.h. mit 90% Sicherheit ist die maximale Längendifferenz der Abfallstücke in diesem Intervall zu finden.

2.4. Der χ^2 -Anpassungstest für die Gleichverteilung

Wir haben in Abschnitt 1.6 schon den χ^2 -Anpassungstest für die Normalverteilung kennengelernt und dabei gesehen, daß die Nullhypothese ganz allgemein so formuliert werden kann:

H_0 : Die Meßreihe x_1, \dots, x_n entstammt einer Verteilung mit Verteilungsfunktion $F(x)$.

Daher können wir natürlich auch konkret einen χ^2 -Anpassungstest für die stetige Gleichverteilung durchführen. Die spezielle Hypothese ist dann:

H_0 : x_1, \dots, x_n entstammen einer $U(a, b)$ -verteilten Grundgesamtheit, und wird gegen die Alternative

H_1 : Die Grundgesamtheit ist nicht $U(a, b)$ -verteilt zum Niveau α getestet.

Dabei gehen wir genauso vor wie in Abschnitt 1.6.

Beispiel: Wir wollen überprüfen, ob die Körpergröße der männlichen Schüler in der Oberstufe eines Gymnasiums gleichverteilt ist. Wir wählen zunächst zufällig $n = 25$ Schüler aus und messen sie (vgl. Tab. 23).

Nun schätzen wir zunächst die Parameter a und b aus dieser Beobachtungsreihe. Es ist

$$x_{(1)} = 168 \quad \text{und} \quad x_{(n)} = 191.$$

Um zum Niveau $\alpha = 0,05$ die Hypothese

H_0 : Die Körpergröße der Schüler der Oberstufe ist $U(168, 191)$ verteilt gegen die Alternative

$$H_1: \text{Die Körpergröße ist nicht } U(168, 191) \text{ verteilt}$$

Tab. 23: Körpergröße der 25 Schüler

Nr.	1	2	3	4	5	6	7	8	9	10
Größe	177	183	178	168	173	179	175	187	172	191
Nr.	11	12	13	14	15	16	17	18	19	20
Größe	182	174	176	186	169	173	171	185	181	172
Nr.	21	22	23	24	25					
Größe	169	178	180	173	186					

zu testen, unterteilen wir das Intervall $(-\infty, \infty)$ in $k = 4$ Gruppen und berechnen die O_i und E_i (also die beobachteten und erwarteten Anzahlen in jeder Gruppe), vgl. **Tab. 24**.

Tab. 24: Beobachtete und erwartete Körpergrößen

i	1 $(-\infty; 173,5]$	2 $(173,5; 179,5]$	3 $(179,5; 185,5]$	4 $(185,5; \infty)$
O_i	9	7	5	4
E_i	5,975	6,525	6,525	5,975

Es ergibt sich hier z. B.

$$P(X \in (-\infty; 173,5]) = P(X \in [168; 173,5]) = \frac{5,5}{23} = 0,239 = p_1$$

und somit

$$E_1 = n p_1 = 25 \cdot 0,239 = 5,975.$$

Damit ergibt sich

$$T(x) = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} = 2,575 \quad \text{und} \quad \chi^2_{k-1-2; 1-\alpha} = \chi^2_{1; 0,95} = 3,841.$$

Wir können unsere Hypothese H_0 also nicht verwerfen.

Natürlich läßt sich der in Abschnitt 1.6 beschriebene Kolmogoroff-Smirnov-Anpassungstest auch bei einer Gleichverteilung anwenden; er verläuft hier völlig analog, wie dort im Fall einer Normalverteilung vorgeführt.

3. Einige diskrete Verteilungen

3.1. Die Binomialverteilung

Beispiel: Ein Experiment wird 3mal ausgeführt. Die Wahrscheinlichkeit für das Ereignis A wird mit $p = P(A)$ bezeichnet; wir setzen $q = 1 - p$. Wie groß ist die Wahrscheinlichkeit, daß der Experimentator höchstens 2mal das Ereignis A be-

Tab. 25: Mögliche Ereignisse des Gesamtversuchs

mögliche Versuchsausgänge	A A A	A A A	A A A	A A A	A A A	A A A	A A A	A A A
Wahrscheinlichkeit dafür	$p \cdot p \cdot p = p^3$	$p \cdot p \cdot q = p^2 q$	$p \cdot q \cdot p = p^2 q$	$q \cdot p \cdot p = p^2 q$	$q \cdot q \cdot p = pq^2$	$q \cdot p \cdot q = pq^2$	$p \cdot q \cdot q = pq^2$	$q \cdot q \cdot q = q^3$
Anzahl x des Auftretens von A	3	2	2	2	1	1	1	0

obachtet? Wir schreiben uns die möglichen Ereignisse des Gesamtversuches tabellarisch auf, s. **Tab. 25**. Hierbei bedeutet z. B. A A A, daß im ersten und dritten Versuch A, im zweiten A nicht beobachtet wurde. Die Zufallsvariable X gibt an, wie oft das Ereignis A beobachtet wurde.

Es ist

$$\begin{aligned} P(X=3) &= p^3 = \binom{3}{3} p^3 q^{3-3}, \\ P(X=2) &= 3p^2 q = \binom{3}{2} p^2 q^{3-2}, \\ P(X=1) &= 3pq^2 = \binom{3}{1} p^1 q^{3-1} \text{ und} \\ P(X=0) &= q^3 = \binom{3}{0} p^0 q^{3-0}. \end{aligned}$$

Damit berechnen wir als Wahrscheinlichkeit dafür, daß A höchstens 2mal eintritt, gerade

$$P(X \leq 2) = q^3 + 3pq^2 + 3p^2q = 1 - P(X > 2) = 1 - p^3.$$

Eine Zufallsvariable X heißt **binomialverteilt mit den Parametern n und p** ($X \sim B(n, p)$), falls gilt

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} = b(n, p, k) \text{ für } k = 0, 1, 2, \dots, n.$$

Mißt eine Zufallsvariable X den Erfolg eines Experiments

$$X = \begin{cases} 1, & \text{falls Erfolg} \\ 0, & \text{falls Mißerfolg,} \end{cases}$$

so bezeichnet man X als **Bernoulli-Variable**, das Experiment als **Bernoulliexperiment** mit Erfolgswahrscheinlichkeit $p = P(X=1)$. Es ist dann $X \sim B(1, p)$. Der Begriff der **Bernoulli-Verteilung** wird häufig synonym zur **Binomialverteilung** verwendet.

Sind X_i stochastisch unabhängig $B(m_i, p)$ – verteilt für $i = 1, 2, \dots, n$, so ist

$$X_1 + X_2 + X_3 + \dots + X_n \sim B\left(\sum_{i=1}^n m_i, p\right);$$

speziell heißt dies:

Wird ein Bernoulli-Experiment (mit Erfolgswahrscheinlichkeit p) n-mal (unabhängig) hintereinander ausgeführt, so ist die Zahl der Erfolge insgesamt $B(n, p)$ -verteilt.

Es sind

$$EX = np \text{ und } Var X = npq = np(1-p)$$

der Erwartungswert bzw. die Varianz einer mit Parametern n und p binomialverteilten Zufallsvariablen X.

Ist $n p q \geq 9$, so lässt sich die Binomialverteilung durch die Normalverteilung $N(n p, n p q)$ approximieren (Faustregel); es ist dann

$$P(X \leq x) \simeq \Phi\left(\frac{X - np}{\sqrt{npq}}\right).$$

Speziell bei etwas kleinerem n ist es jedoch meist günstiger, folgende Modifikation dieser Approximation zu verwenden:

$$P(x_1 \leq X \leq x_2) \simeq \Phi\left(\frac{x_2 - np + 0,5}{\sqrt{npq}}\right) - \Phi\left(\frac{x_1 - np - 0,5}{\sqrt{npq}}\right).$$

Für kleine n können solche Wahrscheinlichkeiten auch Tabellen entnommen werden, vgl. z.B. Romig (1953).

Außerdem ist oft eine Beziehung zwischen der Binomial- und der F-Verteilung hilfreich bei der Bestimmung von Binomialwahrscheinlichkeiten. Ist $X \sim B(n, p)$ -verteilt und F eine F-verteilte Zufallsvariable mit $2(x+1), 2(n-x)$ Freiheitsgraden, dann gilt

$$P(X < x) = 1 - P\left(F \leq \frac{n-x}{x+1} \cdot \frac{p}{1-p}\right).$$

Beispiel: Zwischen einem Produzenten von Transistoren und einer Firma zur Herstellung von Fernsehgeräten wird vereinbart, daß bei der Eingangskontrolle nicht mehr als 10% der Transistoren defekt sein sollten. Dazu wird aus jeder Lieferung eine Stichprobe von 10 Transistoren entnommen und überprüft.

(a) Wie groß ist die Wahrscheinlichkeit, daß eine Lieferung von Transistoren zurückgewiesen wird, wenn der Lieferant weiß, daß 5% der Transistoren defekt sind?

(b) Wie groß ist die Wahrscheinlichkeit, daß die Lieferung angenommen wird, obwohl sie 20% defekte Transistoren enthält?

Die Zufallsvariable X gebe die Anzahl der defekten Transistoren in der Stichprobe an. Falls $X \leq 1$, so wird die Lieferung angenommen, falls $X > 1$ ($X \geq 2$), so wird die Lieferung zurückgewiesen.

zu (a): Wenn der Lieferant weiß, daß nur 5% defekte Transistoren in der Lieferung sind, so ist $X \sim B(10, 0,05)$ -verteilt, d. h. die Wahrscheinlichkeit, daß eine Lieferung zurückgeht, ist

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - \sum_{i=0}^1 \binom{10}{i} 0,05^i 0,95^{10-i} = 0,0862;$$

zu (b): Hier ist $X \sim B(10, 0,2)$, d. h.

$$P(X \leq 1) = \sum_{i=0}^1 \binom{10}{i} 0,2^i 0,8^{10-i} = 0,1074 + 10 \cdot 0,2 \cdot 0,1342 = 0,3758.$$

Wenn eine Lieferung 20% defekte Transistoren enthält, so wird sie mit einer Wahrscheinlichkeit von 37,58% trotzdem akzeptiert.

In der folgenden Tab. 26 sind die Binomialwahrscheinlichkeiten noch einmal aufgeführt.

Tab. 26: Binomialwahrscheinlichkeiten im Beispiel

k	p = 0,05		p = 0,2	
	P(X = k)	P(X ≤ k)	P(X = k)	P(X ≤ k)
0	0,5987	0,5987	0,1074	0,1074
1	0,3151	0,9138	0,2684	0,3758
2	0,0746	0,9884	0,3020	0,6778
3	0,0105	0,9999	0,2013	0,8791
4	0,0010	0,9999	0,0881	0,9672
5	0,0001	1,0000	0,0264	0,9936
6	0,0000	1,0000	0,0055	0,9991
7			0,0008	0,9999
8			0,0001	1,0000
9			0,0000	1,0000
10				

Ist das p der Binomialverteilung klein und n verhältnismäßig groß (etwa $n \geq 30$, $p < 0,1$), so ist

$$P(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda} \text{ mit } \lambda = p \cdot n,$$

d. h. die Binomialverteilung lässt sich durch eine Poissonverteilung mit dem Parameter $\lambda = pn$ approximieren, vgl. Abschnitt 3.4.

Bei Simulationsstudien benötigt man gelegentlich Realisationen binomialverteilter Zufallsgrößen. Dazu werden auf dem Rechner zunächst Realisationen x_i von $U(0, 1)$ -verteilten Zufallsgrößen erzeugt, vgl. Abschnitt 2.1. A. Daraus erhält man dann durch Bildung von

$$y_i = \begin{cases} 1, & \text{falls } x_i \leq p \\ 0, & \text{falls } x_i > p \end{cases}$$

Realisationen einer $B(1, p)$ -verteilten Zufallsgröße. Die Summe $y = y_1 + \dots + y_n$ kann schließlich als Realisation einer $B(n, p)$ -verteilten Zufallsgröße angesehen werden.

Wir werden nun Punkt- und Intervallschätzungen für den Parameter p einer $B(n, p)$ -verteilten Grundgesamtheit angeben. Später werden wir dann auch Hypothesen über diesen Parameter testen. Bzgl. des Vergleichs der Parameter mehrerer Binomialverteilungen sei auf Abschnitt 1.3 bzw. 5.1.2 B in Kap. VII verwiesen.

3.1.1. Punkt- und Intervallschätzung des Parameters p

Als **Punktschätzer für die unbekannte Wahrscheinlichkeit p** können wir natürlich die relative Häufigkeit des Eintretens eines gewissen Ereignisses A bei der Durchführung von n unabhängigen Versuchen betrachten; wir schätzen also p durch den Quotienten

$$\hat{p} = \frac{\text{Anzahl der Versuche, in denen das Ereignis A eingetreten ist}}{\text{Anzahl der gesamten Versuche}} = \frac{m}{n}.$$

Konfidenzintervalle zum Niveau $1 - \alpha$ für p lassen sich **für große Anzahlen n** von Versuchen ($np(1-p) \geq 9$) sehr leicht angeben, da der zentrale Grenzwertsatz, vgl. Abschnitt 9 in Kap. II, anwendbar ist:

$$Z = \frac{\frac{m}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \left(= \frac{m - np}{\sqrt{np(1-p)}} \right)$$

ist dann nämlich approximativ $N(0, 1)$ -verteilt.

Das Konfidenzintervall läßt sich nun aus der Gleichung

$$\begin{aligned} P\left(u_{\alpha/2} \leq \frac{\frac{m}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq u_{1-\alpha/2}\right) &= 1 - \alpha, \text{ d. h.} \\ P\left(p - u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq \frac{m}{n} \leq p + u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}\right) &= 1 - \alpha \end{aligned}$$

bestimmen zu $[p_1; p_2]$ mit

$$p_1 = \frac{2m + u_{1-\alpha/2}^2 - u_{1-\alpha/2} \sqrt{u_{1-\alpha/2}^2 + 4m\left(1 - \frac{m}{n}\right)}}{2(n + u_{1-\alpha/2}^2)}$$

$$\text{und } p_2 = \frac{2m + u_{1-\alpha/2}^2 + u_{1-\alpha/2} \sqrt{u_{1-\alpha/2}^2 + 4m\left(1 - \frac{m}{n}\right)}}{2(n + u_{1-\alpha/2}^2)}.$$

Ist die Zahl n zu klein, so ist die Approximation der Binomial- durch die Normalverteilung nicht mehr verwendbar. Trotzdem läßt sich mit Hilfe der **Pearson-Clopper-Werte** ein $1 - \alpha$ -Konfidenzintervall für p angeben.

Die Zufallsvariable X , die die Häufigkeit eines Ereignisses A beschreibt, ist binomialverteilt, so daß gilt:

$$P(X = m) = \binom{n}{m} p^m (1-p)^{n-m}$$

$$\text{und } P(X \geq m) = \sum_{k=m}^n \binom{n}{k} p^k (1-p)^{n-k},$$

wobei n die Gesamtzahl der unabhängigen Versuche ist.

Wir suchen nun ein Intervall, daß den wahren Parameter p für die Wahrscheinlichkeit eines Ereignisses mit der Wahrscheinlichkeit $1 - \alpha$ überdeckt.

Ist im konkreten Fall des Vorliegens einer Meßreihe m die Häufigkeit des Ereignisses, so müssen wir Intervallgrenzen p_1 und p_2 so bestimmen, daß gilt:

$$P(X \geq m) = \sum_{k=m}^n \binom{n}{k} p_1^k (1-p_1)^{n-k} = \frac{\alpha}{2},$$

$$P(X \leq m) = \sum_{k=0}^m \binom{n}{k} p_2^k (1-p_2)^{n-k} = \frac{\alpha}{2}.$$

Da gilt $P(X < x) = 1 - P\left(F \leq \frac{n-x}{x+1} \cdot \frac{p}{1-p}\right)$ mit $F \sim F_{2(x+1), 2(n-x)}$,

ergibt sich – unter Verwendung der im Anhang vertafelten Quantile der F-Verteilung – das gesuchte $1 - \alpha$ -Konfidenzintervall für p zu

$[p_1; p_2]$ mit

$$p_1 = \frac{m F_{2m, 2(n-m+1); \alpha/2}}{n-m+1 + m F_{2m, 2(n-m+1); \alpha/2}} \text{ und}$$

$$p_2 = \frac{(m+1) F_{2(m+1), 2(n-m); 1-\alpha/2}}{n-m+(m+1) F_{2(m+1), 2(n-m); 1-\alpha/2}}.$$

Diese Größen p_1 und p_2 heißen auch **Pearson-Clopper-Werte**.

Beispiel: (a) Aus einem Produktionslos werden $n = 1000$ Glühbirnen zufällig ausgewählt und auf ihre Funktionsfähigkeit überprüft. Dabei stellt man fest, daß 19 Birnen defekt sind.

Eine Punktschätzung für die Wahrscheinlichkeit p , daß eine Glühbirne aus dem Produktionslos defekt ist, ist also

$$\frac{19}{1000} = 0,019 = \hat{p}.$$

Da hinreichend viele Beobachtungen ($n = 1000$) vorliegen, können wir ein Konfidenzintervall mit Hilfe der Normalverteilungsapproximation berechnen. Als Niveau wählen wir hier $1 - \alpha = 0,95$.

Es ergibt sich mit $u_{1-\alpha/2} = 1,96$:

$$p_1 = \frac{41,842 - 17,354}{2007,683} = 0,012 \text{ und}$$

$$p_2 = \frac{41,842 + 17,354}{2007,683} = 0,029.$$

Das gesuchte $1 - \alpha$ -Konfidenzintervall für p ist also

$$[0,012; 0,029].$$

(b) Aus der Schraubenproduktion eines Betriebes werden zufällig 500 Schrauben entnommen. Die Schrauben müßten eine Gewindelänge von mindestens 2 cm haben. Man will nun ein Konfidenzintervall zum Niveau $1 - \alpha = 0,9$ für die Wahrscheinlichkeit p , daß eine Schraube kürzer ist als 2 cm, angeben. In der konkreten Stichprobe vom Umfang $n = 500$ waren 11 Schraubengewinde kürzer als 2 cm. Wir bestimmen nun zunächst p_1 und p_2 :

$$\begin{aligned}
 p_1 &= \frac{m F_{2m, 2(n-m+1); \alpha/2}}{n-m+1 + m F_{2m, 2(n-m+1); \alpha/2}} = \frac{11 \cdot F_{22, 980; 0.05}}{490 + 11 F_{22, 980; 0.05}} \\
 &= \frac{11 \cdot 0.6}{490 + 11 \cdot 0.6} = \frac{6.6}{496.6} = 0.0133; \\
 p_2 &= \frac{(m+1) F_{2(m+1), 2(n-m); 1-\alpha/2}}{n-m+(m+1) F_{2(m+1), 2(n-m); 1-\alpha/2}} = \frac{12 \cdot F_{24, 978; 0.95}}{489 + 12 F_{24, 978; 0.95}} \\
 &= \frac{12 \cdot 1.52}{479 + 12 \cdot 1.52} = 0.0367.
 \end{aligned}$$

Das gesuchte 0,9-Konfidenzintervall ist also

$$[p_1; p_2] = [0.0133; 0.0367]$$

3.1.2. Testen von Hypothesen über den Parameter p

Wir werden hier zunächst Tests von Hypothesen über den Parameter p durchführen, wenn die Anzahl n der Versuche groß genug ist, um eine Approximation der Binomialverteilung durch die Normalverteilung zu rechtfertigen. Eine Faustregel besagt, daß dies der Fall ist, falls gilt $n p(1-p) \geq 9$. Wie schon in Abschnitt 1.5 verweisen wir auch hier bzgl. der Darstellung sequentieller Testverfahren auf Abschnitt 6.

Wir wollen nun für die einseitigen Fragestellungen

$$H_{01}: p \leq p_0 \text{ gegen } H_{11}: p > p_0$$

$$\text{und } H_{02}: p \geq p_0 \text{ gegen } H_{12}: p < p_0$$

sowie für die zweiseitige Fragestellung

$$H_{03}: p = p_0 \text{ gegen } H_{13}: p \neq p_0$$

Niveau- α -Tests vorstellen.

Für $p = p_0$ ist die Prüfgröße

$$Z = \frac{m - np_0}{\sqrt{np_0(1-p_0)}},$$

die wir schon in 3.1.1 kennengelernt haben, approximativ standardnormalverteilt (m bezeichnet die Häufigkeit des interessierenden Ereignisses). Wir müssen also die Nullhypothese H_{01} verwerfen, falls gilt

$$Z > u_{1-\alpha}$$

und H_{02} verwerfen im Falle

$$Z < u_\alpha.$$

Im zweiseitigen Testproblem wird schließlich die Hypothese H_{03} verworfen, falls

$$|Z| > u_{1-\alpha/2}.$$

Beispiel: Kommen wir noch einmal auf Beispiel (a) aus Abschnitt 3.1.1 zurück. Wir wollen zum Niveau $\alpha = 0.05$ die Hypothese $H_{03}: p = 0.03$ gegen die Alternative

$H_{13}: p \neq 0,03$ testen. Es ist

$$|Z| = \left| \frac{19 - 1000 \cdot 0,03}{\sqrt{1000 \cdot 0,03 \cdot 0,97}} \right| = |-2,039| = 2,039$$

und $u_{1-\alpha/2} = u_{0,975} = 1,960$.

Die Nullhypothese, daß 3% der Glühbirnen im Produktionslos defekt sind, muß also verworfen werden.

Ist n klein, so machen wir uns die Tatsache zunutze, daß die zur Realisation m gehörige Zufallsvariable M unter $p = p_0$ nach $B(n, p_0)$ -verteilt ist. Wir verwerfen H_{01} , wenn m größer als $c_{1-\alpha}$ ist, wobei sich $c_{1-\alpha}$ bestimmt als der kleinste Wert γ mit

$$\sum_{i=\gamma+1}^n \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha.$$

Wir verwerfen H_{02} , wenn m kleiner als c_α ist, wobei c_α bestimmt wird durch den größten Wert γ mit

$$\sum_{i=0}^{\gamma-1} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha.$$

Im zweiseitigen Testproblem schließlich verwerfen wir die Nullhypothese H_{03} , wenn m größer ist als $c_{1-\alpha/2}$ oder kleiner als $c_{\alpha/2}$.

3.1.3. Bestimmung des Stichprobenumfangs n beim Testen von Hypothesen über den Parameter p einer Binomialverteilung bei vorgegebenen Fehlern 1. und 2. Art

Wollen wir

$$H_0: p \leq p_0 \text{ gegen } H_1: p > p_0$$

zum Niveau α testen und dabei einen Fehler 2. Art β an einer Stelle $p_1 > p_0$ absichern, so müssen wir als Stichprobenumfang

$$n \geq \left(\frac{\sqrt{p_0(1-p_0)} u_{1-\alpha} + \sqrt{p_1(1-p_1)} u_{1-\beta}}{p_1 - p_0} \right)^2$$

wählen. Die gleiche Formel zur Bestimmung von n gilt für die Fragestellung

$$H_0: p \geq p_0 \text{ gegen } H_1: p < p_0$$

bei vorgegebenem β -Fehler an einer Stelle $p_1 < p_0$. Im zweiseitigen Testproblem

$$H_0: p = p_0 \text{ gegen } H_1: p \neq p_0$$

zum Niveau α bestimmt man den Stichprobenumfang n aus

$$n \geq \left(\frac{\sqrt{p_0(1-p_0)} u_{1-\alpha/2} + \sqrt{p_1(1-p_1)} u_{1-\beta}}{p_1 - p_0} \right)^2,$$

wenn man an einer Stelle $p_1 \neq p_0$ einen Fehler 2. Art β vorgibt; vgl. auch Abschnitt 1.5.2 und Kapitel VI.

Beispiel: Wir wollen zum Niveau $\alpha = 0,05$

$$H_0: p \leq 0,3 \text{ gegen } H_1: p > 0,3$$

testen und dabei an der Stelle $p_1 = 0,5$ einen Fehler $\beta = 0,05$ 2. Art vorgeben. Um dann die beiden Fehlerwahrscheinlichkeiten α und β nicht zu überschreiten, müssen wir eine Stichprobe vom Umfang

$$\begin{aligned} n &\geq \left(\frac{\sqrt{p_0(1-p_0)} u_{1-\alpha} + \sqrt{p_1(1-p_1)} u_{1-\beta}}{p_1 - p_0} \right)^2 \\ &= \left(\frac{\sqrt{0,3 \cdot 0,7} u_{0,95} + \sqrt{0,5 \cdot 0,5} u_{0,95}}{0,5 - 0,3} \right)^2 \\ &= \left(\frac{\sqrt{0,21} \cdot 1,645 + \sqrt{0,25} \cdot 1,645}{0,2} \right)^2 = 62,1 \end{aligned}$$

aus der Grundgesamtheit ziehen.

3.2. Die hypergeometrische Verteilung

Beispiel: Aus einem Sortiment von $N = 8$ Dioden werden $n = 3$ zufällig (**ohne Zurücklegen**) gezogen. Es ist bekannt, daß $M = 5$ Dioden des Sortiments defekt sind. Wie groß ist die Wahrscheinlichkeit genau $m = 2$ defekte unter den 3 gezogenen Dioden zu finden?

Bezeichnet X die Anzahl der defekten Dioden in der Stichprobe, so ist mit Hilfe der Kombinatorik, vgl. Abschnitt 3 in Kap. II:

$$P(X = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}} = \frac{\binom{5}{2} \binom{3}{1}}{\binom{8}{3}} = \frac{10 \cdot 3}{56} = 0,5357.$$

Diese Verteilung heißt **hypergeometrische Verteilung** mit Parametern (N, M, n) .

Eine Zufallsvariable X heißt **hypergeometrisch verteilt mit Parametern N, M , und n** , $X \sim H(N, M, n)$, falls

$$P(X = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}} = h(N, M, n, m) \text{ gilt, wobei}$$

$$m = \max(0, n + M - N), \dots, \min(M, n).$$

Es sind

$$EX = \frac{M \cdot n}{N} \text{ und } Var X = \frac{N-n}{N-1} n \left(1 - \frac{M}{N}\right) \frac{M}{N}$$

Erwartungswert und Varianz einer $H(N, M, n)$ -verteilten Zufallsvariablen X .

Man bezeichnet $\frac{M}{N} = p$ auch als **Anteilswert**. Ausgedrückt durch den Anteilswert p ist

$$EX = n \cdot p \text{ und } Var X = \frac{N-n}{N-1} \cdot n \cdot p \cdot q, \text{ mit } q = 1 - p.$$

Hätten wir im obigen **Beispiel** die Dioden mit **Zurücklegen** gezogen, so wäre die Anzahl Y der defekten Dioden $B\left(n, \frac{M}{N}\right)$ -verteilt gewesen. In diesem Fall würde sich

$$P(Y = m) = \binom{n}{m} \left(\frac{M}{N}\right)^m \left(1 - \frac{M}{N}\right)^{n-m} = \binom{3}{2} \left(\frac{5}{8}\right)^2 \left(1 - \frac{5}{8}\right)^1 = 3 \cdot 0,14648 \\ = 0,4395$$

ergeben.

Wird der Umfang N der Grundgesamtheit immer größer (für konstanten Anteilswert p), so spielt die Frage, ob mit oder ohne Zurücklegen gezogen wurde, immer weniger eine Rolle, d. h. es gilt:

$$h(N, M, n, m) \approx b(n, p, m) \quad \text{mit } p = \frac{M}{N}$$

(Approximation der hypergeometrischen Verteilung durch die Binomialverteilung).

Für große N besitzt der Ausdruck $\frac{N-n}{N-1}$ (**Endlichkeitskorrektur**) bei der Varianz der hypergeometrischen Verteilung praktisch keine Bedeutung mehr; die Formel geht über in die Formel für die Varianz der $B(n, p)$ -Verteilung.

3.2.1. Punktschätzungen für die Hypergeometrische Verteilung

Ein erwartungstreuer Schätzer für den Anteilwert p einer hypergeometrischen Verteilung ist – wie bei der Binomialverteilung – die relative Häufigkeit des Ereignisses A , dessen Anteil an der Grundgesamtheit p ist; d. h.

$$\hat{p} = \frac{m}{n} = \frac{\text{Anzahl der Versuche bei denen } A \text{ eintritt}}{\text{Gesamtzahl der Versuche}}$$

Dieser Schätzer hat die Varianz

$$\text{Var } \hat{p} = \frac{N-n}{N-1} \cdot \frac{1}{n} \cdot p(1-p).$$

Wie wir wissen, ist die hypergeometrische Verteilung für Grundgesamtheiten, deren Umfang N sehr groß ist, mit der Binomialverteilung nahezu identisch. Genauso nähert sich die Varianz des Schätzers \hat{p} für großes N der Varianz des Anteilsschätzers bei der Binomialverteilung.

Beispiel: Wir wollen anhand einer Stichprobe von $n = 25$ Glühbirnen den Anteil an defekten Birnen in einem Produktionslos mit $N = 700$ Glühbirnen schätzen. In der Stichprobe finden wir $m = 2$ defekte Glühbirnen. Damit ergibt sich der Anteilsschätzer zu

$$\hat{p} = \frac{m}{n} = \frac{2}{25} = 0,08.$$

Wir schätzen also den Anteil defekter Glühbirnen im Produktionslos mit 8%, was gerade 56 defekten Glühbirnen entspricht.

Wir wollen jetzt ein ganz anderes Problem betrachten, nämlich die Schätzung des unbekannten Umfangs N einer Grundgesamtheit.

Beispiel: Um den ökologischen Zustand eines Sees beurteilen zu können, möchte man den Fischbestand in diesem See kennen: Natürlich können die Fische in diesem See nicht gezählt werden, so daß die Anzahl N der Fische geschätzt werden muß. Dabei geht man so vor:

Man fängt z. B. $M = 500$ Fische, markiert sie und wirft sie dann in den See zurück. Nach einer gewissen Zeit – man kann annehmen, daß sich die Fischpopulation dann wieder vermischt hat – fängt man $n = 100$ Fische und stellt die Anzahl der in dieser Stichprobe markierten Fische – etwa $m = 5$ – fest. Um einen Schätzer für N zu bekommen, muß die Wahrscheinlichkeit

$$\frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}} = \frac{\binom{500}{5} \binom{N-500}{95}}{\binom{N}{100}}$$

für das Auftreten dieser Stichprobe dann maximiert werden. Die Anzahl N der Fische im See kann durch den Wert \hat{N} , der dieses Maximum liefert, abgeschätzt werden. In diesem Beispiel wird das Maximum für

$$\hat{N} = 10000$$

angenommen.

Das obige Beispiel ist eine Anwendung der in Kapitel III beschriebenen Maximum-Likelihood-Methode. Allgemein wird die Funktion

$$\frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}$$

bzgl. N maximal für $M \frac{n}{m}$. Dieses Maximum kann man natürlich nicht als Schätzer wählen, da N immer ganzzahlig ist. Man wählt deshalb als Maximum-Likelihood-Schätzer für den Umfang N einer Grundgesamtheit die in der Nähe davon liegende ganze Zahl \hat{N} .

Dieser Schätzer ist nicht erwartungstreu. Trotzdem ist er ein zu empfehlender Schätzer, da **kein erwartungstreuer Schätzer existiert**.

3.3. Die Multinomialverteilung

Befinden sich in einer Grundgesamtheit vom Umfang N genau k unterschiedliche Sorten von Elementen ($k \geq 2$), so können wir uns fragen: Wie groß ist die Wahrscheinlichkeit (beim Ziehen mit Zurücklegen bzw. bei (unendlich) großer Grundgesamtheit), vgl. 3.2., genau x_1 Elemente von der ersten Sorte, x_2 von der zweiten Sorte etc. in einer Stichprobe vom Umfang n zu haben?

Bezeichnet die Zufallsvariable X_i die Anzahl der Elemente der i -ten Sorte in der Stichprobe, p_i den Anteil der i -ten Sorte in der Grundgesamtheit, so gilt:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} \cdot p_1^{x_1} \cdot \dots \cdot p_k^{x_k}$$

mit $x_i = 0, 1, 2, \dots, n$, $i = 1, \dots, k$, $\sum_{i=1}^k x_i = n$, $\sum_{i=1}^k p_i = 1$.

Die Verteilung des Zufallsvektors (X_1, \dots, X_k) heißt **Multinomial- (auch Polynomial-)Verteilung mit den Parametern n, p_1, \dots, p_k** , (kurz: $X = (X_1, \dots, X_k) \sim M(n, p_1, \dots, p_k)$).

Als Spezialfall ergibt sich für $k = 2$ gerade die Binomialverteilung. Ist der Umfang der Grundgesamtheit N klein, so ist analog wie bei dem Paar Binomial – Hypergeometrische Verteilung, als Verallgemeinerung der hypergeometrischen die **multidimensionale hypergeometrische Verteilung**, vgl. Morgenstern (1968), angebracht.

Für die Multinomialverteilung gilt

$$E X = (n p_1, \dots, n p_k) \quad \text{mit } X = (X_1, \dots, X_k) \quad \text{und}$$

$$\text{Cov}(X_i, X_j) = \begin{cases} n \cdot p_i(1 - p_i), & \text{falls } i = j \\ -n \cdot p_i \cdot p_j, & \text{falls } i \neq j \end{cases}$$

Für die i -te Randverteilung von X , d. h. für die Verteilung von X_i gilt $X_i \sim B(n, p_i)$.

Beispiel: (a) Beim Ausgang eines Versuches sind k verschiedene Ereignisse A_1, \dots, A_k möglich. Der Versuch wird n -mal unabhängig wiederholt. Dann ist der Vektor $X = (X_1, \dots, X_k) \sim M(n, p_1, \dots, p_k)$, wobei X_i zählt, wie oft das Ereignis A_i eingetreten ist und $P(A_i) = p_i$. Allgemein gilt: Sind X_1, \dots, X_n (stochastisch) unabhängig mit $X_j \sim M(n_j, p_1, \dots, p_k)$, so ist

$$\sum_{j=1}^n X_j \sim M\left(\sum_{j=1}^n n_j, p_1, \dots, p_k\right).$$

(b) Bei einem Fragebogen wird nach dem Alter der befragten Person gefragt. Das Alter ist in Klassen eingeteilt: 0–20 Jahre, 20–40 Jahre, 40–60 Jahre, älter als 60 Jahre. Der Anteil der Bevölkerung an der i -ten Klasse beträgt p_i mit $i = 1, 2, 3, 4$ und $\sum_{i=1}^4 p_i = 1$.

Es werden 1000 Personen von einem Interviewer befragt. Wie groß ist die Wahrscheinlichkeit, daß höchstens 10% der Befragten bis zu 20 Jahre und höchstens 10% älter als 60 Jahre waren?

Ist $X_i^l = 1$ oder $= 0$, je nachdem ob die l -te Person in der i -ten Altersklasse liegt oder nicht, so ist $X^l = (X_1^l, \dots, X_4^l) \sim M(1, p_1, \dots, p_4)$.

Die Zufallsvariable $X = \sum_{l=1}^{1000} X^l$ ist dann $M(1000, p_1, \dots, p_4)$ -verteilt.

Gesucht ist also:

$$\begin{aligned}
 P(X_1 \leq 100, X_4 \leq 100) &= P(X_1 \leq 100, X_2 + X_3 = 1000 - X_1 - X_4, \\
 &\quad X_4 \leq 100) \\
 &= \sum_{i=0}^{100} \sum_{j=0}^{100} P(X_1 = i, X_2 + X_3 = 1000 - i - j, \\
 &\quad X_4 = j) \\
 &= \sum_{i=0}^{100} \sum_{j=0}^{100} \frac{1000!}{i! \cdot j!} p_1^i \cdot p_4^j (p_2 + p_3)^{1000-i-j}.
 \end{aligned}$$

3.3.1. Konfidenzbereich für die Multinomialverteilung

Wir wollen nun einen Konfidenzbereich zum Niveau $1 - \alpha$ für die Parameter p_1, \dots, p_k einer $M(n, p_1, \dots, p_k)$ -Verteilung bestimmen. Da es sich hier um einen Konfidenzbereich für einen Parametervektor handelt, treten einige Schwierigkeiten auf. Wir gehen aus von der Gleichung

$$P(p_{1i} \leq p_i \leq p_{2i}; i = 1, \dots, k) = 1 - \alpha$$

und erhalten für $i = 1, \dots, k$ schließlich (approximativ)

$$\begin{aligned}
 p_{1i} &= \frac{\chi_{k-1; 1-\alpha}^2 + 2r_i - \sqrt{\chi_{k-1; 1-\alpha}^2 \left(\chi_{k-1; 1-\alpha}^2 + 4r_i \frac{n - r_i}{n} \right)}}{2(n + \chi_{k-1; 1-\alpha}^2)} \\
 \text{und } p_{2i} &= \frac{\chi_{k-1; 1-\alpha}^2 + 2r_i + \sqrt{\chi_{k-1; 1-\alpha}^2 \left(\chi_{k-1; 1-\alpha}^2 + 4r_i \frac{n - r_i}{n} \right)}}{2(n + \chi_{k-1; 1-\alpha}^2)}.
 \end{aligned}$$

Dabei bezeichnet n die Anzahl der Versuche, r_i die absolute beobachtete Häufigkeit des i -ten interessierenden Ereignisses und $\chi_{v, \alpha}^2$ das α -Quantil der χ^2 -Verteilung.

Beispiel: Eine Urne enthält Kugeln in den Farben rot, grün, blau und gelb ($k = 4$). Man möchte nun anhand einer Stichprobe vom Umfang $n = 200$ einen Konfidenzbereich zum Niveau $1 - \alpha = 0,95$ für die Wahrscheinlichkeiten p_1, \dots, p_4 , daß eine gezogene Kugel rot (p_1), grün (p_2), blau (p_3) bzw. gelb (p_4) ist, angeben. Von den 200 gezogenen Kugeln waren 35 rot ($r_1 = 35$), 74 grün ($r_2 = 74$), 22 blau ($r_3 = 22$) und 69 gelb ($r_4 = 69$).

Es ergibt sich

$$\begin{aligned}
 p_{11} &= \frac{\chi_{3; 0,95}^2 + 2 \cdot 35 - \sqrt{\chi_{3; 0,95}^2 \left(\chi_{3; 0,95}^2 + 4 \cdot 35 \cdot \frac{200 - 35}{200} \right)}}{2(200 + \chi_{3; 0,95}^2)} \\
 &= \frac{7,815 + 70 - \sqrt{7,815(7,815 + 140 \cdot 0,825)}}{2(200 + 7,815)} \\
 &= \frac{77,815 - \sqrt{963,71}}{415,63} = \frac{46,77}{415,63} = 0,113,
 \end{aligned}$$

$$p_{21} = \frac{77,815 + \sqrt{963,71}}{415,63} = \frac{108,86}{415,63} = 0,262,$$

$$p_{12} = \frac{7,815 + 148 - \sqrt{7,815 \left(7,815 + 296 \cdot \frac{200 - 74}{200} \right)}}{415,63}$$

$$= \frac{155,815 - \sqrt{1518,42}}{415,63} = \frac{116,85}{415,63} = 0,281,$$

$$p_{22} = \frac{155,815 + \sqrt{1518,42}}{415,63} = \frac{194,78}{415,63} = 0,469,$$

$$p_{13} = \frac{7,815 + 44 - \sqrt{7,815 \left(7,815 + 88 \cdot \frac{200 - 22}{200} \right)}}{415,63}$$

$$= \frac{51,815 - \sqrt{673,15}}{415,63} = \frac{25,87}{415,63} = 0,062,$$

$$p_{23} = \frac{51,815 + \sqrt{673,15}}{415,63} = \frac{77,76}{415,63} = 0,187,$$

$$p_{14} = \frac{7,815 + 138 - \sqrt{7,815 \left(7,815 + 276 \cdot \frac{200 - 69}{200} \right)}}{415,63}$$

$$= \frac{145,815 - \sqrt{1473,87}}{415,63} = \frac{107,42}{415,63} = 0,258 \quad \text{und}$$

$$p_{24} = \frac{145,815 + \sqrt{1473,87}}{415,63} = \frac{184,21}{415,63} = 0,443.$$

Der gesuchte Konfidenzbereich zum Niveau 0,95 ist also

$$\left[\begin{pmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \end{pmatrix}, \begin{pmatrix} p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \end{pmatrix} \right] = \left[\begin{pmatrix} 0,113 \\ 0,281 \\ 0,062 \\ 0,258 \end{pmatrix}, \begin{pmatrix} 0,262 \\ 0,469 \\ 0,187 \\ 0,443 \end{pmatrix} \right].$$

Bzgl. des Vergleichs mehrerer Multinomialverteilungen sei auf Abschnitt 5.1.2.C in Kap. VII verwiesen.

3.4. Die Poissonverteilung

Die Poissonverteilung heißt auch die **Verteilung der seltenen Ereignisse**. Tritt ein Ereignis relativ selten ein, so findet die Poissonverteilung Verwendung bei der Beschreibung der Anzahl der Fälle, in denen das Ereignis eintritt. Beispiele für Poisson-verteilte Zufallsvariable sind somit etwa die Anzahl der radioaktiven Zerfälle in

einem festen Zeitintervall, die Anzahl von Isolationsfehlern bei Kupferdraht oder die Anzahl der Unkrautsamen im Getreide.

Eine Zufallsvariable X heißt **Poisson-verteilt mit Parameter λ** ($X \sim Po(\lambda)$), falls gilt

$$P(X \leq k) = \sum_{i=0}^k \frac{\lambda^i}{i!} e^{-\lambda} = p(\lambda, k) \quad \text{mit } k = 0, 1, 2, \dots$$

Erwartungswert und Varianz dieser Verteilung sind

$$E X = \lambda \quad \text{und} \quad \text{Var } X = \lambda.$$

Der Poissonsche Grenzwertsatz, vgl. Kap. II, besagt, daß eine Binomialverteilung mit Parametern n und p sich für wachsendes n immer mehr einer Poissonverteilung $Po(\lambda)$ mit Parameter $\lambda = np$ nähert. Ist λ hinreichend groß (etwa $\lambda \geq 9$), so läßt sich die Poissonverteilung ihrerseits durch die Normalverteilung $N(\lambda, \lambda)$ approximieren:

$$P(X \leq k) \simeq \Phi\left(\frac{k - \lambda}{\sqrt{\lambda}}\right).$$

Zur Berechnung von **Poissonwahrscheinlichkeiten** ist manchmal auch folgende Beziehung verwendbar. Ist X eine $Po(\lambda)$ -verteilte Zufallsvariable und χ^2 eine χ^2_{2x} -verteilte Zufallsvariable, so gilt

$$P(X < x) = 1 - P(\chi^2 \leq 2\lambda).$$

Sind X_i , für $i = 1, \dots, n$, unabhängige Poisson-verteilte Zufallsvariable mit Parametern λ_i , so ist die Summe $X_1 + X_2 + \dots + X_n$ poisson-verteilt mit Parameter $\lambda_1 + \lambda_2 + \dots + \lambda_n$.

Beispiel: (vgl. Bortkiewicz (1898), Das Gesetz der kleinen Zahlen, Leipzig).

In 10 Regimentern der preußischen Armee starben in 20 Jahren ($n = 10 \cdot 20 = 200$) 122 Soldaten $\left(\sum_{i=0}^{\infty} i \cdot n_i = 122 \right)$ an den Folgen eines Huftrittes, s. Tab. 27.

Wir betrachten als Zufallsvariable X die Anzahl der Soldaten, die im Laufe eines Jahres pro Regiment durch Huftritt ums Leben kamen. Die durchschnittliche To-

Tab. 27: Häufigkeit der Todesfälle an den Folgen eines Huftrittes in 10 preußischen Regimentern in 20 Jahren

Anzahl i der Todesfälle pro Jahr pro Regiment	Anzahl n_i der Jahre mit i beobachteten Todesfällen	Häufigkeit $h_i = n_i/n$	Wahrscheinlichkeit p_i nach der Poisson-Verteilung	Theoretische Anzahl $p_i n$ der Jahre mit i Todesfällen (gerundet)
0	109	0,545	0,544	109
1	65	0,325	0,331	66
2	22	0,110	0,101	20
3	3	0,015	0,021	4
4	1	0,005	0,003	1
≥ 5	0	0,000	0,000	0
Total	$n = 200$	1,000	1,000	200

desrate (pro Regiment) ist dann $\lambda = 122/200 = 0,61$. Setzt man diesen Wert in die Poissonverteilung ein, so ergeben sich die theoretischen Werte p_i der dritten Spalte.

3.4.1. Punkt- und Intervallschätzung für den Parameter λ einer $Po(\lambda)$ -Verteilung

Für eine $Po(\lambda)$ -verteilte Zufallsvariable X gilt

$$P(X < x) = \sum_{k=0}^{x-1} \frac{e^{-\lambda} \lambda^k}{k!} = 1 - P(\chi^2 \leq 2\lambda),$$

wobei x eine natürliche Zahl ist und wobei die Größe χ^2 verteilt ist gemäß einer χ^2_{2x} -Verteilung. Aus dieser Beziehung zwischen Poisson- und χ^2 -verteilten Größen ergibt sich als **Punktschätzer für λ** der Wert (wobei x Realisation von X)

$$\frac{1}{2} \chi^2_{2x+1; 0,5} (\hat{=} 0,5 \cdot (0,5 \text{ Quantil der } \chi^2_{2x+1}\text{-Verteilung})).$$

Gleichfalls aus dieser Beziehung läßt sich ein $1 - \alpha$ -Konfidenzintervall für den Parameter λ einer $Po(\lambda)$ -verteilten Zufallsgröße X bestimmen. Die Grenzen λ_1 und λ_2 eines solchen Konfidenzintervalls ergeben sich aus der eingangs genannten Beziehung zur χ^2 -Verteilung zu:

$$\lambda_1 = \frac{1}{2} \chi^2_{2x; \alpha/2} \text{ und } \lambda_2 = \frac{1}{2} \chi^2_{2x+2; 1-\alpha/2},$$

so daß sich das gesuchte Konfidenzintervall schreiben läßt als

$$[\lambda_1; \lambda_2] = \left[\frac{1}{2} \chi^2_{2x; \alpha/2}; \frac{1}{2} \chi^2_{2x+2; 1-\alpha/2} \right].$$

Beispiel: (entnommen aus A. Hald (1965))

Wir wollen Aussagen über die Verunreinigung von Kleesaatgut mit Flachs machen. Daher säen wir eine Probe von 100 g aus und zählen die Anzahl x der aufgegangenen Flachssamen (da hier n (= Anzahl der Samen) groß und p (= Wahrscheinlichkeit, daß es sich um einen aufgegangenen Flachssamen handelt) sehr klein ist, liegt hier eine Poissonverteilung vor). Es ergab sich speziell, daß die Anzahl x der aufgegangenen Flachssamen 4 war.

Eine Punktschätzung für den Parameter λ dieser Poissonverteilung ist also

$$\frac{1}{2} \chi^2_{2 \cdot 4 + 1; 0,5} = \frac{1}{2} \cdot 8,3 = 4,15$$

und ein Konfidenzintervall zum Niveau 0,9 für den Parameter λ ist

$$\left[\frac{1}{2} \chi^2_{8; 0,05}; \frac{1}{2} \chi^2_{10; 0,95} \right] = \left[\frac{1}{2} \cdot 2,7; \frac{1}{2} \cdot 18,3 \right] = [1,35; 9,15].$$

3.4.2. Test über den Parameter λ einer $Po(\lambda)$ -Verteilung

Wie schon erwähnt, ist eine $Po(\lambda)$ -Verteilung bei großem λ durch eine $N(\lambda, \lambda)$ -Verteilung approximierbar. Davon wollen wir nun Gebrauch machen.

Ist die Zufallsvariable X $Po(\lambda)$ -verteilt, so ist natürlich die standardisierte Zufallsvariable

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

annähernd $N(0, 1)$ -verteilt, so daß wir zur Überprüfung der Hypothesen

$$H_{01}: \lambda \leq \lambda_0 \quad \text{gegen} \quad H_{11}: \lambda > \lambda_0,$$

$$H_{02}: \lambda \geq \lambda_0 \quad \text{gegen} \quad H_{12}: \lambda < \lambda_0,$$

$$\text{und} \quad H_{03}: \lambda = \lambda_0 \quad \text{gegen} \quad H_{13}: \lambda \neq \lambda_0$$

den schon bekannten Einstichproben-Gaußtest verwenden können. Die Prüfgröße ist natürlich

$$Z = \frac{X - \lambda_0}{\sqrt{\lambda_0}}$$

und die Hypothese H_{01} wird verworfen zum Niveau α , falls

$$Z > u_{1-\alpha}$$

ist. Die Hypothesen H_{02} und H_{03} werden verworfen zum Niveau α , wenn gilt:

$$Z < u_\alpha \quad \text{bzw.} \quad |Z| > u_{1-\alpha/2}.$$

Beispiel: (vgl. Student (1907), On the error of counting with a haemacytometer, Biometrika 5, S. 351–360)

Wir können davon ausgehen, daß die Anzahl von Hefezellen in einer Suspension Poisson-verteilt ist und wollen nun eine Hypothese über die erwartete Anzahl von Hefezellen in einer Volumeneinheit der Suspension testen. Zugrundegelegt wird dabei eine Stichprobe vom Umfang $n = 400$ Volumeneinheiten, in denen die Hefezellen gezählt wurden, vgl. **Tab. 28**.

Tab. 28: Hefezellen in der Suspension

Anzahl der Hefezellen k	Anzahl der Volumeneinheiten n_k mit k Hefezellen
0	0
1	20
2	43
3	53
4	86
5	70
6	54
7	37
8	18
9	10
10	5
11	2
12	2
> 12	0

Wir wollen testen, ob die Anzahl von Hefezellen pro Volumeneinheit der Suspension $Po(\lambda)$ mit $\lambda \geq 5 = \lambda_0$ verteilt ist. Es ist die hier vorliegende Realisation der

Zufallsvariablen X gerade

$x = \text{Gesamtzahl der Hefezellen in } 400 \text{ Volumeneinheiten} = 1872.$

Testen wir zunächst die Nullhypothese

$H_{02}: \lambda \geq 5$ gegen $H_{12}: \lambda < 5$ zum Niveau 0,05.

Unter dieser Nullhypothese ist die Gesamtzahl X von Hefezellen in 400 Einheiten der Suspension Po $(400 \cdot \lambda_0)$ -verteilt, so daß sich

$$Z = \frac{1872 - 400 \cdot 5}{\sqrt{400 \cdot 5}} = -2,862$$

ergibt. Somit muß wegen

$$u_\alpha = u_{0,05} = -1,645$$

die Nullhypothese verworfen werden. Nun wollen wir zum Niveau 0,02 die Hypothesen

$H_{03}: \lambda = 5$ gegen $H_{13}: \lambda \neq 5$

testen. Hier ergibt sich wiederum

$$Z = -2,862 \text{ also } |Z| = 2,862,$$

so daß wir wegen $u_{0,99} = 2,33$ auch diese Nullhypothese verwerfen müssen.

3.4.3. Der χ^2 -Anpassungstest für die Poissonverteilung

Natürlich können wir beim χ^2 -Anpassungstest der Hypothese

H_0 : die Beobachtungen x_1, \dots, x_n entstammen einer Po(λ)-Verteilung gegen die Alternative

H_1 : sie entstammen keiner Po(λ)-Verteilung auf das in 1.6.A (Schritte 1.–4.) vorgestellte Verfahren zurückgreifen.

Beispiel: Wir wollen zum Niveau $\alpha = 0,05$ mit Hilfe des χ^2 -Anpassungstests überprüfen, ob die Anzahl der Anrufe in der Telefonzentrale eines Hotels im Verlauf von 15 min Po(3)-verteilt ist. Beobachtet wurde die Zahl der Anrufe in $n = 50$ Zeitintervallen à 15 min, vgl. Tab. 29.

Tab. 29: Anzahl der Anrufe in einer Telefonzentrale

Anzahl der Anrufe k	Anzahl der Zeitintervalle n_k mit k Anrufern
0	1
1	6
2	8
3	10
4	6
5	6
6	7
7	1
8	1
9	2
10	2
> 10	0

Tab. 30: Beobachtete und erwartete Anrufe in 5 Intervallen

i	O _i	E _i
1 ($-\infty; 1]$)	7	9,95
2 (1; 2]	8	11,2
3 (2; 3]	10	11,2
4 (3; 4]	6	8,4
5 (4; $\infty)$	19	9,25

Wir bilden hier $m = 5$ Intervalle. In **Tab. 30** sind die beobachteten und erwarteten Werte in den einzelnen Intervallen eingetragen.

Hier ist etwa

$$\begin{aligned} p_1 &= P(X \in (-\infty, 1]) = P(X = 0, X = 1) = \frac{e^{-\lambda} \lambda^0}{0!} + \frac{e^{-\lambda} \lambda^1}{1!} \\ &= \frac{e^{-3} \cdot 3^0}{1} + \frac{e^{-3} \cdot 3^1}{1} = 0,050 + 0,149 = 0,199 \end{aligned}$$

und

$$\begin{aligned} E_1 &= n p_1 = 50 \cdot 0,199 = 9,95, \\ \text{oder } p_3 &= P(X \in (2, 3]) = P(X = 3) = \frac{e^{-\lambda} \cdot \lambda^3}{3!} = \frac{e^{-3} \cdot 3^3}{6} = 0,224 \end{aligned}$$

und $E_3 = n p_3 = 50 \cdot 0,224 = 11,2$.

Der Wert der Teststatistik ergibt sich dann in diesem Fall zu

$$T(x) = \sum_{i=1}^5 \frac{(O_i - E_i)^2}{E_i} = 12,88.$$

Weiter ergibt sich mit Hilfe der im Anhang zu findenden Tafel der Quantile der χ^2 -Verteilung

$$\chi^2_{5-1; 0,95} = 9,5 < T(x),$$

so daß wir die Nullhypothese zu diesem Niveau verwerfen müssen.

Auf den Vergleich der Parameter mehrerer Poissonverteilungen wird im Abschnitt 5.1.2.A von Kap.VII eingegangen

4. Einige Lebensdauerverteilungen

In diesem Abschnitt sollen beispielhaft einige Verteilungen vorgestellt werden, die sich vor allem bei der Modellierung von Lebensdauern in der Zuverlässigkeitstheorie und in der Biologie/Medizin bewährt haben, vgl. auch Kap. XIII.

Bevor wir uns den konkreten Verteilungen zuwenden, werden zunächst einige Begriffe erläutert, die hier von speziellem Interesse sind.

Unter der **Lebensdauer** eines Objektes versteht man die Zeit zwischen Betriebsbeginn (Geburt) und Ausfall (Tod) des Objektes. Den Betriebsbeginn nehmen wir hier stets als Zeitpunkt $x = 0$ an.

Beschreibt nun die Zufallsvariable X die Lebensdauer eines Objektes und ist $F_X(x)$ die zugehörige Verteilungsfunktion, so ist die Wahrscheinlichkeit dafür, daß das Objekt den Zeitpunkt $x \geq 0$ überlebt, gerade

$$P(X > x) = 1 - F_X(x) = \bar{F}_X(x);$$

diese Wahrscheinlichkeit nennt man auch die **Überlebenswahrscheinlichkeit**.

Hat ein Objekt den Zeitpunkt x_0 überlebt, so ist es vielfach von Interesse, die Wahrscheinlichkeit zu kennen, mit der das Objekt auch den Zeitpunkt $x_0 + x$ überlebt. Diese (unter x_0) bedingte **Überlebenswahrscheinlichkeit** ist natürlich gerade

$$\begin{aligned} P(X > x_0 + x | X > x_0) &= \frac{P(X > x_0 + x)}{P(X > x_0)} = \frac{\bar{F}_X(x_0 + x)}{\bar{F}_X(x_0)} \\ &= \bar{F}_X(x_0 + x | x_0). \end{aligned}$$

Die Zeit zwischen dem festen Zeitpunkt x_0 und dem Ausfall eines Objektes nennt man auch seine **Restlebensdauer** bzgl. x_0 .

Natürlich ist

$$\begin{aligned} F_X(x_0 + x | x_0) &= P(X \leq x_0 + x | X > x_0) \\ &= 1 - \bar{F}_X(x_0 + x | x_0) \end{aligned}$$

gerade die **Verteilungsfunktion (Verteilung) der Restlebensdauer** bzgl. des festen Zeitpunktes x_0 .

Eine wichtige Kenngröße von Lebensdauerverteilungen ist ihre **Ausfallrate** oder **Hazardrate**. Sie ist definiert als

$$\begin{aligned} r(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t < X \leq t + \Delta t | X > t) \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \cdot \frac{P(t < X \leq t + \Delta t)}{P(X > t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \cdot \frac{F_X(t + \Delta t) - F_X(t)}{1 - F_X(t)} \\ &= \frac{1}{1 - F_X(t)} \cdot \frac{dF}{dt} \\ &= \frac{1}{1 - F_X(t)} \cdot f_X(t), \end{aligned}$$

wobei $f_X(t)$ die Dichtefunktion der Lebensdauerverteilung bezeichnet. $r(t) \cdot \Delta t$ gibt also für kleine Zeitintervalle Δt in etwa die Wahrscheinlichkeit an, daß ein Objekt mit Lebensdauerverteilung $F_X(t)$ nach Erreichen des Lebensalters t innerhalb der nachfolgenden Zeitspanne Δt ausfällt bzw. stirbt, vgl. auch Abschnitt 8 in Kapitel I, wo bereits die empirische Ausfallrate betrachtet wurde.

Die Ausfallrate $r(t)$ ist deshalb besonders wichtig, da sie die Lebensdauerverteilung $F_X(t)$ vollständig charakterisiert. Es ist nämlich mit

$$r(t) = \frac{f_X(t)}{1 - F_X(t)} = - \frac{d(\ln(1 - F_X(t)))}{dt}$$

und somit wegen

$$[\ln(1 - F_X(\xi))]_{\xi=0}^t = - \int_0^t r(\xi) d\xi \quad \text{und} \quad F_X(0) = 0$$

gerade

$$F_X(t) = 1 - e^{- \int_0^t r(\xi) d\xi}.$$

Weitere Kenngrößen von Lebensdauerverteilungen sind natürlich, wie bei allen anderen Verteilungen, der Erwartungswert, der in diesem Zusammenhang auch **mittlere Lebensdauer** heißt, die Varianz, der Median, die Schiefe und der Exzeß, vgl. Kapitel II.

4.1. Die Exponentialverteilung

Die Exponentialverteilung eignet sich zur Modellierung der Lebensdauer von Objekten, die nicht altern. Die Verteilungsfunktion einer mit Parameter $\lambda > 0$ exponentialverteilten Zufallsvariablen X ist

$$F_X(x) = 1 - e^{-\lambda x} \quad \text{für } x \geq 0;$$

somit ergibt sich

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{für } x \geq 0$$

als Dichte der **Exponentialverteilung mit Parameter λ** ($Ex(\lambda)$). Die Verteilungsfunktion und die Dichte einer $Ex(\lambda)$ -Verteilung sind bereits im Abschnitt 6 bzw. 7 des Kapitels II graphisch dargestellt.

Die **Überlebenswahrscheinlichkeit**, also die Wahrscheinlichkeit, daß ein Objekt mit $Ex(\lambda)$ -verteilter Lebensdauer einen Zeitpunkt x überlebt, ist natürlich gerade

$$P(X > x) = \bar{F}_X(x) = 1 - F_X(x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

und die unter x_0 bedingte Überlebenswahrscheinlichkeit ist gerade

$$\bar{F}_X(x_0 + x | x_0) = \frac{\bar{F}_X(x_0 + x)}{\bar{F}_X(x_0)} = \frac{e^{-\lambda(x_0+x)}}{e^{-\lambda x_0}} = e^{-\lambda x},$$

d. h. die **bedingte Überlebenswahrscheinlichkeit** ist unabhängig vom Alter. Dies ist der Grund dafür, daß die Exponentialverteilung auch **gedächtnislose Verteilung** genannt wird; natürlich bedingt dies, daß eine vorzeitige Erneuerung noch nicht ausgestrahlter Objekte im Fall exponentialverteilter Lebensdauern sinnlos ist.

Hiermit ist auch die Verteilung der Restlebensdauer im Falle der Exponentialverteilung unabhängig vom Bezugszeitpunkt x_0 , d. h. es ist

$$F_X(x_0 + x | x_0) = F_X(x) = 1 - e^{-\lambda x} \quad \text{für jedes } x_0 \geq 0.$$

Erwartungswert und Varianz einer $Ex(\lambda)$ -verteilten Zufallsvariablen X sind

$$EX = \frac{1}{\lambda} \quad \text{und} \quad \text{Var } X = \frac{1}{\lambda^2}.$$

Die erwartete mittlere Lebensdauer eines Objektes entspricht somit gerade dem Kehrwert des Parameters λ der Verteilung seiner Lebensdauer. Unabhängig vom Parameter λ ist die Schiefe der Exponentialverteilung 2 und der Exzeß 6. Ferner ist der Median einer $Ex(\lambda)$ -Verteilung $\frac{1}{\lambda} \ln 2$.

Die **Ausfallrate** der Exponentialverteilung ist

$$r(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda,$$

wegen der Gedächtnislosigkeit der Verteilung also zu jedem Zeitpunkt t konstant gleich dem Parameter λ .

Die **Erzeugung exponentialverteilter Zufallsvariablen** zur Durchführung von Simulationsstudien ist recht einfach, denn man muß lediglich gleichverteilte Zufallszahlen erzeugen:

Ist die Zufallsvariable U auf $[0; 1]$ gleichverteilt, so ist $X = -\frac{1}{\lambda} \ln U$ gerade $Ex(\lambda)$ -verteilt.

Neben ihrer Verwendung als Ausfallverteilung, z. B. für die Lebensdauern von Dioden oder Transistoren, findet die Exponentialverteilung, die in engem Zusammenhang zur Poissonverteilung steht, Verwendung bei der Beschreibung von Wartezeiten zwischen zwei Ereignissen.

Wir wollen nun zunächst Punkt- und Intervallschätzungen sowie Tests für den Parameter λ einer $Ex(\lambda)$ -Verteilung behandeln. Anschließend wollen wir noch die aus Abschnitt 1.6 bekannten Anpassungstests – den χ^2 -Test und den Kolmogoroff-Smirnov-Test – im Falle von Exponentialverteilungshypothesen anwenden.

4.1.1. Punkt- und Intervallschätzung für den Parameter einer $Ex(\lambda)$ -Verteilung

Wie wir schon gesehen haben, ist der Erwartungswert der $Ex(\lambda)$ -Verteilung gerade $\frac{1}{\lambda}$, so daß sich als **Punktschätzer für λ** die Größe

$$\hat{\lambda} = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i} = \frac{1}{\bar{x}},$$

wobei x_1, \dots, x_n die Lebensdauern von n beobachteten Objekten aus einer interessierenden Grundgesamtheit von Objekten mit exponentialverteilter Lebensdauer sind, anbietet. Ein diesem Punktschätzer entsprechendes **Konfidenzintervall zum Niveau $1 - \alpha$ für den Parameter λ** einer Exponentialverteilung ist gerade:

$$\left[\frac{\chi^2_{2n; \alpha/2}}{2 \sum_{i=1}^n x_i}; \frac{\chi^2_{2n; 1-\alpha/2}}{2 \sum_{i=1}^n x_i} \right].$$

Beispiel: (a) Bei der Untersuchung von $n = 12$ Glühbirnen ergaben sich die Lebensdauern [in Std.] aus **Tab. 31**. Daraus ergibt sich

Tab. 31: Lebensdauern der Glühbirnen

i	1	2	3	4	5	6	7	8	9	10	11	12
Lebensdauer	635	840	721	610	805	585	913	468	1079	876	715	988

$$\bar{x} = 769,58$$

und als Punktschätzung für die Ausfallrate λ der Glühbirnen

$$\frac{1}{\bar{x}} = 0,0013.$$

Ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ für die Ausfallrate λ der Glühbirnen ist gerade

$$\left[\frac{\chi^2_{24; 0,025}}{2 \sum_{i=1}^{12} x_i}; \frac{\chi^2_{24; 0,975}}{2 \sum_{i=1}^{12} x_i} \right] = \left[\frac{12,4}{2 \cdot 9235}; \frac{39,4}{2 \cdot 9235} \right] \\ = [0,00067; 0,0021].$$

Bisher haben wir die Schätzung des Parameters λ nur dann durchgeführt, wenn der Versuch so lange dauerte, bis das letzte beobachtete System (Objekt) ausgefallen war. Nun wollen wir jedoch davon ausgehen, daß der Versuch nur eine festgelegte Zeit T dauert.

In solchen Fällen sind dann, nach Ablauf der Zeit T , m der beobachteten n Systeme ausgefallen. Wollen wir aufgrund dessen einen Punktschätzer für λ angeben, so ist das etwas schwierig. Eine recht gute Schätzung der Aufallrate λ liefert in Fällen, wo $m \geq 10$ und $\frac{m}{n} \leq 0,1$ ist, die Punktschätzung

$$\lambda_T = \frac{m}{n - \frac{m}{2}} \cdot \frac{1}{T}.$$

Bei dieser Schätzung verzichtet man also völlig auf die Berücksichtigung der Lebensdauern x_1, \dots, x_m der zum Zeitpunkt T bereits ausgefallenen Systeme.

Der – zwar nicht erwartungstreue – Maximum-Likelihood-Schätzer ist bei abgebrochener Lebensdauer-Prüfung die Größe

$$\lambda_{ML} = \frac{m}{\sum_{j=1}^m x_j + (n - m)x_m}.$$

Wenn wir n Systeme während einer festgewählten Zeit T beobachten, ist die Anzahl M der nach Ablauf von T ausgefallenen Systeme gerade $B(n, 1 - e^{-\lambda T})$ -verteilt.

Daher liegt es nahe, zur Konstruktion von $1 - \alpha$ -Konfidenzintervallen für λ , die schon in 3.1.1 hergeleiteten Konfidenzintervalle für den Parameter p einer $B(n, p)$ -Verteilung zu verwenden.

Sind p_1 und p_2 die Grenzen des Konfidenzintervall für $p = 1 - e^{-\lambda T}$, so ist natürlich das $1 - \alpha$ -Konfidenzintervall für λ gegeben als

$$\left[\frac{1}{T} \ln(1 - p_1)^{-1}; \frac{1}{T} \ln(1 - p_2)^{-1} \right].$$

Wir müssen also lediglich p_1 und p_2 bestimmen und unterscheiden dabei natürlich wieder – wie bei der Binomialverteilung – ob es sich um „große“ oder „kleine“ Stichprobenumfänge n handelt; vgl. Abschnitt 3.1.1.

Beispiel: (b) In einem Zeitraum von $T = 20$ Tagen beobachtet man $n = 500$ Glühbirnen, deren Lebensdauern exponentialverteilt sind mit unbekannter Ausfallrate λ . Nach Ablauf der 20 Tage stellt man fest, daß $m = 17$ Glühbirnen defekt sind (große Stichprobe).

Wir bestimmen ein Konfidenzintervall zum Niveau $1 - \alpha = 0,9$ für den Parameter λ der Verteilung der Lebensdauern der Glühbirnen. Es ist

$$u_{\alpha/2} = u_{0,05} = -1,645 \quad \text{und}$$

$$\begin{aligned} p_1 &= \frac{34 + 2,706 - 1,645 \sqrt{2,706 + 68 \left(1 - \frac{17}{500}\right)}}{2(500 + 2,706)} \\ &= \frac{36,706 - 1,645 \sqrt{68,394}}{1005,412} = \frac{36,706 - 13,604}{1005,412} = 0,023, \\ p_2 &= \frac{36,706 + 13,604}{1005,412} = 0,050. \end{aligned}$$

Damit ergibt sich das 0,9-Konfidenzintervall für λ zu:

$$\begin{aligned} \left[\frac{1}{T} \ln(1 - p_1)^{-1}; \frac{1}{T} \ln(1 - p_2)^{-1} \right] &= \left[\frac{1}{20} \ln(0,977)^{-1}; \frac{1}{20} \ln(0,95)^{-1} \right] \\ &= \left[\frac{1}{20} \ln 1,024; \frac{1}{20} \ln 1,053 \right] \\ &= \left[\frac{0,023}{20}; \frac{0,051}{20} \right] \\ &= [0,00115; 0,00255]. \end{aligned}$$

(c) Wir wollen ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ für den Parameter λ der exponentialverteilten Lebensdauern von Silicium-Transistoren bestimmen. 35 Tage lang beobachten wir $n = 30$ solcher Silicium-Elemente und stellen fest, daß nach Ablauf der 35 Tage $m = 3$ Elemente defekt sind (kleine Stichprobe).

Wir bestimmen zunächst p_1 und p_2 . Es ist

$$F_{2m, 2(n-m+1); \alpha/2} = F_{6, 56; 0,025} = \frac{1}{F_{56, 6; 0,975}} = \frac{1}{4,99} = 0,200 \quad \text{und}$$

$$F_{2(m+1), 2(n-m); 1-\alpha/2} = F_{8, 54; 0,975} = 2,43.$$

Daraus ergibt sich:

$$p_1 = \frac{3 \cdot 0,2}{30 - 3 + 1 + 3 \cdot 0,2} = \frac{0,6}{28,6} = 0,021,$$

$$p_2 = \frac{4 \cdot 2,43}{30 - 3 + 4 \cdot 2,43} = \frac{9,72}{36,72} = 0,265.$$

Und das gesuchte Konfidenzintervall für λ ist:

$$\begin{aligned} \left[\frac{1}{T} \ln(1-p_1)^{-1}; \frac{1}{T} \ln(1-p_2)^{-1} \right] &= \left[\frac{1}{35} \ln(0,979)^{-1}; \frac{1}{35} \ln(0,735)^{-1} \right] \\ &= \left[\frac{1}{35} \ln 1,021; \frac{1}{35} \ln 1,361 \right] \\ &= \left[\frac{0,021}{35}; \frac{0,308}{35} \right] \\ &= [0,0006; 0,0088]. \end{aligned}$$

Einige weitere Schätzverfahren im Zusammenhang mit der Exponentialverteilung findet man im Abschnitt 3 des Kap. XIII.

4.1.2. Tests von Hypothesen über den Parameter einer $Ex(\lambda)$ -Verteilung

Wie schon in Abschnitt 4.1.1 betrachten wir auch bei den Tests über den Parameter λ der $Ex(\lambda)$ -Verteilung die Wahrscheinlichkeit p für den Ausfall eines Systems mit exponentieller Lebensdauer während eines festen Zeitraumes T . Beobachtet man dann n Systeme, so ist die Zufallsvariable M , die die Anzahl der Ausfälle in T beschreibt, $B(n, p)$ mit $p = 1 - e^{-\lambda T}$ verteilt.

Ist die Anzahl der beobachteten Systeme groß (etwa $ne^{-\lambda T}(1 - e^{-\lambda T}) \geq 9$), so kann diese Binomialverteilung durch die Normalverteilung approximiert werden:

$$\frac{M - n(1 - e^{-\lambda T})}{\sqrt{n e^{-\lambda T}(1 - e^{-\lambda T})}} \sim N(0, 1).$$

Zum Überprüfen der einseitigen Fragestellungen (zum Niveau α)

$$H_{01}: \lambda \leq \lambda_0 \quad \text{gegen} \quad H_{11}: \lambda > \lambda_0$$

$$\text{und} \quad H_{02}: \lambda \geq \lambda_0 \quad \text{gegen} \quad H_{12}: \lambda < \lambda_0$$

sowie des zweiseitigen Problems

$$H_{03}: \lambda = \lambda_0 \quad \text{gegen} \quad H_{13}: \lambda \neq \lambda_0$$

zur Ausfallrate der Systeme, können wir also den Gauß-Test, den wir schon in Kap. II kennengelernt haben, verwenden, wobei die Prüfgröße gerade

$$Z = \frac{M - n(1 - e^{-\lambda_0 T})}{\sqrt{n e^{-\lambda_0 T} (1 - e^{-\lambda_0 T})}}$$

ist.

Wir müssen also die Hypothese H_{01} verwerfen, wenn

$$Z > u_{1-\alpha}$$

ist und H_{02} verwerfen, falls gilt:

$$Z < u_\alpha.$$

Die Hypothese H_{03} des zweiseitigen Testproblems verwerfen wir, falls

$$|Z| > u_{1-\alpha/2}.$$

Beispiel: Kommen wir noch einmal auf das Beispiel (b) aus dem Abschnitt über die Punkt- und Intervallschätzung für λ zurück. Wir wollen nun die Hypothese, ob die Ausfallrate λ der Glühbirnen zum 1%-Niveau signifikant größer als 0,001 ist, testen, d. h. wir testen (zum Niveau $\alpha = 0,01$):

$$H_{01}: \lambda \leq 0,001 \quad \text{gegen} \quad H_{11}: \lambda > 0,001.$$

Dazu berechnen wir zunächst den Wert der Prüfgröße

$$Z = \frac{17 - 500(1 - e^{-0,001 \cdot 20})}{\sqrt{500 e^{-0,001 \cdot 20} (1 - e^{-0,001 \cdot 20})}} = \frac{7,099}{3,115} = 2,279.$$

Das Quantil der Standardnormalverteilung für $\alpha = 0,01$ ist

$$u_{1-\alpha} = u_{0,99} = 2,326,$$

so daß wir unsere Nullhypothese nicht verwerfen können.

Natürlich kann man auch dann einen Test zum Niveau α über den Parameter λ einer exponentialverteilten Grundgesamtheit durchführen, wenn man den Versuch nicht nach einer festen Beobachtungszeit T abbricht. Man wartet solange, bis jedes Versuchsobjekt aus der Grundgesamtheit ausgefallen ist. Wählt man als Prüfgröße dann die Summe der Lebensdauern x_1, \dots, x_n aller Beobachtungsobjekte, so kann man die Hypothese

$$H_0: \lambda \leq \lambda_0$$

gegen die Alternative

$$H_1: \lambda > \lambda_0$$

wie folgt zum Niveau α testen. Man verwirft die Hypothese H_0 , falls gilt:

$$\sum_{i=1}^n x_i < c_{1-\alpha},$$

wobei $c_{1-\alpha}$ das $(1 - \alpha)$ -Quantil der Erlang-n-Verteilung ist, vgl. auch Abschnitt 4.4.

Die Verteilungsfunktion der Erlang-n-Verteilung mit Parameter λ ist

$$F(x) = 1 - e^{-\lambda x} \sum_{i=1}^n \frac{(\lambda x)^{i-1}}{(i-1)!} \quad \text{für } x \geq 0,$$

so daß also gerade

$$c_{1-\alpha} = F^{-1}(1-\alpha)$$

gilt.

Für Sequentialtests bei der Exponentialverteilung sei auf Abschnitt 6 verwiesen.

4.1.3. Der χ^2 -Anpassungstest für die Exponentialverteilung

Erinnern wir uns noch einmal an den in Abschnitt 1.6 eingeführten χ^2 -Anpassungstest. Wir wollen nun mit Hilfe desselben die Hypothese

H_0 : die Beobachtungen x_1, \dots, x_n entstammen einer $Ex(\lambda)$ -Verteilung gegen die Alternative

H_1 : sie stammen nicht aus einer $Ex(\lambda)$ -Verteilung zum Niveau α testen.

Wir können hier genau die Schritte 1.–4. aus Abschnitt 1.6 wiederholen und den Test durchführen.

Beispiel: Bei der Untersuchung der Lebensdauern von 87 Kühlaggregaten bis zum ersten Ausfall ergaben sich [in Jahren] die in Tab. 32 angegebenen Lebensdauern.

Tab. 32: Lebensdauern der Kühlaggregate (in Jahren)

0,05	0,06	0,06	0,08	0,11	0,13	0,15	0,16
0,20	0,22	0,24	0,25	0,25	0,28	0,31	0,34
0,37	0,42	0,43	0,47	0,51	0,51	0,53	0,59
0,60	0,61	0,63	0,68	0,75	0,76	0,76	0,79
0,87	0,88	0,88	0,92	0,99	1,12	1,16	1,18
1,22	1,27	1,35	1,38	1,39	1,42	1,45	1,49
1,53	1,69	1,74	1,81	1,83	1,87	1,92	1,93
2,07	2,09	2,15	2,22	2,24	2,36	2,39	2,41
2,47	2,49	2,53	2,64	2,69	2,83	2,90	3,21
3,25	3,49	3,61	3,80	3,88	4,37	4,58	4,62
5,29	5,68	6,02	6,23	6,71	7,82	9,93	

Tab. 33: Anzahlen der Kühlaggregate in den Intervallen

I_i	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
[0; 1/3]	15	13,35	0,204
(1/3; 2/3]	12	11,31	0,042
(2/3; 1]	10	9,57	0,019
(1; 1,5]	11	11,67	0,038
(1,5; 2]	8	9,09	0,131
(2; 4]	21	20,24	0,029
(4; 7]	8	9,15	0,145
(7; ∞)	2	2,62	0,147

Wir testen zum Niveau $\alpha = 0,1$ nun $H_0: X \sim \text{Ex}(0,5)$ gegen $H_1: X \not\sim \text{Ex}(0,5)$. Dabei wählen wir die Intervalle

$$I_1 = [0; 1/3], I_2 = (1/3; 2/3], I_3 = (2/3; 1], I_4 = (1; 1,5],$$

$$I_5 = (1,5; 2], I_6 = (2; 4], I_7 = (4; 7], I_8 = (7; \infty].$$

In Tab. 33 sind nun die beobachteten und erwarteten Anzahlen von Kühlaggrenzen mit Lebensdauern im jeweiligen Intervall angegeben. Dabei ergibt sich z. B. wegen

$$\begin{aligned} p_3 &= P(X \in (2/3; 1]) = F(1) - F(2/3) = (1 - e^{-\lambda \cdot 1}) - (1 - e^{-\lambda \cdot 2/3}) \\ &= -e^{-0,5 \cdot 1} + e^{-0,5 \cdot 2/3} \\ &= -e^{-0,5} + e^{-0,33} \\ &= 0,11 \end{aligned}$$

$$E_3 = n p_3 = 87 \cdot 0,11 = 9,57$$

und wegen

$$\begin{aligned} p_8 &= P(X \in (7; \infty)) = F(\infty) - F(7) = 1 - (1 - e^{-\lambda \cdot 7}) = e^{-7 \cdot 0,5} \\ &= 0,0302 \end{aligned}$$

$$E_8 = n p_8 = 87 \cdot 0,0302 = 2,62.$$

$$\text{Es ist also } T(x) = \sum_{i=1}^8 \frac{(O_i - E_i)^2}{E_i} = 0,754 \text{ und } \chi^2_{k-1, 1-\alpha} = \chi^2_{7, 0,9} = 12,02.$$

Die Hypothese der $\text{Ex}(0,5)$ -Verteilung der Lebensdauern unserer Kühlaggregate kann also nicht verworfen werden.

4.1.4. Der Kolmogoroff-Smirnov-Anpassungstest für die Exponentialverteilung

Wie im Falle der Normalverteilung ergeben sich für die Frage, ob eine konkrete Stichprobe x_1, \dots, x_n mit stetiger Verteilungsfunktion $F(x)$ aus einer Exponentialverteilung stammt, zwei Testprobleme, nämlich

1. H_0 : Die Grundgesamtheit ist $\text{Ex}(\lambda_0)$ -verteilt (mit bekannten λ_0)

gegen

H_1 : Die Grundgesamtheit ist nicht $\text{Ex}(\lambda_0)$ -verteilt,

und

2. H'_0 : Die Grundgesamtheit ist $\text{Ex}(\lambda)$ -verteilt (mit unbekanntem λ)

gegen

H'_1 : Die Grundgesamtheit ist nicht $\text{Ex}(\lambda)$ -verteilt.

Wie in Abschnitt 1.6 berechnet man zunächst die empirische Verteilungsfunktion $S_n(x)$ zu den Daten x_1, \dots, x_n , und verwendet im ersten Fall die Teststatistik

$$\sqrt{n} D_n, \text{ mit } D_n = \sup_x |S_n(x) - F_0(x)|, \text{ wobei hier } F_0(x) = 1 - e^{-\lambda_0 x},$$

und man kann auch dieselben kritischen Werte $d_{n;1-\alpha}$ aus der dortigen Tab. 14 heranziehen. Man verwirft hier die Hypothese H_0 , falls gilt

$$\sqrt{n} D_n > d_{n;1-\alpha}.$$

Natürlich sind auch hier die in Abschnitt 1.6 beschriebenen einseitigen Tests analog, unter Verwendung der nicht betraglichen Abstände

$$D_n^+ = \sup_x (S_n(x) - F_0(x)) \text{ und } D_n^- = \sup_x (F_0(x) - S_n(x))$$

sowie der kritischen Werte $d_{n;1-2\alpha}$, anwendbar.

Im zweiten Fall schätzen wir zunächst λ durch $\frac{1}{\bar{x}}$ und erhalten analog zum Normalverteilungstest die Teststatistik

$$\sqrt{n} L_n^{\exp}, \text{ mit } L_n^{\exp} = \sup_x |S_n(x) - (1 - e^{-x/\bar{x}})|,$$

deren kritische Werte nicht vom unbekannten Parameter λ abhängen. Lilliefors (1969) hat erstmals Simulationsstudien zur Ermittlung der kritischen Werte durchgeführt. Weitere von Stephens simulierte Quantile finden sich in Pearson/Hartley (1972), vgl. auch Durbin (1975).

Die Hypothese H'_0 auf Vorliegen einer Exponentialverteilung wird hier verworfen, wenn gilt

$$\sqrt{n} L_n^{\exp} > l_{n;1-\alpha}^{\exp},$$

wobei die Quantile $l_{n;1-\alpha}^{\exp}$ der Tab. 34 entnommen werden können (nächstgrößeres Quantil bei nicht vertafeltem n).

Tab. 34: Kritische Werte $l_{n;1-\alpha}^{\exp}$ zum Test auf nicht spezifizierte Exponentialverteilung

n	5	8	10	20	30	> 30
$l_{n;0.90}^{\exp}$	0,91	0,93	0,94	0,96	0,97	0,98
$l_{n;0.95}^{\exp}$	0,99	1,02	1,03	1,05	1,06	1,08
$l_{n;0.99}^{\exp}$	1,15	1,20	1,21	1,24	1,26	1,28

Beispiel: Wir wollen mit Hilfe des Kolmogoroff-Smirnov-Tests überprüfen, ob die Lebensdauer von Kühlaggregaten exponentialverteilt ist. Dabei werden die Lebensdauern einer zufälligen Stichprobe mit $n = 16$ Kühlaggregaten festgestellt. Diese Lebensdauern sind in Tab. 35 zu finden.

Tab. 35: Lebensdauer von 16 Kühlaggregaten (in Jahren)

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Lebensdauer	0,14	0,27	0,43	0,68	0,81	1,14	1,45	1,82	2,36	2,53	2,90	3,45	4,51	5,12	5,68	7,84

Wir prüfen zunächst zum Niveau $\alpha = 0,05$ die Hypothese

H_0 : Die Lebensdauer von Kühlaggregaten ist $Ex(0,5)$ -verteilt.

Dazu stellen wir zunächst **Tabelle 36** auf.

Tab. 36: Arbeitstabelle zum Kolmogoroff-Smirnov-Test bei spezifiziertem Parameter $\lambda = 0,5$

x_i	$S_n(x_i)$	$F_0(x_i) = 1 - e^{-0,5x_i}$	$F_0(x_i) - S_n(x_{i-1})$	$S_n(x_i) - F_0(x_i)$
0,14	0,0625	0,0676	0,0676	
0,27	0,1250	0,1263	0,0638	
0,43	0,1875	0,1935	0,0685	
0,68	0,2500	0,2882	0,1007	
0,81	0,3125	0,3330	0,0830	
1,14	0,3750	0,4345	0,1220	
1,45	0,4375	0,5157	0,1407	
1,82	0,5000	0,5975	0,1600	< 0
2,36	0,5625	0,6927	0,1927	
2,53	0,6250	0,7178	0,1553	
2,90	0,6875	0,7654	0,1404	
3,45	0,7500	0,8218	0,1343	
4,51	0,8125	0,8951	0,1451	
5,12	0,8750	0,9227	0,1102	
5,68	0,9375	0,9416	0,0666	
7,84	1,0000	0,9802	0,0427	0,0198

Der vierten Spalte aus Tabelle 36 entnehmen wir als größten Wert, vgl. auch die Ausführungen in Abschnitt 1.6, $D_n^- = 0,1927$ und der fünften Spalte $D_n^+ = 0,0198$. Somit ist $D_n = \max(D_n^+, D_n^-) = D_n^-$ und wir erhalten für die Teststatistik den Wert

$$\sqrt{n} D_n = \sqrt{16} \cdot 0,1927 = 0,7708,$$

der kleiner ist als der kritische Wert

$$d_{n; 1-\alpha} = d_{16, 0, 95} = 1,31,$$

so daß wir die Hypothese H_0 nicht verwerfen können.

Nun wollen wir noch die Durchführung des Tests über

H'_0 : Die Lebensdauer der Kühlaggregate ist exponentialverteilt
(mit unbekanntem Parameter λ)

demonstrieren. Hier müssen wir zunächst den Parameter λ aus der Stichprobe schätzen. Es ergibt sich hierfür der Wert

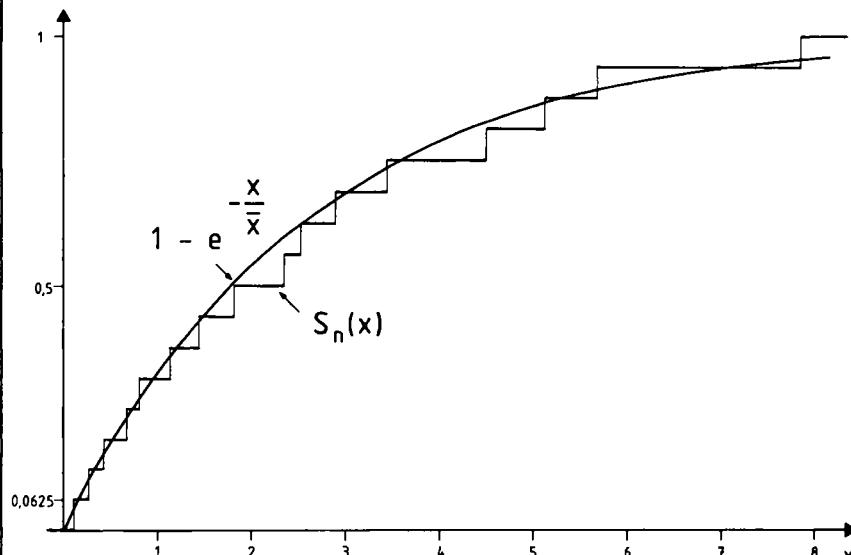
$$\frac{1}{\bar{x}} = \frac{1}{\frac{1}{16} \sum_{i=1}^{16} x_i} = 0,39.$$

Die benötigten Hilfsgrößen sind in **Tabelle 37** eingetragen und in **Abb. 19** ist der Sachverhalt veranschaulicht.

Wir entnehmen als größten Wert der letzten beiden Spalten in Tabelle 37 die Zahl 0,1016 und die Prüfgröße ergibt sich zu

Tab. 37: Arbeitstabelle zum Kolmogoroff-Smirnov-Test bei nicht spezifiziertem Parameter λ

x_i	$S_n(x_i)$	$1 - e^{-0,39x_i}$	$1 - e^{-0,39x_i} - S_n(x_{i-1})$	$S_n(x_i) - 1 + e^{-0,39x_i}$
0,14	0,0625	0,0531	0,0531	0,0094
0,27	0,1250	0,0999	0,0374	0,0251
0,43	0,1875	0,1544	0,0294	0,0331
0,68	0,2500	0,2329	0,0454	0,0171
0,81	0,3125	0,2709	0,0209	0,0416
1,14	0,3750	0,3589	0,0464	0,0161
1,45	0,4375	0,4319	0,0569	0,0056
1,82	0,5000	0,5083	0,0708	
2,36	0,5625	0,6016	0,1016	< 0
2,53	0,6250	0,6272	0,0647	
2,90	0,6875	0,6773	0,0523	0,0102
3,45	0,7500	0,7396	0,0521	0,0104
4,51	0,8125	0,8277	0,0777	< 0
5,12	0,8750	0,8642	0,0517	0,0108
5,68	0,9375	0,8909	0,0159	0,0466
7,84	1,0000	0,9530	0,0155	0,0470

**Abb. 19:** Empirische Verteilungsfunktion und $Ex\left(\frac{1}{\bar{x}}\right)$ -Verteilung zu den Daten im Beispiel

$$\sqrt{n} L_n^{\text{exp}} = \sqrt{16} \cdot 0,1016 = 0,4064.$$

Da nach Tab. 34 der kritische Wert gegeben ist als

$$I_{n;1-\alpha}^{\text{exp}} = I_{16;0,95}^{\text{exp}} = 1,05,$$

können wir die Hypothese H'_0 also nicht verwerfen und somit ist mit diesem Test

aus den Daten keine zum Niveau $\alpha = 0,05$ signifikante Abweichung von einer Exponentialverteilung zu erkennen.

4.2. Die Weibullverteilung

Die Weibullverteilung wurde erstmals zur Beschreibung von Materialermüdungserscheinungen, vgl. Weibull (1939), verwandt. Außerdem erwies sie sich als günstig bei der Beschreibung der Ausfälle von Elektronenröhren, Kugellagern usw.

Eine Zufallsvariable X heißt weibullverteilt mit den Parametern $\alpha > 0$ und $\beta > 0$, falls gilt

$$\alpha X^\beta \sim \text{Ex}(1);$$

dann ergibt sich die Verteilungsfunktion der Zufallsvariablen X zu

$$F_X(x) = 1 - e^{-\alpha x^\beta} \quad \text{für } x > 0$$

und ihre Dichte ist

$$f_X(x) = \alpha \beta x^{\beta-1} e^{-\alpha x^\beta} \quad \text{für } x > 0, \quad \text{vgl. auch Abb. 20.}$$

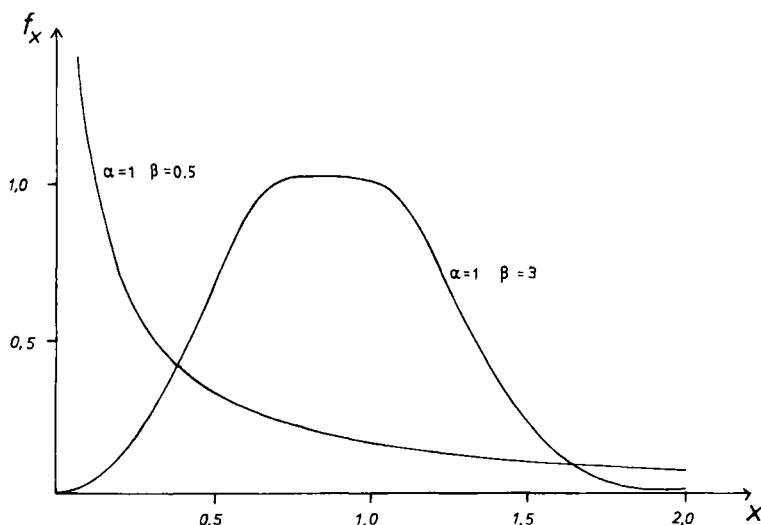


Abb. 20: Dichtefunktion der Weibullverteilung mit a) $\alpha = 1, \beta = 3$; b) $\alpha = 1, \beta = 0,5$

Ist die Lebensdauer X eines Objekts weibullverteilt mit Parametern α und β , so ist die Überlebenswahrscheinlichkeit des Objekts

$$P(X > x) = 1 - F_X(x) = \bar{F}_X(x) = e^{-\alpha x^\beta}$$

und die bedingte Überlebenswahrscheinlichkeit (= 1 – Verteilung der Restlebensdauer) eines Objekts, daß zum Zeitpunkt x_0 noch lebt, ist

$$\begin{aligned} \bar{F}(x_0 + x | x_0) &= P(X > x + x_0 | X > x_0) = \frac{\bar{F}_X(x_0 + x)}{\bar{F}_X(x_0)} \\ &= \frac{e^{-\alpha(x_0+x)^\beta}}{e^{-\alpha x_0^\beta}} = e^{-\alpha((x_0+x)^\beta - x_0^\beta)}. \end{aligned}$$

Erwartungswert und Varianz einer weibullverteilten (mit Parametern α, β) Zufallsvariablen X sind

$$EX = \alpha^{-1/\beta} \Gamma\left(\frac{1}{\beta} + 1\right) \text{ und } \text{Var } X = \alpha^{-2/\beta} \left(\Gamma\left(\frac{2}{\beta} + 1\right) - \left(\Gamma\left(\frac{1}{\beta} + 1\right) \right)^2 \right)$$

und der Median von X ist $\left(\frac{\ln 2}{\alpha}\right)^{1/\beta}$, wobei $\Gamma(z)$ die Gammafunktion, vgl. Abschnitt 1, bezeichnet.

Die Ausfallrate $r(t)$ der Weibull (α, β) -Verteilung ergibt sich zu

$$r(t) = \alpha \beta t^{\beta-1}.$$

Für $\beta > 1$ wächst die Ausfallrate also monoton (speziell für $\beta = 2$ ergibt sich eine Rayleigh-Verteilung mit Parameter $1/\alpha$), für $\beta = 1$ ist sie konstant (die Verteilung ist dann eine Exponentialverteilung mit Parameter α) und für $\beta < 1$ ist die Ausfallrate monoton fallend, vgl. hierzu auch Abb. 21.

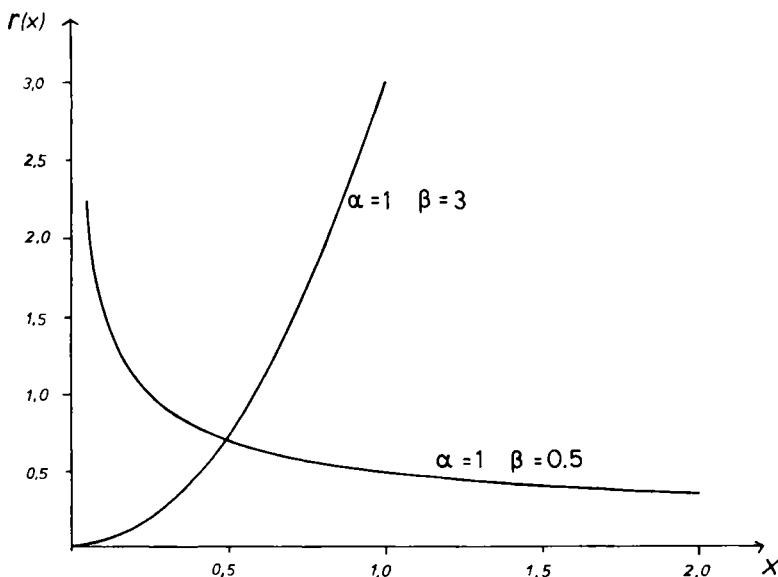


Abb. 21: Ausfallrate der Weibullverteilung mit a) $\alpha = 1, \beta = 3$; b) $\alpha = 1, \beta = 0,5$

Die Erzeugung weibullverteilter Zufallsvariablen zur Durchführung von Simulationsstudien ist recht einfach, wenn eine auf dem Intervall $[0; 1]$ gleichverteilte Zufallsvariable U zur Verfügung steht. Es ist nämlich gerade

$$X = \exp\left(\frac{1}{\beta} \ln\left(-\frac{1}{\alpha} \ln U\right)\right)$$

weibullverteilt mit den Parametern α und β .

4.2.1. Schätzen der Parameter α und β der Weibullverteilung

Die Schätzer nach der Maximum-Likelihood-Methode für die Parameter α und β können nur iterativ berechnet werden. Man bestimmt (z. B. mit Hilfe des Newton-Verfahrens oder seinen Modifikationen) die Nullstelle der Funktion

$$g(\beta) = \frac{\sum_{i=1}^k t_i^\beta \ln t_i + (n-k)t_s^\beta \ln t_s}{\sum_{i=1}^k t_i^\beta + (n-k)t_s^\beta} - \frac{1}{\beta} - \frac{1}{k} \sum_{i=1}^k \ln t_i.$$

Dabei ist n die Anzahl der Objekte mit weibullverteilter Lebensdauer T in der Stichprobe und t_i die Lebensdauer des i -ten Objekts. Beobachtet werden die n Objekte entweder in einem festen Zeitraum von 0 bis t_s , – dann ist k die Anzahl der in diesem Zeitintervall ausgefallenen Objekte –, oder bis zum Zeitpunkt $t_s = t_k$ des Ausfalls des k -ten Objektes; die Zahl k wird dann vor Beginn des Experiments festgelegt.

Ist

$$g(\hat{\beta}) = 0,$$

so ist $\hat{\beta}$ ein Schätzer für β und der Schätzer für den Parameter α wird als

$$\hat{\alpha} = k \left(\sum_{i=1}^k t_i^{\hat{\beta}} + (n-k)t_s^{\hat{\beta}} \right)^{-1}$$

bestimmt.

Schätzer für die Parameter α und β bei Prüfung bis zum Ausfall des letzten der n Objekte sind in Kapitel XIII zu finden.

4.3. Die IDB-Verteilung (Hjorth-Verteilung)

Mit der Exponential- und Weibull-Verteilung können, wie wir in Abschnitt 4.1 und 4.2 gesehen haben, konstante, fallende und steigende Ausfallraten modelliert werden.

Häufig hat jedoch die Ausfallrate einen badewannenförmigen Verlauf: Zunächst fallen viele Objekte aus (Kinderkrankheiten), dann ist die Ausfallrate eine gewisse Zeit recht konstant und schließlich mehren sich die Ausfälle aufgrund von Alterungserscheinungen.

Eine Verteilung, die auch – neben konstanten, fallenden und steigenden Ausfallraten – die Modellierung badewannenförmiger Ausfallraten erlaubt, ist die von Hjorth (1980) eingeführte IDB-Verteilung (Increasing, Decreasing, Bathtub-shaped). Die Verteilungsfunktion einer mit Parametern α , β und γ IDB-verteilten Zufallsvariablen X ist

$$F_X(t) = 1 - \frac{e^{-\alpha t^{2/2}}}{(1 + \beta t)^{\gamma/\beta}}, \quad \text{für } t \geq 0.$$

Die Dichte dieser Verteilung ist dann

$$f_X(t) = \frac{\alpha t(1 + \beta t) + \gamma}{(1 + \beta t)^{\gamma/\beta + 1}} e^{-\alpha t^{2/2}}, \quad \text{vgl. Abb. 22.}$$

Erwartungswert und Varianz der Verteilung von X lassen sich nicht direkt angeben

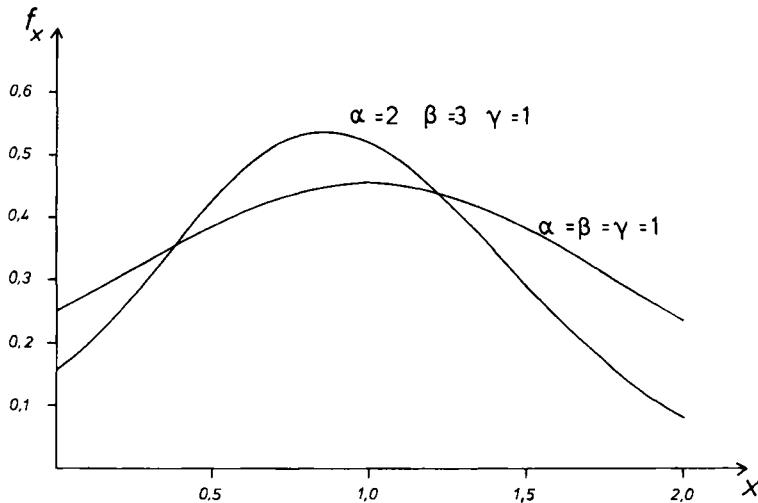


Abb. 22: Dichtefunktion der Hjorth-Verteilung mit a) $\alpha = \beta = \gamma = 1$; b) $\alpha = 2, \beta = 3, \gamma = 1$

und müssen numerisch bestimmt werden aus

$$\begin{aligned} E(X) &= \int_0^\infty \frac{e^{-\alpha t^2/2}}{(1 + \beta t)^{\gamma/\beta}} dt \quad \text{und} \\ \text{Var}(X) &= \frac{2}{\beta^2} \int_0^\infty \frac{t}{(1 + t)^{\gamma/\beta}} e^{-\alpha t^2/2} dt - \left(\frac{1}{\beta} \int_0^\infty (1 + t)^{-\gamma/\beta} e^{-\alpha t^2/2} dt \right)^2. \end{aligned}$$

Die Wahrscheinlichkeit, daß ein Objekt mit IDB(α, β, γ)-verteilter Lebensdauer X den Zeitpunkt t überlebt (Überlebenswahrscheinlichkeit) ist

$$P(X > t) = \bar{F}_X(t) = 1 - F_X(t) = \frac{e^{-\alpha t^2/2}}{(1 + \beta t)^{\gamma/\beta}} \quad \text{für } t \geq 0.$$

Hat das Objekt schon bis zum Zeitpunkt t_0 gelebt, so ist die (**bedingte**) Überlebenswahrscheinlichkeit durch

$$P(X > t_0 + t | X > t_0) = \frac{\bar{F}_X(t_0 + t)}{\bar{F}_X(t_0)} = \left(\frac{1 + \beta t_0}{1 + \beta t_0 + \beta t} \right)^{\gamma/\beta} \cdot e^{-\alpha((t_0 + t)^2 - t_0^2)/2}$$

gegeben. Die **Ausfallrate** der IDB(α, β, γ)-Verteilung schließlich ist

$$r(t) = \frac{f_X(t)}{\bar{F}_X(t)} = \alpha t + \frac{\gamma}{1 + \beta t}, \quad \text{vgl. Abb. 23.}$$

Damit ergeben sich als Spezialfälle der IDB(α, β, γ)-Verteilung für $\alpha = \beta = 0$ die Exponential-Verteilung mit Parameter γ und für $\gamma = 0$ die Rayleigh-Verteilung mit Parameter $1/2\alpha$ – also eine spezielle Weibullverteilung.

Die Ausfallrate ist monoton fallend für $\alpha = 0$, und **badewannenförmig** für $0 < \alpha < \beta\gamma$.

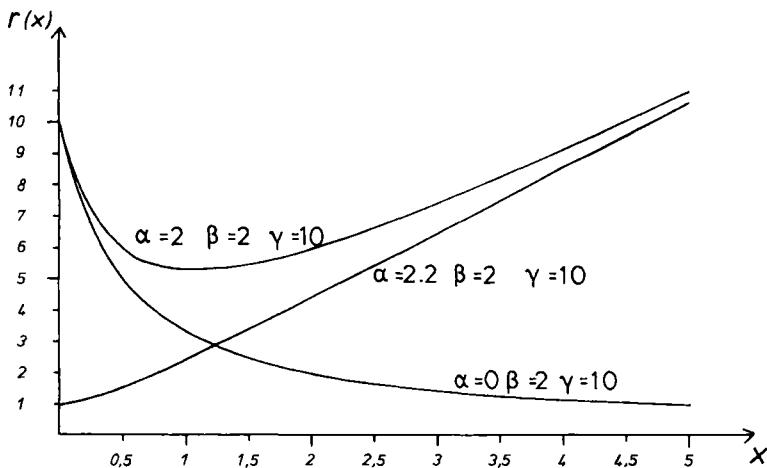


Abb.23: Ausfallrate der Hjorth-Verteilung mit a) $\alpha = 2,2, \beta = 2, \gamma = 10$; b) $\alpha = 0, \beta = 2, \gamma = 10$; c) $\alpha = 2, \beta = 2, \gamma = 10$

4.3.1. Schätzen der Parameter α , β und γ der IDB-Verteilung

Die Maximum-Likelihood-Schätzer für die Parameter α , β und γ einer IDB-Verteilung lassen sich nur numerisch bestimmen. Werden n Objekte mit $IDB(\alpha, \beta, \gamma)$ -verteilter Lebensdauer in einem Experiment beobachtet und bezeichnet man deren Lebensdauern, also die Zeit bis zu ihrem Ausfall, mit t_1, \dots, t_n , so werden α , β und γ durch Lösen des Systems

$$\begin{aligned} \sum_{i=1}^n t_i \left(\left(\alpha t_i + \frac{\gamma}{1 + \beta t_i} \right)^{-1} - \frac{1}{2} t_i \right) &= 0 \\ \sum_{i=1}^n \left[\left(\alpha t_i + \frac{\gamma}{1 + \beta t_i} \right)^{-1} \cdot (1 + \beta t_i)^{-1} - \frac{1}{\beta} \ln(1 + \beta t_i) \right] &= 0 \\ \sum_{i=1}^n \left(\frac{1}{\beta^2} \ln(1 + \beta t_i) - t_i [(\beta + \beta^2 t_i)^{-1} + \left(\alpha t_i + \frac{\gamma}{1 + \beta t_i} \right)^{-1} \cdot (1 + \beta t_i)^{-2}] \right) &= 0 \end{aligned}$$

bestimmt.

4.4. Die Erlang-n-Verteilung

Die **Erlang-n-Verteilung** spielt bei der Modellierung von Lebensdauern eine Rolle insofern, als die Summe von n unabhängig mit Parameter λ exponentialverteilten Zufallsvariablen Erlang-n-verteilt ist mit Parameter $\lambda > 0$ und Stufenzahl n . (n ist natürlich eine positive ganze Zahl.)

Besteht also ein System aus n Komponenten mit $Ex(\lambda)$ -verteilter Lebensdauer, von denen zu Beginn nur eine Komponente arbeitet, die zweite zu arbeiten anfängt, wenn die erste ausgefallen ist, usw., so ist die Lebensdauer des Systems Erlang-n-verteilt mit Parameter λ , falls es erst dann ausfällt, wenn auch die letzte der n Komponenten defekt ist, vgl. auch Kap. XIII.

Die Verteilungsfunktion einer mit Parameter λ Erlang-n-verteilten Zufallsvariablen X ist

$$F_X(x) = 1 - e^{-\lambda x} \sum_{i=1}^n \frac{(\lambda x)^{i-1}}{(i-1)!}, \quad \text{für } x \geq 0,$$

und die Dichte von X ist durch

$$f_X(x) = \lambda \cdot \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}$$

gegeben.

Natürlich ist die Zufallsvariable X gerade $E(X|\lambda)$ -verteilt, wenn $n = 1$ ist, d. h. wenn das System aus einem einzigen Element besteht.

Die Überlebenswahrscheinlichkeit, d. h. die Wahrscheinlichkeit, daß ein System mit Erlang-n(λ)-verteilter Lebensdauer einen Zeitpunkt x überlebt, ist

$$\bar{F}_X(x) = P(X > x) = 1 - F_X(x) = e^{-\lambda x} \sum_{i=1}^n \frac{(\lambda x)^{i-1}}{(i-1)!}, \quad \text{für } x \geq 0.$$

Ist das System zum Zeitpunkt x_0 noch nicht ausgefallen, so ist die bedingte Überlebenswahrscheinlichkeit, ($= 1 - V$ -Verteilung der Restlebensdauer), gerade

$$\bar{F}_X(x_0 + x|x_0) = \frac{\bar{F}_X(x_0 + x)}{\bar{F}_X(x_0)} = e^{-\lambda x} \frac{\sum_{i=1}^n (\lambda(x_0 + x))^{i-1}/(i-1)!}{\sum_{i=1}^n (\lambda x_0)^{i-1}/(i-1)!}.$$

Der Erwartungswert und die Varianz einer mit Parameter λ und Stufenzahl n Erlang-n-verteilten Zufallsvariablen X sind

$$E X = \frac{n}{\lambda} \quad \text{und} \quad \text{Var } X = \frac{n}{\lambda^2}$$

und die Ausfallrate der Verteilung der Zufallsvariablen X ist

$$r(x) = \frac{f_X(x)}{\bar{F}_X(x)} = \frac{\lambda(\lambda x)^{n-1}/(n-1)!}{\sum_{i=1}^n (\lambda x)^{i-1}/(i-1)!}.$$

5. Nichtparametrische Test- und Schätzmethoden im Ein-Stichproben-Fall

5.1. Konfidenzintervalle und Tests für Quantile

Wie wir in Kap. II gesehen haben, ist das α -Quantil einer theoretischen Verteilung mit Verteilungsfunktion $F(x)$ gerade definiert als

$$\xi_\alpha = F^{-1}(\alpha).$$

Wir wollen nun ein Konfidenzintervall für solche Quantile – ohne etwas über die konkrete Verteilung zu wissen – aus einer Beobachtungsreihe x_1, \dots, x_n angeben. Dann wollen wir aufgrund von Beobachtungen x_1, \dots, x_n Tests über die zugehöri-

gen α -Quantile (ohne Kenntnis der theoretischen Verteilung) angeben. Sowohl für die Bestimmung der Konfidenzintervalle als auch für die Durchführung von Tests muß die Zahl der Beobachtungen recht groß (Faustregel: $\alpha(1 - \alpha)n > 9$) sein. Bei kleinerer Zahl von Beobachtungen muß ein exaktes Verfahren benutzt werden, was hier nicht vorgestellt werden soll, vgl. aber David (1970).

5.1.1. Ein Konfidenzintervall für Quantile

Ordnen wir unsere Beobachtungen zunächst der Größe nach

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(r)} \leq \dots \leq x_{(s)} \leq \dots \leq x_{(n)},$$

so können wir ein $1 - \alpha$ -Konfidenzintervall für das p -Quantil ξ_p der zugehörigen Grundgesamtheit wie folgt bestimmen. Wir berechnen zunächst

$$r^* = np - u_{1-\alpha/2} \sqrt{np(1-p)} \quad \text{und}$$

$$s^* = np + u_{1-\alpha/2} \sqrt{np(1-p)}.$$

Dabei bezeichnet u_α das α -Quantil der $N(0, 1)$ -Verteilung. Die Werte r^* und s^* werden dann auf die jeweils nächstgrößere natürliche Zahl r bzw. s aufgerundet. Das gesuchte $1 - \alpha$ -Konfidenzintervall für ξ_p ist dann

$$[x_{(r)}; x_{(s)}].$$

Wollen wir etwa speziell ein Konfidenzintervall zum Niveau $1 - \alpha$ für den Median $\xi_{0,5}$ der Grundgesamtheit bestimmen, so sind r und s gerade durch Aufrunden der Werte

$$r^* = \frac{n}{2} - u_{1-\alpha/2} \sqrt{\frac{n}{4}} = \frac{1}{2}(n - u_{1-\alpha/2} \sqrt{n}) \quad \text{und}$$

$$s^* = \frac{n}{2} + u_{1-\alpha/2} \sqrt{\frac{n}{4}} = \frac{1}{2}(n + u_{1-\alpha/2} \sqrt{n})$$

gegeben.

Beispiel: Wir bestimmen ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ für das $0,2$ -Quantil $\xi_{0,2}$ und den Median $\xi_{0,5}$ der Verteilung der Größe von 3jährigen Kindern. Gezogen wird eine Stichprobe vom Umfang $n = 30$. Die Größen der 30 gemessenen Kinder sind (der Größe nach geordnet [in cm]) in Tab. 38 angegeben

Tab. 38: Größe der 30 Kinder

88,3	93,5	97,0	98,2	100,3
89,7	94,8	97,5	98,7	103,5
91,6	94,9	97,6	98,7	103,8
91,9	96,2	97,6	99,3	104,5
92,4	96,4	98,0	99,8	105,1
93,1	96,8	98,1	99,8	107,6

Damit ergibt sich das Konfidenzintervall für $\xi_{0,2}$ mit Hilfe von:

$$\begin{aligned} r^* &= np - u_{1-\alpha/2} \sqrt{np(1-p)} = 30 \cdot 0,2 - 1,96 \sqrt{30 \cdot 0,2 \cdot 0,8} \\ &= 6 - 1,96 \sqrt{4,8} = 6 - 4,29 = 1,71 \quad (\text{d. h. } r = 2) \text{ und} \end{aligned}$$

$$s^* = np + u_{1-\alpha/2} \sqrt{np(1-p)} = 6 + 4,29 = 10,29 \quad (\text{d. h. } s = 11) \text{ zu}$$

$$[x_{(r)}; x_{(s)}] = [x_{(2)}; x_{(11)}] = [89,7; 96,4].$$

Das Konfidenzintervall für den Median $\xi_{0,5}$ ergibt sich wegen

$$r^* = \frac{1}{2}(n - u_{1-\alpha/2} \sqrt{n}) = \frac{1}{2}(30 - 1,96 \sqrt{30}) = \frac{1}{2}(30 - 10,74) = 9,63$$

(d. h. $r = 10$) und

$$s^* = \frac{1}{2}(n + u_{1-\alpha/2} \sqrt{n}) = \frac{1}{2}(30 + 10,74) = 20,37 \quad (\text{d. h. } s = 21) \text{ zu}$$

$$[x_{(r)}; x_{(s)}] = [x_{(10)}; x_{(21)}] = [96,2; 98,7].$$

5.1.2. Tests für Quantile

Wollen wir nun testen, ob das p -Quantil ξ_p der unseren Beobachtungen zugrundeliegenden Verteilung in einem bestimmten Verhältnis zu einem vorgegebenen Wert z steht, so können wir folgende **Ordnungsstatistiken** zum Testen benutzen:

$$T_1(z) = |\{x_i | x_i < z\}| \quad \text{bzw.} \quad T_2(z) = |\{x_i | x_i \leq z\}|.$$

Die Werte, die bei Vorliegen einer konkreten Stichprobe von $T_1(z)$ bzw. $T_2(z)$ angenommen werden, können wir dann mit den Werten r^* bzw. s^* (vgl. Abschnitt 5.1.1) vergleichen, also $T_1(z)$ mit r^* und $T_2(z)$ mit s^* .

Die konkreten Entscheidungssituationen sind damit natürlich

$$\text{für } H_{01}: \xi_p \leq z \quad \text{gegen} \quad H_{11}: \xi_p > z:$$

Verwerfe die Nullhypothese zum Niveau $\alpha/2$, falls gilt: $T_1(z) < r^*$;

$$\text{für } H_{02}: \xi_p \geq z \quad \text{gegen} \quad H_{12}: \xi_p < z:$$

Verwerfe die Nullhypothese zum Niveau $\alpha/2$, falls gilt: $T_2(z) > s^*$ und

$$\text{für } H_{03}: \xi_p = z \quad \text{gegen} \quad H_{13}: \xi_p \neq z:$$

Verwerfe die Nullhypothese zum Niveau α , falls gilt: $T_1(z) < r^*$ oder $T_2(z) > s^*$.

Die zweiseitige Hypothese H_{03} wird also gerade dann verworfen, wenn z außerhalb des $1 - \alpha$ -Konfidenzintervalls für ξ_p liegt.

Beispiel: Kommen wir noch einmal auf unser Beispiel aus 5.1.1 zurück. Wir wollen zum Niveau $\alpha = 0,05$ die Hypothese

H_{03} : Der Median der Grundgesamtheit ist genau 1 m ($\xi_{0,5} = 100$)
gegen die Alternative

$$H_{13}: \text{Der Median ist nicht } 1 \text{ m } (\xi_{0,5} \neq 100)$$

testen. Die Größen r^* und s^* hatten wir schon berechnet:

$$r^* = 9,63, s^* = 20,37.$$

Wir benötigen nun noch die Werte der beiden Ordnungsstatistiken:

$$T_1(z) = |\{x_i | x_i < 100\}| = 24, T_2(z) = |\{x_i | x_i \leq 100\}| = 24$$

Da $T_2(z) > s^*$ ist, müssen wir die Hypothese zu diesem Niveau verwerfen.

5.2. Nichtparametrische Toleranzintervalle

Wir wollen in diesem Abschnitt die Bestimmung von approximativen Toleranzintervallen zur Sicherheitswahrscheinlichkeit $1 - \alpha$ mit minimaler Trefferwahrscheinlichkeit p aus n Beobachtungen x_1, \dots, x_n , die einer beliebigen Grundgesamtheit entstammen, behandeln.

Solche nichtparametrischen Toleranzintervalle beruhen auf den geordneten Beobachtungswerten

$$\dot{x}_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

Weiterhin definiert man

$$x_{(0)} = -\infty \quad \text{und} \quad x_{(n+1)} = \infty,$$

wenn man sich für einseitige Toleranzintervalle interessiert.

a) Wir suchen zu vorgegebenem α und p , sowie zwei selbstgewählten natürlichen Zahlen r und m den notwendigen Stichprobenumfang n . Man bestimmt diesen Stichprobenumfang n aus

$$n \approx \frac{1}{4} \chi^2_{2(r+m); 1-\alpha} \cdot \frac{1+p}{1-p} + \frac{1}{2}(r+m-1).$$

Das gesuchte Toleranzintervall ist dann

$$[x_{(r)}; x_{(n-m+1)}].$$

Praktiker wählen häufig $r = 1$ und $m = 1$, was die Berechnung erleichtert. Allerdings wird dieses Toleranzintervall unter Umständen sehr lang.

Beispiel: Ein Produkt, zu dem eine passende Verpackung hergestellt werden muß, variiert in der Länge. Um die Verpackung nun möglichst exakt passend herstellen zu können, muß ein Toleranzintervall für die Länge des Produkts bestimmt werden.

Wählen wir einmal $\alpha = 0,05$ und $p = 0,9$ sowie $m = 2$, $r = 1$, so ergibt sich die benötigte Beobachtungszahl n zu

$$\begin{aligned} n &\approx \frac{1}{4} \chi^2_{2(r+m); 1-\alpha} \cdot \frac{1+p}{1-p} + \frac{1}{2}(r+m-1) = \frac{1}{4} \chi^2_{6; 0,95} \frac{1,90}{0,10} + \frac{1}{2} \cdot 2 \\ &= \frac{1}{4} \cdot 12,592 \cdot 19 + 1 = 60,812. \end{aligned}$$

Um die gewünschten Sicherheiten p und $1 - \alpha$ einzuhalten, müssen wir also eine Stichprobe vom Umfang $n = 61$ ziehen. Das gesuchte Toleranzintervall ist dann gerade

$$[x_{(1)}; x_{(60)}].$$

Wollen wir nun ein einseitiges Toleranzintervall

$$[-\infty; x_{(n-m+1)}]$$

mit $\alpha = 0,1$ und $p = 0,7$ sowie $m = 3$ bestimmen, so ergibt sich (wegen $r = 0$)

$$n \approx \frac{1}{4} \chi^2_{2 \cdot 3; 0,9} \cdot \frac{1,7}{0,3} + \frac{1}{2} \cdot 2 = \frac{1}{4} \cdot 10,6 \cdot 5,667 + 1 = 16,08.$$

Wir müssen hier nun eine Stichprobe vom Umfang $n = 17$ ziehen.

b) Wir suchen zu vorgegebenem α , n , m und r die minimale Trefferwahrscheinlichkeit p .

Es ist hier dann

$$p \approx \frac{n - a - b}{n + a - b}$$

wobei gilt

$$a = \frac{1}{4} \chi^2_{2(r+m); 1-\alpha} \text{ und } b = \frac{1}{2}(r+m-1).$$

Beispiel: Wir wählen $\alpha = 0,1$, $n = 50$, $m = 5$ und $r = 7$. Dann ist die minimale Trefferwahrscheinlichkeit zum Sicherheitsniveau $1 - \alpha = 0,9$ des Toleranzintervalls

$$[x_{(7)}; x_{(46)}] \text{ gerade}$$

$$p = \frac{50 - a - b}{50 + a - b} = \frac{50 - 8,3 - 5,5}{50 + 8,3 - 5,5} = \frac{36,2}{52,8} = 0,686.$$

c) Wollen wir zu vorgegebenem p, α und n die Größen r und m (wir setzen hier einmal $r = m$) bestimmen, so können wir dies mit Hilfe der Ungleichung

$$4\left(n + \frac{1}{2} - r\right) \cdot \frac{1-p}{1+p} \geq \chi^2_{4r; 1-\alpha}$$

tun, indem wir die Tabelle der χ^2 -Verteilung zu Hilfe nehmen. Wir bestimmen das größte r , welches diese Ungleichung noch erfüllt.

Beispiel: Wir wählen $\alpha = 0,05$, $p = 0,7$ und $n = 50$. Wir müssen nun für verschiedene Werte r die Ungleichung überprüfen: Es ergibt sich für

$$r = 4: 32,824 > 26,3 = \chi^2_{16; 0,95}$$

$$r = 5: 32,118 > 31,4 = \chi^2_{20; 0,95}$$

$$r = 6: 31,412 \not> 36,4 = \chi^2_{24; 0,95}.$$

Das gesuchte Toleranzintervall ist also

$$[x_{(5)}; x_{(50-5+1)}] = [x_{(5)}; x_{(46)}].$$

d) Wollen wir nun zu vorgegebenem p, α und n ein linksseitiges Toleranzintervall bestimmen, ist also $r = 0$, so suchen wir das größte m , daß die Ungleichung

$$(4n - 2m + 2) \cdot \frac{1-p}{1+p} \geq \chi^2_{2m; 1-\alpha}$$

erfüllt.

Beispiel: Setzen wir einmal voraus, daß gelten soll $\alpha = 0,05$, $p = 0,7$ und $n = 50$, so ergibt sich für

$$m = 9: 32,471 > 28,9 = \chi^2_{18; 0,95}$$

$$m = 10: 32,118 > 31,4 = \chi^2_{20; 0,95}$$

$$m = 11: 31,765 \not> 33,9 = \chi^2_{22; 0,95}.$$

Das gesuchte einseitige Toleranzintervall ist also wegen $m = 10$ gerade:

$$[-\infty; x_{(50-10+1)}] = [-\infty; x_{(41)}].$$

e) Nun wollen wir noch zu einem Intervall mit vorgegebener minimaler Trefferwahrscheinlichkeit p und vorgegebenem Stichprobenumfang n die eingehaltene Sicherheitswahrscheinlichkeit $1 - \alpha$ bestimmen. Es ergibt sich hier gerade, daß man die Sicherheitswahrscheinlichkeit erhält, wenn man die Gleichung

$$\chi^2_{2(r+m); 1-\alpha} = (4n - 2(r+m-1)) \cdot \frac{1-p}{1+p}$$

in α löst.

Beispiel: Für das Toleranzintervall

$$[x_{(5)}; x_{(50-8+1)}] = [x_{(5)}; x_{(43)}]$$

mit minimaler Trefferwahrscheinlichkeit $p = 0,65$ und $n = 50$, $r = 5$ sowie $m = 8$ liegt die Sicherheitswahrscheinlichkeit $1 - \alpha$ wegen

$$\chi^2_{26; 1-\alpha} = (200 - 24) \cdot \frac{0,35}{1,65} = 37,33$$

zwischen 0,90 und 0,95, wie sich aus der χ^2 -Tafel im Anhang ergibt (genau: 37,33 ist das 0,93-Quantil der χ^2 -Verteilung mit 26 Freiheitsgraden).

5.3. Konfidenzstreifen für eine unbekannte Verteilungsfunktion

Ist das Merkmal einer Grundgesamtheit verteilt gemäß einer Verteilungsfunktion $F(x)$, die unbekannt ist, so läßt sich mit dem hier beschriebenen Verfahren ein Konfidenzstreifen angeben, der mit einer vorgegebenen Sicherheit $1 - \alpha$ die unbekannte Verteilungsfunktion $F(x)$ beinhaltet.

Hierzu beschafft man sich eine Stichprobe x_1, \dots, x_n aus der interessierenden Grundgesamtheit und erstellt für diese die empirische Verteilungsfunktion $S_n(x)$. Mit den kritischen Werten $d_{n; 1-\alpha}$ zum Kolmogoroff-Smirnov-Anpassungstest aus der Tabelle 14 in Abschnitt 1.6 bestimmen wir dann die obere Grenzfunktion

$$G_{ob}(x) = S_n(x) + \frac{1}{\sqrt{n}} d_{n; 1-\alpha}$$

und die untere Grenzfunktion

$$G_{unt}(x) = S_n(x) - \frac{1}{\sqrt{n}} d_{n; 1-\alpha}.$$

Mit Wahrscheinlichkeit $1 - \alpha$ gilt dann, daß die unbekannte Verteilungsfunktion $F(x)$ für alle x der Grundgesamtheit zwischen den Grenzfunktionen $G_{ob}(x)$ und $G_{unt}(x)$ und natürlich zwischen 0 und 1 liegt, auch wenn die Grenzfunktionen diesen Bereich überschreiten. Die Breite des Konfidenzstreifens

$$G_{ob} - G_{unt} = \frac{2}{\sqrt{n}} d_{n; 1-\alpha}$$

wird mit wachsendem Stichprobenumfang n immer geringer, da sich $d_{n; 1-\alpha}$ stabilisiert, und somit wird die wahre Verteilungsfunktion $F(x)$ zunehmend enger einge-

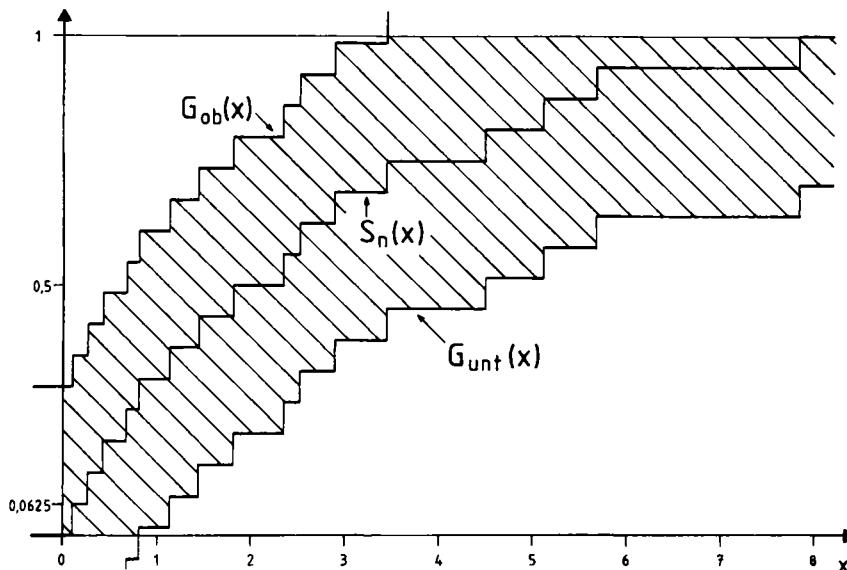


Abb. 24: Empirische Verteilungsfunktion und Konfidenzstreifen für die unbekannte Verteilungsfunktion in der Grundgesamtheit

schlossen. Ist die wahre Verteilungsfunktion $F(x)$ diskret, so ist dieser Konfidenzstreifen konservativ, d. h. breiter als nötig.

Beispiel: Aufgrund der $n = 16$ Beobachtungen aus Tabelle 35 in Abschnitt 4.1.4 wollen wir einen Konfidenzstreifen zum Niveau $1 - \alpha = 0,90$ für die unbekannte Verteilungsfunktion $F(x)$ der Lebensdauer einer Grundgesamtheit von Kühlaggregaten bestimmen. Dazu haben wir in Abb. 24 die empirische Verteilungsfunktion $S_n(x)$ der 16 Kühlagggregate eingezeichnet. Zudem finden sich in der Abbildung 24 die obere und untere Grenzfunktion $G_{ob}(x)$ und $G_{unt}(x)$. Es ist

$$\frac{1}{\sqrt{n}} \cdot d_{n; 1-\alpha} = \frac{1}{\sqrt{16}} \cdot 1,19 = 0,2975;$$

somit natürlich

$$G_{unt}(x) = S_n(x) - 0,2975 \text{ und } G_{ob}(x) = S_n(x) + 0,2975,$$

d. h. der Konfidenzstreifen, der in Abb. 24 schraffiert eingezeichnet ist, hat die Breite 0,5950. Da Lebensdauern natürlich stets positiv sind, konnten wir den Konfidenzstreifen hier auch bei $x = 0$ abschneiden.

Hätten wir eine größere Stichprobe von Kühlagggregaten, etwa die 87 Aggregate aus Tab. 32 in Abschnitt 4.1, so wäre der Konfidenzstreifen natürlich sehr viel schmäler geworden. Er hätte wegen

$$\frac{1}{\sqrt{n}} \cdot d_{87; 0,90} = \frac{1}{\sqrt{87}} \cdot 1,23 = 0,132$$

lediglich die Breite 0,264.

5.4. Nichtparametrische Einstichproben – Lokationsvergleiche und Tests auf Trend

In diesem Abschnitt wollen wir noch einige verteilungsfreie Tests im Einstichprobenproblem vorstellen: den Zeichentest, den Vorzeichenrangtest und zwei verschiedene Tests auf Trend. Bei allen diesen Tests wird vorausgesetzt, daß die Beobachtungen x_1, \dots, x_n , auf denen die Durchführung der Tests basiert, aus einer Grundgesamtheit mit stetiger Verteilung entstammen und unabhängig voneinander sind. (Der an der allgemeinen Theorie der nichtparametrischen Tests interessierte Leser sei auf Hajek/Sidak (1967), Wittig/Nölle (1970), Lehmann (1975) verwiesen.)

5.4.1. Der Zeichentest

Wir wollen uns zunächst mit dem Zeichentest, der aufgrund von n Beobachtungen x_1, \dots, x_n Hypothesen über einen Median $\xi_{0,5}$ der zugehörigen Grundgesamtheit testet, beschäftigen, da dieser wohl am leichtesten zu handhaben ist. Allerdings geht hier auch am meisten Information, die aufgrund der Beobachtungen vorliegt, verloren. Man will beim Zeichentest folgende Testprobleme betrachten:

$$H_{01}: \xi_{0,5} \leq \xi_0 \quad \text{gegen} \quad H_{11}: \xi_{0,5} > \xi_0,$$

$$H_{02}: \xi_{0,5} \geq \xi_0 \quad \text{gegen} \quad H_{12}: \xi_{0,5} < \xi_0$$

und

$$H_{03}: \xi_{0,5} = \xi_0 \quad \text{gegen} \quad H_{13}: \xi_{0,5} \neq \xi_0.$$

Bzgl. des Skalenniveaus der Grundgesamtheit wird hier lediglich vorausgesetzt, daß man feststellen kann, ob eine Beobachtung größer, kleiner oder gleich ξ_0 ist.

Berechnen wir zunächst die transformierten Beobachtungswerte

$$x'_i = x_i - \xi_0 \quad \text{für } i = 1, \dots, n$$

und setzen dann

$$y_i = \begin{cases} 1, & \text{falls } x'_i > 0 \\ 0, & \text{falls } x'_i \leq 0 \end{cases}$$

so ist $\sum_{i=1}^n y_i$ für $\xi_{0,5} = \xi_0$ gerade $B(n, 1/2)$ -verteilt.

Wir können also unsere Hypothesen mit Hilfe des in Abschnitt 3.1 vorgestellten Tests für die Binomialverteilung prüfen, d. h.: Ist n groß (etwa $n \geq 20$), so verwenden wir als Prüfgröße, die unter $\xi_{0,5} = \xi_0$ approximativ standardnormalverteilte Größe

$$Z = \frac{\sum_{i=1}^n y_i - n \cdot \frac{1}{2}}{\sqrt{n \cdot \frac{1}{2} \left(1 - \frac{1}{2}\right)}} = \frac{\sum_{i=1}^n y_i - \frac{n}{2}}{\frac{1}{2} \cdot \sqrt{n}}.$$

Sind Bindungen vorhanden, d. h. hier, es sind transformierte Beobachtungen gleich Null, so ersetzt man die y_i durch

$$\tilde{y}_i = \begin{cases} 1, & \text{falls } x'_i > 0 \\ \frac{1}{2}, & \text{falls } x'_i = 0 \\ 0, & \text{falls } x'_i < 0. \end{cases}$$

Ist n klein, so verwenden wir direkt die unter $\xi_{0,5} = \xi_0$ $B(n, 1/2)$ -verteilte Prüfgröße

$$T = \sum_{i=1}^n y_i.$$

Beispiel: Wir wollen mit Hilfe des Zeichentests Hypothesen über das Alter von Frauen bei der Geburt des ersten Kindes prüfen. Es werden zufällig $n = 36$ Mütter ausgewählt und nach diesem Merkmal befragt.

Wir testen dann zum Niveau $\alpha = 0,05$ die Hypothese

$$H_{01}: \xi_{0,5} \leq 25 \quad \text{gegen} \quad H_{11}: \xi_{0,5} > 25.$$

Die Ergebnisse unserer Befragung (und die transformierten Werte) sind in Tab. 39 eingetragen.

Da n hier recht groß ist, können wir die Approximation der Binomialverteilung durch die Normalverteilung benutzen. Wir berechnen also den Wert

$$Z = \frac{\sum_{i=1}^n y_i - \frac{1}{2} \cdot n}{\frac{1}{2} \cdot \sqrt{n}} = \frac{20 - 18}{\frac{1}{2} \cdot \sqrt{36}} = \frac{2}{3} = 0,667.$$

Dieser ist kleiner als das 0,95-Quantil

$$u_{1-\alpha} = u_{0,95} = 1,645$$

der Standardnormalverteilung, so daß wir die Nullhypothese „ $\xi_{0,5} \leq 25$ “ nicht verwerfen können.

5.4.2. Der Vorzeichenrangtest nach Wilcoxon

Im Gegensatz zum Zeichentest werden beim **Vorzeichenrangtest nach Wilcoxon (Wilcoxon-Signed-Rank-Test)** Hypothesen über das Symmetriezentrum einer Grundgesamtheit anhand von Rangzahlen getestet, die den Beobachtungen zugeordnet werden. Hier wird also schon wesentlich mehr Information aus den Beobachtungen gezogen. Stammen die Beobachtungen x_1, \dots, x_n , die unabhängig voneinander sein müssen, aus einer Gesamtheit mit Verteilungsfunktion F , so wird getestet, ob die Verteilung symmetrisch um einen Wert ζ_0 ist, d. h. ob für alle y gilt $F(\zeta_0 - y) = 1 - F(\zeta_0 + y)$.

Genau wie beim Zeichentest bilden wir zunächst die transformierten Beobachtungen

$$x'_i = x_i - \zeta_0.$$

(Ist $x'_i = 0$, so wird dieser Beobachtungswert bei der Berechnung der unten erwähnten Teststatistik nicht berücksichtigt.)

Ohne Berücksichtigung der Vorzeichen dieser Werte x'_1, \dots, x'_n vergeben wir

Tab. 39: Alter von Frauen bei der Geburt des ersten Kindes

i	Alter x_i	x_i'	y_i
1	30,6	5,6	1
2	17,8	-7,2	0
3	22,8	-2,2	0
4	29,7	4,7	1
5	40,1	15,1	1
6	33,5	8,5	1
7	21,6	-3,4	0
8	24,9	-0,1	0
9	25,1	0,1	1
10	31,4	6,4	1
11	37,8	12,8	1
12	18,3	-6,7	0
13	32,4	7,4	1
14	30,6	5,6	1
15	27,5	2,5	1
16	19,4	-5,6	0
17	23,7	-1,3	0
18	31,8	6,8	1
19	24,2	-0,8	0
20	20,7	-4,3	0
21	27,4	2,4	1
22	29,3	4,3	1
23	18,2	-6,8	0
24	40,1	15,1	1
25	30,8	5,8	1
26	32,7	7,7	1
27	25,2	0,2	1
28	24,3	-0,7	0
29	17,9	-7,1	0
30	20,8	-4,2	0
31	32,6	7,6	1
32	35,9	10,9	1
33	21,4	-3,6	0
34	36,1	11,1	1
35	20,0	-5,0	0
36	23,5	-1,5	0

dann Rangzahlen: 1 für den kleinsten Wert, ..., n für den größten Wert. Treten Bindungen auf, so werden die zugehörigen Rangzahlen gemittelt (**Midranks**); vgl. Beispiel.

Diesen Rangzahlen wird dann das Vorzeichen der zugehörigen Werte x'_1, \dots, x'_n zugewiesen; die so entstehenden Rangstatistiken werden mit \tilde{R}_i bezeichnet. Als Teststatistik für den Vorzeichenrangtest berechnen wir dann die Größe

$$T^+ = \sum (\text{positive Rangzahlen } \tilde{R}_i) = \sum_{i=1}^n c_i \tilde{R}_i \text{ mit } c_i = \begin{cases} 0, & \text{falls } \tilde{R}_i < 0 \\ 1, & \text{falls } \tilde{R}_i > 0 \end{cases}$$

und vergleichen diese mit den kritischen Werten $w_{n,y}$ aus Tabelle 40.

Tab. 40: Kritische Werte $w_{n,y}$ des Vorzeichenrangtests von Wilcoxon

n	$w_{n; 0,01}$	$w_{n; 0,025}$	$w_{n; 0,05}$	$w_{n; 0,10}$	$w_{n; 0,90}$	$w_{n; 0,95}$	$w_{n; 0,975}$	$w_{n; 0,99}$
4	0	0	0	1	8	9	10	10
5	0	0	1	3	11	13	14	14
6	0	1	3	4	16	17	19	20
7	1	3	4	6	21	23	24	26
8	2	4	6	9	26	29	31	33
9	4	6	9	11	33	35	38	40
10	6	9	11	15	39	43	45	48
11	8	11	14	18	47	51	54	57
12	10	14	18	22	55	59	62	66
13	13	18	22	27	63	68	72	77
14	16	22	26	32	72	78	82	88
15	20	26	31	37	82	88	93	99
16	24	30	36	43	92	99	105	111
17	28	35	42	49	103	110	117	124
18	33	41	48	56	114	122	129	137
19	38	47	54	63	126	135	142	151
20	44	53	61	70	139	148	156	165

Wollen wir nun zum Niveau α die Hypothese

$$H_{01}: F \text{ ist symmetrisch um } \zeta \leq \zeta_0$$

testen, so müssen wir H_{01} verwerfen, falls gilt

$$T^+ > w_{n; 1-\alpha};$$

falls wir

$$H_{02}: F \text{ ist symmetrisch um } \zeta \geq \zeta_0$$

zum Niveau α testen, verwerfen wir die Hypothese H_{02} im Falle

$$T^+ < w_{n; \alpha}.$$

Im zweiseitigen Testproblem zum Niveau α

$$H_{03}: F \text{ ist symmetrisch um } \zeta = \zeta_0$$

schließlich verwerfen wir die Hypothese, falls

$$T^+ > w_{n; 1-\alpha/2} \text{ oder } T^+ < w_{n; \alpha/2}$$

ist.

Für große Stichprobenumfänge n lässt sich eine Approximation durch die Standardnormalverteilung verwenden. Da der Erwartungswert und die Varianz der Teststatistik T^+ auf dem Rand der Hypothesen gerade

$$E T^+ = \frac{n(n+1)}{4}, \quad \text{Var } T^+ = \frac{n(n+1)(2n+1)}{24}$$

sind, ist die Statistik

$$T^* = \frac{T^+ - E T^+}{\sqrt{\text{Var } T^+}} = \frac{T^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

für große n approximativ standardnormalverteilt, so daß wir in solchen Fällen vom Gaußtest Gebrauch machen können. Diese Approximation gilt allerdings nur, falls in den x'_1, \dots, x'_n keine Bindungen auftreten. Treten in den x'_1, \dots, x'_n Bindungen auf, so ist die Approximation an die $N(0, 1)$ -Verteilung besser mit

$$T^* = \frac{T^+ - n(n+1)/4}{\sqrt{\frac{1}{24} \left[n(n+1)(2n+1) - \frac{1}{2} \sum_{j=1}^g t_j(t_j-1)(t_j+1) \right]}},$$

wobei g die Anzahl der gebundenen Gruppen und t_j die Anzahl der Elemente der j -ten gebundenen Gruppe ist, das heißt: Haben die x'_1, \dots, x'_n insgesamt g verschiedene Werte ($g = n$, falls keine Bindungen vorliegen), so nennt man alle x'_i , die jeweils den gleichen absoluten (hier dem Betrage nach, weil die Ränge auf die Beträge verteilt werden) Wert haben, eine gebundene Gruppe. Die Anzahl der x'_i in der j -ten Gruppe ist t_j . Ist $g = n$, treten also keine Bindungen auf, so ist $t_1 = t_2 = \dots = t_g = 1$.

In obigen Entscheidungsvorschriften ersetzt man dann die Größe T^+ durch T^* und $w_{n; \gamma}$ durch das γ -Quantil u_γ der Standardnormalverteilung. Man verwirft etwa H_{03} , wenn

$T^* > u_{1-\alpha/2}$ oder $T^* < u_{\alpha/2}$, d. h. wenn $|T^*| > u_{1-\alpha/2}$, ist.

Beispiel: In einem psychologischen Test werden verschiedene Fragen gestellt und die Antworten mit Punktzahlen bewertet. Wir wollen anhand von 11 Testergebnissen zum Niveau $\alpha = 0,05$ die Hypothese

H_{03} : Die Verteilungsfunktion F der Punktzahlen ist symmetrisch um $\zeta = 61$ testen. Die Ergebnisse x_1, \dots, x_{11} des Tests, die transformierten Werte x'_i sowie die zugehörigen Rangzahlen sind **Tabelle 41** zu entnehmen:

Tab. 41: Arbeitstabelle zum Vorzeichenrangtest von Wilcoxon im Beispiel

i	x_i	x'_i	$R_i = \tilde{R}_i $	\tilde{R}_i
1	72	11	$(10 + 11)/2 = 10,5$	10,5
2	55	-6	$(2 + 3 + 4)/3 = 3$	-3
3	67	6	$(2 + 3 + 4)/3 = 3$	3
4	53	-8	$(6 + 7 + 8)/3 = 7$	-7
5	69	8	$(6 + 7 + 8)/3 = 7$	7
6	71	10	9	9
7	55	-6	$(2 + 3 + 4)/3 = 3$	-3
8	68	7	5	5
9	65	4	1	1
10	72	11	$(10 + 11)/2 = 10,5$	10,5
11	69	8	$(6 + 7 + 8)/3 = 7$	7

Die x'_i zerfallen in folgende gebundene Gruppen, dem Betrage nach geordnet:

$x'_9, x'_2 x'_3 x'_7; x'_8; x'_4 x'_5 x'_{11}; x'_6; x'_1 x'_{10},$

so daß sich zum einen hieraus die Rangzahlen $|\tilde{R}_i|$ der Tabelle 41 ergeben; zum anderen ergibt sich, daß die Zahl der gebundenen Gruppen $g = 6$ ist mit den Gruppengrößen $t_1 = 1, t_2 = 3, t_3 = 1, t_4 = 3, t_5 = 1, t_6 = 2$. Aus Tabelle 41 erhält man

so als Wert für die Prüfgröße

$$\begin{aligned} T^+ &= \sum_{i=1}^{11} c_i \tilde{R}_i = 1 \cdot 10,5 + 0 \cdot (-3) + 1 \cdot 3 + 0 \cdot (-7) + 1 \cdot 7 + 1 \cdot 9 + 0 \cdot (-3) \\ &\quad + 1 \cdot 5 + 1 \cdot 1 + 1 \cdot 10,5 + 1 \cdot 7 = 53. \end{aligned}$$

Da nun mit den kritischen Werten aus Tabelle 40 gilt:

$$w_{n; \alpha/2} = w_{11; 0,025} = 11 \leq 53 = T^+ \leq 54 = w_{11; 0,975} = w_{n; 1 - \alpha/2},$$

können wir die Hypothese H_0 zum 5%-Niveau nicht verworfen. Allerdings müßte H_0 beim 10%-Niveau verworfen werden. Nun führen wir noch das approximative Verfahren unter Verwendung der Prüfgröße T^* vor. Es ist

$$\begin{aligned} T^* &= \frac{53 - 11 \cdot 12/4}{\sqrt{\frac{1}{24}[11 \cdot 12 \cdot 23 - \frac{1}{2}(1 \cdot 0 \cdot 2 + 3 \cdot 2 \cdot 4 + 1 \cdot 0 \cdot 2 + 3 \cdot 2 \cdot 4 + 1 \cdot 0 \cdot 2 + 2 \cdot 1 \cdot 3)]}} \\ &= \frac{20}{\sqrt{\frac{1}{24}[3036 - 27]}} = \frac{20}{11,20} = 1,79. \end{aligned}$$

Somit können wir auch bei Verwendung der Approximation wegen

$$|T^*| = 1,79 < 1,96 = u_{0,975} = u_{1 - \alpha/2}$$

die Nullhypothese H_0 nicht verworfen.

5.4.3. Tests auf Trend

A. Der Test von Cox und Stuart

Kommen wir nun zum **Test auf Trend von Cox und Stuart**. Liegt eine Beobachtungsreihe x_1, \dots, x_n vor, die zeitabhängig ist, so möchte man überprüfen, ob ein Trend in den Beobachtungen vorhanden ist. Wir bestimmen zunächst die Größe

$$m = \begin{cases} \frac{n}{2}, & \text{falls } n \text{ gerade} \\ \frac{n+1}{2}, & \text{falls } n \text{ ungerade} \end{cases}$$

und bilden dann die Differenzen

$$y_i = (x_{i+m} - x_i) \quad \text{für } i = \begin{cases} 1, \dots, m & \text{falls } n \text{ gerade} \\ 1, \dots, m-1 & \text{falls } n \text{ ungerade.} \end{cases}$$

Wählen wir dann als Teststatistik

$$T = \text{„Anzahl der positiven } y_i\text{“},$$

vernachlässigen alle y_i mit $y_i = 0$ und setzen

$$1 = \text{„Anzahl der } y_i \neq 0\text{“}, \text{ so können wir die Hypothesen}$$

- H_{01} : $P(y_i > 0) \leq P(y_i < 0)$ gegen H_{11} : $P(y_i > 0) > P(y_i < 0)$
 [\Leftrightarrow (kein Aufwärtstrend gegen Aufwärtstrend)],
- H_{02} : $P(y_i > 0) \geq P(y_i < 0)$ gegen H_{12} : $P(y_i > 0) < P(y_i < 0)$
 [\Leftrightarrow (kein Abwärtstrend gegen Abwärtstrend)] und
- H_{03} : $P(y_i > 0) = P(y_i < 0)$ gegen H_{13} : $P(y_i > 0) \neq P(y_i < 0)$
 [\Leftrightarrow (kein Trend gegen Trend)]

mit Hilfe des Zeichentests für 1 Beobachtungen überprüfen (vgl. Abschnitt 5.4.1).

Beispiel: Zwei Jahre lang werden jeden Monat die Regenfälle in einem bestimmten Gebiet gemessen. Die 24 beobachteten Werte [in mm] sind in **Tabelle 42** zusammengefaßt.

Tab. 42: Monatliche Regenfälle

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	52,4	37,4	41,2	71,5	51,3	21,6	17,7	21,2	41,3	32,0	53,0	61,4

i	13	14	15	16	17	18	19	20	21	22	23	24
x_i	47,8	42,0	33,2	41,0	29,5	28,1	17,4	21,5	47,4	51,3	49,5	53,7

Wir wollen nun mit Hilfe des Tests auf Trend von Cox und Stuart überprüfen, ob die Regenfälle in beiden Jahren einen zum 10%-Niveau signifikanten Trend aufweisen, d.h. wir testen

$$H_{03}: \text{kein Trend} \quad \text{gegen} \quad H_{13}: \text{Trend}.$$

Dazu bestimmen wir zunächst

$$m = \frac{n}{2} = 12$$

und die Differenzen y_i für $i = 1, \dots, 12$, die in **Tab. 43** zu finden sind.

Tab. 43: Differenzen y_i im Beispiel

i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	-4,6	4,6	-8	-30,5	-21,8	6,5	-0,3	0,3	6,1	19,3	-3,5	-7,7

Hieraus ergibt sich

$$l = m = 12 \quad \text{und} \quad T = 5.$$

Nach den Entscheidungsregeln des Zeichentests müssen wir die Hypothese verwerfen, falls gilt:

$$T < r^* \quad \text{oder} \quad T > 1 - r^*$$

$$\text{mit} \quad r^* = \frac{1}{2} (1 - u_{1-\alpha/2} \sqrt{12}) = \frac{1}{2} (12 - 1,645 \sqrt{12}) = 3,151.$$

Die Hypothese „es besteht kein Trend“ kann also nicht verworfen werden.

B. Der Test nach Mann

Kommen wir noch zu einem anderen **Einstichproben-Trendtest**, der auf **Mann** zurückgeht. Wir beobachten zu n aufeinander folgenden Zeitpunkten x_1, \dots, x_n gewisse Merkmalswerte y_1, \dots, y_n . Der **Mann-Test** überprüft dann anhand der Beobachtungen y_1, \dots, y_n **Hypothesen über einen Trend** in der Beobachtungsreihe (**Zeitreihe**) mittels der Teststatistik

$$C = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(y_j - y_i) = \sum_{i=1}^n \Sigma_i,$$

wobei sgn für die Vorzeichenfunktion „Signum“ steht:

$$\text{sgn}(y_j - y_i) = \begin{cases} 1, & > 0 \\ 0, & \text{falls } y_j - y_i = 0 \\ -1, & < 0 \end{cases}$$

Wollen wir die Hypothese

$$H_{01}: \text{kein Aufwärtstrend} \quad \text{gegen} \quad H_{11}: \text{Aufwärtstrend}$$

zum Niveau α testen, so müssen wir H_{01} verwerfen, falls gilt

$$C > K_{n; 1-\alpha}.$$

Dabei bezeichnet $K_{n; 1-\alpha}$ das $1 - \alpha$ -Quantil der Kendallschen K-Statistik, die z. B. bei Hollander/Wolfe (1973) ausführlich vertafelt ist. Einige kritische Werte sind in **Tabelle 44** angegeben; ansonsten kann man untenstehende Approximation verwenden.

Tab. 44: Kritische Werte $K_{n; 1-\alpha}$ der Kendallschen K-Statistik

n	$K_{n; 0,995}$	$K_{n; 0,99}$	$K_{n; 0,975}$	$K_{n; 0,95}$	$K_{n; 0,90}$
4	6	6	6	6	5
5	10	9	8	8	7
6	14	13	11	9	7
7	17	15	13	11	9
8	21	19	16	14	11
9	25	23	19	16	13
10	29	26	23	19	15
15	52	47	40	34	27
20	80	72	59	49	38
25	110	100	84	71	56
30	145	131	111	93	73
35	183	165	139	117	88
40	222	200	169	142	111

Testen wir zum Niveau α die Nullhypothese

$H_{02}: \text{kein Abwärtstrend} \quad \text{gegen} \quad H_{12}: \text{Abwärtstrend},$
so verwerfen wir die Nullhypothese, falls

$$C < -K_{n; 1-\alpha}.$$

Das zweiseitige Testproblem

$$H_{03}: \text{kein Trend} \quad \text{gegen} \quad H_{13}: \text{Trend}$$

führt zur Ablehnung der Nullhypothese zum Niveau α , falls gilt

$$C < -K_{n; 1-\alpha/2} \quad \text{oder} \quad C > K_{n; 1-\alpha/2}.$$

Ist die Anzahl der Beobachtungen „groß“, so ist die transformierte Teststatistik

$$C^* = \frac{C}{\sqrt{n(n-1)(2n+5)/18}}$$

auf dem Rand der Nullhypotesen approximativ $N(0, 1)$ -verteilt, so daß wir dann zur Entscheidung unserer Testprobleme nicht mehr C mit $K_{n; 1-\alpha}$ vergleichen, sondern C^* mit dem Quantil $u_{1-\alpha}$ der Standardnormalverteilung.

Beispiel: (vgl. Vital Statistics – Special Reports 47 (8) (1960), “Illegitimate births in the United States 1938–1957”.)

Im Staat Kansas wurde in verschiedenen Jahren die Anzahl der unehelich geborenen Kinder festgestellt, vgl. **Tabelle 45**.

Tab. 45: Anzahl unehelicher Geburten im US-Staat Kansas

i	1	2	3	4	5
Jahr x_i	1940	1945	1947	1950	1955
y_i (uneheliche Kinder)	318	593	564	530	694

Wir wollen mit Hilfe des Trend-Tests von Mann überprüfen, ob die Anzahl der unehelich geborenen Kinder während dieser Zeit signifikant angestiegen ist, d.h. wir überprüfen zum Niveau $\alpha = 0,05$ die Nullhypothese

$$H_0: \text{kein Aufwärtstrend} \quad \text{gegen} \quad H_{11}: \text{Aufwärtstrend}$$

Berechnen wir zunächst den Wert der Teststatistiken C und C^* , vgl. **Tabelle 46**.

Tab. 46: Arbeitstabelle zum Beispiel des Tests auf Trend nach Mann

(i, j)	$y_j - y_i$	$\text{sgn}(y_j - y_i)$
(1, 2)	275	1
(1, 3)	246	1
(1, 4)	212	1
(1, 5)	376	1
(2, 3)	-29	-1
(2, 4)	-63	-1
(2, 5)	101	1
(3, 4)	-34	-1
(3, 5)	130	1
(4, 5)	164	1

Damit ergibt sich

$$C = \sum_{i=1}^4 \sum_{j=i+1}^5 \text{sgn}(y_j - y_i) = 4$$

bzw.

$$C^* = \frac{4}{\sqrt{5 \cdot 4 \cdot 15/18}} = \frac{4}{4,08} = 0,98.$$

Das 0,95-Quantil der Kendallschen K-Statistik ist $K_{5,0,95} = 8$, so daß wir die Nullhypothese nicht verwerfen können. Obwohl nur 5 Beobachtungen vorliegen und die Approximation durch die $N(0, 1)$ -Verteilung somit nicht sehr gut sein wird, wollen wir auch diese Vorgehensweise vorführen.

Es ist $u_{1-\alpha} = u_{0,95} = 1,645$ und somit

$$C^* > u_{1-\alpha},$$

d. h. die Hypothese H_0 wird nicht verworfen.

6. Sequentielle Quotiententests

Möchte man bei der Anwendung von Tests, die in den vorangegangenen Abschnitten besprochen wurden, bei Einhaltung eines vorgegebenen Niveaus α , und damit **Kontrolle des Fehlers 1. Art**, auch noch den **Fehler 2. Art** – zumindest für spezielle Alternativen – begrenzen, so hat man dafür Sorge zu tragen, daß der **Stichprobenumfang** einen – durch die gestellten Anforderungen festgelegten – Wert nicht unterschreitet (vgl. Abschnitt 1 und 3). Legt man den Stichprobenumfang nun nicht von vornherein fest, sondern überlegt sich nach jeder Prüfung eines Stichprobenelementes, ob die bis dahin vorliegenden Daten ausreichen, um eine genügend fundierte Entscheidung zwischen der Hypothese und der Alternative zu treffen, oder ob erst noch weitere Daten erhoben werden müssen, so kann man durch eine geeignete Vorgehensweise – bei gleichen Güteeigenschaften – im Durchschnitt mit wesentlich geringeren Stichprobenumfängen auskommen, als wenn man schon vor der Gewinnung der Daten deren benötigte Anzahl endgültig festgelegt hätte. Diese sogenannten **sequentiellen Tests** sind Gegenstand dieses Abschnittes. Zunächst wird ein allgemeines Verfahren zur Konstruktion sequentieller Tests mit vorgegebenen Güteeigenschaften vorgestellt. Es folgen dann konkrete Anwendungen dieses Verfahrens zum Testen der Parameter der Binomial-, Normal- und Exponentialverteilung.

Die diskrete Zufallsvariable X nehme den Wert x mit der Wahrscheinlichkeit $P_\theta(X = x)$ an. Es soll

$$H_0: \theta = \theta_1 \quad \text{gegen} \quad H_1: \theta = \theta_2$$

getestet werden.

Der Fehler 1. Art, also die Entscheidung für H_1 obwohl H_0 richtig ist, soll nur mit Wahrscheinlichkeit α auftreten, was gleichbedeutend ist mit einem Testniveau α , und der Fehler 2. Art, d. h. die Entscheidung für H_0 bei Vorliegen von H_1 , soll höchstens mit der Wahrscheinlichkeit β auftreten. Bezeichnen nun x_1, \dots, x_k die Ergebnisse der ersten k Versuche, so ist die Wahrscheinlichkeit für das Auftreten gerade dieser Werte bei Gültigkeit der Hypothese H_0 gegeben durch

$$P_{\theta_1}(X = x_1) \cdot \dots \cdot P_{\theta_1}(X = x_k)$$

und berechnet sich unter der Alternative H_1 zu

$$P_{\theta_2}(X = x_1) \cdot \dots \cdot P_{\theta_2}(X = x_k).$$

Der sogenannte **sequentielle Quotiententest (sequential probability ratio test =**

SPRT) beruht nun auf dem Vergleich dieser beiden Wahrscheinlichkeiten. Ist die Wahrscheinlichkeit dafür, daß die beobachtete Stichprobe bei Vorliegen von H_0 vorliegt, groß im Vergleich zu der, die sich bei der Gültigkeit von H_1 ergibt, so entscheidet man für H_0 , während man in der umgekehrten Situation für H_1 entscheidet. Sind dagegen die beiden Wahrscheinlichkeiten „ungefähr gleich“, so muß noch weiter beobachtet werden. Mit

$$T_k(x_1, \dots, x_k) = \frac{P_{\theta_2}(X = x_1) \cdot \dots \cdot P_{\theta_2}(X = x_k)}{P_{\theta_1}(X = x_1) \cdot \dots \cdot P_{\theta_1}(X = x_k)}$$

lässt sich das wie folgt formulieren.

Gebe zwei Zahlen A und B vor mit $A < 1 < B$ und beobachte solange, wie

$A < T_k < B$ gilt;

sobald diese Bedingung nicht mehr erfüllt ist, entscheide für die Hypothese H_0 oder die Alternative H_1 und zwar für H_0 , falls $T_k \leq A$ und für H_1 , falls $T_k \geq B$. Soll der so durchgeführte Test die angegebenen Fehlerwahrscheinlichkeiten einhalten, so muß man

$$A = \frac{\beta}{1 - \alpha} \quad \text{und} \quad B = \frac{1 - \beta}{\alpha}$$

wählen.

Ist X nicht wie oben eine diskrete, sondern eine stetige Zufallsvariable mit der Wahrscheinlichkeitsdichte f_{θ_1} unter der Hypothese H_0 und f_{θ_2} unter der Alternative H_1 , so setzt man

$$T_k(x_1, \dots, x_k) = \frac{f_{\theta_2}(x_1) \cdot \dots \cdot f_{\theta_2}(x_k)}{f_{\theta_1}(x_1) \cdot \dots \cdot f_{\theta_1}(x_k)}$$

und behält die obige Vorgehensweise bei. Man entscheidet also für H_0 , falls $T_k \leq A$, für H_1 , falls $B \leq T_k$ und beobachtet weiter, solange $A < T_k < B$ gilt. Dabei ist wieder

$$A = \frac{\beta}{1 - \alpha} \quad \text{und} \quad B = \frac{1 - \beta}{\alpha}$$

zu setzen, um die geforderten Güteigenschaften zu erhalten.

6.1. Der sequentielle Quotiententest für die Binomialverteilung

Ein Experiment, das nur die beiden Ausgänge „Erfolg“ und „Mißerfolg“ haben kann, läßt sich durch eine Zufallsvariable X, die die Werte Null und Eins annimmt, beschreiben; dabei soll die Eins für „Erfolg“ und die Null für „Mißerfolg“ stehen. Zunächst nehmen wir an, daß für die Erfolgswahrscheinlichkeit p nur die beiden Werte p_1 und p_2 mit $p_1 < p_2$ in Frage kommen und betrachten das Testproblem

$$H_0: p = p_1 \quad \text{gegen} \quad H_1: p = p_2.$$

Die Entscheidung für H_0 oder H_1 soll mit Hilfe eines sequentiellen Quotiententests getroffen werden. Dazu berechnen wir zuerst die Statistik T_k . Wir haben

$$P_{p_1}(X = x_i) = p_1^{x_i} (1 - p_1)^{1 - x_i} = \begin{cases} p_1 & , \text{ falls } x_i = 1 \\ 1 - p_1 & , \text{ falls } x_i = 0 \end{cases}$$

und $P_{p_2}(X = x_i) = p_2^{x_i} (1 - p_2)^{1 - x_i} = \begin{cases} p_2 & , \text{ falls } x_i = 1 \\ 1 - p_2 & , \text{ falls } x_i = 0. \end{cases}$

Damit erhält man für

$$\begin{aligned} T_k(x_1, \dots, x_k) &= \frac{p_2^{x_1} (1 - p_2)^{1 - x_1} \cdot \dots \cdot p_2^{x_k} (1 - p_2)^{1 - x_k}}{p_1^{x_1} (1 - p_1)^{1 - x_1} \cdot \dots \cdot p_1^{x_k} (1 - p_1)^{1 - x_k}} \\ &= \left(\frac{p_2}{p_1}\right)^{x_1 + \dots + x_k} \cdot \left(\frac{1 - p_2}{1 - p_1}\right)^{k - x_1 - \dots - x_k}. \end{aligned}$$

Um das Testverfahren einfacher handhaben zu können, ersetzen wir die in jedem Schritt zu überprüfenden Ungleichungen

$$T_k \leq A, A < T_k < B, B \leq T_k$$

durch gleichwertige, aber leichter zu verifizierende Ungleichungen. So ist

$$A < T_k < B \quad \text{gleichbedeutend mit}$$

$$\ln A < \ln T_k < \ln B \quad \text{und wegen}$$

$$\ln T_k = m_k \ln \frac{p_2}{p_1} + (k - m_k) \ln \frac{1 - p_2}{1 - p_1},$$

wobei $m_k = x_1 + \dots + x_k$ die Anzahl der „Erfolge“ in den ersten k Versuchen ist (m_k ist damit binomialverteilt mit den Parametern k und p), auch äquivalent zu

$$\frac{\ln \frac{1 - p_1}{1 - p_2}}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}} \cdot k + \frac{\ln A}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}} < m_k < \frac{\ln \frac{1 - p_1}{1 - p_2}}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}} \cdot k + \frac{\ln B}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}}.$$

Mit den Bezeichnungen

$$a = \frac{-\ln A}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}} = \frac{-\ln \frac{\beta}{1 - \alpha}}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}},$$

$$b = \frac{\ln B}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}} = \frac{\ln \frac{1 - \beta}{\alpha}}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}} (= a \quad \text{für } \alpha = \beta) \text{ und}$$

$$c = \frac{\ln \frac{1 - p_1}{1 - p_2}}{\ln \frac{p_2}{p_1} \cdot \frac{1 - p_1}{1 - p_2}}$$

lässt sich das Testverfahren nun wie folgt beschreiben. Man zählt jeweils die Anzahl m_k der Erfolge in den ersten k Versuchen und entscheidet für H_0 , wenn zum ersten Mal gilt:

$$m_k \leq ck - a;$$

ist zum ersten Mal

$$m_k \geq ck + b,$$

so wird H_1 angenommen, und solange

$$ck - a < m_k < ck + b$$

gilt, beobachtet man weiter.

Wir haben diesen Test hergeleitet zum Prüfen der Hypothese $H_0: p = p_1$ gegen die Alternative $H_1: p = p_2$. Meistens liegt jedoch das folgende Testproblem vor:

$$H_0: p \leq p_1 \quad \text{gegen} \quad H_1: p > p_1.$$

Wenden wir das obige Testverfahren in dieser Situation an, so wird das vorgegebene Niveau α eingehalten, und außerdem ist die Wahrscheinlichkeit für den Fehler 2. Art durch den Wert β beschränkt, wenn die wahre Erfolgswahrscheinlichkeit p größer oder gleich p_2 ist. Bezeichnen wir die Wahrscheinlichkeit die Nullhypothese anzunehmen bei Vorliegen von p , also die Operationscharakteristik an der Stelle p (vgl. Kap. VI), mit $L(p)$, ($1 - L(p)$ ist dann die Gütfunktion), so gilt also

$$1 - L(p) \leq \alpha, \quad \text{für } p \leq p_1$$

$$\text{und} \quad L(p) \leq \beta, \quad \text{für } p \geq p_2.$$

Genauer ist

$$L = L(p) = \frac{1 - B^h}{A^h - B^h} = \frac{1 - \left(\frac{1 - \beta}{\alpha}\right)^h}{\left(\frac{\beta}{1 - \alpha}\right)^h - \left(\frac{1 - \beta}{\alpha}\right)^h},$$

wobei h durch die Beziehung

$$p = \frac{1 - \left(\frac{1 - p_2}{1 - p_1}\right)^h}{\left(\frac{p_2}{p_1}\right)^h - \left(\frac{1 - p_2}{1 - p_1}\right)^h}$$

festgelegt wird.

Den Verlauf der Funktion L in Abhängigkeit von p erhält man einfach, indem man zu vorgegebenen Werten von h die Wertepaare $(p, L(p))$ berechnet und in einem Koordinatensystem darstellt. Dabei ist für $h = 1$:

$$p = p_1 \quad \text{und} \quad L = 1 - \alpha$$

während sich für $h = -1$ die Werte

$$p = p_2 \quad \text{und} \quad L = \beta$$

ergeben. Mit dieser Funktion L lässt sich nun in Abhängigkeit von p der zu erwartende Stichprobenumfang \hat{n} berechnen zu

$$\hat{n} = \frac{L \ln A + (1-L) \ln B}{p \ln \frac{p_2}{p_1} + (1-p) \ln \frac{1-p_2}{1-p_1}} = \frac{L \ln \frac{\beta}{1-\alpha} + (1-L) \ln \frac{1-\beta}{\alpha}}{p \ln \frac{p_2}{p_1} + (1-p) \ln \frac{1-p_2}{1-p_1}}$$

$$= \left[\frac{(1-2L) \ln \frac{1-\beta}{\alpha}}{p \ln \frac{p_2}{p_1} + (1-p) \ln \frac{1-p_2}{1-p_1}} \quad \text{für } \alpha = \beta \right], \quad \text{falls } p \neq c \text{ ist.}$$

Für $p = p_1$ ergibt sich

$$\hat{n} = \frac{(1-\alpha) \ln \frac{\beta}{1-\alpha} + \alpha \ln \frac{1-\beta}{\alpha}}{p_1 \ln \frac{p_2}{p_1} + (1-p_1) \ln \frac{1-p_2}{1-p_1}} \quad \text{und für } p = p_2$$

$$\hat{n} = \frac{\beta \ln \frac{\beta}{1-\alpha} + (1-\beta) \ln \frac{1-\beta}{\alpha}}{p_2 \ln \frac{p_2}{p_1} + (1-p_2) \ln \frac{1-p_2}{1-p_1}}.$$

Ist $p = c = \frac{\ln \frac{1-p_1}{1-p_2}}{\ln \frac{p_2}{p_1} \frac{1-p_1}{1-p_2}}$,

so wird \hat{n} maximal, und es ergibt sich

$$\hat{n} = \frac{-\ln \frac{\beta}{1-\alpha} \cdot \ln \frac{1-\beta}{\alpha}}{\ln \frac{1-p_1}{1-p_2} \cdot \ln \frac{p_2}{p_1}} \left(= \frac{\left(\ln \frac{1-\beta}{\alpha} \right)^2}{\ln \frac{1-p_1}{1-p_2} \cdot \ln \frac{p_2}{p_1}} \quad \text{für } \alpha = \beta \right).$$

Kennt man a, b und c , so läßt sich \hat{n} wie folgt bestimmen:

$$\hat{n} = \begin{cases} \frac{ab}{c(1-c)}, & \text{für } p = c \\ \frac{-La + (1-L)b}{p(1-c) - (1-p)c} = \frac{b - (a+b)L}{p-c}, & \text{für } p \neq c. \end{cases}$$

Beispiel: Zwei Medikamente I und II sollen miteinander verglichen werden. Man weiß, daß das Mittel I, das schon längere Zeit auf dem Markt ist, eine Erfolgswahrscheinlichkeit $p_1 = 70\%$ hat. Das neu entwickelte Mittel II, dessen Erfolgswahrscheinlichkeit p noch unbekannt ist, soll, wenn es besser ist als das alte, dieses ersetzen. Wir haben also

$$H_0: p \leq p_1 \quad \text{gegen} \quad H_1: p > p_1$$

zu testen.

Die Einhaltung eines Niveaus α garantiert uns, daß die (falsche) Entscheidung, H_1 anzunehmen, obwohl H_0 richtig ist, also fälschlicherweise die Überlegenheit des neuen Medikamentes anzunehmen, höchstens mit der Wahrscheinlichkeit α , z. B. $\alpha = 1\%$, getroffen wird. Andererseits möchten wir eine deutliche Verbesserung, etwa eine Erhöhung der Erfolgswahrscheinlichkeit um 10%, auch wirklich entdecken. Die Wahrscheinlichkeit H_0 anzunehmen, obwohl die Erfolgswahrscheinlichkeit p des Mittels II um mindestens 10% größer ist als die des Mittels I, also $p \geq p_1 + 10\% = 80\% = p_2$, darf deshalb nicht zu groß werden, d. h. einen vorgegebenen Wert β , z. B. $\beta = 1\%$, nicht überschreiten.

Um den zu erwartenden Stichprobenumfang n möglichst klein zu halten, benutzen wir das oben beschriebene sequentielle Testverfahren. Das Medikament II wird demnach Probanden verabreicht, und es wird festgestellt, ob die Behandlung erfolgreich gewesen ist oder nicht. Bezeichnet dann m_k die Anzahl der Erfolge bei den ersten k Versuchspersonen, so entscheidet man

$$\text{für } H_0, \text{ falls } m_k \leq ck - a,$$

$$\text{für } H_1, \text{ falls } m_k \geq ck + b$$

und behandelt eine weitere Person, wenn

$$ck - a < m_k < ck + b.$$

Dabei müssen die Konstanten a , b und c folgendermaßen gewählt sein, wenn die geforderten Güteigenschaften eingehalten werden sollen:

$$a = \frac{-\ln \frac{\beta}{1-\alpha}}{\ln \frac{p_2}{p_1} \frac{1-p_1}{1-p_2}} = \frac{-\ln \frac{0,01}{0,99}}{\ln \frac{0,8}{0,7} \cdot \frac{0,3}{0,2}} = 8,53,$$

$$b = a = 8,53 \quad (\text{wegen } \alpha = \beta) \quad \text{und}$$

$$c = \frac{\ln \frac{1-p_1}{1-p_2}}{\ln \frac{p_2}{p_1} \cdot \frac{1-p_1}{1-p_2}} = \frac{\ln \frac{0,3}{0,2}}{\ln \frac{0,8}{0,7} \cdot \frac{0,3}{0,2}} = 0,75.$$

Die Versuchsauswertung läßt sich auch mit einem graphischen Verfahren vornehmen. Man zeichnet in ein Koordinatensystem die beiden durch

$$ck - a \quad \text{und} \quad ck + b$$

festgelegten Geraden ein (vgl. Abb. 25) und trägt in diese Abbildung nach jeder Behandlung einen Punkt mit den Koordinaten k und m_k ein. Solange nur Punkte zwischen den beiden Geraden eingetragen werden, muß weiter behandelt werden, doch sobald eine der beiden Geraden überschritten wird, kann eine endgültige Entscheidung getroffen werden, und zwar für H_1 beim Überschreiten der oberen Geraden und für H_0 beim Überschreiten der unteren Geraden.

Zum Schluß wollen wir uns noch ein Bild über die zu erwartenden Stichproben-

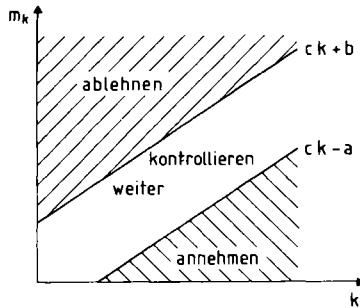


Abb. 25: Annahme- und Ablehnungsbereich für den sequentiellen Quotiententest bei der Binomialverteilung

umfänge \hat{n} und den Verlauf der Operationscharakteristik L machen. Dazu berechnen wir für verschiedene Werte von h jeweils

$$L = \frac{1 - \left(\frac{1 - \beta}{\alpha}\right)^h}{\left(\frac{\beta}{1 - \alpha}\right)^h - \left(\frac{1 - \beta}{\alpha}\right)^h} = \frac{1 - \left(\frac{0,99}{0,01}\right)^h}{\left(\frac{0,01}{0,99}\right)^h - \left(\frac{0,99}{0,01}\right)^h} = \frac{1 - 99^h}{\left(\frac{1}{99}\right)^h - 99^h},$$

sowie das zugehörige

$$p = \frac{1 - \left(\frac{1 - p_2}{1 - p_1}\right)^h}{\left(\frac{p_2}{p_1}\right)^h - \left(\frac{1 - p_1}{1 - p_2}\right)^h} = \frac{1 - \left(\frac{0,2}{0,3}\right)^h}{\left(\frac{0,8}{0,7}\right)^h - \left(\frac{0,2}{0,3}\right)^h} = \frac{1 - \left(\frac{2}{3}\right)^h}{\left(\frac{8}{7}\right)^h - \left(\frac{2}{3}\right)^h}$$

und werten dann noch den Ausdruck (für $h \neq 0$, d. h. $p \neq c$)

$$\begin{aligned} \hat{n} &= \frac{L \ln \frac{\beta}{1 - \alpha} + (1 - L) \ln \frac{1 - \beta}{\alpha}}{p \ln \frac{p_2}{p_1} + (1 - p) \ln \frac{1 - p_2}{1 - p_1}} = \frac{L \ln \frac{0,01}{0,99} + (1 - L) \ln \frac{0,99}{0,01}}{p \ln \frac{0,8}{0,7} + (1 - p) \ln \frac{0,2}{0,3}} \\ &= \frac{L (\ln 1 - \ln 99) + (1 - L) (\ln 99 - \ln 1)}{p \ln \frac{0,8}{0,7} + (1 - p) \ln \frac{0,2}{0,3}} = \frac{(1 - 2L) \ln 99}{p \ln \frac{0,8}{0,7} + (1 - p) \ln \frac{0,2}{0,3}} \end{aligned}$$

aus, vgl. **Tab. 47** sowie **Abb. 26, 27**.

Ist $p = c = 0,75$, so erhält man den maximal zu erwartenden Stichprobenumfang

$$\hat{n} = \frac{\left(\ln \frac{1 - \beta}{\alpha}\right)^2}{\ln \frac{1 - p_1}{1 - p_2} \cdot \ln \frac{p_2}{p_1}} = \frac{\left(\ln \frac{0,99}{0,01}\right)^2}{\ln \frac{0,3}{0,2} \cdot \ln \frac{0,8}{0,7}} = 390.$$

Tab. 47: Operationscharakteristik und erwarteter Stichprobenumfang \hat{n} eines sequentiellen Tests für den Parameter p einer Binomialverteilung

h	L	p	\hat{n}
-2	0,0001	0,84	95
-1	0,01	0,80	175
-0,5	0,09	0,78	284
0	0,50	0,75	390
0,5	0,91	0,73	272
1	0,99	0,70	160
2	0,9999	0,64	79

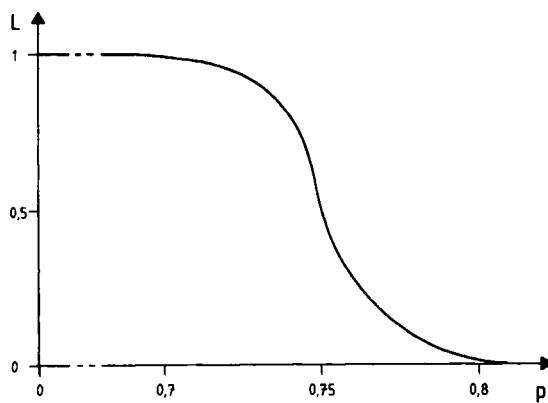


Abb. 26: Operationscharakteristik für einen Stichprobenplan

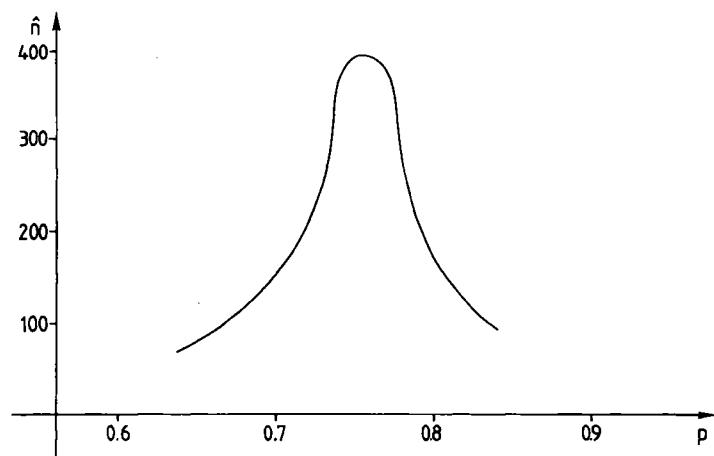


Abb. 27: Erwarteter Stichprobenumfang

6.2. Sequentieller Quotiententest für den Erwartungswert einer Normalverteilung

Eine normalverteilte Zufallsvariable X mit bekannter Varianz σ^2 und unbekanntem Erwartungswert μ hat die Wahrscheinlichkeitsdichte

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Es soll

$$H_0: \mu = \mu_1 \quad \text{gegen} \quad H_1: \mu = \mu_2,$$

wobei $\mu_1 < \mu_2$ ist, getestet werden.

Bezeichnen wir die beobachteten Werte mit x_1, x_2, \dots, x_k , so ergibt sich für die Testgröße des sequentiellen Quotiententests

$$\begin{aligned} T_k(x_1, \dots, x_k) &= \frac{f_{\mu_2}(x_1) \cdot \dots \cdot f_{\mu_2}(x_k)}{f_{\mu_1}(x_1) \cdot \dots \cdot f_{\mu_1}(x_k)} \\ &= \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-\mu_2)^2}{2\sigma^2}} \cdot \dots \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_k-\mu_2)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-\mu_1)^2}{2\sigma^2}} \cdot \dots \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_k-\mu_1)^2}{2\sigma^2}}} \\ &= \frac{e^{\frac{-x_1^2 + 2x_1\mu_2 - \mu_2^2}{2\sigma^2}} \cdot \dots \cdot e^{\frac{-x_k^2 + 2x_k\mu_2 - \mu_2^2}{2\sigma^2}}}{e^{\frac{-x_1^2 + 2x_1\mu_1 - \mu_1^2}{2\sigma^2}} \cdot \dots \cdot e^{\frac{-x_k^2 + 2x_k\mu_1 - \mu_1^2}{2\sigma^2}}} \\ &= e^{\frac{(\mu_2 - \mu_1)(x_1 + \dots + x_k) + \frac{\mu_1^2 - \mu_2^2}{2\sigma^2}k}{2\sigma^2}}. \end{aligned}$$

Anstatt diese Testgröße mit A und B zu vergleichen, kann man genauso gut

$$\frac{(\mu_2 - \mu_1)}{\sigma^2}(x_1 + \dots + x_k) + \frac{\mu_1^2 - \mu_2^2}{2\sigma^2}k$$

mit $\ln A$ und $\ln B$ vergleichen. Setzt man nun noch

$$a = \frac{-\ln A}{\frac{\mu_2 - \mu_1}{\sigma^2}} = -\frac{\sigma^2 \ln \frac{\beta}{1-\alpha}}{\mu_2 - \mu_1},$$

$$b = \frac{\ln B}{\frac{\mu_2 - \mu_1}{\sigma^2}} = \frac{\sigma^2 \ln \frac{1-\beta}{\alpha}}{\mu_2 - \mu_1} (= a, \text{ falls } \alpha = \beta) \quad \text{und}$$

$$c = -\frac{\frac{\mu_1^2 - \mu_2^2}{2\sigma^2}}{\frac{\mu_2 - \mu_1}{\sigma^2}} = \frac{\mu_1 + \mu_2}{2},$$

so besteht das Testverfahren darin, der Reihe nach die Summen $\sum_{i=1}^k x_i$ mit

$$ck - a = \frac{\mu_1 + \mu_2}{2}k - \frac{-\sigma^2 \ln \frac{\beta}{1-\alpha}}{\mu_2 - \mu_1} \quad \text{und}$$

$$ck + b = \frac{\mu_1 + \mu_2}{2}k + \frac{\sigma^2 \ln \frac{1-\beta}{\alpha}}{\mu_2 - \mu_1}$$

zu vergleichen. Solange

$$ck - a < \sum_{i=1}^k x_i < ck + b$$

gilt, beobachtet man weiter, doch sobald eine der beiden Grenzen überschritten wird, kommt es zu endgültigen Entscheidungen. Dieses Verfahren lässt sich auch zur Prüfung der Hypothese

$$H_0: \mu \leq \mu_1 \quad \text{gegen} \quad H_1: \mu > \mu_1$$

heranziehen. Dabei wird das Niveau für alle $\mu \leq \mu_1$ eingehalten und der Fehler 2. Art ist durch β beschränkt, wenn $\mu \geq \mu_2$ der wahre Parameter ist. Bezeichnet $L(\mu)$ die Operationscharakteristik an der Stelle μ , also die Wahrscheinlichkeit bei Vorliegen von μ , die Hypothese H_0 anzunehmen, so heißt das

$$\begin{aligned} 1 - L(\mu) &\leq \alpha, \quad \text{für } \mu \leq \mu_1 \quad \text{und} \\ L(\mu) &\leq \beta, \quad \text{für } \mu \geq \mu_2. \end{aligned}$$

Den genaueren Verlauf der Operationscharakteristik erhält man aus

$$L = L(\mu) = \frac{1 - B^h}{A^h - B^h} \quad \text{mit} \quad h = \frac{\mu_1 + \mu_2 - 2\mu}{\mu_2 - \mu_1}, \quad \text{also}$$

$$L = L(\mu) = \frac{1 - \left(\frac{1-\beta}{\alpha}\right)^{\frac{\mu_1 + \mu_2 - 2\mu}{\mu_2 - \mu_1}}}{\left(\frac{\beta}{1-\alpha}\right)^{\frac{\mu_1 + \mu_2 - 2\mu}{\mu_2 - \mu_1}} - \left(\frac{1-\beta}{\alpha}\right)^{\frac{\mu_1 + \mu_2 - 2\mu}{\mu_2 - \mu_1}}}.$$

Damit berechnet sich nun der mittlere Stichprobenumfang für

$$\mu \neq \frac{\mu_1 + \mu_2}{2} \quad \text{zu}$$

$$\begin{aligned}\hat{n} &= \sigma^2 \cdot \frac{L \ln A + (1-L) \ln B}{\frac{1}{2}(2(\mu_2 - \mu_1)\mu + \mu_1^2 - \mu_2^2)} \\ &= \sigma^2 \cdot \frac{L \ln \frac{\beta}{1-\alpha} + (1-L) \ln \frac{1-\beta}{\alpha}}{\frac{1}{2}(2(\mu_2 - \mu_1)\mu + \mu_1^2 - \mu_2^2)} \\ &= \sigma^2 \cdot \left(\frac{(1-2L) \ln \frac{1-\beta}{\alpha}}{\frac{1}{2}(2(\mu_2 - \mu_1)\mu + \mu_1^2 - \mu_2^2)}, \text{ für } \alpha = \beta \right).\end{aligned}$$

Ist $\mu = \frac{\mu_1 + \mu_2}{2}$, so erhält man für \hat{n} den (maximalen) Wert

$$\hat{n} = \sigma^2 \frac{-\ln \frac{\beta}{1-\alpha} \ln \frac{1-\beta}{\alpha}}{(\mu_2 - \mu_1)^2} \left(= \sigma^2 \left(\frac{\ln \frac{1-\beta}{\alpha}}{\mu_2 - \mu_1} \right)^2 \text{ für } \alpha = \beta \right).$$

Setzt man nun die Konstanten a, b und c ein, so ergibt sich kurz

$$\hat{n} = \begin{cases} \frac{b - (a+b)L}{\mu - c}, & \text{für } \mu \neq c \\ \frac{a \cdot b}{\sigma^2}, & \text{für } \mu = c. \end{cases}$$

6.3. Sequentieller Quotiententest für eine Exponentialverteilung

Ist X eine Zufallsvariable mit der Wahrscheinlichkeitsdichte

$$f_\lambda(x) = \lambda e^{-\lambda x},$$

also eine exponentialverteilte Zufallsvariable, so ist ein sequentieller Test für

$$H_0: \lambda = \lambda_1 \text{ gegen } H_1: \lambda = \lambda_2$$

gesucht, wobei mit λ_2 der größere von den beiden Werten bezeichnet werden soll. Das Niveau α soll eingehalten werden und der Fehler 2. Art durch β beschränkt sein. Da Lebensdauern oft näherungsweise als exponentialverteilt angenommen werden können, lässt sich das Testproblem auch so formulieren. Es soll geprüft werden, ob die mittlere Lebensdauer, z. B. einer bestimmten Sorte von Glühbirnen, den Wert $\frac{1}{\lambda_1}$ oder $\frac{1}{\lambda_2}$ annimmt. $\frac{1}{\lambda}$ ist der Erwartungswert der Exponentialverteilung. Dazu werden nacheinander die Lebensdauern x_1, x_2, \dots einzelner Stücke festgestellt (da diese Vorgehensweise im allgemeinen wenig praktikabel ist, werden wir später noch eine Modifikation dieses Verfahrens kennenlernen). Zur Durchführung des Tests benötigen wir noch die Teststatistik

$$\begin{aligned} T_k(x_1, \dots, x_k) &= \frac{f_{\lambda_2}(x_1) \cdot \dots \cdot f_{\lambda_2}(x_k)}{f_{\lambda_1}(x_1) \cdot \dots \cdot f_{\lambda_1}(x_k)} \\ &= \frac{\lambda_2 e^{-\lambda_2 x_1} \cdot \dots \cdot \lambda_2 e^{-\lambda_2 x_k}}{\lambda_1 e^{-\lambda_1 x_1} \cdot \dots \cdot \lambda_1 e^{-\lambda_1 x_k}} \\ &= \left(\frac{\lambda_2}{\lambda_1}\right)^k e^{-(\lambda_2 - \lambda_1)(x_1 + \dots + x_k)}. \end{aligned}$$

Anstatt nun T_k mit $A = \frac{\beta}{1-\alpha}$ und $B = \frac{1-\beta}{\alpha}$ zu vergleichen, lässt sich als Entscheidungskriterium auch der Vergleich von

$$k \ln \frac{\lambda_2}{\lambda_1} - (\lambda_2 - \lambda_1) \cdot \sum_{i=1}^k x_i$$

mit $\ln A$ und $\ln B$ heranziehen. Setzen wir

$$\begin{aligned} a &= \frac{\ln B}{\lambda_2 - \lambda_1} = \frac{\ln \frac{1-\beta}{\alpha}}{\lambda_2 - \lambda_1}, \\ b &= \frac{-\ln A}{\lambda_2 - \lambda_1} = \frac{-\ln \frac{\beta}{1-\alpha}}{\lambda_2 - \lambda_1} (= a, \text{ für } \alpha = \beta) \text{ und} \\ c &= \frac{\ln \frac{\lambda_2}{\lambda_1}}{\lambda_2 - \lambda_1}, \end{aligned}$$

so nehmen wir die Hypothese H_0 an, wenn zum ersten Mal

$$\sum_{i=1}^k x_i \leq ck - a$$

gilt. Ist dagegen zum ersten Mal

$$\sum_{i=1}^k x_i \geq ck + b,$$

so führt das zur Annahme von H_1 und bei

$$ck - a < \sum_{i=1}^k x_i < ck + b$$

muß weiter beobachtet werden.

Nun zur angekündigten Modifikation des Testverfahrens. Man unterwirft gleichzeitig n Glühbirnen einer Lebensdauerprüfung und registriert zu jedem Zeitpunkt t die Anzahl k der ausgefallenen Birnen. Dabei werden zwei Methoden unterschieden: bei Methode I werden die ausgefallenen Stücke sofort durch neue ersetzt, während dies bei der Methode II unterbleibt. Die Gesamtbrenndauer T aller Glühbirnen bis zum Zeitpunkt t (bei Methode I ist $T = nt$

und bei der Methode II ist $T = \sum_{i=1}^k x_i + (n - k)t$ wird nun laufend mit $ck - a$ und $ck + b$ verglichen. Solange T innerhalb dieser Grenzen liegt, muß weiter beobachtet werden, doch sobald $T \leq ck - a$ gilt, entscheidet man für H_0 und bei $T \geq ck + b$ für die Alternative H_1 . Bei Methode II kann es passieren, daß noch keine Entscheidung getroffen wurde, wenn alle Birnen ausgefallen sind. Dann nimmt man in der Regel H_0 an. Das Niveau α wird so zwar noch eingehalten, aber die Wahrscheinlichkeit für den Fehler 2. Art überschreitet den Wert β . Bei genügend großem n tritt diese Situation allerdings selten auf, und die Wahrscheinlichkeit, die Alternative H_1 fälschlicherweise zu verwerfen, ist nur geringfügig größer als β .

Bisher sind wir von dem Testproblem

$$H_0: \lambda = \lambda_1 \quad \text{gegen} \quad H_1: \lambda = \lambda_2$$

ausgegangen. Mit der oben dargestellten Vorgehensweise läßt sich aber auch

$$H_0: \lambda \leq \lambda_1 \quad \text{gegen} \quad H_1: \lambda > \lambda_1$$

testen. Es wird auch hier das Niveau α eingehalten, und außerdem ist für alle $\lambda \geq \lambda_2$ der Fehler 2. Art durch β beschränkt. Bezeichnet wieder $L = L(\lambda)$ die Operationscharakteristik des Testverfahrens, so heißt das

$$1 - L(\lambda) \leq \alpha, \quad \text{für } \lambda \leq \lambda_1$$

$$\text{und} \quad L(\lambda) \leq \beta, \quad \text{für } \lambda \geq \lambda_2.$$

Den genauen Verlauf der Operationscharakteristik erhält man durch

$$L = \frac{1 - B^h}{A^h - B^h} = \frac{1 - \left(\frac{1 - \beta}{\alpha}\right)^h}{\left(\frac{\beta}{1 - \alpha}\right)^h - \left(\frac{1 - \beta}{\alpha}\right)^h};$$

dabei ist h nun durch die Beziehung

$$\lambda = \frac{1 - \left(\frac{\lambda_2}{\lambda_1}\right)^h}{(\lambda_1 - \lambda_2) \cdot h}$$

festgelegt. Der folgende Ausdruck \hat{n} gibt die mittlere Anzahl der Ausfälle an, die man beobachten muß, um zu einer endgültigen Entscheidung zu gelangen.

$$\text{Für} \quad \ln \frac{\lambda_2}{\lambda_1} \neq \frac{\lambda_2 - \lambda_1}{\lambda}$$

ergibt sich

$$\hat{n} = \frac{L \ln A + (1 - L) \ln B}{\ln \frac{\lambda_2}{\lambda_1} - \frac{\lambda_2 - \lambda_1}{\lambda}}$$

$$\begin{aligned}
 &= \frac{L \ln \frac{\beta}{1-\alpha} + (1-L) \ln \frac{1-\beta}{\alpha}}{\ln \frac{\lambda_2}{\lambda_1} - \frac{\lambda_2 - \lambda_1}{\lambda}} \\
 &\left(= \frac{(1-2L) \ln \frac{1-\beta}{\alpha}}{\ln \frac{\lambda_2}{\lambda_1} - \frac{\lambda_2 - \lambda_1}{\lambda}}, \text{ falls } \alpha = \beta \right).
 \end{aligned}$$

Ist $\ln \frac{\lambda_2}{\lambda_1} = \frac{\lambda_2 - \lambda_1}{\lambda}$,

so erhält man

$$\hat{n} = \frac{\ln \frac{\beta}{1-\alpha} \ln \frac{1-\beta}{\alpha}}{\left(\ln \frac{\lambda_2}{\lambda_1} \right)^2} \left(= \left(\frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{\lambda_2}{\lambda_1}} \right)^2, \text{ für } \alpha = \beta \right).$$

Setzt man nun die Konstanten a , b und c ein, so erhält man

$$\hat{n} = \begin{cases} \frac{b - (a+b)L}{c - \frac{1}{\lambda}}, & \text{für } c \neq \frac{1}{\lambda} \\ \frac{a \cdot b}{c^2}, & \text{für } c = \frac{1}{\lambda} \end{cases}$$

Wesentlich interessanter ist aber im allgemeinen die Frage nach der nötigen Beobachtungsdauer. Dazu überlegen wir uns zunächst einmal, wie groß die gesamte Brenndauer aller Glühbirnen bis zum Zeitpunkt einer endgültigen Entscheidung ist. Da die erwartete Lebensdauer einer einzelnen Birne $\frac{1}{\lambda}$ beträgt, beträgt die erwartete Gesamt-Brenndauer

$$T_{\text{ges}} = \frac{1}{\lambda} \hat{n}.$$

Für die Methode I lässt sich daraus die durchschnittlich notwendige Beobachtungszeit T_{beob} einfach ermitteln. Es gilt z.B. für $c \neq 1/\lambda$

$$T_{\text{beob}} = \frac{1}{n\lambda} \hat{n} = \frac{L \ln \frac{\beta}{1-\alpha} + (1-L) \ln \frac{1-\beta}{\alpha}}{n \lambda \ln \frac{\lambda_2}{\lambda_1} - n(\lambda_2 - \lambda_1)}$$

$$\left(= \frac{(1 - 2L) \ln \frac{1 - \beta}{\alpha}}{n \lambda \ln \frac{\lambda_2}{\lambda_1} - n(\lambda_2 - \lambda_1)} \quad \text{für } \alpha = \beta \right).$$

Beispiel: Es soll geprüft werden, ob die Lebensdauer von Glühbirnen kleiner oder größer als 1000 (Stunden) ist. Dazu testen wir

$$H_0: \lambda \leq \lambda_1 = \frac{1}{1000} \quad \text{gegen} \quad H_1: \lambda > \lambda_1 = \frac{1}{1000}.$$

Die Einhaltung des Niveaus $\alpha = 0,05$ sichert uns, daß die Wahrscheinlichkeit, eine zu kleine Lebensdauer anzunehmen, sich also für H_1 zu entscheiden, obwohl H_0 richtig ist, höchstens $\alpha = 0,05$ sein kann. Um dagegen sicherzustellen, daß eine zu kleine Lebensdauer (etwa kleiner als 500 Stunden) mit genügend großer Sicherheit erkannt wird, fordern wir noch, daß die Wahrscheinlichkeit fälschlicherweise, H_0 anzunehmen, wenn in Wirklichkeit die zu erwartende Lebensdauer höchstens $\frac{1}{\lambda} = 500$ (Stunden) beträgt, durch $\beta = 0,05$ beschränkt ist. Nach diesen Überlegungen können wir die Konstanten a , b und c berechnen.

$$a = \frac{\ln \frac{1 - \beta}{\alpha}}{\lambda_2 - \lambda_1} = \frac{\ln \frac{0,95}{0,05}}{\frac{1}{500} - \frac{1}{1000}} = \frac{\ln 19}{\frac{1}{1000}} = 1000 \cdot \ln 19 = 2944,44,$$

$$b = a = 2944,44 \quad (\text{wegen } \alpha = \beta) \quad \text{und}$$

$$c = \frac{\ln \frac{\lambda_2}{\lambda_1}}{\lambda_2 - \lambda_1} = \frac{\ln \frac{1000}{500}}{\frac{1}{500} - \frac{1}{1000}} = 1000 \cdot \ln 2 = 693,15.$$

Wir beladen nun gleichzeitig $n = 100$ Glühbirnen und ersetzen jeweils die ausgebrannten Birnen durch neue. Die Glühlampen haben dann bis zum Zeitpunkt t insgesamt $T = nt$ Stunden gebrannt. Es muß also solange beobachtet werden, wie

$$ck - a = 693,15k - 2944,44 < 100t < 693,15k + 2944,44 = ck + b$$

gilt. Ist zum ersten Mal

$$100t \leq 693,15k - 2944,44,$$

so entscheidet man sich für H_0 , und bei

$$693,15k + 2944,44 \leq 100t$$

entscheidet man sich für die Alternative H_1 . Wir wollen uns jetzt noch überlegen, wie groß der durchschnittliche Zeitaufwand bei der Durchführung dieses Tests ist.

Dazu bestimmen wir zunächst für einige Lebensdauern $\frac{1}{\lambda}$ die erwartete Anzahl \hat{n} ausgestellter Glühbirnen, die wir beobachten müssen, um zu einer endgültigen Entscheidung zu gelangen. \hat{n} ist am größten, wenn gilt

$$\ln \frac{\lambda_2}{\lambda_1} = \frac{\lambda_2 - \lambda_1}{\lambda},$$

also wenn

$$\frac{1}{\lambda} = \frac{\ln \frac{\lambda_2}{\lambda_1}}{\lambda_2 - \lambda_1} = c = 693,15.$$

Wegen $\alpha = \beta = 0,05$ erhält man dann hieraus

$$\hat{n} = \left(\frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{\lambda_2}{\lambda_1}} \right)^2 = \left(\frac{\ln \frac{0,95}{0,05}}{\ln \frac{1000}{500}} \right)^2 = \left(\frac{\ln 19}{\ln 2} \right)^2 = 18.$$

Für $\lambda = \lambda_1 = \frac{1}{1000}$ ergibt sich

$$\begin{aligned} \hat{n} &= \frac{(1 - 2L) \ln \frac{1-\beta}{\alpha}}{\ln \frac{\lambda_2}{\lambda_1} - \frac{\lambda_2 - \lambda_1}{\lambda}} = \frac{(1 - 2 \cdot 0,95) \ln \frac{0,95}{0,05}}{\ln \frac{1000}{500} - 1000 \left(\frac{1}{500} - \frac{1}{1000} \right)} \\ &= \frac{-0,9 \ln 19}{\ln 2 - 1} = 8,6, \end{aligned}$$

und für $\lambda = \lambda_2 = 1/500$ ergibt sich

$$\hat{n} = \frac{(1 - 2 \cdot 0,05) \ln \frac{0,95}{0,05}}{\ln \frac{1000}{500} - 500 \left(\frac{1}{500} - \frac{1}{1000} \right)} = \frac{0,9 \ln 19}{\ln 2 - 0,5} = 13,7.$$

Die durchschnittliche Testdauer bei einer tatsächlich zu erwartenden Lebensdauer von

$$\frac{1}{\lambda} = c = 693,15 \text{ (Stunden)}$$

ergibt sich zu

$$T_{\text{beob}} = \frac{1}{n\lambda} \hat{n} = \frac{1}{100} \cdot 693,15 \cdot 18 = 124,8 \text{ (Stunden)}.$$

$$\text{Für } \frac{1}{\lambda} = \frac{1}{\lambda_1} = 1000 \text{ (Stunden)}$$

ergibt sich

$$T_{\text{beob}} = \frac{1}{n\lambda} \hat{n} = \frac{1}{100} \cdot 1000 \cdot 8,6 = 86,0 \text{ (Stunden)}$$

und für

$$\frac{1}{\lambda} = \frac{1}{\lambda_2} = 500 \text{ (Stunden)}$$

ergibt sich

$$T_{\text{beob}} = \frac{1}{n\lambda} \hat{n} = \frac{1}{100} \cdot 500 \cdot 13,7 = 68,5 \text{ (Stunden)}.$$

Man sieht, daß die durchschnittlichen Testdauern erheblich kürzer sind als die erwarteten Lebensdauern der einzelnen Glühbirnen, so daß mit diesem Verfahren in verhältnismäßig kurzer Zeit – auch bei relativ langlebigen Erzeugnissen – eine Lebensdauerprüfung vorgenommen werden kann. Eine weitere Reduzierung der Beobachtungsdauer läßt sich durch die Erhöhung der Anzahl n der gleichzeitig zu belastenden Glühlampen erreichen.

Im Beispiel würde sich bei einer Anzahl von $n = 500$ Glühbirnen, einer erwarteten Lebensdauer

$$\frac{1}{\lambda} = \frac{1}{0,001} = 1000 \quad \text{und} \quad \hat{n} = 8,6$$

die Testdauer zu

$$T_{\text{beob}} = \frac{1}{n\lambda} \hat{n} = \frac{1}{500} \cdot 1000 \cdot 8,6 = 17,2 \text{ (Stunden)}$$

berechnen, wohingegen für $n = 100$ noch eine Testdauer von 86,0 Stunden nötig war.

Kapitel V:

Aspekte der Datengewinnung – Stichprobentheorie, Meßfehler, Ausreißertests, Datentransformationen, Versuchsplanung, Klinische Versuche, Skalierung

Bevor man statistische Verfahren überhaupt anwenden kann, müssen Daten gewonnen werden, aufgrund derer eine Aussage möglich ist.

Dabei stellt sich zunächst die Frage nach der Datenbeschaffung. Hierzu liefert die Stichprobentheorie, die in den Abschnitten 1, 2 und 3 ausführlich behandelt wird, wichtige Anhaltspunkte.

Stehen schließlich Rohdaten zur Verfügung, so sollte man sich fragen, ob diese Daten fehlerbehaftet sind (Meßfehler), ob Ausreißer darunter sind, die das abzuleitende Ergebnis verfälschen können, und ob diese Daten so geartet sind, daß man ein geeignetes statistisches Verfahren darauf anwenden kann, oder ob sie zunächst geeignet transformiert werden müssen. Alle erwähnten Probleme und Fragestellungen werden in Abschnitt 4 behandelt.

In Abschnitt 5 werden wir uns mit der Planung komplizierterer Versuche beschäftigen, die dann etwa mit Methoden der Varianzanalyse, vgl. Kap. XI, ausgewertet werden können.

Spezielle Probleme, die bei der Planung und Auswertung klinischer Versuche auftreten, werden dann in Abschnitt 6 angerissen und an einem Beispiel verdeutlicht.

Schließlich wird in Abschnitt 7 auf die Skalierung von Merkmalsausprägungen und Testergebnissen eingegangen; es werden dort verschiedene Probleme angesprochen und Verfahren vorgestellt, die es etwa erlauben, Bewertungsskalen für psychologische Tests (z. B. Intelligenztests) usw. zu erstellen.

1. Abriß der klassischen Stichprobentheorie am Beispiel der Inventur auf Stichprobenbasis

Die wichtigste Zielsetzung der Stichprobentheorie ist es, Schlüsse von einer „geogenen Stichprobe“ auf eine interessierende **Grundgesamtheit** zu ziehen. Diese Schlüsse beziehen sich auf bestimmte Größen in der Grundgesamtheit, die von besonderem Interesse sind, die aber nur aus der exakten Kenntnis aller Elemente der Gesamtheit heraus anzugeben sind. In vielen Fällen sind solche exakten Kenntnisse gar nicht zu gewinnen oder es würde einen übermäßig hohen Zeit- oder Kostenaufwand bedeuten, sich solche Kenntnisse zu verschaffen. Aus diesen Gründen hat die Stichprobentheorie ihre Legitimation gefunden, und so hat sich in den letzten Jahren diese Theorie immer weiter entwickelt und neue Anwendungsmöglichkeiten erfahren.

Die Vorteile und Anwendungsmöglichkeiten der Stichprobentheorie sollen an einem Beispiel erläutert werden, der **Inventur auf Stichprobenbasis**.

Beispiel: Ein Unternehmen, das auf dem Elektronikmarkt tätig ist, besitzt ein Lager, in dem alle fertiggestellten Artikel des Betriebes vorhanden sind. Die rechtlichen Bestimmungen verpflichten jeden Vollkaufmann (bzw. an seiner Stelle stehen-

de Gruppen oder Institutionen) jedes Jahr an einem festgelegten Tag eine Inventur durchzuführen, d. h. eine Bestandsaufnahme aller Vermögensgegenstände vorzunehmen. Dazu gehört auch eine explizite Bestimmung des Wertes der vorhandenen Lagerbestände.

Bei großen Unternehmen bzw. großen Lagern wird die Erstellung des Lagerwertes durch gesamte körperliche Bestandsaufnahme zu einer großen Belastung zeitlicher und finanzieller Art (bis zu 8% des Lagerwertes). Dies hatte seine Auswirkung auf die steuerlich-rechtlichen Vorschriften für die Inventur, so daß in einer Novellierung des Handelsgesetzes (§ 39 Absatz 2a HGB) die Möglichkeit einer Inventur mit Hilfe anerkannter mathematisch-statistischer Methoden anstelle der Vollinventur vorgesehen wurde.

Die im weiteren Verlauf angegebenen Berechnungen sind nicht aus einer real existierenden Stichprobenuntersuchung entnommen. Sie basieren auf einer Simulationsstudie, die aus den für solche Probleme typischen Modellvoraussetzungen, welche noch erläutert werden, erstellt wurde.

1.1. Die Stichprobe

Wie bereits erwähnt, soll mit Hilfe der Stichprobentheorie ein Schluß von der Stichprobe auf die Grundgesamtheit gezogen werden. Dazu müssen einige Begriffe näher erläutert werden. Betrachtet wird zunächst die interessierende Grundgesamtheit, bestehend aus den Einheiten E_1, \dots, E_N , d. h. eine **Grundgesamtheit vom Umfang N**.

Beispiel: Im Inventurbeispiel besteht die Grundgesamtheit aus den N Artikeln, aus denen das Lager gebildet wird. Dabei ist zu beachten, daß ein Artikel mehrmals vorhanden sein kann. Es ist deshalb günstiger, die Bezeichnung Artikelposition zu wählen, wobei in einer Position eine gewisse Anzahl gleichartiger Artikel zusammengefaßt ist. Der Index i der Position E_i ist hierbei gut als Artikelnummer interpretierbar, der Umfang N der Grundgesamtheit könnte als „Größe“ des Sortiments bezeichnet werden. Das zugrundegelegte Lager der Elektronikfirma soll einen Lagerumfang von $N = 18\,000$ Artikeln besitzen. Diese Artikel bilden das gesamte Sortiment der Firma. Jeder Artikel ist mit einer der Nummern eindeutig gekennzeichnet, die von 00001 bis 18 000 reichen.

Jedes Element E_i der Grundgesamtheit ist mit einem **Merkmalswert** behaftet, der mit Y_i bezeichnet wird.

Beispiel: Für eine Artikelposition E_i ist der Merkmalswert Y_i definiert als der Wert dieser Position. Dieser ist gegeben durch das Produkt der vorhandenen Stückzahl X_i mit dem Preis P_i .

Beispielsweise hat der Artikel mit der Nummer 16 765, es handelt sich hierbei um einen Transistor, einen Stückpreis von 1,61 DM. Da am Stichtag (z. B. 1.1.1981) 318 Stück dieses Transistors im Lager sind, ist die Artikelposition „Transistor“ mit $318 \cdot 1,61 \text{ DM} = 511,98 \text{ DM}$ zu bewerten. Mit den eingeführten Bezeichnungen bedeutet dies:

$$\begin{aligned} P_{16765} &= 1,61 \text{ DM} \\ X_{16765} &= 318 \\ Y_{16765} &= 511,98 \text{ DM}. \end{aligned}$$

Charakteristisch für die Merkmalswerte Y_i ist, daß sie vor Beginn einer Untersu-

chung unbekannt sind. So ist es das Ziel der Untersuchung, eine Aussage über diese Merkmalswerte zu machen, z. B. die **Summe**

$$Y = \sum_{i=1}^N Y_i$$

oder den **Durchschnitt**

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

anzugeben.

Beispiel: Ziel der Inventur ist es, den Lagerwert, d. h. die Summe Y der Positions-werte anzugeben.

Es ist natürlich klar, daß eine exakte Bestimmung solcher Größen nur mit allen Merkmalswerten geschehen kann, wozu man eine **Total- oder Vollerhebung** durchführen müßte. Beschränkt man sich allerdings auf die Erhebung einiger Merkmals-werte, zieht man also eine **Stichprobe**, so ist man auf eine „Schätzung“ angewiesen. Man spricht immer dann von einer **Stichprobe vom Umfang n** , wenn man sich bei der Untersuchung von Merkmalswerten auf eine Auswahl von $n < N$ Einheiten aus der interessierenden Grundgesamtheit beschränkt. Eine erhobene Stichprobe ist dann eine Gruppe (y_1, \dots, y_n) von Realisationen der Merkmalswerte Y_1, \dots, Y_N .

Beispiel: Zieht man aus N Artikelpositionen eine Stichprobe vom Umfang n , so ist y_i der Wert der i -ten Position in der Stichprobe. Man beachte, daß i ein Laufindex ist. Es ist deshalb wohl zu unterscheiden zwischen Y_i und y_i : Y_i bezeichnet den Merkmalswert der i -ten Einheit in der Grundgesamtheit, y_i den Wert der i -ten Einheit in der Stichprobe. Im allgemeinen gilt also $Y_i \neq y_i$.

Ziel der Stichprobenerhebung ist es, eine Aussage über eine interessierende Größe (etwa Y) der Grundgesamtheit zu machen. Mit den vorausgegangenen Bezeichnungen ist dieses Problem in folgender Weise beschreibbar. Man sagt, „es ist ein **Schätzer** \hat{Y} für die unbekannte Größe Y gesucht“. Die Konstruktion eines solchen \hat{Y} , das eine Zufallsvariable darstellt, hängt von dem jeweiligen Stichprobenverfah-ren ab.

1.2. Überlegungen und Vorgehensweisen bei Stichprobenerhebungen

Jede Durchführung einer Stichprobenerhebung ist mit einer vorherigen sorgfälti-gen Planung verbunden. Dabei ist vor allem folgendes zu beachten.

Zunächst ist eine inhaltliche und formale Spezifikation des Problems vorzuneh-men und in eindeutiger Weise die betrachtete **Grundgesamtheit abzugrenzen**. Daß dies nicht immer so einfach ist, wie es bei dem betrachteten Inventurproblem er-scheint, zeigt sich schon daran, daß sich z. B. im Bereich der Marktforschung viele Abhandlungen mit solchen Fragen beschäftigen. Hat man nun solche Spezifikatio-nen vorgenommen, so ergeben sich oft daraus Hinweise auf die spezielle **Art der Erhebungstechnik**, die man als nächstes festlegt.

Im folgenden Schritt tritt das Problem der **Größe des Stichprobenumfangs** auf, was eng mit der Frage nach der Genauigkeit der Untersuchung verknüpft ist. Bei der Genauigkeitsfrage sind dabei von vornherein zwei Größen zu berücksichtigen: die **Sicherheitswahrscheinlichkeit** $1 - \alpha$ sowie der **relative Zufallsfehler e**. Die Sicher-heitswahrscheinlichkeit $1 - \alpha$ hat dabei die Bedeutung, daß bei statistischen Aussa-

gen (z. B. dem Angeben von Schätzern) immer ein Irrtum möglich ist, wenn eine „ungünstige“ Stichprobe gezogen wurde. Die Wahrscheinlichkeit $1 - \alpha$ gibt dann an, daß in durchschnittlich $100 \cdot \alpha$ von 100 Fällen eine solche „ungünstige“ Stichprobe möglich ist. Dies muß in Kauf genommen werden. Der Zufallsfehler e dagegen ist eine von vornherein angenommene Fehlergrenze für die Schätzwerte. Er entspricht etwa der Genauigkeit ϵ eines Konfidenzintervalls zum Niveau $1 - \alpha$, bezogen auf das arithmetische Mittel. (Man vergleiche hierzu die Ausführungen im Abschnitt 1.4.2 des Kap. IV über die Einhaltung prozentualer Genauigkeiten bei der Angabe eines Konfidenzintervalls für den Mittelwertparameter einer Normalverteilung.)

Beispiel: Während in vielen Fällen die Genauigkeit einer Untersuchung von umgangssprachlichen Grenzen abhängt, ist im Falle der Stichprobeninventur die Genauigkeit durch die bereits erwähnte Gesetzesvorschrift festgelegt. Die anzuwendende Methode soll keine größeren Fehler machen als die bislang durchgeföhrte Vollinventur. Dabei hat sich aber gezeigt, daß Schwankungen bis zu 2% um den wahren Inventurwert vorkommen können. Deshalb erscheint ein relativer Zufallsfehler von $e = 0,02$ als angebracht. Eine übliche Sicherheitswahrscheinlichkeit wäre $1 - \alpha = 0,95$.

Die Angabe von $1 - \alpha$ und e ermöglicht nun eine von der Art der Erhebung abhängige Berechnung des Stichprobenumfangs. Dabei gehen meist unbekannte Größen der Grundgesamtheit mit ein.

Um sich diesem Dilemma zu entziehen, führt man eine sogenannte **Vorerhebung** durch, schätzt mit den Ergebnissen dieser Erhebung die unbekannten Größen und verwendet diese Schätzer zur Berechnung des eigentlichen Stichprobenumfangs.

Ohne Berücksichtigung blieben bisher Kostenüberlegungen, die bei Stichproben erhebungen eine wesentliche Rolle spielen und teilweise mit in die Berechnungen eingehen. Dabei gilt es meist den Konflikt zu lösen, daß Genauigkeit Kosten mit sich bringt, Kostensparnis aber zu weniger brauchbaren Ergebnissen führt.

Beispiel: Wie man diesen Konflikt löst, ist im Einzelfall zu erörtern, im Falle der Inventur allerdings ist dieses Problem weniger interessant. Das liegt daran, daß die Stichprobeninventur eine echte Alternative zur Vollinventur darstellt. Somit stellt sich nur die Frage nach der Größe der Zeit- und Kostensparnis, nicht aber nach der Einhaltung einer Kostengrenze.

Sind alle Überlegungen abgeschlossen, so kann die Erhebung durchgeführt und mit deren Ergebnissen weitergearbeitet werden.

1.3. Verteilungsannahmen bei Stichprobenerhebungen

In der Stichprobentheorie erweist sich der Gebrauch der Normalverteilung von großer Bedeutung. Dies hat seine Begründung in einem Satz, der in Analogie zum zentralen Grenzwertsatz, vgl. Abschnitt 9 in Kap. II, unter bestimmten Voraussetzungen die angenäherte Normalverteilung der Schätzer \hat{Y} bzw. \bar{y} aus Abschnitt 1.4 folgert. Diese Voraussetzungen sind bei genügend großen Stichprobenumfängen (etwa $n > 60$) als hinreichend genau erfüllt zu betrachten.

Hat man also eine Schätzfunktion der Art

$$\hat{Y} = N \frac{1}{n} \sum_{i=1}^n y_i \quad \text{oder} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

so kann man annehmen, daß diese Zufallsvariable normalverteilt ist, selbst wenn die Y_i in der Grundgesamtheit wesentlich anders verteilt sind.

Beispiel: Im Inventurproblem ist die Grundgesamtheit sicher nicht normalverteilt. Hier, wie auch bei anderen Gesamtheiten im Wirtschaftsbereich, liegt eine linkschiefe Verteilung vor, vgl. Abschnitt 4.5 und 5 in Kap. I. Dies hat sich in der sogenannten **20–80-Regel** niedergeschlagen, die besagt, daß nur 20% der Waren eines Lagers 80% des Wertes für sich beanspruchen. Dies erklärt sich durch die unterschiedliche „Besetzung“ eines Lagers, in dem i. a. viele Artikel einer niedrigen und wenige einer hohen Preiskategorie zugeordnet werden. So ist für ein Lager eine Lognormal- oder auch eine Dreiecksverteilung, vgl. Abschnitt 1.2 und 2 in Kap. IV, ein sinnvoller Ansatz.

Um die Verteilung der Warenwerte im Lager zu veranschaulichen, kann man sich auch der Konzentrationsmessung nach Lorenz (Lorenzkurve), vgl. Kap. I, Abschnitt 6, bedienen. Hierbei werden die kumulierten Wertanteile auf der Ordinate und die nach dem Wert geordneten anteiligen kumulierten Artikelpositionen auf der Abszisse abgetragen. Ein Punkt auf der so entstandenen Kurve sagt im Inventurproblem dann aus, daß p % der Artikelpositionen q % des Warenwertes verkörpern. Als Beispiel für eine solche Lorenzkurve diene die aus **Abb. 1**, welche die Lagerstruktur eines real existierenden Werkstattlagers aufzeigt, vgl. auch Wienhold (1978).

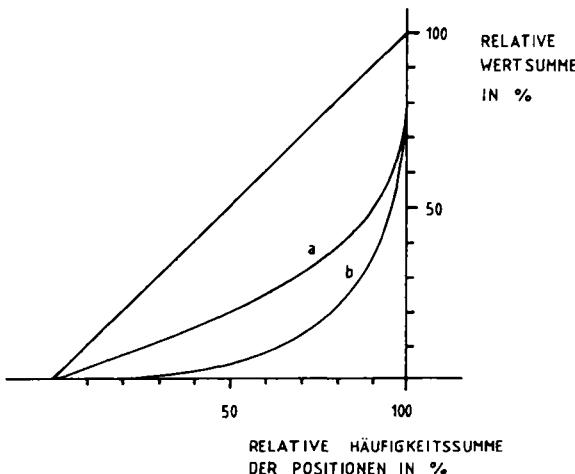


Abb. 1: Lorenzkurven für Daten (a) und nach der 20 – 80-Regel (b)

1.4. Die einfache Zufallsauswahl

Aus den N Elementen der interessierenden Grundgesamtheit soll eine Stichprobe vom Umfang n gezogen werden. Die Anzahl der Möglichkeiten, dies zu tun, ist

$$\binom{N}{n} = \frac{N!}{n!(N-n)!},$$

falls die N Elemente der Grundgesamtheit unterscheidbar sind, vgl. Kap. II. Wählt

man die n Elemente nun so aus der Grundgesamtheit aus, daß jede der $\binom{N}{n}$ möglichen Stichproben die gleiche Auswahlwahrscheinlichkeit hat, so spricht man von einer **einfachen Zufallsstichprobe (SRS, „Simple Random Sampling“)**.

Numeriert man die Elemente E_i von 1 bis N durch und zieht n Zufallszahlen aus 1 bis N , so führt man eine einfache Zufallsauswahl durch, wenn man die Elemente mit den gezogenen Nummern in die Stichprobe aufnimmt.

Bezeichnet Π_i die Wahrscheinlichkeit, daß E_i in die Stichprobe gelangt, bzw. Π_{ij} die Wahrscheinlichkeit, daß E_i und E_j in die Stichprobe gelangen, so gilt bei einer einfachen Zufallsauswahl

$$\Pi_i = \frac{n}{N} \quad \text{und} \quad \Pi_{ij} = \frac{n(n-1)}{N(N-1)}.$$

Jedes Element der Grundgesamtheit hat somit die gleiche Wahrscheinlichkeit in die Stichprobe zu gelangen. (Dies ist nur ein notwendiges Kriterium für die einfache Zufallsauswahl.) Die vorliegende Definition der einfachen Zufallsauswahl weicht von der ab, die nur fordert, daß

$$\Pi_i = \text{constant} \quad \text{für alle } i = 1, \dots, N.$$

Daß diese Definitionen nicht identisch sind, mag folgendes Beispiel zeigen.

Beispiel: $N = 4$, $n = 2$ und $P(E_1, E_2) = \frac{1}{2}$, $P(E_3, E_4) = \frac{1}{2}$.

Hier gilt $\Pi_i = \frac{1}{2}$ für alle i , aber z. B. $P(E_1, E_3) = 0$. Daraus folgt, daß keine einfache Zufallsauswahl vorliegt (die Einheiten der Grundgesamtheit sind gekoppelt, z. B. Ehepaare).

Das Ziel der Erhebung ist es, Schätzer für die Summe, den Durchschnitt und die Varianz der Merkmalswerte in der Grundgesamtheit

$$Y = \sum_{i=1}^N Y_i, \quad \bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i = \frac{1}{N} Y \quad \text{und} \quad S_Y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

zu berechnen. Man beachte, daß hier bei der „Varianz“ S_Y^2 durch $N-1$ geteilt wird und nicht durch N , wie man eigentlich erwartet hätte. Diese Festlegung von S_Y^2 , die wir in der Stichprobentheorie trotzdem mit Varianz bezeichnen wollen, bringt uns bei den zu betrachtenden Stichproben aus endlichen Grundgesamtheiten eine Vereinfachung der Formeln und Beziehungen.

Beispiel: Im Inventurproblem bedeuten diese Größen folgendes:

Y ist der Gesamtlagerwert, \bar{Y} der durchschnittliche Positionswert, S_Y^2 die Varianz der Positionswerte in der Grundgesamtheit.

Bei der einfachen Zufallsauswahl gilt:

$$\hat{Y} = N\bar{y} = N \frac{1}{n} \sum_{i=1}^n y_i$$

ist ein erwartungstreuer Schätzer für Y .

Zur erwartungstreuen Schätzung von S_Y^2 verwendet man die Stichprobenvarianz

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

und schließlich kann die Varianz des Schätzers \hat{Y}

$$\text{Var } \hat{Y} = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) S_y^2$$

durch

$$\widehat{\text{Var } \hat{Y}} = N^2 \left(1 - \frac{n}{N}\right) \frac{1}{n} s_y^2$$

erwartungstreu geschätzt werden.

In den beiden letzten Formeln tritt der Term $\left(1 - \frac{n}{N}\right) = (1 - f)$ auf; dieser heißt auch **Endlichkeitskorrektur**.

Ist der **Auswahlsatz**

$$f = \frac{n}{N} < 0,05,$$

so ist die Endlichkeitskorrektur in der Regel vernachlässigbar.

Wie in Abschnitt 1.3 bereits erwähnt, ist es möglich, für \hat{Y} bzw. \bar{y} approximativ die Normalverteilung mit Mittelwert Y bzw. \bar{Y} anzunehmen. Aus dieser Annahme lassen sich Konfidenzintervalle für Y und \bar{Y} berechnen. Es ist annähernd

$$\left[\bar{y} - \frac{u_{1-\alpha/2} s_y}{\sqrt{n}} \sqrt{1-f}; \bar{y} + \frac{u_{1-\alpha/2} s_y}{\sqrt{n}} \sqrt{1-f} \right]$$

ein $1 - \alpha$ Konfidenzintervall für den mittleren Merkmalswert \bar{Y} in der Grundgesamtheit und

$$\left[\hat{Y} - \frac{N}{\sqrt{n}} u_{1-\alpha/2} s_y \sqrt{1-f}; \hat{Y} + \frac{N}{\sqrt{n}} u_{1-\alpha/2} s_y \sqrt{1-f} \right]$$

ein Konfidenzintervall für die Summe Y der Merkmalswerte. Ist n klein, so sind die Quantile $u_{1-\alpha/2}$ der Standardnormalverteilung durch die der t-Verteilung mit $n - 1$ Freiheitsgraden, $t_{n-1; 1-\alpha/2}$, zu ersetzen. Diese Quantile sind für verschiedene α im Anhang vertafelt.

Hat man sich zur Durchführung der Stichprobenerhebung als einfache Zufallsauswahl entschieden, so muß zunächst der Stichprobenumfang n bestimmt werden.

Aus einer Vorerhebung schätzt man dazu zunächst die unbekannten Größen S_y^2 und \bar{Y} durch s_y^2 und \tilde{y} , d.h. durch empirische Varianz und empirisches Mittel in der Vorerhebung.

Will man nun erreichen, daß mit einer Sicherheitswahrscheinlichkeit $1 - \alpha$ der Zufallsfehler höchstens e beträgt, so erhält man als erste Näherung für den Stichprobenumfang n die Größe

$$n_0 = \left(\frac{u_{1-\alpha/2} \tilde{s}_y}{e \tilde{y}} \right)^2.$$

Ist der Umfang der Grundgesamtheit N bekannt, so wählt man dann als Stichprobenumfang

$$n = \frac{n_0}{1 + \frac{n_0}{N}};$$

Ist N unbekannt oder $\frac{n_0}{N}$ sehr klein, so setzt man direkt $n = n_0$.

Mit diesem Verfahren zur Bestimmung des Stichprobenumfanges n hat man auch die Varianz $\text{Var } \hat{Y}$ des Schätzers für Y unter Kontrolle, denn es ist

$$\text{Var } \hat{Y} = \left(\frac{eY}{u_{1-\alpha/2}} \right)^2.$$

Nicht immer ist das Verfahren, den Stichprobenumfang mit Schätzern einer Vorerhebung zu bestimmen, durchführbar. Dies kann daran liegen, daß eine Vorerhebung prinzipiell nicht möglich ist bzw. Kostenüberlegungen eine Rolle spielen. Deshalb ist es notwendig, andere Verfahren bereitzustellen, die eine Abschätzung der Standardabweichung der Grundgesamtheit ermöglichen.

Schwarz (1975, S. 82ff) gibt dazu einige wesentliche Verfahren an, die auf einfachen Verteilungsannahmen über die Merkmalswerte Y_i in der Grundgesamtheit beruhen.

Beispiel: Eine Verteilungsannahme für das Inventurproblem wäre z. B. die bereits erwähnte Lognormal-Verteilung oder auch die Dreiecksverteilung.

Voraussetzung für die Abschätzung von S_Y ist dabei die Angabe einer oberen und einer unteren Schranke für die zu untersuchenden Merkmalswerte, die in der Regel immer anzugeben sind.

Beispiel: Bei der vorliegenden Inventur liegen beispielsweise Preise von 0,01 DM bis ca. 1000,- DM vor. Daraus ergibt sich für die Lagerposition (nach Erfahrung der Lagerverwaltung) ein kleinstes Y_i mit $Y_{i\min} = 5$ DM und ein größtes mit $Y_{i\max} = 10000$ DM. Unterstellt man eine Dreiecksverteilung (einseitig) mit Spannweite $R = Y_{i\max} - Y_{i\min}$, so erhält man eine Standardabweichung von

$$\tilde{s}_y = \frac{R}{\sqrt{18}} \approx 0,236 \cdot 9995 \text{ DM} \approx 2358,82 \text{ DM}.$$

Beispiel: Die allgemeine Vorgehensweise bei einer Stichprobeninventur mit Hilfe der einfachen Zufallsauswahl ist somit die folgende:

Zunächst bestimmt man die Grundgesamtheit und deren Umfang N (die Größe des Sortiments). Über rechtliche und inhaltliche Vorschriften erfolgt die Angabe von Sicherheitswahrscheinlichkeit und relativem Zufallsfehler, etwa $1 - \alpha = 0,95$ und $e = 0,02$. Vorerhebung oder Verteilungsannahmen über die Grundgesamtheit führen zu einer Abschätzung der Standardabweichung, die mit den Genauigkeitsforderungen den notwendigen Stichprobenumfang n ergibt. Dann ermittelt man n gleichverteilte Zufallszahlen zwischen 1 und N . Diese Zahlen geben die Artikelnummern an, deren Positionen ausgewählt werden sollen. Auswahl bedeutet dabei, Ermittlung der Positionsmenge x_i mit Preis p_i und Errechnen des Merkmalswertes $y_i = p_i x_i$. Mit diesen Werten sind dann die angegebenen Schätzer und Konfidenzintervalle zu berechnen.

Für das angesprochene Unternehmen auf dem Elektronikmarkt sind folgende Berechnungen denkbar:

Aus den $N = 18\,000$ Artikeln des Lagers soll eine Stichprobe vom Umfang $n = 1050$ durch einfache Zufallsauswahl gezogen werden. (Inwieweit durch diesen Stichprobenumfang die Genauigkeitsforderungen erfüllt werden, wird später erläutert.) Nach Auswahl der 1050 Einheiten und Auszählung der einzelnen Stückmengen konnte ein Stichprobenmittel von

$$\bar{y} = 523,30 \text{ DM}$$

festgestellt werden.

Das führt zu einer erwartungstreuen Schätzung des Gesamtlagerwertes Y von

$$\hat{Y} = 9\,419\,400 \text{ DM}.$$

Um die „Güte“ dieser Schätzung zu beurteilen, ist es notwendig Varianzen und Konfidenzintervalle zu berechnen. Für die Stichprobenvarianz ist errechnet worden:

$$s_y^2 = 61\,675,89 \text{ DM}^2.$$

Setzt man diesen Wert in die Schätzformel für die Varianz ein, so erhält man:

$$\widehat{\text{Var}} \hat{Y} = 1,792\,1251 \cdot 10^{10} \text{ DM}^2.$$

Dieser Wert ist natürlich schlecht zu interpretieren, und so berechnet man das Konfidenzintervall für den gesuchten Inventarwert Y . Mit $\alpha = 0,05$ ergibt sich folgendes Intervall:

$$[9\,157\,014,30; 9\,681\,785,80].$$

Mit einer Wahrscheinlichkeit von $1 - \alpha = 0,95$ liegt der gesuchte Inventarwert in diesem Intervall.

Bei der oben erläuterten Vorgehensweise wurde der Stichprobenumfang n aus der vorgeschriebenen Genauigkeit errechnet. Mit den hier durchgeföhrten Berechnungen bietet sich die umgekehrte Vorgehensweise an, und man ermittelt den relativen Zufallsfehler als:

$$\hat{e} = \frac{9681785,80}{9419400} - 1 = 0,0278559.$$

Die Bezeichnung \hat{e} statt e wurde gewählt, da zur Berechnung nicht die wahren Werte der Grundgesamtheit verwendet wurden, sondern die Schätzwerte aus der Stichprobe.

Man sieht also, daß der Umfang $n = 1050$ nicht ausreichend ist, um einen relativen Zufallsfehler kleiner 2% zu garantieren. Trotzdem kann man das Ergebnis als eine zufriedenstellende Schätzung des Gesamtlagerwertes bezeichnen.

So einfach wie sich diese Vorgehensweise hier darstellt, ist sie allerdings nicht. Der mathematisch einfachen Struktur stehen praktische Bedenken gegenüber.

Beispiel: Ein Lager ist nach Preisen und nicht nach Artikelnummern sortiert. Deshalb ist es schwierig, die Zufallsauswahl streng durchzuführen, was aber notwendig ist, um die Exaktheit des Verfahrens nicht zu gefährden.

Solche und ähnliche Probleme werden in Abschnitt 3 besprochen.

1.5. Geschichtete Zufallsauswahl

Bei der Durchführung einer einfachen Zufallsauswahl stellt sich oft heraus, daß die Varianzen der Schätzer übermäßig groß sind. Dies ist oft darin begründet, daß die Grundgesamtheit sehr inhomogen aufgebaut ist. Meist besitzt man ein solches Wissen bzw. diesbezügliche Vermutungen und ist somit bestrebt, dies durch die Art der Stichprobenerhebung zu berücksichtigen.

Beispiel: Im Inventurproblem besteht das a-priori-Wissen über die Grundgesamtheit in der bereits in Abschnitt 1.3 erwähnten 20–80-Regel. Das bedeutet, daß sich auf Grund der starken Inhomogenität bei einer einfachen Zufallsauswahl hohe Varianzen und somit hohe Stichprobenumfänge und Kosten ergeben würden.

Diese Überlegungen führen zur Durchführung einer **geschichteten Zufallsauswahl**. Diese Art der Auswahl trägt der Struktur der Grundgesamtheit in der Form Rechnung, daß die Grundgesamtheit in **Schichten** aufgeteilt wird, die in sich möglichst homogen, untereinander aber heterogen sein sollten. Diese Forderungen wären am besten erfüllt, wenn die Schichtung nach den Merkmalswerten Y_i durchgeführt würde. Eine solche Schichtung ist aber nicht möglich, da die Merkmalswerte unbekannt sind und zum Teil erst noch erhoben werden müssen. Deshalb muß im Einzelfall eine andere, im Sinnzusammenhang mit Y_i stehende **Schichtungsvariable** bestimmt werden.

Beispiel: Die Wahl der Schichtungsvariable ist ein viel diskutiertes Problem in der Stichprobentheorie. Im vorliegenden Fall der Inventur bietet sich als Schichtungsvariable der Preis P_i eines Artikels an. Dies hat zwei Gründe: erstens besteht der direkte Zusammenhang $Y_i = P_i X_i$ zum eigentlichen Merkmalswert, dem Positionswert Y_i , und zweitens ist der Aufbau des Lagers nach „Preiskategorien“ denkbar, d.h. daß z.B. eine getrennte Lagerung von „Pfennigartikeln“ und anderen Artikeln besteht.

Hat man sich für eine geschichtete Zufallsauswahl entschieden, so stellt sich die Frage nach Größe und Anzahl der Schichten. Oft ergibt sich die Antwort auf diese Frage direkt aus dem Zusammenhang, nämlich dann, wenn die Schichtungsvariable qualitative Werte annimmt. Bevorzugte Variablen sind beispielsweise in der Marktforschung geographische und demographische Variablen o. ä. Bei quantitativen Schichtungsvariablen ist es meist wesentlich schwieriger, das Problem zu lösen.

An die Frage der Schichtenanzahl können etwa die Fragen

- um wieviel wird die Varianz kleiner, wenn die Schichtenanzahl größer wird?
 - um wieviel wird die Untersuchung bei größerer Schichtenanzahl teurer?
- geknüpft werden.

Die exakte Beantwortung einer solchen Frage ist sehr problemabhängig und muß im Einzelfall immer wieder neu entschieden werden. Wie die sogenannte **Kumulationsregel von Dalenius und Hodges** zur Schichtenbildung, haben sich allerdings einige zu beachtende Faustregeln herausgebildet, die sich besonders an die erwähnten Fragestellungen anlehnen.

Beispiel: Das Problem der Schichtenbildung bzw. -anzahl in der Stichprobeneinventur wird so gelöst, daß das Lager in ca. 10–14 Schichten aufgeteilt wird, die bei zunehmendem Preis immer breiter werden.

Im vorliegenden Beispiel sollen 13 Schichten gebildet werden. Die Zuordnung in die Schichten soll nach dem Stückpreis P_i wie in **Tab. 1** erfolgen.

Tab. 1: Schichtenbildung im Beispiel

Schicht	Stückpreis von ... bis unter ... DM		
1	0	-	1
2	1	-	2
3	2	-	3
4	3	-	4
5	4	-	5
6	5	-	7
7	7	-	9
8	9	-	12
9	12	-	20
10	20	-	30
11	30	-	50
12	50	-	100
13	100 und größer		

In der Stichprobentheorie ist es üblich, bei geschichteten Auswahlverfahren folgende Schreibweise zu benutzen:

L Anzahl der Schichten,

N_h Umfang der h -ten Schicht in der Grundgesamtheit,

$N = \sum_{h=1}^L N_h$ Umfang der Grundgesamtheit,

Y_{hi} Merkmalswert der i -ten Einheit in der h -ten Schicht,

$Y_h = \sum_{i=1}^{N_h} Y_{hi}$ Summe der Merkmalswerte in der h -ten Schicht,

$Y = \sum_{h=1}^L Y_h$ Summe der Merkmalswerte in der Grundgesamtheit,

$W_h = \frac{N_h}{N}$ Schichtgewicht

und für die entsprechenden Größen in der Stichprobe:

n_h Stichprobenumfang in der h -ten Schicht,

$n = \sum_{h=1}^L n_h$ Stichprobenumfang,

y_{hi} Merkmalswert der i -ten gezogenen Einheit in der h -ten Schicht

$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ durchschnittlicher Merkmalswert der Stichprobenelemente in der h -ten Schicht.

Beispiel: Für unser Beispiel heißt dies: $L = 13$. Die Umfänge N_h der h -ten Schicht sind bekannt, vgl. **Tab. 2**.

Tab. 2: Umfänge der 13 Schichten

h	N_h
1	2742
2	2777
3	2112
4	1512
5	1190
6	1669
7	1162
8	1190
9	1711
10	816
11	646
12	354
13 (= L)	119
insgesamt	18000

Bei einer geschichteten Stichprobe zieht man aus den L Schichten jeweils eine Zufallsstichprobe vom Umfang n_h . Dabei ist es notwendig, daß diese L Zufallsstichproben unabhängig voneinander sind. Ist jede einzelne Zufallsauswahl eine einfache Zufallsstichprobe, so ist

$$\hat{Y} = \sum_{h=1}^L N_h \bar{y}_h$$

ein erwartungstreuer Schätzer für die Summe Y der Merkmalswerte in der Grundgesamtheit; dieser hat die Varianz

$$\text{Var } \hat{Y} = \sum_{h=1}^L N_h^2 \frac{1}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_h^2,$$

die durch

$$\widehat{\text{Var } \hat{Y}} = \sum_{h=1}^L N_h^2 \frac{1}{n_h} \left(1 - \frac{n_h}{N_h}\right) s_h^2$$

geschätzt werden kann. Hierbei bezeichnet

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$$

die wahre Varianz in der h -ten Schicht und

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$$

einen Schätzer, nämlich die Stichprobenvarianz der h -ten Schicht, hierfür.

Natürlich kann jede einzelne Zufallsauswahl auch anders erfolgen, etwa auch eine geschichtete Auswahl sein. Allgemein ist dann

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h, \quad \text{Var } \hat{Y} = \sum_{h=1}^L \text{Var } \hat{Y}_h \quad \text{und} \quad \widehat{\text{Var } \hat{Y}} = \sum_{h=1}^L \widehat{\text{Var } \hat{Y}_h}.$$

Beispiel: Für das in unserem Beispiel zugrundegelegte Lager sind die Beobachtungen aus **Tab. 3** gemacht worden.

Tab. 3: Beobachtungen im Beispiel

h	n _h	ȳ _h	s _h ²
1	156	289,14	14879,30
2	168	423,05	7574,83
3	119	479,46	5279,76
4	81	472,73	6008,46
5	64	514,35	11515,24
6	96	692,06	33699,55
7	67	474,40	26602,23
8	66	624,14	57143,62
9	93	626,74	85686,40
10	68	731,67	142708,99
11	44	779,21	151454,02
12	21	792,48	131794,88
13	7	947,56	28967,99

Damit sind die Berechnungen aus **Tab. 4** möglich.

Tab. 4: Schätzungen aus den Daten der Tabelle 3

h	Ŷ _h	Var Ŷ _h
1	792821,88	6,7632256 · 10 ⁸
2	1174809,85	3,2667325 · 10 ⁸
3	1012619,52	1,8675336 · 10 ⁸
4	714767,76	1,6049798 · 10 ⁸
5	612076,50	2,4108954 · 10 ⁸
6	1155048,14	9,2159037 · 10 ⁸
7	551252,80	5,0520017 · 10 ⁸
8	742726,60	11,580761 · 10 ⁸
9	1072352,10	25,506888 · 10 ⁸
10	597042,72	12,809559 · 10 ⁸
11	503369,66	13,386195 · 10 ⁸
12	280537,92	7,3982114 · 10 ⁸
13	112759,64	0,55155053 · 10 ⁸
insgesamt	9 322 185,13	1,0141444 · 10 ¹⁰

Das heißt, ein erwartungstreuer Schätzer für den Lagerwert Y ist durch

$$\hat{Y} = 9\,322\,185,13 \text{ DM}$$

gegeben. Die Varianz dieses Schätzers wird durch

$$\widehat{\text{Var}} \hat{Y} = 1,0141444 \cdot 10^{10} \text{ DM}^2$$

geschätzt.

Hat man sich für eine Schichtung entschieden, so muß im nächsten Schritt die Art der Aufteilung bestimmt werden. Das soll bedeuten, daß ein Stichprobenumfang n

in die Schichtstichprobenumfänge n_h aufgeteilt werden muß. Hier werden vor allem die folgenden drei Aufteilungsmöglichkeiten verwandt:

- die optimale Aufteilung oder Neyman-Tschuprow-Aufteilung,
- die proportionale Aufteilung,
- die Aufteilung nach Auswahl der Stichprobe.

1.5.1. Die optimale Aufteilung (Neyman-Tschuprow-Aufteilung)

Man spricht von einer **optimalen Aufteilung** des Gesamtstichprobenumfangs in die Stichprobenumfänge n_h , wenn diese Umfänge so gewählt werden, daß man bei gegebener Varianz des Schätzers die Kosten der Untersuchung bzw. bei gegebenen Kosten die Varianz des Schätzers minimiert, vgl. auch Schneeberger (1984).

Bezeichnet man mit C die Gesamtkosten einer Erhebung, die sich aus fixen Kosten c_0 (z. B. Kosten der Einstellung des Lagerbetriebs während der Erhebung) und variablen Kosten pro erhobener Einheit zusammensetzen, so ist die einfachste und wohl gebräuchlichste Kostenfunktion

$$C = c_0 + \sum_{h=1}^L c_h n_h,$$

wobei c_h die Kosten pro erhobener Einheit in der h -ten Schicht bezeichnet; diese werden für alle Einheiten einer Schicht als gleich angenommen.

Bei so strukturierten vorgegebenen Gesamtkosten C soll nun erreicht werden, daß die Varianz des Schätzers \bar{Y} für die Summe der Merkmalswerte in der Grundgesamtheit minimal wird.

Dies wird erreicht, wenn man als Stichprobenumfang der h -ten Schicht für $h = 1, \dots, L$

$$n_h = \frac{n N_h S_h / \sqrt{c_h}}{\sum_{k=1}^L N_k S_k / \sqrt{c_k}}$$

mit

$$n = \frac{(C - c_0) \sum_{h=1}^L N_h S_h / \sqrt{c_h}}{\sum_{h=1}^L N_h S_h \cdot \sqrt{c_h}}$$

wählt, wobei natürlich $n_h \leq N_h$ für $h = 1, \dots, L$ gelten muß. Beide Formeln enthalten allerdings noch die unbekannten Schichtvarianzen S_h^2 der Grundgesamtheit. Um diese zu erhalten wird mit Hilfe einer Vorerhebung eine Varianzschätzung durchgeführt, so daß die Schätzer, ähnlich wie bei der Berechnung von n bei der einfachen Zufallsauswahl, eingesetzt werden können, oder man geht den Weg der Ausnutzung von Verteilungsannahmen zur Abschätzung der unbekannten Schichtvarianzen.

Oft erweist sich, daß annähernd gilt

$$c_1 = c_2 = \dots = c_L = c,$$

d.h. daß die Kosten pro Erhebungseinheit in allen Schichten gleich c sind.

Beispiel: Die Durchführung der Lagerinventur ist in den letzten Jahren weiter funktionalisiert worden. Aus diesem Grund kann man mitunter annehmen, daß

Kostenunterschiede in der Inventur von Schicht zu Schicht nur in sehr geringem Maße bei der Durchführung einer Stichprobenerhebung auftreten.

Vereinfachend ergibt sich dann mit

$$C = c_0 + nc,$$

als optimaler Umfang der Stichprobe aus der h-ten Schicht für $h = 1, \dots, L$

$$n_h = \frac{n N_h S_h}{\sum_{k=1}^L N_k S_k} \quad \text{mit} \quad n = \frac{1}{c} (C - c_0).$$

Als numerisches Beispiel für eine Neyman-Tschuprow-Aufteilung kann das bereits erwähnte Datenmaterial verwendet werden. Dabei wird die Stichprobe vom Umfang 1050 als Vorerhebung interpretiert, um Schätzer für die Schichtvarianz zu erhalten.

Unterstellt wird folgende lineare Kostenfunktion:

$$100000 = 37028,45 + n \cdot 22,37,$$

d. h. die Gesamtkosten der Stichprobeninventur sollen $C = 100000$ DM betragen, es liegen fixe Kosten von $c_0 = 37028,45$ DM vor, und die Erhebung einer Einheit kostet $c = 22,37$ DM.

Daraus berechnet sich ein Stichprobenumfang von

$$n = 2815.$$

Mit Hilfe der vereinfachten Aufteilungsformel können nun die Stichprobenumfänge in den Schichten bestimmt werden; vgl. **Tab. 5**.

Tab. 5: Stichprobenumfänge mittels vereinfachter Aufteilungsformel

h	s_h	$N_h s_h$	n_h
1	121,98	334 469,16	318
2	87,03	241 682,31	230
3	72,66	153 457,92	146
4	77,51	117 195,12	112
5	107,31	127 698,90	122
6	183,57	306 378,33	291
7	163,10	189 522,20	180
8	239,05	284 469,50	271
9	292,72	500 843,92	476
10	377,77	308 260,32	293
11	389,17	251 403,82	239
12	363,04	128 516,16	123
13	170,20	20 251,46	20
insgesamt	-	2964 167,80	2821

Hierbei gilt allerdings nicht mehr $\sum_{h=1}^{13} n_h = n$, da bei allen Berechnungen zur Bestimmung von n_h aufgerundet wurde. Die Genauigkeit wird hierdurch garantiert

und der Stichprobenumfang um 6 erhöht, was nur unwesentlich höhere Kosten verursacht.

Besonders auffällig ist das starke Gewicht, das bei optimaler Aufteilung der neunten Schicht (12–20 DM) zufällt. 9,51% der Grundgesamtheit fallen in diese Schicht, aber 16,87% der Stichprobeneinheiten sollen nach der Neyman-Tschuprow-Aufteilung aus dieser Schicht sein. Die hohe Varianz in der neunten Schicht, die diese Forderung verursacht, ist durch die stark streuende Anzahl der Lagermengen bedingt. Es liegen Stückzahlen (in der Erhebung) von 9 bis 114 vor. Eine Erklärung dafür kann eine unterschiedliche Nachfrage nach Gütern in dieser Schicht sein.

Die erreichte minimale Varianz von \hat{Y} bei einer Neyman-Tschuprow-Aufteilung bei vorgegebenen Gesamtkosten C ist dann, falls $n_h \leq N_h$ für $h = 1, \dots, L$,

$$\begin{aligned} \text{Var}_{\text{opt}} \hat{Y} &= \frac{1}{n} \left(\sum_{h=1}^L N_h S_h \sqrt{c_h} \right) \left(\sum_{h=1}^L N_h S_h / \sqrt{c_h} \right) - \sum_{h=1}^L S_h^2 N_h \\ &= \left[\frac{1}{n} \left(\sum_{h=1}^L N_h S_h \right)^2 - \sum_{h=1}^L S_h^2 N_h \quad \text{für } c = c_1 = \dots = c_L \right]; \end{aligned}$$

diese wird (allerdings nicht erwartungstreu) durch

$$\begin{aligned} \widehat{\text{Var}}_{\text{opt}} \hat{Y} &= \frac{1}{n} \left(\sum_{h=1}^L N_h s_h \sqrt{c_h} \right) \left(\sum_{h=1}^L N_h s_h / \sqrt{c_h} \right) - \sum_{h=1}^L s_h^2 N_h \\ &= \left[\frac{1}{n} \left(\sum_{h=1}^L N_h s_h \right)^2 - \sum_{h=1}^L s_h^2 N_h \quad \text{für } c = c_1 = \dots = c_L \right] \end{aligned}$$

geschätzt.

Beispiel: Verwendet man wiederum die aus der Vorerhebung erhaltenen Varianzen als Schätzgrößen für die Schichtvarianzen, so berechnet sich eine Schätzung für die optimale Varianz nach Neyman-Tschuprow, vgl. Tab. 6, als

$$\widehat{\text{Var}}_{\text{opt}} \hat{Y} = 2,4526349 \cdot 10^9 \text{ DM}^2.$$

Tab. 6: Berechnung der Varianz bei optimaler Aufteilung

h	$s_h^2 \cdot N_h$
1	40 799 041,0
2	21 035 303,0
3	11 150 853,0
4	9 084 791,5
5	13 703 136,0
6	56 244 549,0
7	30 911 791,0
8	68 000 908,0
9	146 609 430,0
10	116 450 536,0
11	97 839 297,0
12	46 655 388,0
13	3 447 190,8
insgesamt	$6,6193221 \cdot 10^8$

Entscheidet man sich bei der Stichprobeninventur für eine geschichtete Zufallsauswahl mit Neyman-Tschuprow-Aufteilung, so kann ein den gesetzlichen Vorschriften genügendes Ergebnis erzielt werden. Dies zeigt nicht nur das in der Literatur immer größer werdende Interesse an diesem Verfahren, sondern auch ein Vergleich von Voll- und Stichprobeninventur, der als Pilotstudie 1978 bei einem Werkstattlager durchgeführt wurde, vgl. Wienhold (1978).

Hierbei wurde ein Lager betrachtet, das nach Artikelpreisen in zehn Schichten zerlegt wurde. Aus den $N = 24476$ Artikelpositionen wurden $n = 3703$ Positionen nach optimaler Aufteilung gezogen. Das entspricht einem Auswahlsatz von ca. 15%, so daß der Rationalisierungseffekt des Verfahrens enorm ist. Der Unterschied zwischen geschätztem Lagerwert und Buchbestand des Lagers betrug 1,168%. Der relative Zufallsfehler lag mit $e = 0,01925$ unter der vom Gesetzgeber vorgeschriebenen Sicherheitsgrenze von 2%.

1.5.2. Die proportionale Aufteilung

Eine optimale Aufteilung hat trotz ihrer varianzminimierenden Eigenschaften einige Nachteile. Diese sind vor allem die schwierige Berechnung der Stichprobenumfänge und die Möglichkeit, daß für einen berechneten Stichprobenumfang gelten kann: $n_h > N_h$.

Eine wesentlich einfachere, sinnvolle Methode ist die der **proportionalen Aufteilung**. Man sagt, daß eine Aufteilung proportional ist, wenn die Auswahlsätze $\frac{n_h}{N_h}$ in allen Schichten gleich dem Auswahlsatz in der Grundgesamtheit sind, d. h. wenn gilt

$$\frac{n_h}{n} = W_h, \quad \text{d. h.} \quad n_h = N_h \cdot \frac{n}{N}.$$

Schätzt man zunächst aus einer Vorerhebung S_h und Y durch \tilde{s}_h und \tilde{y} , so kann daraus der Gesamtstichprobenumfang n zur Einhaltung eines Zufallsfehlers e mit Sicherheitswahrscheinlichkeit $1 - \alpha$ abgeschätzt werden. Als erste Näherung wählt man

$$n_0 = N \cdot \left(\frac{u_{1-\alpha/2}}{e \cdot \tilde{y}} \right)^2 \sum_{h=1}^L N_h \tilde{s}_h^2$$

und bei kleinem Auswahlsatz $n = n_0$ bzw. bei großem Auswahlsatz

$$n = \frac{n_0}{1 + n_0/N}.$$

Die Varianz des Schätzers \hat{Y} bei der proportionalen Aufteilung ist dann

$$\text{Var}_{\text{prop}} \hat{Y} = \frac{N}{n} \sum_{h=1}^L N_h S_h^2 - \sum_{h=1}^L N_h S_h^2$$

und kann durch

$$\widehat{\text{Var}}_{\text{prop}} \hat{Y} = \frac{N}{n} \sum_{h=1}^L N_h \tilde{s}_h^2 - \sum_{h=1}^L N_h \tilde{s}_h^2$$

erwartungstreu geschätzt werden.

| **Beispiel:** Interpretiert man die bislang durchgeführten numerischen Berechnungen

als Ergebnisse einer Vorerhebung, so läßt sich der Stichprobenumfang berechnen, der notwendig ist, um eine gewisse Genauigkeit zu erreichen.

Setzt man z. B. $e = 0,0278559$ und $\alpha = 0,05$ (vgl. die Werte zur einfachen Zufallsauswahl), so erhält man

$$n = 654.$$

Eine Varianzberechnung soll hier für den Fall $n = 2821$ (vgl. optimale Aufteilung) durchgeführt werden. Es ergibt sich

$$\widehat{\text{Var}}_{\text{prop}} \hat{Y} = 3,5616693 \cdot 10^9 \text{ DM}^2.$$

1.5.3. Aufteilung nach Auswahl der Stichprobe

Eine weitere Möglichkeit einer Zufallsauswahl nach Schichten wird durch eine einfache Zufallsstichprobe gegeben. Nach der einfachen Zufallsauswahl vom Umfang m wird jedes Element der Stichprobe seiner Schicht zugeordnet, d. h. die Schichtenbildung erfolgt nachträglich. Durch diese Vorgehensweise ist die Anzahl der Stichprobenelemente in der Schicht nicht mehr von vornherein festgelegt, sondern eine zufällige Größe; daher werden diese Umfänge hier mit m_h , $h = 1, \dots, L$, bezeichnet.

Bei diesem Auswahlverfahren ist

$$\hat{Y} = N\bar{y} = N \sum_{h=1}^L W_h \bar{y}_h = \sum_{h=1}^L N_h \left(\frac{1}{m_h} \sum_{i=1}^{m_h} y_{hi} \right)$$

ein erwartungstreuer Schätzer für die Summe der Merkmalswerte in der interessierenden Grundgesamtheit. Die Varianz dieses Schätzers ist approximativ

$$\text{Var } \hat{Y} \approx N^2 \left(\frac{N-m}{Nm} \sum_{h=1}^L W_h S_h^2 + \frac{1}{m^2} \sum_{h=1}^L (1-W_h) S_h^2 \right).$$

Ein erwartungstreuer Schätzer für $\text{Var } \hat{Y}$ ist

$$\widehat{\text{Var}} \hat{Y} = N^2 \left[\frac{N-m}{m(N-1)} \left(\frac{1}{N} \sum_{h=1}^L \frac{N_h}{m_h} \sum_{j=1}^{m_h} y_{hj}^2 - \bar{y}_{st}^2 + \widehat{\text{Var}} \bar{y}_{st} \right) \right]$$

wobei:

$$\begin{aligned} \bar{y}_{st} &= \frac{1}{N} \sum_{h=1}^L N_h \bar{y}_h \\ &= \frac{1}{N} \sum_{h=1}^L N_h \left(\frac{1}{m_h} \sum_{i=1}^{m_h} y_{hi} \right), \\ \widehat{\text{Var}} \bar{y}_{st} &= \frac{1}{N^2} \sum_{h=1}^L N_h^2 \frac{1}{m_h} \left(1 - \frac{m_h}{N_h} \right) s_h^2. \end{aligned}$$

(Der Index „st“ steht für „stratified“, „geschichtet“.)

Die Schichtenbildung nach Auswahl der Stichprobe besitzt gegenüber anderen Verfahren den Vorteil, daß eine Schichtung nach den Merkmalswerten vorgenommen werden kann. Dies kann von Fall zu Fall eine größere Genauigkeit bringen; diese ist aber explizit nicht berechenbar. Der Nachteil des Verfahrens ist die „doppelte Zufälligkeit“ der Schätzer, da sowohl die erhobenen Merkmalswerte wie auch

die Stichprobenumfänge Zufallsvariablen darstellen. Folglich wird nicht nur die Berechnung der Schätzer und Varianzen kompliziert, sondern es ergibt sich auch ein gewisser Grad der Unzuverlässigkeit, d. h. der Zufallsfehler ist schlecht kontrollierbar. Aus diesen Gründen wird das Verfahren in der Praxis wenig angewandt.

1.5.4. Genauigkeitsvergleiche

Eigentliche Motivation für die Durchführung geschichteter Auswahlverfahren ist die Varianzverringerung der Schätzer, da man eine Anpassung an die Grundgesamtheit vorzunehmen wünscht. Nun ist die eigentliche Zusammensetzung der Grundgesamtheit aber unbekannt, und so ist es interessant zu untersuchen, wie groß die Varianzunterschiede bei den einzelnen Verfahren sind.

Allgemein gilt bei vernachlässigten Auswahlsätzen folgende Beziehung

$$\text{Var}_{\text{opt}} \leq \text{Var}_{\text{prop}} \leq \text{Var}_{\text{SRS}}$$

für die Varianzen von optimaler (opt) und proportionaler Auswahl (prop) sowie einfacher Zufallsauswahl (SRS).

Beispiel: Dieses Ergebnis wird auch an den bisherigen Berechnungen deutlich, bei denen immer das gleiche Datenmaterial verwendet wurde. Diese Berechnungen sollen hier noch einmal zusammengefaßt dargestellt werden.

einfache Zufallsauswahl ($e = 0,0278559$)

$$\begin{aligned} n &= 1050 \\ \hat{Y} &= 9419400 \text{ DM} \\ \text{Var } \hat{Y} &= 17,921251 \cdot 10^9 \text{ DM}^2 \end{aligned}$$

geschichtete Zufallsauswahl (als Vorerhebung)

$$\begin{aligned} n &= 1050 \\ \hat{Y} &= 9322185,13 \text{ DM} \\ \text{Var } \hat{Y} &= 10,141444 \cdot 10^9 \text{ DM}^2 \end{aligned}$$

geschichtete Auswahl (mit Neyman-Tschuprow-Aufteilung)

$$\begin{aligned} n &= 2821 \\ \text{Var}_{\text{opt}} \hat{Y} &= 2,4526349 \cdot 10^9 \text{ DM}^2 \end{aligned}$$

geschichtete Auswahl (proportionale Aufteilung; $e = 0,0278559$)

$$n = 654$$

geschichtete Auswahl (proportionale Aufteilung)

$$\begin{aligned} n &= 2821 \\ \text{Var}_{\text{prop}} \hat{Y} &= 3,5616693 \cdot 10^9 \text{ DM}^2. \end{aligned}$$

Daraus lassen sich Aussagen der folgenden Art ableiten :

- bei gleicher Genauigkeit wird bei der proportional aufgeteilten Stichprobe ein kleinerer Stichprobenumfang n benötigt als bei der einfachen Zufallsauswahl;
- bei gleichem Stichprobenumfang n ist die Varianz der geschichteten Auswahl kleiner als die bei einfacher Zufallsauswahl;

– bei gleichem Stichprobenumfang n ist die Varianz bei optimaler Aufteilung am geringsten.

Betrachtet man die Varianzen bei den verschiedenen Auswahlverfahren genauer, so ergibt sich als Unterschied von optimaler und proportionaler Aufteilung

$$\text{Var}_{\text{prop}} \hat{Y} - \text{Var}_{\text{opt}} \hat{Y} = \frac{N^2}{n} \sum_{h=1}^L (S_h - \bar{S})^2 W_h$$

wobei $\bar{S} = \sum_{h=1}^L W_h \cdot S_h$.

Daraus kann man schließen, daß eine optimale Aufteilung immer dann vorzuziehen ist, wenn große Varianzunterschiede bei den Schichten auftreten.

Beispiel: Diese Erkenntnis rechtfertigt noch einmal die Durchführung einer optimalen Aufteilung im Inventurproblem. Hier war die Schichtungsvariable der Preis. Dabei ist in einem Lagerbetrieb häufig zu beobachten, daß bei Niedrigpreisartikeln Mindestlagerbestände vorhanden sind, so daß in diesen Schichten kleine Varianzen zu erwarten sind, während in den Schichten höherer Preise eine solche Lagerpolitik nicht durchgeführt wird, so daß hohe Varianzen vorliegen.

Im Vergleich von proportionaler Aufteilung und einfacher Zufallsauswahl ergibt sich:

$$\text{Var}_{\text{SRS}} \hat{Y} - \text{Var}_{\text{prop}} \hat{Y} \simeq \frac{N^2}{n} (1-f) \sum_{h=1}^L W_h (\bar{Y}_h - \bar{Y})^2 > 0,$$

falls große Schichten vorliegen, so daß in diesem Fall eine proportionale Aufteilung immer einen Genauigkeitsgewinn erbringt. Die Gültigkeit der verwandten Approximation ist dabei zwar auf große Schichten begrenzt, doch ist in praktischen Fällen fast immer mit einer Varianzverringerung durch proportionale Aufteilung zu rechnen.

1.6. Klumpenstichprobenverfahren

Die Einführung der geschichteten Zufallsauswahl wurde durch die Möglichkeit, im Gegensatz zur einfachen Zufallsauswahl eine Varianzverringerung zu erreichen, motiviert. Tatsächlich haben sich diese Verfahren auch als bestmögliche erwiesen; jedoch treten, besonders bei der Beachtung des der Theorie zugrundeliegenden Zufallsprinzips, bei der Erhebung oft große Schwierigkeiten auf. Einen großen Einfluß auf diese Problematik hat im allgemeinen die Größe der Grundgesamtheit. Es hat sich beispielsweise als unmöglich herausgestellt, die Wohnbevölkerung eines Landes genau zu erfassen und in Schichten aufzuteilen. Diese Überlegungen führen besonders für den Fall großer Gesamtheiten zu der Auswahlmethode mit Hilfe von sogenannten **Klumpen** (*clusters*).

Die Durchführung eines **Klumpenverfahrens** wird folgendermaßen praktiziert. Zunächst wird die Grundgesamtheit in Teile (Klumpen) zerlegt. Nach einem Zufallsprinzip werden dann einige dieser Klumpen ausgewählt. Erhebt man dann alle Einheiten der bislang gewählten Klumpen, so spricht man von einem **einstufigen Verfahren**. zieht man jedoch aus diesen Klumpen nur einige Einheiten zufällig, so spricht man von **mehrstufigen Auswahlverfahren**.

Beispiel: Auch im Inventurproblem sind Klumpenverfahren denkbar. Beispielsweise besitzt der betrachtete Betrieb Verkaufsstellen im ganzen Bundesgebiet.

Grundgesamtheit sind somit alle Artikelpositionen in der Gesamtheit aller Lager des Betriebes. Geht man davon aus, daß die Verkäufe (Abgänge aus dem Lager) in allen Verkaufsstellen ähnlich sind, so sind durch die gleiche Lagerpolitik ähnliche Bestände in den verschiedenen Lagern zu erwarten, und man kann sich auf die Untersuchung einiger Lager beschränken. Dieses „Beschränken“ würde einer Auswahl von Klumpen entsprechen, wobei ein Klumpen der Lagerbestand einer Verkaufsstelle ist. Je nach dem, ob man in den ausgewählten Stellen nun eine Vollinventur oder eine weitere Auswahl vornimmt, liegt dann ein einstufiges oder mehrstufiges Auswahlverfahren vor.

Bei der Durchführung von Klumpenverfahren bedient man sich üblicherweise folgender Terminologie:

K – Anzahl der Klumpen

M_i – Anzahl der Einheiten im i -ten Klumpen, $i = 1, \dots, K$; $\sum_{i=1}^K M_i = N$,

E_{ij} – j -te Einheit im i -ten Klumpen, $i = 1, \dots, K, j = 1, \dots, M_i$,

Y_{ij} – Merkmalswert von E_{ij} .

Die Notwendigkeit eines Klumpenverfahrens ergibt sich meist durch Größe und Struktur der Grundgesamtheit. Somit wird die Frage nach der Festlegung der Anzahl und der Art der Klumpen oft schon durch die zu untersuchende Gesamtheit bestimmt. Beispielsweise sind bei der Grundgesamtheit der Wohnbevölkerung die Gemeinden als Klumpen vorstellbar bzw. beim Inventurproblem die Vertriebsstellen einer Unternehmung. Man sollte bei der Festlegung von Klumpen beachten, daß die Unterschiede zwischen den Klumpen gering sein sollten, innerhalb eines jeden Klumpens aber eine starke Inhomogenität vorliegt.

1.6.1. Einstufige Auswahlverfahren

Gehen alle Einheiten von k ausgewählten Klumpen in die Erhebung ein, so liegt ein einstufiges Auswahlverfahren vor.

Betrachtet man zunächst den Spezialfall, daß $M_i = M$ für alle $i = 1, \dots, K$ gilt, dann sind folgende Aussagen möglich. Werden aus K Klumpen (vom Umfang M) k Klumpen als einfache Zufallsstichprobe ausgewählt, dann ist

$$\hat{Y} = \frac{K}{k} \sum_{i=1}^k Y_i$$

ein erwartungstreuer Schätzer für Y mit der Varianz

$$\begin{aligned} \text{Var } \hat{Y} &= \frac{K^2}{k} \left(1 - \frac{k}{K}\right) \frac{1}{K-1} \sum_{i=1}^K (Y_i - \bar{Y})^2 \\ &= \frac{K^2}{k} \left(1 - \frac{k}{K}\right) \frac{KM-1}{K-1} \cdot S_Y^2 (1 + (M-1)\varrho), \end{aligned}$$

wobei $Y_i = \sum_{j=1}^M Y_{ij}$

die i -te Klumpensumme,

$$\bar{Y} = \frac{1}{K} \sum_{i=1}^K Y_i$$

den durchschnittlichen Klumpenwert,

$$\bar{Y}_e = \frac{1}{M} \bar{Y}$$

den durchschnittlichen Merkmalswert,

$$S_Y^2 = \frac{1}{N-1} \sum_{i=1}^K \sum_{j=1}^M (Y_{ij} - \bar{Y}_e)^2$$

die Varianz in der Gesamtheit und

$$\varrho = \frac{1}{(M-1) \cdot (N-1) \cdot S_Y^2} \sum_{i=1}^K \sum_{\substack{j=1 \\ i \neq j}}^M (Y_{ij} - \bar{Y}_e)(Y_{ij} - \bar{Y}_e)$$

den **Intraklass-Korrelationskoeffizienten** bezeichnen.

Für den Intraklass-Korrelationskoeffizienten ϱ gilt:

$$-\frac{1}{M-1} \leq \varrho \leq 1.$$

Er stellt ein Maß des Zusammenhangs zwischen den beobachteten Merkmalswerten innerhalb eines Klumpens dar. Der Faktor $(1 + (M-1)\varrho)$ in der Formel für $\text{Var } \hat{Y}$ wird als **Varianzaufblähungsfaktor** bezeichnet. Dieser Faktor hat seine Bedeutung darin, daß er den Multiplikator angibt, mit dem bei gleichem Stichprobenumfang die Varianz bei einfacher Zufallsauswahl versehen werden muß, um die Varianz durch das Klumpenverfahren zu erhalten. Daraus ist die Aussage zu gewinnen, daß für $\varrho = 0$, d.h. wenn die Merkmalswerte innerhalb eines Klumpens nicht zusammenhängen, Klumpenverfahren und einfache Zufallsstichprobe die gleiche Genauigkeit liefern.

Beispiel: Im Inventurproblem liegt eine denkbare Anwendungsmöglichkeit für das vorgestellte spezielle einstufige Auswahlverfahren. Der im allgemeinen ziemlich hypothetische Ansatz $M_i = M$, $i = 1, \dots, K$, ist hier nämlich wirklich erfüllt, da alle Verkaufslager das gleiche Sortiment haben.

Weiterhin liegt ein Anwendungsgrund für dieses Verfahren in der regelmäßigen Wiederholung (eine Rechnungsperiode, meist ein Jahr) der Durchführung. Hat man dadurch beispielsweise erfahren, daß $\varrho = 0$ angenommen werden kann, so wird man das Klumpenverfahren der einfachen Zufallsstichprobe vorziehen, da dieses im allgemeinen kostengünstiger sein wird.

Liegen Klumpen unterschiedlicher Größe vor, so ist

$$\hat{Y}_a = \frac{K}{k} \sum_{i=1}^k Y_i$$

ein erwartungstreuer Schätzer für den Gesamtwert Y mit Varianz

$$\text{Var } \hat{Y}_a = \frac{K^2}{k} \left(1 - \frac{k}{K}\right) \cdot \frac{1}{K-1} \sum_{i=1}^K (Y_i - \bar{Y})^2.$$

Dieser Schätzer hat den Nachteil einer sehr großen Varianz, besonders bei extrem schwankenden Klumpengrößen. Aus diesem Grunde wird, vor allem dort, wo ver-

schiedene Klumpengrößen verstärkt auftreten, häufig der nicht erwartungstreue Mittelwert aus den Klumpen zur Schätzung von Y verwendet:

$$\hat{Y}_b = N \frac{1}{k} \sum_{i=1}^k \bar{Y}_i;$$

für die Varianz dieses Schätzers gilt:

$$\text{Var } \hat{Y}_b = N^2 \left(\frac{1}{k} - \frac{1}{K} \right) \cdot \frac{1}{K-1} \sum_{i=1}^k \left(\bar{Y}_i - \frac{1}{K} \sum_{j=1}^K \bar{Y}_j \right)^2.$$

Diese Varianz hängt noch von den Schwankungen der Klumpenmittelwerte ab. Im Gegensatz dazu war bei der Varianz von \hat{Y}_a eine Abhängigkeit von den Merkmalssummen zu beobachten. Eine weitere Möglichkeit der Schätzung von Y wird durch den ebenfalls verzerrten Verhältnisschätzer gegeben:

$$\hat{Y}_c = \frac{N}{\sum_{i=1}^k M_i} \sum_{i=1}^k Y_i.$$

Dieser besitzt zwar eine größere Varianz als \hat{Y}_b , jedoch ist sein mittlerer quadratischer Fehler (MSE), vgl. Kap. III, geringer.

Es gilt:

$$\text{Var } \hat{Y}_c \approx K^2 \left(\frac{1}{k} - \frac{1}{K} \right) \cdot \frac{1}{K-1} \sum_{i=1}^k M_i^2 (\bar{Y}_i - \bar{Y}_c)^2.$$

1.6.2. Mehrstufige Auswahlverfahren

Führt man ein einstufiges Klumpenauswahlverfahren durch, so kann man zwar die Anzahl der auszuwählenden Klumpen, nicht aber den Stichprobenumfang bestimmen. Da man dies allein schon aus Kostenüberlegungen heraus in den meisten Fällen aber muß, ist es nötig, das Verfahren entsprechend abzuwandeln bzw. zu erweitern.

In einem ersten Schritt werden mittels einfacher Zufallsstichprobe k der K Klumpen ausgewählt. Im zweiten Schritt werden aus den jeweils M_i Einheiten des i -ten gezogenen Klumpens, $i = 1, \dots, k$, m_i Einheiten ausgewählt an denen dann die Merkmalswerte bestimmt werden. Diese Merkmalswerte werden mit y_{ij} , $i = 1, \dots, k$, $j = 1, \dots, m_i$ bezeichnet.

Führt man dieses Verfahren durch, so ist

$$\hat{Y} = \frac{K}{k} \sum_{i=1}^k M_i \bar{y}_i$$

ein erwartungstreuer Schätzer für Y . Seine Varianz berechnet sich zu

$$\text{Var } \hat{Y} = K^2 \left(\frac{1}{k} - \frac{1}{K} \right) \cdot S_b^2 + \frac{K}{k} \sum_{i=1}^k M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_{wi}^2,$$

wobei

$$S_b^2 = \frac{1}{K-1} \sum_{i=1}^k (Y_i - \bar{Y})^2$$

die Varianz zwischen (**between**) den Klumpen und

$$S_{wi}^2 = \frac{1}{M_i - 1} \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y}_i)^2$$

die Varianz innerhalb (**within**) des i-ten Klumpens ($i = 1, \dots, k$) ist.

Diese Varianz kann erwartungstreu geschätzt werden, indem man die Stichprobenvarianzen s_b^2 und s_{wi}^2 in die Formel einsetzt:

$$\widehat{\text{Var } \bar{Y}} = K^2 \left(\frac{1}{k} - \frac{1}{K} \right) s_b^2 + \frac{K}{k} \sum_{i=1}^k M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) s_{wi}^2 \text{ mit}$$

$$\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij},$$

$$s_b^2 = \frac{1}{k-1} \sum_{i=1}^k \left(M_i \bar{y}_i - \frac{1}{k} \sum_{j=1}^k M_j \bar{y}_j \right)^2 \quad \text{und}$$

$$s_{wi}^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2.$$

2. Weitere Verfahren der Stichprobentheorie

2.1. Ziehen mit und ohne Zurücklegen

Bei den bisherigen Betrachtungen ging man davon aus, daß nach einem festgelegten Verfahren Einheiten der Grundgesamtheit in die Stichprobe gelangen und ausgewertet werden. Dabei erschien es sinnvoll, eine betrachtete Einheit nicht ein zweites Mal zu berücksichtigen. Man spricht deshalb vom **Ziehen ohne Zurücklegen**. Andererseits ist es aber auch möglich, eine bereits erhobene Einheit ein weiteres Mal zu berücksichtigen, so daß man in diesem Fall vom **Ziehen mit Zurücklegen** spricht.

Wird die Auswahl beispielsweise mit Zufallszahlen durchgeführt, so bedeutet dies, daß im Fall „ohne Zurücklegen“ eine Wiederholung einer Zahl in der Zufallszahlenfolge keine Auswirkung hat, während im Fall „mit Zurücklegen“ eine solche Wiederholung die erneute Erhebung des Merkmalswertes der betreffenden Einheit zur Folge hat.

Generell ist es natürlich nicht verständlich, eine Auswahl mit Zurücklegen durchzuführen, da dann die Gefahr besteht, Einheiten mehrmals zu erfassen, wodurch sich die Erhebungskosten steigern bzw. die Genauigkeit der Aussagen verringert. Dennoch wird die aus diesen Überlegungen unpraktische Auswahl in der Stichprobentheorie behandelt, da sich die dadurch entstehenden Ergebnisse einfacher darstellen lassen und in manchen Fällen ausreichend sind. Entscheidende Unterschiede bei der Durchführung von Auswahlen mit und ohne Zurücklegen erhält man bei Betrachtung der Verteilung der einzelnen Merkmalswerte.

Bei einer einfachen Zufallsauswahl (ohne Zurücklegen) sind die Merkmalswerte y_1, \dots, y_n identisch verteilt. Sie sind nicht stochastisch unabhängig, da durch jeden Zug die Auswahlmöglichkeiten eingeschränkt werden. Liegt eine Zufallsauswahl mit Zurücklegen in der Form vor, daß jede Einheit E_i bei jedem Zug die gleiche Wahrscheinlichkeit hat, gezogen zu werden, so sind die Merkmalswerte unabhängig und identisch verteilt.

Sind y_1, \dots, y_n die beobachteten Merkmalswerte beim Ziehen mit Zurücklegen, so ist – wie beim Ziehen ohne Zurücklegen –

$$\hat{Y} = N\bar{y}$$

ein erwartungstreuer Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit. Die Varianz dieses Schätzers

$$\text{Var } \hat{Y} = \frac{N^2}{n} \left(1 - \frac{1}{N}\right) S_Y^2$$

ist jedoch größer als beim Ziehen ohne Zurücklegen, vgl. Abschnitt 1.4; man nimmt also hier einen Genauigkeitsverlust in Kauf. Dieser wird mit wachsendem Auswahlsatz $\frac{n}{N}$ immer größer. Daraus kann man die Konsequenz ziehen, daß bei Erhebungen mit kleinen Auswahlsätzen eine Auswahl mit Zurücklegen an Stelle der einfachen Zufallsauswahl möglich ist. Dies sollte nicht als allgemeingültige Regel aufgefaßt, sondern als Verfahrensvereinfachung interpretiert werden. Das Ziehen mit Zurücklegen ist somit mehr als Auswertungs- denn als Auswahlmethode anzusehen.

2.2. Schätzen von Anteilen

Bei vielen Stichprobenerhebungen steht nicht – wie bei den bisherigen Ausführungen – die Untersuchung quantitativer Merkmale wie Summen, Mittelwerte im Vordergrund. So kann z. B. von Interesse sein, wieviele Personen einer Grundgesamtheit verheiratet sind oder wieviele Personen öffentliche Verkehrsmittel benutzen, um den Arbeitsplatz zu erreichen. Diese Probleme führen allgemein zur Frage nach einem Anteil in der Grundgesamtheit.

Ist eine Eigenschaft (ein Merkmal) gegeben, z. B. das Merkmal verheiratet zu sein, so daß die Grundgesamtheit durch Besitzen oder Nichtbesitzen dieser Eigenschaft in zwei Klassen zerfällt, und ist M die Anzahl der Einheiten, die dieses Merkmal besitzen, so ist

$$P = \frac{M}{N}$$

der Anteil der Einheiten in der Grundgesamtheit, die das Merkmal besitzen.

Ziel der Erhebung ist es dann, einen solchen Anteil P zu schätzen. Eine sinnvolle Methode dies zu tun, ist die Bestimmung der Anzahl m der Einheiten in einer Stichprobe vom Umfang n , die mit dem Merkmal behaftet sind. Ein erwartungstreuer Schätzer für P ist dann

$$p = \frac{m}{n}.$$

Dieser Schätzer läßt sich in die bisherigen Aussagen einbetten, indem man folgende Schreibweise benutzt:

$$Y_i = \begin{cases} 1, & \text{falls } E_i \text{ das Merkmal besitzt} \\ 0, & \text{sonst.} \end{cases}$$

Hiermit gilt $M = \sum_{i=1}^N Y_i$ und damit $P = \bar{Y}$. Der angegebene Schätzer p für den Anteil P ist somit mit dem Stichprobenmittel \bar{y} identisch.

Dabei ist es möglich, eine gegenüber der bisher z. T. benutzten Normalapproximation spezialisierte Aussage über die Verteilung des Schätzers zu machen. Die Summe der Stichprobenmerkmalswerte ist nämlich hypergeometrischverteilt mit Parametern n , NP und $N(1 - P)$, vgl. Abschnitt 3.3 in Kap. IV. Das ist allerdings nur dann gültig, wenn das Ziehen ohne Zurücklegen erfolgt. Für das Ziehen mit Zurücklegen ergibt sich als Verteilung der Summe der Stichproben-Merkmalswerte eine Binomialverteilung mit Parametern n und P , vgl. Abschnitt 3.1 in Kap. IV.

Die Varianz des Schätzers $p = \bar{y} = \frac{m}{n}$ ist

$$\text{Var } p = \frac{1}{n} \left(\frac{N-n}{N-1} \right) P(1-P)$$

für die Auswahl ohne Zurücklegen und

$$\text{Var } p = \frac{1}{n} P(1-P)$$

für die Auswahl mit Zurücklegen.

In beiden Fällen hängt die Varianz des Schätzers p von dem Term $P(1 - P)$ ab. Betrachtet man diese Abhängigkeit genauer, so zeigt sich, daß sich für einen wahren Wert von $P = \frac{1}{2}$ große, bei $P < 0,05$ bzw. $P > 0,95$ kleine Varianzen des Schätzers p ergeben, vgl. Abb. 2. Anders verhält sich dagegen der Variationskoeffizient des Schätzers, der die Streuung auf die Lage dieses Schätzers bezieht. Es ist

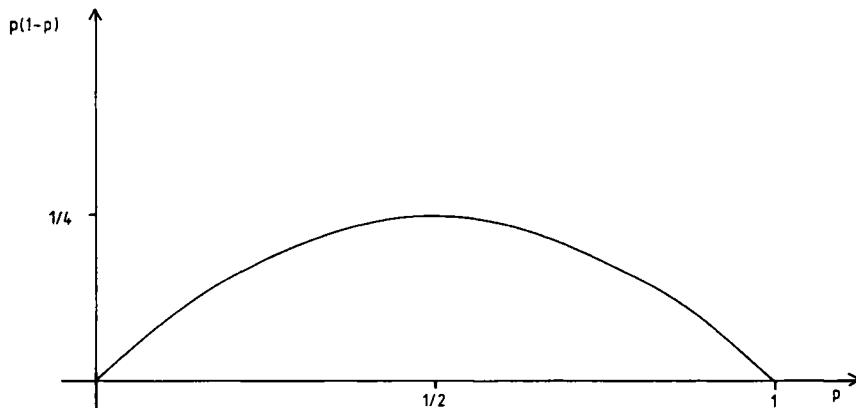


Abb. 2: $p \cdot (1 - p)$ als Funktion von p

$$V(p) = \sqrt{\text{Var } p / Ep}$$

$$= \begin{cases} \sqrt{\frac{1}{n} \left(\frac{N-n}{N-1} \right) \cdot \frac{1-P}{P}} & \text{beim Ziehen ohne Zurücklegen} \\ \sqrt{\frac{1}{n} \cdot \frac{1-P}{P}} & \text{beim Ziehen mit Zurücklegen} \end{cases}$$

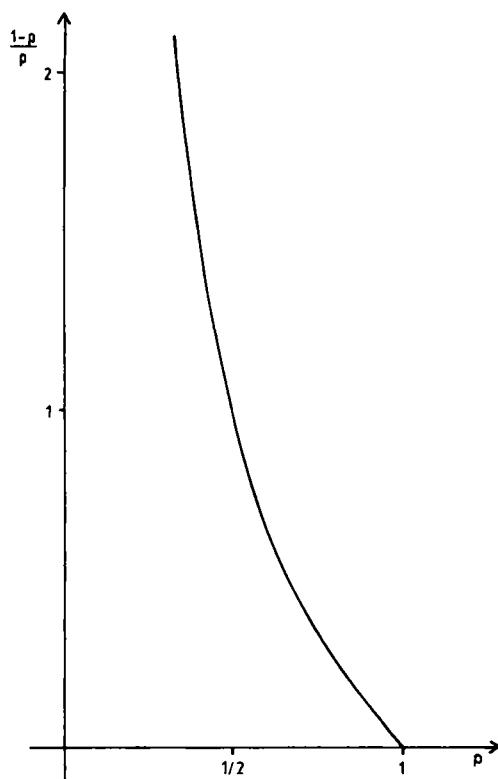


Abb. 3: $\frac{1-p}{p}$ als Funktion von p

Hier ist in beiden Fällen eine Abhängigkeit von $\frac{1-p}{p}$ gegeben. In **Abb. 3** erkennt man, daß der Variationskoeffizient mit steigendem Anteilswert P fällt bzw. bei kleinem P bis ins Unendliche wächst.

Dieses Phänomen bei der Schätzgenauigkeit für Anteile P einer Grundgesamtheit findet z. B. seine Berücksichtigung bei Umfragen zum Wahlverhalten, denn gerade hier treten extrem niedrige Anteile P auf.

2.3. Die systematische Stichprobe

Bei der Durchführung von Stichprobenuntersuchungen ist man immer bestrebt, ein möglichst kostengünstiges Verfahren anzuwenden. Deshalb erscheint es vielfach ungünstig, die erhebungstechnisch komplizierten Verfahren anzuwenden, die eine zufällige Auswahl der Einheiten aus der Grundgesamtheit voraussetzen. Beispielsweise scheitert die Durchführung einer einfachen Zufallsauswahl oft daran, daß es nicht möglich ist, jeder Einheit der Grundgesamtheit eine Zahl von 1 bis N zuzuordnen und dann die speziell für die Stichprobe benötigten Einheiten zu erheben.

Geht man davon aus, daß eine verhältnismäßig inhomogene Grundgesamtheit vorliegt, so bieten sich alternativ zur Zufallsauswahl auch **systematische Stichproben** an. Eine Vorgehensweise bei systematischen Stichprobenerhebungen wird im

Folgenden beschrieben. Ist $N = n \cdot k$ (k ganzzahlig), so ist eine Zerlegung der Grundgesamtheit in k Segmente der Größe n möglich. Diese Segmente können z. B. konstruiert werden, indem man dem i -ten Segment die Einheiten $E_1, E_{i+k}, \dots, E_{i+(n-1)k}$ der Grundgesamtheit zuordnet. Wählt man nun als Stichprobe eines dieser k Segmente zufällig aus, so spricht man von einer systematischen Stichprobe mit Stichprobenintervall k . Praktisch beschafft man sich zunächst eine Zufallszahl z zwischen 1 und k und bildet dann eine Stichprobe aus den Elementen $E_z, E_{z+k}, \dots, E_{z+(n-1)k}$, d.h. man wählt jede k -te Einheit aus.

Die systematische Stichprobe ist keine einfache Zufallsauswahl, da z. B. die Einheiten E_1, \dots, E_n nicht die Stichprobe bilden können.

Die Wahrscheinlichkeit, daß die i -te Einheit E_i in die Stichprobe gelangt, ist hier wie bei der einfachen Zufallsauswahl

$$\pi_i = \frac{1}{k} = \frac{n}{N},$$

jedoch ist die Wahrscheinlichkeit, daß E_i und E_j zur Stichprobe gehören, bei der systematischen Stichprobe mit Stichprobenintervall k gerade

$$\pi_{ij} = \begin{cases} \frac{1}{k} = \frac{n}{N}, & \text{falls } \frac{i-j}{k} \text{ eine ganze Zahl ist} \\ 0, & \text{sonst} \end{cases}$$

Eine systematische Stichprobe hat sicherlich erhebungstechnisch große Vorzüge, jedoch wird sie nur in den seltensten Fällen den an eine Untersuchung geknüpften Genauigkeitsforderungen gerecht; denn bezeichnet man die Summe der Merkmalswerte im z -ten Segment der Grundgesamtheit mit

$$T_z = \sum_{j=0}^{n-1} Y_{z+kj}, \quad \text{für } z = 1, \dots, k,$$

so ist

$$\hat{Y} = k \cdot T_z \left(\approx \frac{N}{n} \sum_{i=1}^n y_i \right)$$

ein erwartungstreuer Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit mit der Varianz

$$\text{Var } \hat{Y} = k \sum_{i=1}^k \left(T_i - \frac{\bar{Y}}{k} \right)^2.$$

Führt man nun einen Vergleich zur einfachen Zufallsauswahl durch, so zeigt sich Folgendes. Systematische Stichproben erweisen sich dann als besser, wenn eine lineare Anordnung der Grundgesamtheit vorliegt und die Systematik dieser Anordnung angepaßt ist. Das bedeutet, daß auch der Stichprobenumfang an die lineare Anordnung angepaßt werden muß, was eine andere Vorgehensweise bei der Bestimmung des Stichprobenumfangs impliziert als üblich.

Beispiel: Zur Veranschaulichung dieses Sachverhalts diene die Abb. 4, die eine Verteilung der Merkmalswerte in der Grundgesamtheit darstellen soll.

Die Grundgesamtheit vom Umfang $N = 9$, wie in Abb. 4, wurde in $k = 3$ Segmente unterteilt, um eine systematische Stichprobe vom Umfang $n = 3$ zu ziehen

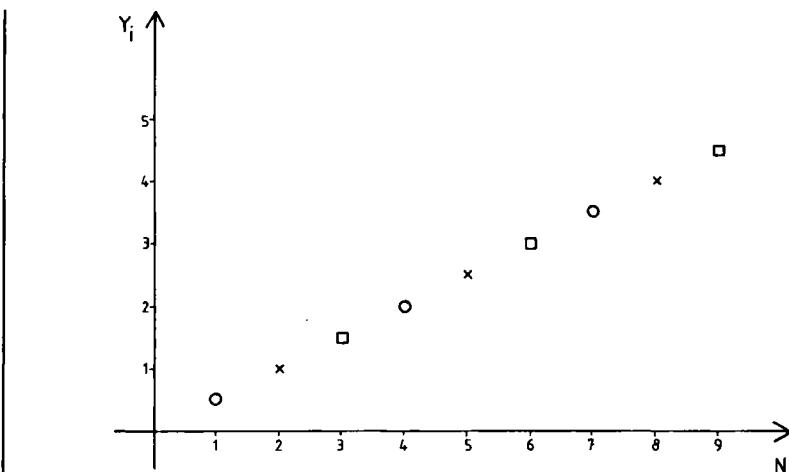


Abb. 4: Eine Verteilung der Merkmalswerte in der Grundgesamtheit, vgl. Textbeispiel

(gleich gekennzeichnete Punkte gehören einem Segment an). Für diese Art der Auswahl ist, vgl. Abb. 4,

$$\text{Var } \bar{y} = 0,16667 \quad (= \frac{1}{N^2} \text{Var } \hat{Y} = \frac{3}{81} ((6 - 7,5)^2 + 0 + (9 - 7,5)^2));$$

bei einer einfachen Zufallsauswahl würde sich hingegen

$$\text{Var } \bar{y} = 0,41666 \quad (= \frac{1}{N^2} \cdot N^2 \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \cdot \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 = \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{8} \cdot 15)$$

ergeben.

Natürlich ist es mitunter äußerst schwierig, eine gute Anpassung an die Grundgesamtheit zu erreichen bzw. überhaupt eine Anordnung in der Grundgesamtheit zu erkennen. Die systematische Stichprobe ist somit zwar ein äußerst praktikables, aber mitunter schwierig auszuwertendes Auswahlverfahren.

2.4. Stichproben mit ungleichen Auswahlwahrscheinlichkeiten

Bei den bisher vorgestellten Verfahren war die Wahrscheinlichkeit Π_i , daß die Einheit E_i in die Stichprobe gelangt, für alle $i = 1, \dots, N$ gleich. Es sind jedoch Erhebungen denkbar, bei denen dies nicht sinnvoll ist.

Beispiel: In einem Betrieb soll eine Stichprobenbefragung der Arbeiter vorgenommen werden. Dabei sollen gleich viel Frauen wie Männer berücksichtigt werden. Da 75% der Arbeiter Männer sind, ist die Wahrscheinlichkeit für einen Mann, in die Stichprobe zu gelangen, kleiner als für eine Frau. Bei der Schätzung müssen deshalb die Merkmalswerte der Männer ein entsprechend höheres Gewicht (hier das Dreifache) erhalten.

Dieses Beispiel zeigt, daß bei „Unterrepräsentation“ einer Gruppe in der Stichprobe eine höhere Wichtung der Merkmalswerte dieser Gruppe vorgenommen werden muß. Eine Möglichkeit, bei einer Erhebung sowie deren Auswertung ungleiche

Auswahlwahrscheinlichkeiten zu berücksichtigen, ist das **PPS-Verfahren („Probability Proportional to Size“-Verfahren)**. Sei P_i die Wahrscheinlichkeit, daß E_i beim einmaligen Ziehen in die Stichprobe gelangt, wobei gilt:

$$\sum_{i=1}^N P_i = 1; \quad P_i > 0 \quad \text{für alle } i = 1, \dots, N.$$

Weiterhin sei für jede Einheit der Grundgesamtheit ein Zusatzmerkmal (= „size“) $X_i, i = 1, \dots, N$ bekannt. Ein PPS-Verfahren mit Zurücklegen liegt dann vor, wenn

$$P_i = \frac{X_i}{\sum_{j=1}^N X_j} \quad \text{für alle } i = 1, \dots, N$$

gilt.

Beispiel: Als Beispiel für ein solches Verfahren wird häufig die Ertragsschätzung von Getreidefeldern durch Stichproben benannt. Hierbei sind N Areale verschiedener Größe X_i mit bestimmten Erträgen Y_i als Grundgesamtheit gegeben. Gesucht ist ein Schätzwert für den Gesamtertrag \bar{Y} .

Auf die mitunter äußerst schwierigen Prozeduren zur Realisierung einer PPS-Auswahl wollen wir hier nicht eingehen; vgl. etwa Hauser (1979).

Beim PPS-Verfahren mit Zurücklegen ist

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{p_i}$$

mit den Merkmalswerten y_i und der Auswahlwahrscheinlichkeit p_i der in die Stichprobe aufgenommenen n Elemente, ein erwartungstreuer Schätzer für die Summe der Merkmalswerte mit Varianz

$$\text{Var } \hat{Y} = \frac{1}{n} \sum_{i=1}^n P_i \left(\frac{Y_i}{P_i} - \bar{Y} \right)^2.$$

Diese kann durch

$$\widehat{\text{Var } \hat{Y}} = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{y_i}{p_i} - \frac{1}{n} \sum_{j=1}^n \frac{y_j}{p_j} \right)^2$$

erwartungstreu geschätzt werden.

Setzt man $P_i = \frac{1}{N}$ für alle $i = 1, \dots, N$ so erhält man die für das Ziehen mit Zurücklegen bekannten Ergebnisse für einfache Zufallsstichproben.

Berücksichtigt man die spezielle Definition der P_i beim PPS-Verfahren, so erhält man

$$\begin{aligned} \text{Var } \hat{Y} &= \frac{1}{n} \sum_{i=1}^N X_i \left(\frac{Y_i}{X_i} - R \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \left(X_i X_j \left(\frac{Y_i}{X_i} - \frac{Y_j}{X_j} \right)^2 \right), \end{aligned}$$

mit $X = \sum_{i=1}^N X_i$ und $R = \frac{Y}{X}$.

Die Anwendung eines PPS-Verfahrens hat nur Sinn, wenn damit ein Genauigkeitsgewinn erzielt werden kann. Dieser wird umso größer, d.h. die Varianz wird umso kleiner, je „proportionaler“ der Zusammenhang von X_i und Y_i ist. Man muß darauf achten, daß diese Verbesserung nur für die Proportionalität, nicht aber für eine Korrelation, vgl. Kap. IX, oder Linearität gilt.

Ob ein Genauigkeitsgewinn im Vergleich zur einfachen Zufallsstichprobe erzielt werden kann, ist aus folgender Aussage zu entnehmen. Die Varianz im PPS-Verfahren ist kleiner als die bei der einfachen Zufallsauswahl, dann und nur dann wenn

$$\sum_{i=1}^N (X_i - \bar{X}) \frac{Y_i^2}{X_i} > 0$$

gilt.

2.5. Die Formel von Horwitz-Thompson

In Abschnitt 2.4 wurde ein Verfahren vorgestellt, in dem ungleiche Auswahlwahrscheinlichkeiten der Einheiten beim Ziehen mit Zurücklegen berücksichtigt wurden. Für das Ziehen ohne Zurücklegen entwickelten Horwitz/Thompson (1952) einen allgemeinen Ansatz, der hier kurz vorgestellt werden soll.

Betrachtet wird ein beliebiges Auswahlverfahren, bei dem jede Einheit höchstens einmal in die Stichprobe gelangen kann, und jede Stichprobe mit positiver Auswahlwahrscheinlichkeit genau n Einheiten enthält. Ist Π_i die Wahrscheinlichkeit dafür, daß Einheit E_i , Π_{ij} die Wahrscheinlichkeit dafür, daß E_i und E_j gleichzeitig in die Stichprobe vom Umfang n gelangen, so ist für $\Pi_i > 0$, $i = 1, \dots, N$,

$$\hat{Y} = \sum_{i=1}^n \frac{1}{\Pi'_i} y_i,$$

mit den Merkmalswerten y_i und den Auswahlwahrscheinlichkeiten Π'_i der in die Stichprobe aufgenommenen Einheiten, ein erwartungstreuer Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit mit der Varianz (für $\Pi_{ij} > 0$, für $i, j = 1, \dots, N$) (**Horwitz-Thompson-Varianzformel**)

$$\begin{aligned} \text{Var } \hat{Y} &= \sum_{i=1}^N \frac{\Pi_i(1-\Pi_i)}{\Pi_i^2} Y_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\Pi_{ij} - \Pi_i \Pi_j}{\Pi_i \Pi_j} Y_i Y_j \\ &= \sum_{i=1}^{N-1} \sum_{j=i+1}^N (\Pi_i \Pi_j - \Pi_{ij}) \left(\frac{Y_i}{\Pi_i} - \frac{Y_j}{\Pi_j} \right)^2; \end{aligned}$$

diese kann durch

$$\begin{aligned} \widehat{\text{Var } \hat{Y}} &= \sum_{i=1}^n \frac{1 - \Pi'_i}{(\Pi'_i)^2} y_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{\Pi'_{ij} - \Pi'_i \Pi'_j}{\Pi'_{ij} \Pi'_i \Pi'_j} y_i y_j \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{\Pi'_i \Pi'_j - \Pi'_{ij}}{\Pi'_{ij}} \left(\frac{y_i}{\Pi'_i} - \frac{y_j}{\Pi'_j} \right)^2 \end{aligned}$$

erwartungstreugeschätzt werden, wobei Π'_{ij} die Π_{ij} entsprechende Wahrscheinlichkeit der gezogenen Stichprobeneinheiten darstellt.

Diese Formeln zeigen unmittelbar einen Nachteil der hier gezeigten Vorgehensweise, denn es ist möglich, daß Horwitz-Thompson-Schätzer für Varianzen negativ

werden. Trotzdem bleiben diese Formeln von fundamentaler Bedeutung, da ein großer Teil denkbarer Vorgehensweisen bei Stichprobenerhebungen damit abgedeckt wird.

2.6. Verhältnis-, Differenzen- und Regressionsschätzung, gebundene und freie Hochrechnung

In den bislang geschilderten Verfahren wurden die Schätzgrößen nur aus den Merkmalswerten der erhobenen Einheiten berechnet. Nun sind dies aber oft nicht die einzigen Informationen, die man über die unbekannte, interessierende Größe in der Grundgesamtheit besitzt; vielmehr hat man noch zusätzliche sogenannte **a-priori-Informationen** über die Grundgesamtheit. Diese können z. B. die wahren Werte einer Totalerhebung vergangener Jahre sein, was beispielsweise bei Bevölkerungsgesamtheiten oder auch bei der Inventur vorkommt. Das Bestreben, die a-priori-Informationen auszunutzen, hat sich in den Verfahren der Verhältnis-, Differenzen- und Regressionsschätzung niedergeschlagen.

2.6.1. Die Verhältnisschätzung

Bei der Verhältnisschätzung geht man davon aus, daß bei jeder Einheit E_i zusätzlich zum Merkmalswert Y_i eine Hilfsvariable X_i gemessen werden kann. Setzt man eine einfache Zufallsstichprobe mit Werten (x_i, y_i) , $i = 1, \dots, n$ voraus, so ist folgender **Verhältnisschätzer** denkbar:

$$\hat{Y}_v = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = X \hat{R}$$

mit der bekannten Hilfsvariablensumme $X = \sum_{i=1}^N x_i$ und

$$\hat{R} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i},$$

wobei keine a-priori-Information eingeht.

\hat{R} ist ein Schätzer für das Verhältnis von Y zu X , d. h. \hat{R} schätzt die unbekannte Größe $R = \frac{Y}{X}$. Der Schätzer \hat{R} ist nicht erwartungstreu für R , er ist verzerrt.

Die Verzerrung von \hat{R} , der **Bias**

$$B(\hat{R}) = E(\hat{R}) - R,$$

kann in folgender Weise berechnet werden:

$$B(\hat{R}) = - \frac{1}{E(\bar{x})} \text{Kov}(\hat{R}, \bar{x}).$$

Bei einer einfachen Zufallsstichprobe ist es möglich, eine Annäherung an die Varianz von \hat{R} anzugeben. Dies geschieht mit der approximativen Angabe des mittleren quadratischen Fehlers

$$\begin{aligned} \text{MSE}(\hat{R}) &= E(\hat{R} - R)^2 \\ &\simeq \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{S_Y^2 + R^2 S_X^2 - 2\varrho S_X S_Y R}{\bar{X}^2} \end{aligned}$$

wobei

$$S_X^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

die Varianz der Hilfsvariablen in der Grundgesamtheit,

$$S_Y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

die Varianz des Merkmalswertes in der Grundgesamtheit und

$$\varrho = \frac{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{S_X^2 S_Y^2}}$$

den Korrelationskoeffizienten zwischen den Merkmalswerten Y_i und der Hilfsvariablen X_i in der Grundgesamtheit angibt.

Um dem Unterschied der Vorgehensweisen beim Schätzen auch sprachlich gerecht zu werden, hält die Stichprobentheorie folgende Formulierung bereit. Erfolgt die Berechnung eines Schätzers für Y über \hat{R} , so heißt dies **gebundene Hochrechnung**, wird \hat{R} nicht berücksichtigt, so spricht man von **freier Hochrechnung**.

Die Wahl der Hilfsvariablen X_i sollte von dem Gedanken geleitet werden, daß ein ziemlich starker Zusammenhang zwischen ihr und den eigentlich interessierenden Merkmalswerten besteht. Dieser Zusammenhang ist wesentlich für die auftretenden Unterschiede in der Genauigkeit bei freier und gebundener Hochrechnung. Liegt in etwa eine Beziehung der Art $Y_i = A + BX_i$ vor, so ist der Schätzer aus der freien Hochrechnung besser als der aus der gebundenen Hochrechnung, wenn

$$\frac{\bar{X}^2 \cdot \text{MSE}\left(\frac{1}{\bar{X}}\right)}{S_X^2} > \frac{1 - \frac{n}{N}}{n} \cdot \frac{B^2}{A^2}, \text{ mit } \text{MSE}\left(\frac{1}{\bar{X}}\right) = E\left(\frac{1}{\bar{X}} - \frac{1}{\bar{X}}\right)^2,$$

gilt. Das bedeutet, daß die gebundene Hochrechnung besser ist, wenn Proportionalität vorliegt.

Eine andere Vergleichsmöglichkeit ergibt sich durch folgende Aussage.

Die Varianz aus der gebundenen Hochrechnung ist i. a. besser als die aus der freien, wenn gilt:

$$\frac{S_X}{\bar{X}} / \frac{S_Y}{\bar{Y}} < 2\varrho.$$

Haben X und Y einen annähernd gleichen Variationskoeffizienten, so ist die gebundene Hochrechnung vorzuziehen, wenn die Korrelation ϱ zwischen X - und Y -Variablen größer als 0,5 ist.

Zur Schätzung des mittleren quadratischen Fehlers $\text{MSE}(\hat{R})$ kann man folgen-

de, nicht erwartungstreue Funktion verwenden:

$$\widehat{\text{MSE}(\hat{R})} = \frac{1}{n} \cdot \frac{1}{\bar{x}^2} \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y} - \hat{R}(x_i - \bar{x}))^2.$$

Die Schätzung von R durch \hat{R} hat den erwähnten Nachteil einer Verzerrung. Führt man statt der einfachen Zufallsauswahl ein modifiziertes Auswahlverfahren durch, so kann der gleiche Schätzer unverzerrt gebildet werden. Ein solches Verfahren besteht darin, die Stichprobe mit einer **Wahrscheinlichkeit proportional zur aggregierten Größe (PPAS: Probability Proportional to Aggregated Size)** zu ziehen. Am einfachsten geschieht dies dadurch, daß man für das erste Element der Stichprobe als a-priori-Auswahlwahrscheinlichkeit $\frac{x_1}{X}$ nimmt und die restlichen $(n-1)$ Elemente der Stichprobe durch einfache Zufallsauswahl (Ziehen ohne Zurücklegen) zieht.

Die Wahrscheinlichkeit, daß als Stichprobe (x_1, \dots, x_n) gezogen wird, ist dann gerade gleich

$$\frac{\sum_{i=1}^n x_i}{\binom{N-1}{n-1} X},$$

so daß obiger Schätzer \hat{R} erwartungstreu (unverzerrt) für R ist:

$$E\hat{R} = \sum_{*} \frac{\sum_{i=1}^n x_i}{\binom{N-1}{n-1} X} \cdot \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{1}{\binom{N-1}{n-1} X} \sum_{*} \sum_{i=1}^n y_i = \frac{Y}{X} = R,$$

wobei \sum_{*} die Summation über alle möglichen Stichproben bezeichnen soll. Es ergibt sich ferner, daß durch

$$\hat{Y}_v = \hat{R} \cdot X$$

ein erwartungstreuer Schätzer für Y gegeben ist mit der Varianz

$$\text{Var} \hat{Y}_v = \frac{X}{\binom{N-1}{n-1}} \cdot \sum_{*} \frac{\left(\sum_{i=1}^n y_i \right)^2}{\sum_{i=1}^n x_i} - Y^2.$$

Ein erwartungstreuer Schätzer für diese Varianz, der allerdings auch negative Werte annehmen kann, ist

$$\widehat{\text{Var} \hat{Y}_v} = \hat{Y}_v^2 - X \cdot \frac{\sum_{i=1}^n y_i^2 + 2 \frac{N-1}{n-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n y_i y_j}{\sum_{i=1}^n x_i}.$$

Unter einigen den Vergleich erleichternden Zusatzvoraussetzungen läßt sich zeigen, daß die Varianz des Schätzers \hat{Y}_V bei dem zuletzt betrachteten PPAS-Auswahlverfahren kleiner ist als bei der davor betrachteten einfachen Zufallsauswahl, vgl. Raj (1968).

2.6.2. Die Differenzenschätzung

Wie in Abschnitt 2.6.1 erläutert, ist der Verhältnisschätzer dann dem Schätzer aus der freien Hochrechnung vorzuziehen, wenn ein proportionaler Zusammenhang zwischen Merkmalswerten Y_i und X_i vorliegt. Besteht in etwa ein additiver Zusammenhang der Form $Y_i = X_i + A$, so bietet sich der **Differenzenschätzer** an:

$$\hat{Y}_D = N(\bar{y} - \bar{x}) + X$$

ist ein erwartungstreuer Schätzer für Y , falls eine einfache Zufallsstichprobe vorliegt. Für die Varianz gilt mit den Bezeichnungen aus 2.6.1:

$$\text{Var } \hat{Y}_D = N^2 \left(1 - \frac{n}{N}\right) \frac{1}{n} (S_Y^2 + S_X^2 - 2\varrho S_X S_Y).$$

Eine erwartungstreue Schätzung für $\text{Var } \hat{Y}_D$ wird durch

$$\widehat{\text{Var } \hat{Y}_D} = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) \frac{1}{n-1} \sum_{i=1}^n (y_i - x_i - (\bar{y} - \bar{x}))^2$$

angegeben. Diese Aussagen gelten natürlich nur, wenn die Hilfsvariablensumme X bekannt ist.

Der Differenzenschätzer \hat{Y}_D ist dem üblichen Schätzer $\hat{Y} = N\bar{y}$ immer dann überlegen, wenn

$$\frac{1}{2} < \varrho \cdot \frac{S_Y}{S_X}$$

gilt.

Beispiele für eine solche Vorgehensweise sind die Schätzung von Ernteerträgen Y mit Hilfe der Vorjahreserträge X oder auch die hierzu etwas modifizierten Verfahren der Hochrechnungen bei Wahlen.

2.6.3. Die Regressionschätzung

Als Mischung von proportionalem und additivem Zusammenhang von X_i und Y_i ist auch ein in etwa erfüllter linearer Zusammenhang der Form $Y_i = BX_i + A$ denkbar. Der aus dieser Überlegung resultierende lineare Regressionsschätzer, vgl. Kap. X, für den Gesamtmittelwert \bar{Y} hat folgende Gestalt:

$$\hat{Y}_R = \bar{y}_R = \bar{y} + b(\bar{X} - \bar{x}).$$

Dabei ist \bar{y} das Stichprobenmittel bei der einfachen Zufallsauswahl und b ein Schätzwert für die Steigung B in der Regressionsbeziehung.

Liegt b von vornherein fest, sagen wir $b = b_0$, d.h.

$$\bar{y}_R = \bar{y} + b_0(\bar{X} - \bar{x}),$$

so ist \bar{y}_R erwartungstreu für \bar{Y} . Die Varianz berechnet sich dann mit den Bezeichnungen aus 2.6.1 als

$$\text{Var } \bar{y}_R = \frac{1 - \frac{n}{N}}{n} (S_Y^2 - 2 b_0 S_{YX} + b_0^2 S_X^2).$$

Als erwartungstreue Schätzung dafür kann

$$\widehat{\text{Var } \bar{y}_R} = \frac{1 - \frac{n}{N}}{n} (s_y^2 - 2 b_0 s_{yx} + b_0^2 s_x^2)$$

verwandt werden, wobei natürlich

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{und}$$

$$s_{yx} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$$

die empirischen Varianzen und die empirische Kovarianz sind.

Die Varianz $\text{Var } \bar{y}_R$ wird minimal, wenn

$$b_0 = B = \frac{S_{YX}}{S_X^2} = \frac{\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

gilt, also b_0 der (lineare) Regressionskoeffizient ist, vgl. Kap. X; dabei gilt

$$\text{Var}_{\min} \bar{y}_R = \frac{1 - \frac{n}{N}}{n} \left(S_Y^2 - \frac{S_{YX}^2}{S_X^2} \right).$$

Steht b nicht von vornherein fest, sondern wird mit Hilfe der kleinsten Quadrate, vgl. Kap. X, aus den Stichprobenwerten geschätzt, so gilt bei einfacher Zufallsauswahl approximativ

$$\text{Var } \bar{y}_R = \frac{1 - \frac{n}{N}}{n} S_Y^2 (1 - \rho^2).$$

Dabei sollte der Stichprobenumfang n so groß sein, daß

$$\sqrt{n} (b - B) (\bar{x} - \bar{X})$$

vernachlässigbar klein wird; ein $n \geq 40$ dürfte hier notwendig sein.

2.7. Zweiphasige Problemstellungen

In den in Abschnitt 2.6 geschilderten Verfahren ging man davon aus, daß die Werte der Hilfsvariablen X_i schon vorher für die ganze Grundgesamtheit bekannt sind. Selbst wenn dies nicht der Fall ist, wird man trotzdem die Hilfsvariable mitberück-

sichtigen, wenn z. B. die Erhebung von X_i wesentlich kostengünstiger ist als die von Y_i .

Bei einem **zweiphasigen Auswahlverfahren** zieht man zunächst eine einfache Zufallsstichprobe vom Umfang n' und beobachtet an den erobtenen Einheiten die Werte der Hilfsvariablen. Diesen Vorgang bezeichnet man als **erste Phase**.

Aus dieser Stichprobe wird dann eine weitere einfache Zufallsauswahl vom Umfang $n \leq n'$ getroffen und an diesen nun erhaltenen Einheiten beide Variablen beobachtet. Dieser Auswahlvorgang wird als **zweite Phase** bezeichnet.

Im Gegensatz zur bisherigen Schreibweise wird jetzt hier (und nur hier) eine andere Schreibweise benutzt, um den Ergebnissen aus erster bzw. zweiter Phase eine unterschiedliche Form zu geben. Großbuchstaben werden nun für die Schätzwerte der ersten, Kleinbuchstaben für die Schätzwerte der zweiten Phase verwendet. Liegen in beiden Phasen einfache Zufallsstichproben vor, so ist ein erwartungstreuer Schätzer für das Gesamtmittel gegeben durch:

$$\hat{\mu} = (\bar{y} - \bar{x}) + \bar{X};$$

die Varianz von $\hat{\mu}$ berechnet sich (in bisheriger Schreibweise) als:

$$\text{Var } \hat{\mu} = \left(\frac{1}{n} - \frac{1}{N} \right) S_Y^2 - \left(\frac{1}{n} - \frac{1}{n'} \right) S_X (2\varrho S_Y - S_X).$$

Das hier beschriebene Verfahren lässt sich auch auf andere Auswahlen als die einfache Zufallsauswahl anwenden. Dies soll hier nicht erläutert werden. Eine ausführliche Beschreibung von geänderten und erweiterten zweiphasigen Auswahlen findet man etwa bei Raj (1968) und Cochran (1972).

3. Probleme bei der praktischen Durchführung einer Erhebung

Will man mit den in Abschnitt 1 und 2 geschilderten Verfahren eine praktische Erhebung durchführen, so wird man schnell mit Problemen konfrontiert, deren Lösung nicht durch die besprochenen Verfahren erfolgen kann. Dies hat nichts mit der mangelnden Güte dieser Verfahren zu tun, sondern liegt daran, daß manche der in der Theorie gemachten Voraussetzungen praktisch nicht erfüllt sind.

Die Kenntnis der möglichen Probleme erlaubt es aber oft, durch geeignete Planung diese Voraussetzungen anzunähern und dann, wenn auch manchmal in abgewandelter Form, die theoretischen Verfahren anzuwenden.

In diesem Rahmen sollen einige Fragestellungen angesprochen werden. Dabei ist weniger Gewicht auf Vollständigkeit und Exaktheit der Lösungsvorschläge gelegt, sondern es soll darauf aufmerksam gemacht werden, in welcher Art und in welchem Umfang Probleme auftreten können.

3.1. Die Abgrenzung der Grundgesamtheit

Am Beginn einer Erhebung steht immer eine genaue Definition der Zielvariablen. Die Aussagen beziehen sich auf eine Grundgesamtheit. Somit ist zunächst die zu betrachtende Grundgesamtheit festzulegen und geeignet abzugrenzen. Eine solche Abgrenzung erscheint zunächst trivial, ist doch das Ziel und damit die Gesamtheit vorgegeben. Allerdings treten hier in der Praxis doch häufig Schwierigkeiten auf.

Beispiel:

- (a) Es soll eine Untersuchung über die Gewohnheiten von Rauchern durchgeführt werden, bei der der durchschnittliche Zigarettenkonsum pro Tag geschätzt werden soll. Die Grundgesamtheit wird dann aus der Menge aller Raucher bestehen, die jedoch nicht eindeutig festlegbar ist, da eine Definition des Begriffs „Raucher“ unterschiedlich erfolgen kann. So kann man einen Raucher dadurch definieren, daß er mehr als eine gewisse Anzahl Zigaretten pro Tag raucht. Eine andere Möglichkeit besteht darin, einen Befragten als Raucher zu qualifizieren, indem man ihn fragt: „Rauchen Sie?“. Wird diese Frage mit „Ja“ beantwortet, so wird er der Gesamtheit zugeordnet, andernfalls nicht.
- (b) Ein anderes Beispiel für die vorliegende Problematik ist bei Bevölkerungsumfragen zu finden. Will man z. B. eine Aussage über die Bewohner einer Stadt machen, so können folgende Abgrenzungsprobleme auftreten. Ist als Bewohner derjenige zu bezeichnen, der seinen Hauptwohnsitz in der Stadt hat oder zählen alle gemeldeten Personen zur Wohnbevölkerung? Die Antwort auf diese Frage ist natürlich mit dem Erhebungsgegenstand verbunden. Geht man davon aus, daß eine Person, die mit Nebenwohnsitz gemeldet ist, sich nicht so häufig in der Stadt aufhält, so wäre die Frage nach den Kaufgewohnheiten eines Bewohners vielleicht eher mit der Gesamtheit „mit Hauptwohnsitz gemeldet“ in Verbindung zu bringen. Gilt die Frage den durchschnittlichen Mietkosten eines Bewohners, so wären alle gemeldeten Personen der Grundgesamtheit zuzurechnen.

Diese Beispiele zeigen, daß je nach Problemstellung der Erhebung eine verbal gegebene Gesamtheit verschieden definiert werden kann. Das führt nicht nur dazu, daß die konkreten Gesamtheiten unterschiedlich zusammengesetzt sein können, sondern auch, daß deren Umfänge unterschiedlich sind. Weiterhin sind diese auch noch unbekannt, und so ist dem ursprünglich „trivialen“ Problem der Abgrenzung der Grundgesamtheit eine wichtige Bedeutung zugekommen.

3.2. Endliche und unendliche sowie fiktive Grundgesamtheiten

In vielen Abhandlungen zur Stichprobentheorie werden Varianzformeln und die Angabe von Stichprobenumfängen für zwei verschiedene Fälle angegeben. Diese Fälle betreffen die endliche und die unendliche Grundgesamtheit.

Bisher haben wir immer den Fall der **endlichen Grundgesamtheit** betrachtet. Dabei wird der Umfang N der Grundgesamtheit als bekannt vorausgesetzt, und durch jedes Ziehen einer Einheit verändert sich die Zusammensetzung der Grundgesamtheit. Ist dies nicht der Fall, d. h. verändert die Erhebung einer Einheit die Grundgesamtheit nicht, ist die Grundgesamtheit also „unerschöpflich“, so liegt eine **unendliche Grundgesamtheit** vor.

Durch diese Definition fällt eine Erhebung durch Ziehen mit Zurücklegen in gewisser Weise auch unter den Begriff der unendlichen Gesamtheit, doch ist letztere Begriffsbildung eigentlich für unüberschaubar große Gesamtheiten gedacht. Beispiele für unendliche Gesamtheiten findet man häufig bei biologischen Untersuchungen. Will man z. B. eine Aussage über das Vorkommen einer bestimmten Insektenart in einer Region machen, so ist die Anzahl der Insekten sicherlich endlich, aber doch so unübersichtlich, daß eine unendliche Gesamtheit vorausgesetzt wird.

Die Auswirkungen bei der Berücksichtigung solcher Gesamtheiten sind allerdings gering, denn die neu entstehenden Berechnungsformeln, z. B. für den Stich-

probenumfang, sind meist in Bezug auf die Formeln für endliche Gesamtheiten nur leicht abgewandelt. So wird dann die Endlichkeitskorrektur nicht mehr berücksichtigt, was aber ohnehin bei kleinen Auswahlsätzen nur in komplizierteren Erhebungen nötig ist.

Eine weitere Betrachtungsweise bei Grundgesamtheiten beschreibt das Begriffs-paar „konkret“ und „fiktiv“. Dabei geht man von folgender Überlegung aus. Liegt eine Stichprobe vor, die eindeutig aus einer Gesamtheit erhoben wurde, so spricht man von einer real existierenden, **konkreten Gesamtheit**. Hat dagegen die vorliegende Stichprobe keinen festen Bezug zu einer Gesamtheit, d. h. sind die erhobenen Daten nicht zuzuordnen, so spricht man von einer **fiktiven Gesamtheit**.

Beispiel: Bei einer landwirtschaftlichen Untersuchung soll ein Vergleich von Düngemitteln durchgeführt werden. Dafür liegen Daten über Getreideerträge von Feldparzellen vor. Diese Daten entsprechen einer Stichprobe aus der nicht existierenden, fiktiven Gesamtheit aller mit Düngemittel behandelten Feldstücke.

Die Betrachtung fiktiver Gesamtheiten ist mehr von theoretischer als von praktischer Bedeutung. Sie gibt die Möglichkeit, vorliegende Daten als Stichprobe zu interpretieren, um dann Verfahren wie beispielsweise Tests anzuwenden. Zu den eigentlichen Verfahren der Stichprobentheorie, die für konkrete Erhebungen gedacht sind, besteht somit keine wesentliche Verbindung.

3.3. Auswahltechniken und Erhebungsprobleme

Bei der theoretischen Bearbeitung von Auswahlverfahren wurde bislang eine Zufallsstichprobe vorausgesetzt. Im Falle der einfachen Zufallsauswahl konnte diese „realisiert“ werden, indem man n Zufallszahlen aus $1, \dots, N$ bestimmte und die entsprechenden Einheiten der Grundgesamtheit in die Stichprobe einbezog.

Bei der praktischen Durchführung einer solchen Zufallsauswahl ergeben sich aber bei fast allen Untersuchungen erhebliche Schwierigkeiten bei der Realisierung der „Zufälligkeit“. Das liegt hauptsächlich daran, daß die betrachtete Grundgesamtheit meist gar nicht in geordneter Form vorliegt. Dieses Problem kann an folgendem Beispiel erläutert werden.

Beispiel: Es soll eine Untersuchung über die Wohnbevölkerung der Bundesrepublik Deutschland gemacht werden, die somit die Grundgesamtheit darstellt. Nach Ermittlung eines notwendigen Stichprobenumfangs n stellt sich die Frage, wie eine einfache Zufallsstichprobe realisiert werden kann. Es treten nämlich folgende Schwierigkeiten auf.

Die Bezeichnung jedes Bewohners mit einer Zahl von $1, \dots, N$ ist nicht möglich, zumal eine genaue Liste über alle Bewohner nicht existiert. Setzt man trotzdem voraus, daß eine „Durchnumerierung“ aller Bundesbürger möglich wäre, so ergibt sich die Schwierigkeit, einen durch Zufall ermittelten Bürger in die Erhebung aufzunehmen. Dies kann vielerlei Gründe haben. Vielleicht ist die Person an einer Aufnahme in die Erhebung nicht interessiert oder sie ist nicht aufzufinden.

Man sieht, daß die Durchführung einer einfachen Zufallsauswahl nicht immer unproblematisch ist. Dies tritt besonders häufig bei Befragungen auf, also bei Erhebungen, in denen bestimmte Personenkreise eine Grundgesamtheit bilden. Das hat beispielsweise in der Marktforschung zu einer Reihe von Verfahren geführt, die sich nur noch mit der Repräsentativität von Erhebungen bzw. Befragungen befassen. Doch auch auf anderen Gebieten hat sich herausgestellt, daß die theoretisch einfach

zu handhabende einfache Zufallsauswahl nur in den wenigsten Fällen anwendbar ist. So wurden einerseits andere Stichprobenverfahren (z. B. Klumpenstichprobe, systematische Stichprobe) entwickelt, die eine solche Erhebungsproblematik besser berücksichtigen, andererseits werden aber auch modifizierte Auswahltechniken betrachtet.

Eine Auswahltechnik, die immer noch praktiziert wird, von der aber auf jeden Fall abzuraten ist, ist die **Auswahl aufs Geradewohl**. Diese Auswahl führt fast immer zu Verzerrungen, die auch bei genauer Kenntnis des betrachteten Sachverhaltes nicht auszuschließen sind. Wie solche Verzerrungen zustande kommen, zeigt folgendes Beispiel.

Beispiel: Nach Kriegsende wollte die amerikanische Militärregierung eine Untersuchung über den Gesundheitszustand der deutschen Stadtbevölkerung durchführen und stellte zu diesem Zweck an Plätzen, Straßenecken etc. Waagen auf, um damit das durchschnittliche Gewicht der Stadtbewohner zu ermitteln. Gewogen wurden Personen, die an den Waagen vorbeikamen und sich freiwillig zur Verfügung stellten. Der durch diese Vorgehensweise ermittelte Durchschnittswert ist in keiner Weise repräsentativ, was folgende Überlegungen verdeutlichen. Die Wahrscheinlichkeit, daß eine Person, die viel unterwegs ist, in die Erhebung kommt, ist wesentlich größer als die anderer Personen. Die so entstandenen Gruppen haben aber sicherlich kein gleiches Durchschnittsgewicht, da z. B. Kranke oder durch Hunger Geschwächte weniger oft oder überhaupt nicht unterwegs sind. Auch war die Messung freiwillig, wobei man wiederum davon ausgehen kann, daß sich dies auf das Ergebnis auswirkt. Eine weitere Ergebnisverfälschung ist dadurch zu erwarten, daß an einigen Wiegestellen, wie an Bahnhöfen, auch Nichtstädter erfaßt wurden, so daß hier vielleicht vornehmlich schwergewichtige Landbewohner berücksichtigt wurden.

Alles in allem kann man bei der Auswahl aufs Geratewohl nie mit zuverlässigen Schätzern rechnen, da eine große Anzahl von nichtkontrollierbaren Einflüssen diese aufs stärkste beeinflussen.

Eine Auswahltechnik, die dem Zufallsprinzip schon wesentlich näher kommt, ist die **systematische Auswahl**, die in Abschnitt 2.3 erläutert wurde. Dabei sind verschiedene Techniken möglich. Liegt beispielsweise eine durchnumerierte Kartei vor, so kann, wie in Abschnitt 2.3 beschrieben, jede k-te Karteikarte entnommen werden. Ist keine Numerierung vorhanden, so ist die Auswahl durch Abstandsmessungen (alle ... cm) möglich.

Systematische Stichproben sind zwar aufgrund der zufälligen Auswahl der ersten Einheit Zufallsstichproben, jedoch können durch eine systematische Struktur der Grundgesamtheit große Ergebnisverzerrungen auftreten. Dies tritt besonders bei den systematischen Karteiauswahlverfahren auf, da Karteien oft nach bestimmten Gesichtspunkten sortiert sind. Kennt man solche Gesichtspunkte, dann kann man meist durch eine Modifikation des Auswahlverfahrens einer Verzerrung entgegenwirken.

Um dem systematischen Fehler bei der Auswahl aus Karteien aus dem Wege zu gehen, werden auch die **Auswahl nach Namen oder Geburtstagen** vorgeschlagen. Hierbei geht man davon aus, daß Personen mit gleichem Anfangsbuchstaben des Nachnamens bzw. mit gleichem Geburtstag zusammengefaßt einen repräsentativen Querschnitt über alle in der Kartei erfaßten Personen bilden. Hinzu kommt, daß

Personen mit gleichen Anfangsbuchstaben wesentlich leichter aus der Kartei herauszusuchen sind.

Inwieweit der Grundsatz der Repräsentativität gewahrt bleibt, ist auch hier genau zu überdenken. Es ist sicherlich verständlich, daß z. B. der Buchstabe „Y“ zu keiner repräsentativen Auswahl führt, da mit „Y“ hauptsächlich Ausländer in die Untersuchung eingehen würden. Böltken (1976) berichtet, daß man aber z. B. mit den Buchstaben „B“ und „L“ durchaus positive Erfahrungen gemacht hat.

Ähnliche Bedenken kann man auch bei der Auswahl nach Geburtstagen einwenden. Nicht jeder Tag des Jahres ist gleichermaßen besetzt. Ausländer weisen oft ihre Geburtstage an markanten Tagen aus (1.1.; 1.7.), was an den für unsere Verhältnisse unzulänglich arbeitenden Behörden der Heimatländer liegt. Weiterhin können religiöse Enthaltsamkeit zur Unterrepräsentation von solchen Religionen zugehörigen Personengruppen in bestimmten Zeiten führen.

Solche Überlegungen führen dann zu sogenannten **Pseudoauswahlverfahren**, deren Ziel es ist, die Repräsentativität einer Stichprobe für die Grundgesamtheit zu erreichen. Interessante Aspekte zu solchen Verfahren wurden beispielsweise von Schach/Schach (1978, 1979) beschrieben.

3.4. Probleme im Zusammenhang mit Befragungen

3.4.1. Fragestellung und Fragebogen

Die erste grundsätzliche Unterscheidung von Fragen ist die zwischen offenen und geschlossenen Fragen. Eine **offene Frage** läßt beliebige, auch textmäßig längere Antworten zu. Bei der Auswertung der Antworten ist man dann vor das Problem gestellt, diese Antworten zu kategorisieren, um sie weiterverarbeiten zu können. Eine **geschlossene Frage** hingegen läßt nur einige vorgeschriebene Antworten zu, die entweder schriftlich formuliert sind (**externe geschlossene Fragen**) oder aber in der Frage selbst mitformuliert werden (**interne geschlossene Fragen**); hierbei müssen die angebotenen Antwortalternativen überschaubar, erschöpfend und gut gegeneinander abgegrenzt sein.

Stimuliert eine scheinbar offene Frage zu einer bestimmten Antwort, so sollte man sie besser gleich geschlossen stellen. Sind bei geschlossenen Fragen **Mehrfachnennungen** möglich, so ist dies im Grunde gleichbedeutend damit, daß verschiedene „Ja-Nein“-Fragen gestellt werden.

Eine **suggestive Fragestellung** („Antwort in den Mund legen“, Stereotypie, Argumente für eine bestimmte Antwort in der Frage) sollte man nach Möglichkeit vermeiden. Eine **Manipulation** der Antworten ist natürlich ebenfalls unzulässig. Sie entsteht dadurch, daß man z. B.

- scheinbar offene Fragen mit impliziten Antwortmöglichkeiten formuliert;
- nur eine Antwort zuläßt, wenn im Prinzip Mehrfachnennungen möglich sind;
- eine lange Antwortliste vorliest, denn dann bleiben nur die letzten Antwortalternativen im Gedächtnis.

Ein guter **Fragebogen** muß den Eindruck einer wirklichen Gesprächssituation erwecken. Daher muß er gründlich, möglichst von mehreren Personen, die sich zunächst einen Überblick über den gesamten Problemkreis verschaffen, vorbereitet werden.

Die Fragen müssen exakt und neutral formuliert werden, die Antwortalternativen müssen klar und ausgewogen sein.

Der Fragebogen sollte nach Möglichkeit keine offenen Fragen enthalten, dagegen sollten zur Kontrolle der sinnvollen Beantwortung ruhig einige sinngemäß gleiche Fragestellungen eingebaut werden.

Antwortalternativen der Form „weiß nicht“ oder „keine Angabe“ sollte man vermeiden, da diese häufig als Fluchtweg benutzt werden.

Spricht eine Frage ein heikles Thema an, so sollte man sie durch das Einflechten „harmloser“ Antwortalternativen entschärfen, um so dem Problem der Nichtbeantwortung, vgl. Abschnitt 3.4.3, auszuweichen.

Wesentlich bei der Erstellung eines Fragebogens ist eine klare graphische Aufmachung. Auch die Reihenfolge der Fragen und der jeweiligen Antwortalternativen ist wichtig: Bei mehreren möglichen Antworten etwa zeigen Erfahrungen, daß viele Leute dazu neigen, die erste oder die letzte Alternative anzukreuzen.

Stellt man im Rahmen eines Fragebogens sogenannte **Bildfragen**, so sind die zu beurteilenden Bilder möglichst gleichartig aufzumachen, da sonst Antwortverzerrungen auftreten.

Vor dem Beginn einer Befragung mittels Fragebogen sollte in einem **Prätest (Pre-Test)** überprüft werden, ob er wirklich geeignet ist, die Zielsetzung der Befragung zu erreichen.

3.4.2. Typen von Befragungen

Man unterscheidet zunächst einmal grob zwischen **Interview-Umfragen** und **postalischen Befragungen** (auf Telefonbefragungen, die eine Mischform beider Typen darstellen, wollen wir hier nicht näher eingehen). Die Nachteile einer Interview-Umfrage, nämlich

- die Aufhebung der Anonymität des Befragten;
- die Abhängigkeit der Antworten vom jeweiligen Interviewer;
- der zeitliche Antwortdruck

treten bei einer postalischen Befragung nicht auf. Dafür bringt diese dann eine Reihe anderer Nachteile mit sich:

- die Situation der Befragung ist nicht kontrollierbar;
- es gibt keine Möglichkeit, Anweisungen und Zusatzauskünfte zu geben, so daß von vornherein nur sehr einfache Fragen gestellt werden können;
- die Antworten sind endgültig und können nicht nachträglich präzisiert werden; sie sind nicht spontan;
- die Reihenfolge der Fragenbeantwortung ist nicht nachvollziehbar.

Zudem ist bei postalischen Befragungen die Rücklaufquote, besonders bei umfangreichen Fragebögen, oft gering. Erfahrungswerte schwanken hier zwischen 10%–80% Rücklaufquote. Jedoch gibt es verschiedene Möglichkeiten, diese Quote zu erhöhen. Zum einen erreicht man gewisse Erfolge schon allein dadurch, daß man dem Fragebogen ein **Begleitschreiben** beilegt, daß nicht zu persönlich, aber auch nicht allzu unpersönlich gehalten ist; zum anderen kann man nach einiger Zeit **Erinnerungsschreiben** versenden. Eine weitere Möglichkeit zur Erhöhung der Rücklaufquote ist es, „Belohnungen“ für die Rücksendung des Fragebogens auszusetzen.

zen. Man könnte mit dem Fragebogen etwa eine Tombola verbinden: Aus allen Rücksendern werden per Los einige ausgewählt, die ein Sachgeschenk erhalten.

Bei der postalischen Umfrage ist es stets möglich, die in die Umfrage einzubeziehenden Personen aus einer Adresskartei zufällig auszuwählen. Bei Interview-Umfragen ist ein solches **Random-Sampling** nicht immer günstig. Zum einen bedingt es oft lange Anfahrtswege des Interviewers, der zudem die betreffende Person oftmals nicht beim ersten Besuch antrifft; viele trifft er auch nach mehrmaligen Versuchen nicht an, so daß am Ende eine Überrepräsentation von Menschen, die häufig zu Hause sind (Rentner, Hausfrauen), in der Stichprobe vorhanden ist.

Daher werden Interview-Umfragen häufig basierend auf **Quotenverfahren** durchgeführt: Jedem Interviewer wird gesagt, wieviele Personen z. B. welchen Geschlechts, Alters, Berufs er befragen muß. Solche Merkmale heißen dann **Quotierungsmerkmale**. Natürlich ist die Auswahl der Quotierungsmerkmale vor einer Untersuchung genau zu überdenken; die Quoten sollten natürlich so festgelegt werden, daß sie repräsentativ sind für die Verhältnisse in der Grundgesamtheit, auf die sich eine Aussage beziehen soll. Vielmals erweist es sich auch als günstig, verschiedene Quotierungsmerkmale miteinander zu koppeln, denn so können Verzerrungen in der Stichprobe gegenüber der interessierenden Grundgesamtheit vermieden werden: Bei einer Kopplung der Merkmale Geschlecht und Alter wird vermieden, daß z. B. nur junge Männer und nur alte Frauen befragt werden. Zudem muß bei einem Quotenverfahren natürlich insbesondere darauf geachtet werden, daß sich auch die nicht quotierten Merkmale in der Stichprobe in etwa so einpendeln, wie dies in der Grundgesamtheit der Fall ist. Die Quotenauswahl wird in Abschnitt 3.6 ausführlich behandelt.

Abschließend wollen wir an dieser Stelle noch kurz auf **Panel-Untersuchungen** eingehen. Bei einer Panel-Untersuchung, die am besten mündlich durchgeführt wird, wird derselbe Personenkreis wiederholt befragt, um so kausale Zusammenhänge und Veränderungen aufzudecken. Hier wird vor der ersten Befragung eine Personengruppe ausgewählt, die dann für die nächsten Befragungen möglichst vollständig erhalten bleiben muß. Erfahrungen zeigen jedoch, daß bei der zweiten Befragung etwa 20%–25% Ausfälle in Kauf genommen werden müssen, da Personen nicht angetroffen werden. Bei den folgenden Befragungen werden dann einige dieser Personen jedoch wieder erreicht, so daß es sich nicht um Totalausfälle handelt. Um Ausfälle bei Paneluntersuchungen im Griff zu haben, legt man oft vor Beginn der Untersuchung den Stichprobenumfang größer als nötig fest oder arbeitet mit sogenannten „**statistischen Zwillingen**“. Das sind solche Personen, die etwa gleiche Charakteristika aufweisen. Fällt dann eine Person während der Panel-Untersuchung aus, so wird sie durch ihren „Zwilling“ ersetzt.

3.4.3. Das Problem der Nichtbeantwortung

Ein breites Anwendungsgebiet findet die Stichprobentheorie bei Verfahren innerhalb der Marktforschung. Hier stehen Befragungen zur Meinung und zum Verhalten von Verbrauchern im Vordergrund.

Führt man aber nun die Befragung durch, so muß man feststellen, daß man von einem nicht unerheblichen Teil der Befragten keine Antwort erhält. Dies kann sich in vielerlei Formen äußern. So können befragte Personen eine Antwort verweigern, man trifft zu befragende Personen nicht an oder ein zugeschickter Fragebogen wird nicht zurückgesandt. Gerade bei Fragebogenaktionen tritt eine sehr hohe Quote

von Nichtbeantwortung auf, und so muß man sich fragen, inwieweit ein solches Problem der Nichtbeantwortung Berücksichtigung finden kann.

Es ist verständlich, daß eine Nichtberücksichtigung der Nichtbeantwortung zu Verzerrungen in den zu treffenden Aussagen führen kann, da es möglich ist, daß ein Zusammenhang zwischen dem zu erhebenden Merkmal und dem Tatbestand der Nichtbeantwortung besteht. So ist es z. B. denkbar, daß bei einer Umfrage zum Gesundheitsbewußtsein der Bevölkerung gerade die nicht Gesundheitsbewußten eine Antwort verweigern. Eine ähnliche Situation kann bei einer Umfrage unter Arbeitslosen entstehen. Will man sich nach dem Bemühen, eine neue Arbeitsstelle zu finden, erkundigen, so werden vielleicht die Arbeitsunwilligen ungern eine Antwort geben.

Eine theoretische Lösungsmöglichkeit für das vorliegende Problem ergibt sich durch das **Modell der Auswahl nach Schichten**. Hier geht man von folgender Modellvorstellung aus.

Die Grundgesamtheit ist in zwei Schichten unterteilt, in die Schicht der Beantworter vom Umfang N_1 und in die Schicht der Nichtbeantworter vom Umfang $N_2 = N - N_1$. Dann entnimmt man der Grundgesamtheit eine einfache Zufallsstichprobe vom Umfang n und erhält n_1 Antworten und $(n - n_1) = n_2$ Nichtantworten.

Im nächsten Schritt wird aus der Stichprobe der Nichtbeantworter eine Unterstichprobe vom Umfang

$$u = \frac{n_2}{k}$$

gezogen, d. h. jeder k -te Nichtbeantworter wird nochmals angesprochen. Geht man davon aus, daß alle u Einheiten der Unterstichprobe erhoben werden können, so sind folgende Aussagen möglich.

Ist \bar{y}_{n_1} der Mittelwert der n_1 Antwortenden und \bar{y}_u der Mittelwert der u Antwortenden aus der zweiten Schicht, so ist

$$\bar{y} = \frac{1}{n} (n_1 \bar{y}_{n_1} + n_2 \bar{y}_u)$$

ein erwartungstreuer Schätzer für \bar{Y} mit Varianz

$$\text{Var } \bar{y} = \frac{k-1}{n} W_2 S_{Y_2}^2 + \left(\frac{1}{n} - \frac{1}{N} \right) S_Y^2,$$

wobei $W_2 = N_2/N$ und $S_{Y_2}^2$ Gewicht und Varianz der zweiten Stichprobe, S_Y^2 die Varianz in der Grundgesamtheit darstellen.

Der Nachteil dieser Vorgehensweise ist, daß $\text{Var } \bar{y}$ nicht schätzbar ist, da die Varianz der Nichtbeantworter berücksichtigt werden müßte. Eine Aussage ist aber dennoch möglich. Die Varianz dieser Schätzung ist immer größer als die bei der einfachen Zufallsauswahl.

Obwohl diese Methode auf dem Ziehen einer einfachen Zufallsstichprobe beruht, ist es nötig, eine andere Kalkulation des Stichprobenumfangs anzugeben, da durch das Ziehen einer Unterstichprobe ein wesentlicher Kostenaspekt in die Erhebung eingeht. Dazu seien c_0 die Kosten für eine Befragung (ob stattgefunden oder nicht) auf der ersten Stufe, z. B. die Kosten für das Verschicken eines Fragebogens. c_1

seien die Kosten, die durch Auswertung der Antworten auf der ersten Stufe entstehen und c_2 die Kosten für eine nochmalige Befragung. Dann sind die Gesamtkosten C der Erhebung gegeben durch

$$C = c_0 n + n W_1 c_1 + \frac{n W_2 c_2}{k}, \quad W_1 = \frac{N_1}{N}.$$

Analog zur Problemstellung der optimalen Aufteilung bei einer Schichtenauswahl, vgl. Abschnitt 1.5.1, sollen nun die Kosten C bei vorgegebenem Wert V_0 für die Varianz von \bar{y} minimiert werden. Diese Minimierungsaufgabe führt zu folgenden Ergebnissen:

$$k^2 = \frac{c_2(S_Y^2 - W_2 S_{Y_2}^2)}{S_{Y_2}^2(c_0 + c_1 W_1)} \quad \text{und} \quad n = \frac{N(S_Y^2 + (k-1)W_2 S_{Y_2}^2)}{N \cdot V_0 + S_Y^2}.$$

Um diese Formeln anwenden zu können, setzt man in der Praxis voraus, daß gilt: $S_{Y_2}^2 = S_Y^2$. Damit ergibt sich dann für großes N

$$k \approx \sqrt{\frac{c_2 W_1}{c_0 + c_1 W_1}} \quad \text{und} \quad n \approx n_0(1 + (k-1)W_2),$$

wobei n_0 den Stichprobenumfang darstellt, der benötigt würde, wenn ohne Nichtbeantwortung eine Varianz von V_0 erreicht werden sollte:

$$n_0 = \frac{S_Y^2}{V_0}.$$

3.5. Vergleich zwischen den Schichten

Bei vielen Erhebungen wird eine Schichtung der Grundgesamtheit nicht nur durchgeführt, um eine Verringerung der Varianz zu erreichen, sondern es werden über die Ermittlung eines Schätzers für das Gesamtmittel hinaus noch weitere speziellere Aussagen gewünscht. Besonders interessiert ist man oft an einem Vergleich zwischen den Schichten und so lauten mögliche Fragestellungen:

- Gibt es einen Unterschied im Kaufverhalten von Mann und Frau?
- Haben einkommensstärkere Schichten ein anderes Konsumverhalten als einkommensschwächere, und wie drückt sich das aus?

Will man solche Fragen beantworten, so kann es sein, daß die proportionale und optimale Aufteilung nicht sehr brauchbar sind. Dies trifft besonders dann zu, wenn die einzelnen Schichtumfänge stark voneinander abweichen.

Ein Vergleich von Schätzern, die aufgrund der unterschiedlichen Stichprobenumfänge eine große zufällige Abweichung besitzen, wäre dann kaum sinnvoll. In einem solchen Fall wäre dann der Vergleich von Variationskoeffizienten zu empfehlen.

Besteht aber der Hauptzweck einer Erhebung im Vergleich von Schichten, so ist es besser, von den herkömmlichen Aufteilungen abzugehen und eine den Gegebenheiten in den Schichten angepaßte Aufteilung zu wählen. Pfanzagl (1972) schlägt als Aufteilung dann beispielsweise

$$n_h = n \frac{S_h}{\sum_{i=1}^L S_i}$$

vor. Die Stichprobenumfänge in den Schichten sind nur noch von der Schichtvarianz S_h^2 , nicht mehr von den Schichtanteilen abhängig. Der sich aus dieser Aufteilung ergebende Schätzwert für das Gesamtmittel \bar{Y} ist verzerrt und so ergibt sich ein Konflikt zwischen dem Vergleich von Schichten und der Schätzung des Gesamtmittels. Meistens wird man an beidem interessiert sein und so ist es dann nötig, einen praktisch durchführbaren Kompromiß zu schließen.

Die Fehleranalysen bei geschichteten Stichprobenverfahren bzw. die Aufschlüsselung der Variationsanteile der einzelnen Schichten erfolgt mit Hilfe der sogenannten Modelle II (Modelle mit zufälligen Effekten) der Varianzanalyse, vgl. Kap. XI und auch Elpelt/Hartung (1982).

3.6. Stichprobenverfahren in der Marktforschung

In den bisherigen Ausführungen wurde die **Marktforschung** als Beispiel für die Anwendung von Stichprobenverfahren bereits erwähnt. Tatsächlich sind in diesem Bereich auch sehr viele Anwendungsmöglichkeiten der Stichprobentheorie gegeben. Andererseits treten aber auch sehr viele Stichprobenprobleme auf, die nicht mit den bereits geschilderten Methoden und Verfahren zu analysieren sind. Deshalb sind innerhalb der Marktforschung zusätzliche Modell- und Begriffsbildungen entstanden, die in der herkömmlichen Stichprobentheorie nicht behandelt werden. Einige dieser Verfahren sollen nun hier erläutert werden.

3.6.1. Marktforschung – Zielsetzungen und Problemstellungen

„Die Marktforschung ist ein Instrument der Absatzwirtschaft. In einem unübersichtlichen Markt mit der heute ständig wachsenden Differenzierung des Angebots braucht jede Unternehmensführung die von der Marktforschung ermittelten Daten als Entscheidungsgrundlage für ihre Marktpolitik“, aus Ott (1972). In diesem Sinne sind unter Marktforschung all diejenigen Methoden und Verfahren zu verstehen, die zur Analyse eines Marktes dienen.

Um eine solche Analyse vornehmen zu können, ist es zunächst notwendig, sich Daten zu einem noch näher zu definierenden Untersuchungsgegenstand zu verschaffen. Dabei kann man zwei verschiedene Wege einschlagen, die als

- **primärstatistische Marktforschung** und
- **sekundärstatistische Marktforschung**

bezeichnet werden. Bei der sekundärstatistischen Marktforschung geht man von einem Datenmaterial aus, das unabhängig von diesem Verwendungszweck erhoben wurde. Somit bildet die spezielle Analyse von bestehenden Daten aus statistischen Jahrbüchern der amtlichen Statistik, Statistiken von Verbänden, Instituten und anderen Organisationen die sekundärstatistische Marktforschung. Der an der speziellen Durchführung sekundärstatistischer Auswertungsmethoden interessierte Leser sei verwiesen auf Ott (1972).

Die eigentliche Anwendung von Stichprobenverfahren findet sich in der primärstatistischen Marktforschung wieder. Hierbei wird eine gezielte Erhebung zu einem bestimmten Zweck durchgeführt.

Im Rahmen der Primärerhebungen geht es darum, aktuelle Informationen über die Marktsituation und über die Marktteilnehmer zu gewinnen. Um diese Informationen zu gewinnen, werden Stichproben entnommen. Nun sind aber die Situationen und Zusammenhänge auf einem Markt so kompliziert, daß das Entnehmen von

Stichproben zu einem großen Problem wird. Das beginnt bei der Frage nach der Festlegung der Elemente der Grundgesamtheit, vgl. auch Abschnitt 3.1, geht über Kostenüberlegungen bis hin zu der wohl wichtigsten Frage nach der Repräsentativität von Stichproben.

Eine Stichprobe kann als **repräsentativ** bezeichnet werden, wenn von ihr ein Schluß auf die Gesamtheit erlaubt ist. Dies soll dadurch erreicht werden, daß die Stichprobe ein „verkleinertes Abbild“ der Grundgesamtheit darstellt.

Im Gegensatz zur wahrscheinlichkeitsorientierten Durchführung der Stichprobengewinnung unterscheidet die Marktforschung zwei Methoden, die zu einer repräsentativen Auswahl führen wollen. Dies sind zum einen **Zufallsstichproben** der bekannten Art und zum anderen **Beurteilungsstichproben**, bei denen auf Grund bewußter Überlegungen und gezielter Auswahlmaßnahmen ein repräsentatives Bild geschaffen werden soll. Warum diese unterschiedlichen Verfahren zum Einsatz kommen, mögen folgende **Beispiele** erläutern.

Beispiel:

- (a) Es soll eine Untersuchung in der Kfz-Branche durchgeführt werden. Ziel der Untersuchung ist es beispielsweise, den Bedarf an Sportwagen zu erkunden, um darauf die Produktion abzustellen. Durch die Zentralkartei in Flensburg mit ihren monatlichen Meldungen und Nachweisen liegen Unterlagen vor, die es erlauben, jeden Autobesitzer – und diese bilden die zugrundeliegende Grundgesamtheit – zu erfassen und einzurichten. Eine Auswahl nach dem Zufallsprinzip ist somit durchführbar.
- (b) Will man dagegen die Nachfragesituation bei alkoholischen Getränken analysieren, etwa um ein vorhandenes Angebot zu erweitern oder zu verkleinern, so ist eine zufällige Auswahl kaum noch zu realisieren. Die Gesamtheit aller Alkoholkonsumenten soll dann von einem Personenkreis nach der Beurteilung des Marktforschers repräsentiert werden.

Die Hauptunterschiede von Zufalls- und Beurteilungsstichproben sind in **Tab. 7** zusammengefaßt, vgl. Menges/Skala (1973), Hauser (1979).

Tab. 7: Unterschiede zwischen Zufalls- und Beurteilungsstichproben

Zufallsstichproben	Beurteilungsstichproben
erfordern eingehende Planung	sind einfach zu planen
eine bestimmte Erhebungseinheit muß in die Stichprobe gelangen (Substitutionen sind nicht zugelassen)	Substitutionen der Erhebungseinheiten sind zugelassen
sind relativ teuer	sind relativ billig
da sie auf der Wahrscheinlichkeitsrechnung beruhen, ist eine genaue Fehlerrechnung möglich	da sie nicht auf der Wahrscheinlichkeitsrechnung beruhen, ist keine Fehlerrechnung möglich (der Fehler muß ohne statistische Mittel „beurteilt“ werden)
besitzen eine theoretische Fundierung	sind nicht theoretisch zu fundieren
sind eindeutig „besser“	sind eindeutig „schlechter“

Die in der Marktforschung vorkommenden Zufallsstichproben sind die einfache Zufallsstichprobe, die geschichtete Auswahl und Klumpenauswahlen. Vor allem aber systematische Stichproben sind weit verbreitet.

Da wir uns in vorherigen Abschnitten ausführlich mit diesen Zufallsstichproben beschäftigt haben, sollen sie an dieser Stelle nicht nochmals behandelt werden. Der nachfolgende Abschnitt beschäftigt sich deshalb nur noch mit den für die Marktforschung vielleicht besonders typischen Beurteilungsstichproben.

3.6.2. Beurteilungsstichproben in der Marktforschung

Beurteilungsstichproben (**Auswahl nach Gutdünken, judgement sampling**) sind in der Markt- und Meinungsforschung häufig angewandte Auswahlverfahren. Hierbei wird versucht, durch Kenntnisse über die Struktur der Grundgesamtheit, eine repräsentative Auswahl zu treffen.

Wesentlicher Unterschied zur Auswahl aufs Geratewohl und die Rechtfertigung dieser Verfahren besteht in der Tatsache, daß ein Stichprobenplan vorliegt, der der Forderung nach Repräsentativität entsprechen soll. Diese Repräsentativität ist aber keinesfalls gesichert; es sind keine signifikanten Aussagen möglich. Dennoch haben Beurteilungsstichproben immer wieder brauchbare Ergebnisse ergeben, so daß sie entgegen der Forderung mancher Theoretiker eine weite Verbreitung gefunden haben.

Innerhalb der Beurteilungsstichproben unterscheidet man üblicherweise folgende Verfahren:

- die gezielte Auswahl „typischer“ Fälle (**typische Auswahl**)
- die Auswahl nach dem Konzentrationsprinzip („**cut-off-Verfahren**“)
- die **Quotenauswahl**

die nun beschrieben werden sollen.

A. Typische Auswahl

Bei der typischen Auswahl werden Einheiten bestimmt, die für die Grundgesamtheit als typisch gelten. Dieses „als typisch gelten“ wird vom Forscher anhand geeigneter Auswahlmerkmale bestimmt, die man aus einer hinreichenden Kenntnis über die Grundgesamtheit gewinnt.

Ist beispielsweise die Investitionsbereitschaft von Betrieben einer bestimmten Branche zu analysieren, so könnte man typische Betriebe gerade so auswählen, daß sie die gleiche durchschnittliche Größe aufweisen wie alle Betriebe der Branche insgesamt.

Sind die Kenntnisse über die Grundgesamtheit ausreichend, so können durch diese Auswahltechnik „relativ gute“ Ergebnisse erzielt werden. Dies ist besonders dann der Fall, wenn die zu erhebenden Merkmale (in unserem Fall die Investitionsbereitschaft) mit den die Stichprobe bestimmenden Merkmalen (bei uns die Betriebsgröße) zusammenhängen.

An diesem Punkt setzt aber eine massive Kritik der Methode an. Untersuchungen innerhalb der Marktforschung befassen sich meist mit mehreren Themen. Deshalb ist es äußerst schwierig, einen Zusammenhang zu allen zur Erhebung anstehenden Merkmalen zu sichern.

Weiterhin besteht die Gefahr, daß eine solche, ausschließlich auf Sachkunde

beruhende Auswahltechnik gerade zu den Ergebnissen führt, die man erwartet, da die Bestimmung „typischer“ Einheiten nach subjektiven Kriterien erfolgt. Hinzu kommt die meist subjektive Entscheidung, wenn man eine von mehreren gleichartigen Einheiten auswählen muß.

Das Verfahren der typischen Auswahl liefert deshalb nur „grobe Repräsentativität“. Trotzdem hat es ein gewisses Maß an Verwendbarkeit, besonders bei Pilotstudien, vgl. Böltken (1976).

B. Auswahl nach dem Konzentrationsprinzip

Dieses Verfahren wird vor allem in der amtlichen, besonders der Wirtschaftsstatistik verwandt. Bei dieser Erhebungsform versucht man nur wesentliche Einheiten in die Stichprobe einzubeziehen. Unwesentliche, unbedeutende Einheiten, Einheiten mit wenig Information werden nicht berücksichtigt, sie werden abgeschnitten. Deshalb wird auch oft vom „**cut-off**“-Verfahren gesprochen. Dieses Verfahren hat den Sinn, die Grundgesamtheit einzuengen und übersichtlicher zu machen, ohne daß die Einengung eine Wirkung auf die tatsächliche Grundgesamtheit hat.

Folgendes Beispiel mag das Verfahren verdeutlichen, vgl. Müller (1979). Es soll die Nachfrage von Investitionsgütern in einer bestimmten Branche, z.B. in der Kunststoffindustrie, erforscht werden. Das Nachfragepotential wird hauptsächlich von wenigen Großbetrieben gedeckt; zahlreiche Kleinbetriebe haben nur eine geringe Nachfrage an Investitionsgütern. Es ist nun naheliegend, alle Kleinbetriebe bei der anstehenden Untersuchung nicht zu berücksichtigen, denn die Kosten für die Erhebung einer Einheit (z. B. durch Interview) sind für alle Betriebe gleich groß, und der Informationsgewinn, den man durch Befragung der Kleinbetriebe erhält, ist nur sehr gering. Das Konzentrieren auf die Großbetriebe birgt dabei allerdings die Gefahr, daß dem Problem der Nichtbeantwortung ein viel stärkeres Gewicht zukommt als gewohnt. Dennoch ist die Auswahl nach dem Konzentrationsprinzip in der Investitionsgüter-Marktforschung von hervorragender Wichtigkeit.

Die Auswahlform bringt besonders dann „brauchbare“ Ergebnisse, wenn die Grundgesamtheit bekannt ist und wenn die Konzentration bestimmt werden kann. Auch hier kann natürlich das Problem entstehen, daß bei mehreren zu erhebenden Merkmalen verschiedene, sogar andersartige Konzentrationen auftreten. Hier muß dann eine vom Untersuchungsziel abhängige Auswahl der Einheiten vorgenommen werden.

Ein weiteres Problem dieses Auswahlverfahrens ergibt sich, wenn die Erhebung fortschreitend weitergeführt wird. Dies ist z. B. denkbar, wenn die Marktsituation über längere Zeit beobachtet werden soll. Um gleiche Bedingungen zu erhalten, wählt man dann immer die gleichen Einheiten aus, man hält den **Berichtskreis** konstant. Konjunkturelle Änderungen können aber bewirken, daß sich die Zugehörigkeit zum Berichtskreis verändert. Beispielsweise können Entlassungen dazu führen, daß ein Betrieb nicht mehr als Großbetrieb bezeichnet werden kann oder Neugründungen bzw. Konsolidierung von Kleinbetrieben neue Großbetriebe entstehen lassen. Ist dies der Fall, so muß durch eine Untersuchung der Grundgesamtheit der Berichtskreis neu festgelegt und die ermittelten Merkmalswerte müssen bei nachfolgenden Stichproben korrigiert werden. Der an der Durchführung des Ausgleichs von „**Berichtssprüngen**“ interessierte Leser sei auf Esenwein-Rothe (1976) verwiesen.

C. Quotenauswahl

Die Quotenauswahl, vgl. auch Abschnitt 3.4, ist die wohl bekannteste Art einer Beurteilungsstichprobe, wenn nicht sogar die in der Praxis am meisten verwandte Auswahltechnik überhaupt, vgl. Böltken (1976). Auf jeden Fall ist sie das in der Markt- und Meinungsforschung am meisten angewandte Auswahlverfahren.

Zum besseren Verständnis dieser Methode und den damit zusammenhängenden Frage- und Problemstellungen stellt man sich vor, daß die Stichprobe durch Umfrage mit Hilfe von Interviewern erstellt wird. Der wesentliche Unterschied zu den bisher geschilderten Beurteilungsstichproben besteht darin, daß die Einheiten nicht gezielt ausgesucht werden. Der Interviewer hat vielmehr eine freie Auswahl bei der Erhebung, muß diese aber bestimmten Auflagen – den Quoten – anpassen.

Eine **Quote** ist ein festgelegtes Verhältnis, nach dem bestimmte Ausprägungen eines Merkmals in der Stichprobe enthalten sein sollen. Sind z. B. 20% der Verbraucher eines Produktes weiblich, so sind bei der Festlegung dieser Quote auch 20% der zu befragenden Personen weiblich zu wählen. Führt man eine Quotenanweisung streng durch, so ist der Anteil von Einheiten mit der bestimmten Merkmalsausprägung in der Stichprobe gleich dem in der Grundgesamtheit. Somit kann bezüglich des **Quotierungsmerkmals** ein repräsentatives Abbild der Grundgesamtheit geschaffen werden.

Um ein hohes Maß an Repräsentativität für eine zugrundeliegende Grundgesamtheit zu erhalten, werden im allgemeinen nicht nur eine, sondern mehrere Quoten angegeben. Die Angabe dieser Quoten erfolgt durch den sogenannten **Quotenplan** an den Interviewer. In ihm sind alle Anweisungen enthalten, die der Interviewer zu beachten hat. Er ist nur angehalten, diese Anweisungen zu befolgen, ansonsten ist er frei in der Auswahl der Erhebungseinheiten.

Ein einfaches **Beispiel** mag das bisher Gesagte verdeutlichen. Ein Marktforschungsinstitut wird beauftragt, das Nachfrageverhalten bei Konsumgütern zu analysieren. Man glaubt, daß die zugrundeliegende Grundgesamtheit der Gesamtbevölkerung durch die Quotierungsmerkmale Geschlecht, Alter und Erwerbstätigkeit charakterisiert wird (dies ist natürlich eine sehr vereinfachte Annahme). Auf Grund der letzten Vollerhebung sind für die genannten Merkmalswerte die entsprechenden Anteile in der Grundgesamtheit bekannt. Beispielsweise gilt

- Geschlecht: 48% männlich, 52% weiblich;
- Alter: 20% 15 und jünger; 43% 15 bis 45; 22% 45 bis 65;
15% 65 und älter;
- Erwerbstätigkeit: 40% Erwerbstätige; 2% Arbeitslose;
58% Nicht-Erwerbstätige.

Will man nun eine Stichprobe vom Umfang 2000 nach dem Quotenverfahren erheben, so müssen 960 männliche und 1040 weibliche Personen erhoben werden; 400 unter 15 Jahren, 860 zwischen 15 und 45 Jahren etc.

Die 2000 durchzuführenden Interviews können z. B. auf 100 Interviewer verteilt werden. Jeder Interviewer erhält dann eine Quotenanweisung für 20 durchzuführende Interviews auf der wiederum genau festgelegt ist, wie viele der 20 zu befragenden Personen männlich bzw. weiblich sind, etc. Die Interviewer müssen dann Personen ausfindig machen (zu Hause, auf der Straße, am Arbeitsplatz), die diesen Anweisungen genügen. Nach jedem geführten Interview werden dann die jeweils zu-

treffenden Angaben der gerade interviewten Person gestrichen, wodurch sich der Auswahlspielraum immer weiter einengt. Deshalb ist darauf zu achten, daß zum Schluß keine unmöglichen oder sehr seltenen Merkmalskombinationen verbleiben. Zum Beispiel dürfte es wesentlich schwieriger sein einen (mindestens) 65jährigen Erwerbstägigen ausfindig zu machen als einen Nicht-Erwerbstägigen, der älter als 65 Jahre ist. Dieser Schwierigkeit kann zum einen durch den erfahrenen Interviewer, der eine ausreichende Kenntnis über die Grundgesamtheit hat, oder durch sogenannte **kombinierte Quoten** zumindest teilweise aus dem Weg gegangen werden. Eine kombinierte Quote liegt im Beispiel dann vor, wenn man nicht nur die obigen Quoten festlegt, sondern z.B. noch fordert: 2% Erwerbstägige über 65 Jahre; 0% Arbeitslose über 65 Jahre; 13% Nicht-Erwerbstägige über 65 Jahre. Die Kombination von Quotenmerkmalen erfordert allerdings eine noch größere Kenntnis der zugrundeliegenden Gesamtheit, was zu erheblichen Schwierigkeiten führen kann.

Es wurde bereits angesprochen, daß der Interviewer einen direkten Einfluß auf die Auswahl der Einheiten hat. Dieser Einfluß, der Hauptkritikpunkt am Quotenverfahren, soll möglichst gering gehalten werden. Deshalb wird häufig versucht, relativ schwierige Quoten anzugeben, um somit den Spielraum des Interviewers einzuschränken, so daß er nicht nur Personen befragt, die ihm sympathisch sind, die er leicht erreichen kann etc. Andererseits dürfen die Quoten nicht zu schwierig sein, da sonst der Interviewer ein „**quota fitting**“, ein Anpassen der Befragten an die Quoten vornehmen könnte. Man muß deshalb eine geeignete Quotierung vornehmen, um so die subjektiven Einflüsse des Interviewers weitgehend auszuschalten. Dies gelingt aber nicht immer. So wurde festgestellt, vgl. Noelle (1963), daß Bewohner von Mietshäusern überrepräsentiert sind, Interviewer eher Personen mit höherem sozialen Status befragen oder, vgl. Boeltken (1976), daß informiertere, aufgeschlossene und mobilere Personen stärker berücksichtigt werden. Ein weiterer Nachteil des Verfahrens ergibt sich aus der Frage, nach welchen Merkmalen quotiert werden soll. Auch hier (vgl. die anderen Beurteilungsstichproben) ist es wünschenswert, eine enge Korrelation zu den interessierenden Merkmalen herzustellen. Erstens jedoch fehlen dazu häufig die Informationen aus einer Vollerhebung, und zweitens tritt auch hier das Problem der Mehrthemenbefragungen auf.

Hat man geeignete Merkmale zur Quotierung gefunden, so müssen die Quoten bestimmt werden. Eine Angabe aus einer Vollerhebung ist dabei gar nicht so problemlos, wie sich dies zunächst darstellt. Die Unterlagen zur Erstellung der Quoten müssen nämlich absolut verlässlich sein, sonst tritt eine Verzerrung der Untersuchungsergebnisse schon durch den Quotenplan selbst auf. Was man dabei unter „**Verlässlichkeit**“ verstehen soll, ist schwer zu beantworten. Zumindest ist die Aktualität der Unterlagen von großer Wichtigkeit. Alle diese Nachteile werden aber bewußt in Kauf genommen, da das Quotenverfahren auch eine Reihe von Vorteilen mit sich bringt. Abgesehen davon, daß Quotenstichproben oft die einzige Möglichkeit bieten, eine mehr oder weniger repräsentative Auswahl zu tätigen, sind sie sehr kostengünstig. Neben dieser Wirtschaftlichkeit, die besonders im Gegensatz zu zufälligen Stichproben zum Tragen kommt, haben Quotenverfahren ein gewisses Maß an Schnelligkeit. Dies begründet sich in der schnellen Ausarbeitung der Quotenpläne, im Gegensatz zur relativ komplizierten Erstellung von zufälligen Stichprobenplänen. Somit ist das Quotenverfahren besonders dann geeignet, wenn kurzfristig Untersuchungen vorgenommen werden sollen.

Ein weiterer Vorteil entsteht dadurch, daß Quotenverfahren auf spezielle Personengruppen anwendbar sind. Eine Marktuntersuchung über Rauchgewohnheiten

ist somit möglich, denn eine zufällige Auswahl ist auf Grund eines fehlenden Verzeichnisses aller Raucher nicht anwendbar.

Alles in allem ist die Quotenauswahl zwar ein verbreitetes, aber berechtigterweise umstrittenes Instrument, zu einer Stichprobe zu gelangen. Die Diskussion darüber wird sicherlich noch lange bestehen bleiben, denn einerseits ist eine Repräsentativität keinesfalls gesichert, andererseits werden aber immer wieder sehr gute Ergebnisse mit dem Verfahren erzielt.

3.7. Die Bedeutung der Stichprobenverfahren

Stichprobenverfahren haben in den letzten Jahren immer mehr an Bedeutung gewonnen, dienen sie doch dem Zweck, einen Schluß von einem Teil auf eine Gesamtheit zu ziehen. Dieses ist bei vielerlei Forschungsgegenständen vonnöten, und so ist die Theorie der Stichprobenverfahren zur Lösung vieler praktischer Probleme herangezogen worden.

Die Anwendungsbereiche der Theorie sind vielfältig, doch sollte man sich davor hüten, ein beliebiges, „passendes“ Verfahren in jeder Situation anzuwenden. Die geschilderten Verfahren sind sehr eng an ihre Voraussetzungen gebunden, und so können Fehlplanungen und falsche Anwendungen schnell zu unbefriedigenden und sogar falschen Ergebnissen führen. Demzufolge ist es die Aufgabe jeder eine Stichprobenerhebung durchführenden Person, ein großes Maß an Sorgfalt walten zu lassen und alle zur Verfügung stehenden Mittel finanzieller und theoretischer Art voll auszuschöpfen. In diesem Zusammenhang muß noch einmal gesagt werden, daß die in diesem Abschnitt 3 geschilderten Probleme nur ein geringer Teil derer sind, die in praktischen Erhebungen auftreten können.

4. Theorie der Meßfehler, Ausreißertests, Datentransformationen

Nimmt man irgendwelche Messungen vor, z. B. physikalische oder chemische, so sind diese niemals exakt; das bemerkt man, wenn Messungen wiederholt werden oder wenn verschiedene Personen das Gleiche messen. Es treten dann Schwankungen oder Ungenauigkeiten in den erhaltenen Meßwerten auf.

Beispiel: Zur Verdeutlichung der Problematik sind in **Tab. 8** verschiedene Ergebnisse bei der Messung der Astronomischen Einheit AE (große Halbachse der Erdbahnelipse = mittlerer Abstand Sonne–Erde) aufgeführt; vgl. Youden (1972).

Man weiß natürlich, daß die Astronomische Einheit AE im Beobachtungszeitraum als konstant anzusehen ist, so daß die unterschiedlichen Ergebnisse nur durch Meßfehler entstanden sein können.

In Abschnitt 4.1 wird nun zunächst der **Begriff des Meßfehlers** präzisiert und in Abschnitt 4.2 wird das **Rechnen mit fehlerbehafteten Zahlen** an zwei Beispielen ausführlich erläutert.

Wie kann man nun **Meßungenauigkeiten beschreiben und kontrollieren**, insbesondere auch dann, wenn sich Meßvorgänge über einen längeren Zeitraum hinziehen? Dieser Problematik ist der Abschnitt 4.3 gewidmet.

Bei wiederholten Messungen stellt man oft fest, daß einige Meßwerte sehr weit von allen anderen entfernt liegen. Schätzt man dann den wahren Meßwert durch den Mittelwert aller Meßergebnisse, so können weit vom „Normalen“ abweichende

Tab. 8: Meßergebnisse für die Astronomische Einheit AE

Ort bzw. Durchführende und Jahr der Messung	AE in Millionen Meilen	Schätzung der Schwankung durch den Experimentator	
Newcomb, 1895	93,28	93,20	-93,35
Hinks, 1901	92,83	92,79	-92,87
Noteboom, 1921	92,91	92,90	-92,92
Spencer Jones, 1928	92,87	92,82	-92,91
Spencer Jones, 1931	93,00	92,99	-93,01
Witt, 1933	92,91	92,90	-92,92
Adams, 1941	92,84	92,77	-92,92
Brouwer, 1950	92,977	92,945	-93,008
Rabe, 1950	92,9148	92,9107	-92,9190
Millstone Hill, 1958	92,874	92,873	-92,875
Jodrell Bank, 1959	92,876	92,871	-92,882
S.T.L., 1960	92,9251	92,9166	-92,9335
Jodrell Bank, 1961	92,960	92,958	-92,962
Cal. Tech., 1961	92,956	92,955	-92,957
Soviets, 1961	92,813	92,810	-92,816

Messungen das Ergebnis stark beeinflussen. Daher kann man sich die Frage stellen, ob solche Meßwerte sich vielleicht nur durch Unregelmäßigkeiten im Versuch ergeben haben und sie daher besser nicht in die Berechnungen mit einbezogen werden sollten. In Abschnitt 4.4 werden verschiedene Tests angegeben, die überprüfen, ob ein stark abweichender Meßwert vielleicht ein **Ausreißer** ist. Schließlich werden in Abschnitt 4.5 noch einige **Transformationen** angegeben. Diese dienen z. B. dazu, Meßwerte so zu transformieren, daß die den transformierten Werten zugrundeliegende Verteilung annähernd eine Normalverteilung ist; dies ist wünschenswert, da viele statistische Verfahren auf einer Normalverteilungsannahme basieren.

4.1. Der Meßfehler bei der Datengewinnung

Führt man eine Messung durch, so wird der gemessene Wert fast immer vom wahren Wert abweichen; die Messung ist also mit einem Fehler behaftet. Dieser Fehler zerfällt noch einmal in den systematischen und in den zufälligen Fehler.

Der **systematische Fehler** ist derjenige, der bei gleichen Meßvorgängen immer wieder die gleiche Struktur besitzt.

Beispiel:

- (a) Geht eine Uhr konstant um 1 Promille vor, so ergibt dies bei jeder Messung von 10 Sekunden einen systematischen Fehler von 0,01 Sekunden.
- (b) Zeigt eine Waage ohne jegliche Belastung schon einen positiven Gewichtsstand an, so ist jede Wägung mit einem systematischen Fehler behaftet.

Wie man Größe und Art von systematischen Fehlern ermitteln kann und wie man systematische Fehler eventuell eliminiert, wird in Abschnitt 4.4 behandelt.

Im Gegensatz zum systematischen Fehler beeinflußt der **zufällige oder statistische Fehler** die Meßergebnisse in unterschiedlicher, unvorhersehbarer Weise. Die Größe dieses Fehlers läßt sich jedoch abschätzen, wenn Ergebnisse mehrerer gleichartiger, etwa wiederholter Messungen, zur Verfügung stehen. So stellt sich der zufällige Fehler oft als Summe vieler Elementarfehler (z. B. Ablesefehler, Umge-

bungsveränderungen) dar und kann damit – wegen der Gültigkeit des zentralen Grenzwertsatzes, vgl. Abschnitt 9 in Kap. II – als normalverteilt angenommen werden. Die Ergebnisse x_1, \dots, x_n wiederholter Messungen eines wahren Wertes μ können dann als Realisationen einer $N(\mu^*, \sigma^2)$ -verteilten Zufallsgröße X aufgefaßt werden. Hier ist der zufällige Fehler bei der i-ten Messung gerade

$$\varepsilon_i = x_i - \mu^*,$$

während der systematische Fehler, den man auch **Bias** nennt, gegeben ist durch

$$b = \mu^* - \mu.$$

Damit ergibt sich nun:

$$x_i = \mu + b + \varepsilon_i$$

= wahrer Wert + systematischer Fehler + zufälliger Fehler.

Mit den Meßwerten x_1, \dots, x_n lassen sich $\mu^* = \mu + b$ und σ z. B. durch

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

schätzen; weiterhin können Konfidenzintervalle für μ^* und σ^2 angegeben werden, vgl. Abschnitt 1.3 in Kap. IV. Dadurch läßt sich ein Eindruck von der Größe und der Struktur des zufälligen Fehlers gewinnen.

Über die Größe des systematischen Fehlers b hingegen läßt sich aus den Meßwerten x_1, \dots, x_n allein keine Information ziehen, da dieser für alle n Werte gleich ist. Es muß also dafür Sorge getragen werden, daß ein systematischer Fehler weitestgehend vermieden wird. Dabei hilft z. B. die sorgfältige **Eichung (Kalibration)** bzw. **Justierung** der verwandten Meßgeräte. Kann man annehmen, daß kein systematischer Fehler vorhanden ist, so liefert \bar{x} eine Schätzung für den wahren Wert μ , und ein Konfidenzintervall für μ entspricht dann gerade einem Konfidenzintervall für den Parameter μ^* .

Häufig findet man Angaben etwa der Gestalt

$$\bar{x} \pm s_{\bar{x}} \quad \text{mit} \quad s_{\bar{x}} = \frac{1}{\sqrt{n}} s = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

die nun allerdings nicht besagen, daß der Parameter μ^* und – falls kein systematischer Meßfehler vorhanden ist – auch der wahre Wert μ zwischen $\bar{x} - s_{\bar{x}}$ und $\bar{x} + s_{\bar{x}}$ liegen muß; vielmehr entspricht diese Angabe (zumindest bei großem Stichprobenumfang n) einem Konfidenzintervall (zum Niveau $1 - \alpha \simeq 0,68$) für den Parameter μ^* : Der Parameter μ^* wird mit einer Wahrscheinlichkeit von ca. 68% vom Intervall

$$[\bar{x} - s_{\bar{x}}, \bar{x} + s_{\bar{x}}]$$

überdeckt. Bei Normalverteilungsannahme ist nämlich für

$$s_{\bar{x}} \simeq \sigma / \sqrt{n},$$

vgl. Kap. IV, Abschnitt 1.3.2 und Tabelle im Anhang,

$$[\bar{x} - s_{\bar{x}} u_{1-\alpha/2}; \bar{x} + s_{\bar{x}} u_{1-\alpha/2}]$$

ein approximatives $1 - \alpha$ Konfidenzintervall für μ^* und es ist $u_{1-\alpha/2} = 1$ falls gilt $1 - \alpha \simeq 0,68$; $u_{1-\alpha/2} = u_{0,84} \simeq 1$. Zwischen $\bar{x} - 2s_{\bar{x}}$ und $\bar{x} + 2s_{\bar{x}}$ liegt μ^* dann

mit 95% Sicherheit und zwischen $\bar{x} - 3s_{\bar{x}}$ und $\bar{x} + 3s_{\bar{x}}$ schließlich mit einer Sicherheit von rund 99%; im Intervall

$$[\bar{x} - \frac{2}{3}s_{\bar{x}}, \bar{x} + \frac{2}{3}s_{\bar{x}}]$$

liegt μ^* mit ca. 50% Sicherheit. Die Größe $\frac{2}{3}s_{\bar{x}}$ nennt man auch den **wahrscheinlichen Fehler**.

Wie sich eine Fehlinterpretation solcher Genauigkeitsangaben auswirken kann, zeigt das folgende Beispiel.

Beispiel: (vgl. auch Pfanzagl (1978) und Lustig/Pfanzagl (1957): „Industrielle Qualitätskontrolle, 50 Beispiele aus der Praxis“, Wien.)

Für Metallfolien, die bei der Magnetophonerzeugung Verwendung finden, war eine Stärke von $6 \cdot 10^{-3}$ mm $\pm 0,2 \cdot 10^{-3}$ mm vorgeschrieben. Nachdem eine größere Lieferung dieser Metallfolien übernommen worden war, kam es bei der Montage zu Schwierigkeiten. Wie eine Rückfrage ergab, stand für die Prüfung der Folien kein Meßgerät mit der erforderlichen Genauigkeit zur Verfügung. Deshalb hatte der Meister jeweils 10 Folien zusammengelegt und die Anforderung von $(6 \pm 0,2) \cdot 10^{-3}$ mm auf $(60 \pm 2) \cdot 10^{-3}$ mm abgeändert.

Dieses Vorgehen ist unzulässig. Ist die Stärke der einzelnen Folien nämlich $N(\mu, \sigma^2)$ -verteilt, so ist die Stärke von 10 zusammengelegten Folien verteilt nach $N(10\mu, 10\sigma^2)$ und somit die für die Breite des Streubereiches maßgebende Standardabweichung nur $\sqrt{10}\sigma$. Also lautet die entsprechende Genauigkeitsangabe für das Päckchen der 10 Folien

$$(60 \pm \sqrt{10} \cdot 0,2) \cdot 10^{-3} \text{ mm} = (60 \pm 0,63) \cdot 10^{-3} \text{ mm}.$$

Die Genauigkeit von $(60 \pm 2) \cdot 10^{-3}$ mm wird dagegen von dem Päckchen selbst dann noch eingehalten, wenn die einzelnen Folien einen dreimal so großen Streubereich aufweisen, wie gefordert ist, nämlich noch bei $(6 \pm 0,6) \cdot 10^{-3}$ mm.

Das in obigem Beispiel auftauchende Problem der **Fehlerfortpflanzung** wird ausführlich in Abschnitt 4.2 dieses Kapitels behandelt.

Weitere Konfidenzintervalle und Punktschätzungen sowie Prognose- und Toleranzintervalle sind – auch für den Fall, daß die Meßwerte nicht als Auskommen von normalverteilten zufälligen Größen angenommen werden können – in Kap. IV zu finden.

Hier soll noch erwähnt werden, daß immer, egal wie die Meßwerte verteilt sind, die **Tschebyscheffsche Ungleichung**, vgl. Kap. II, gilt:

$$P(|\bar{X} - \mu^*| > k \cdot \sigma / \sqrt{n}) \leq \frac{1}{k^2}.$$

Die Wahrscheinlichkeit, daß man einen Mittelwert zwischen $\mu^* - k \cdot \sigma / \sqrt{n}$ und $\mu^* + k \cdot \sigma / \sqrt{n}$ erhält, ist also mindestens gleich $1 - 1/k^2$; oder wie man auch anders sagt: Mit einer Wahrscheinlichkeit von mindestens $1 - 1/k^2$ gilt für den Erwartungswert μ^* :

$$\bar{x} - k \cdot \sigma / \sqrt{n} \leq \mu^* \leq \bar{x} + k \cdot \sigma / \sqrt{n}.$$

Hierbei wird man $k > 1$ wählen, denn für $0 < k \leq 1$ erhält man eine triviale Aussage; für $k = 2$ ist μ^* mit einer Wahrscheinlichkeit von mindestens 0,75 in diesem Intervall zu finden und für $k = 3$ mit mindestens ca. 89% Wahrscheinlichkeit.

Wird in konkreten Situationen die Größe σ/\sqrt{n} durch $s_{\bar{x}} = s/\sqrt{n}$ geschätzt, so können sich diese Wahrscheinlichkeiten natürlich noch wesentlich verringern, wenn der Stichprobenumfang n nicht sehr groß ist.

Wir wollen nun einige Maßzahlen vorstellen, die die **Güte von Meßverfahren** beurteilen sollen. Wie schon aus den bisherigen Ausführungen zu entnehmen ist, lässt sich ein Meßvorgang bei zugrundeliegender Normalverteilung durch den wahren Wert μ zusammen mit den Parametern μ^* und σ vollständig charakterisieren.

Dabei beschreibt die Standardabweichung σ die Größe des zufälligen Fehlers. Als Schätzer für diesen Wert kann man z.B.

$$s_c = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - c)^2}$$

verwenden, wobei dieser Ausdruck gerade für $c = \bar{x}$ minimal wird (**Minimumeigenschaft des arithmetischen Mittels**). Um eine Vergleichbarkeit zu erleichtern, wird deshalb üblicherweise der **mittlere Fehler**

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

(**empirische Standardabweichung**, vgl. aber auch Abschnitt 1.3.1 in Kap. IV) als Schätzer verwandt.

Eine weitere Klasse von Schätzern der Standardabweichung σ ist gegeben durch

$$d_c = \frac{1}{n} \sum_{i=1}^n |x_i - c|.$$

In einigen Bereichen wird der **durchschnittliche Fehler**

$$\delta = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

benutzt. Da jedoch d_c nicht durch $c = \bar{x}$, sondern durch den Median $c = \tilde{x}_{0,5}$ minimiert wird (**Minimumeigenschaft des Medians**) ist aus Gründen der Vergleichbarkeit die durchschnittliche absolute Abweichung vom Median

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}_{0,5}|$$

vorzuziehen, vgl. auch Kap. I, Abschnitt 5.

In einigen Bereichen, vor allem in der Technik, stützt man sich bei der Schätzung von σ auf Spannweitenberechnungen, vgl. Abschnitt 1.3 in Kap. IV. Alle diese Schätzungen von σ können als Genauigkeitsmaße dienen.

Die bis jetzt vorgestellten Maßzahlen sind nicht sehr aussagekräftig, wenn nicht zusätzlich der zu messende Wert bzw. eine Schätzung dieses Wertes angegeben wird. So erscheint bei einem Wert $\bar{x} = 10$ ein mittlerer Fehler $s = 1$ recht groß, während bei $\bar{x} = 10000$ das Gegenteil der Fall ist. Diese Tatsache wird (für $x_i > 0$) sowohl durch den **relativen Fehler (Variationskoeffizient, Variationszahl)**

$$v = \frac{s}{\bar{x}} = \frac{\text{Stichprobenstandardabweichung}}{\text{Stichprobenmittelwert}}$$

als auch durch die **Streuzahl**

$$z = \frac{R}{\bar{x}} = \frac{\text{Stichprobenspannweite}}{\text{Stichprobenmittelwert}}$$

berücksichtigt; diese Größen benutzt man häufig auch zur Ermittlung von – zur Einhaltung einer geforderten Genauigkeit bei der Schätzung von μ^* – benötigten Stichprobenumfängen mittels Leitertafeln, vgl. Abschnitt 1.4 in Kap. IV.

Wir wollen nun noch den systematischen Fehler in die Gütebeurteilung mit einbeziehen. Dies ist z. B. wichtig bei der Untersuchung neuer Meßmethoden, da man oft annehmen muß, daß ein systematischer Fehler vorhanden ist. Ein geeignetes Gütemaß ist dann etwa der sogenannte (**relative**) **Gesamtfehler**

$$G = \frac{|\bar{x} - \mu| + 2s}{|\mu|},$$

der zuweilen auch als $\tilde{G} = 100 \cdot G\%$ angegeben wird. Der wahre Wert μ kann dabei durch eine Schätzung ersetzt werden, die man z. B. aus den mittels einer alten, bewährten Methode (die man etwa aus Kostengründen durch die zu beurteilende neue Methode ersetzen möchte) erhaltenen Messungen gewinnt. Hierbei unterstellt man natürlich, daß die „bewährte“ Methode keinen systematischen Fehler hat. Zu beachten ist, daß G für $\mu = 0$ nicht definiert ist und zudem auch bei völlig verschiedenen Sachverhalten möglicherweise gleiche Werte liefert, vgl. Abb. 5 und 6: In

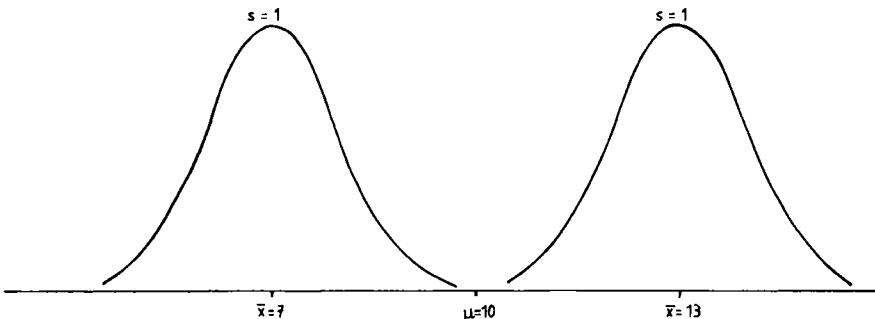


Abb. 5: Verteilungen des zufälligen Fehlers, so daß sich $G = 0,5$ ergibt

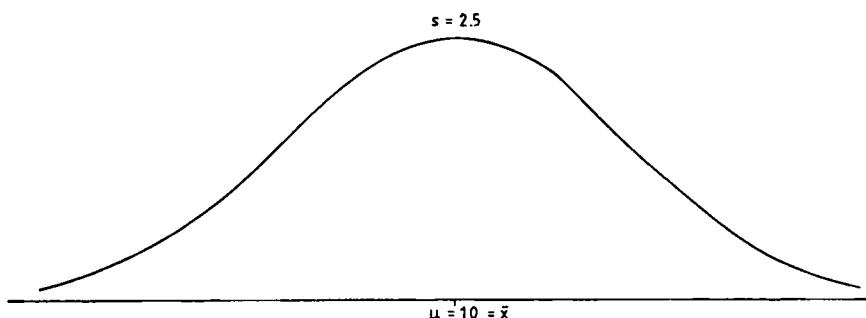


Abb. 6: Eine andere Verteilung des zufälligen Fehlers, die $G = 0,5$ liefert

allen Verteilungssituationen des zufälligen Fehlers ergibt sich $G = 0,5$. In manchen experimentell-analytischen Bereichen ist eine Methode schon als sehr gut anzusehen, wenn $G \leq 0,25$ ist. Ist $G > 0,5$, so ist eine Methode i. a. als unbrauchbar zu betrachten.

4.2. Das Gaußsche Fehlerfortpflanzungsgesetz

In Abschnitt 4.1 haben wir uns mit der Berechnung von Maßen für die Genauigkeit von Meßwerten beschäftigt. Dort gingen wir implizit davon aus, daß die interessierenden Größen direkt gemessen werden konnten (**direkte Beobachtung**). Muß die interessierende Größe aus anderen meßbaren Größen berechnet werden, so spricht man von einer **indirekten Beobachtung** der Größe.

Beispiel:

- (a) Bei Flächen- und Volumenmessungen werden zunächst nur Seiten- bzw. Kantenlängen gemessen. Daraus berechnet man dann Fläche oder Volumen, die somit nur indirekt beobachtet werden.
- (b) Zur Bestimmung spezifischer Gewichte mißt man Masse und Volumen und berechnet daraus indirekt das eigentlich interessierende spezifische Gewicht.

Die Messungen, die zur Berechnung der eigentlich interessierenden Größen vorgenommen werden, sind natürlich fehlerbehaftet. Diese Fehler gehen in die Berechnung ein. Das **Fehlerfortpflanzungsgesetz** nach **Carl Friedrich Gauß** (1777–1855) ist nun ein Instrument zur näherungsweisen Bestimmung des mittleren Fehlers bei indirekten Beobachtungen.

Berechnet sich die interessierende Größe als Funktion $f(X_1, \dots, X_n)$ von n direkt meßbaren Größen X_1, \dots, X_n und ist $f(X_1, \dots, X_n)$ nach den X_i differenzierbar ($\frac{\partial f}{\partial X_i}$), so gilt das Fehlerfortpflanzungsgesetz: Sind für $i, j = 1, \dots, n$, $\mu_i = E(X_i)$ der bekannte Erwartungswert von X_i , $\sigma_i^2 = \text{Var}(X_i)$ die bekannte Varianz von X_i und $\sigma_{ij} = \text{Cov}(X_i, X_j)$ die bekannte Kovarianz von X_i und X_j , so gilt näherungsweise:

$$E(f(X_1, \dots, X_n)) \approx f(\mu_1, \dots, \mu_n) \quad \text{und}$$

$$\text{Var}(f(X_1, \dots, X_n)) \approx \sum_{i=1}^n \left(\frac{\partial f}{\partial X_i} \Big|_{X=\mu} \right)^2 \sigma_i^2 + 2 \cdot \sum_{i < j} \left(\frac{\partial f}{\partial X_i} \Big|_{X=\mu} \right) \left(\frac{\partial f}{\partial X_j} \Big|_{X=\mu} \right) \sigma_{ij}.$$

Hierbei bezeichnet $\frac{\partial f}{\partial X_i} \Big|_{X=\mu}$ den Wert der partiellen Ableitung von $f(X_1, \dots, X_n)$

nach X_i an der Stelle $(X_1, \dots, X_n) = (\mu_1, \dots, \mu_n)$. Die hier angegebene Näherung für die Varianz von $f(X_1, \dots, X_n)$ ist natürlich auch eine solche für das Quadrat des mittleren Fehlers dieser Größe.

Zunächst wollen wir einige einfache Funktionen betrachten. Sind X_1 und X_2 unabhängige Zufallsvariable mit Erwartungswerten μ_1 und μ_2 und Varianzen σ_1^2 und σ_2^2 , so gilt:

(a) für $f(X_1, X_2) = X_1 + X_2$:

$$E(f(X_1, X_2)) = \mu_1 + \mu_2 \quad \text{und}$$

$$\text{Var}(f(X_1, X_2)) = 1^2 \cdot \sigma_1^2 + 1^2 \cdot \sigma_2^2 + 2 \cdot 0 = \sigma_1^2 + \sigma_2^2;$$

(β) für $f(X_1, X_2) = X_1 - X_2$:

$$E(f(X_1, X_2)) = \mu_1 - \mu_2 \quad \text{und}$$

$$\text{Var}(f(X_1, X_2)) = 1^2 \cdot \sigma_1^2 + (-1)^2 \cdot \sigma_2^2 = \sigma_1^2 + \sigma_2^2;$$

(γ) für $f(X_1, X_2) = X_1 \cdot X_2$:

$$E(f(X_1, X_2)) = \mu_1 \cdot \mu_2 \quad \text{und}$$

$$\text{Var}(f(X_1, X_2)) \simeq \mu_2^2 \cdot \sigma_1^2 + \mu_1^2 \cdot \sigma_2^2;$$

(δ) für $f(X_1, X_2) = X_1/X_2$:

$$E(f(X_1, X_2)) \simeq \mu_1/\mu_2 \quad \text{und}$$

$$\text{Var}(f(X_1, X_2)) \simeq \frac{1}{\mu_2^2} \cdot \sigma_1^2 + \frac{\mu_1^2}{\mu_2^4} \cdot \sigma_2^2 = \frac{1}{\mu_2^2} \left(\sigma_1^2 + \frac{\mu_1^2}{\mu_2^2} \cdot \sigma_2^2 \right);$$

(ε) für $f(X) = \ln X$:

$$E(\ln X) \simeq \ln E(X) \quad \text{und}$$

$$\text{Var}(\ln X) \simeq \text{Var}(X)/E(X)^2;$$

die Varianz des natürlichen Logarithmus von X ist also näherungsweise gleich dem Quadrat des Variationskoeffizienten von X (vgl. auch Abschnitt 4.5C).

Diese einfachen Ergebnisse werden uns in den folgenden Beispielen noch von Nutzen sein. An den Fehlerformeln läßt sich auch gut erkennen, welche Meßgröße auf den Gesamtfehler den größten Einfluß ausübt, so daß eine Verringerung der Varianz dieser Größe am wirkungsvollsten ist.

Beispiel:

(c) In einer Eigenheim-Siedlung haben jeweils ein großes und ein kleines Grundstück die gemeinsame Grenze X_1 , vgl. Abb. 7. Auf dem jeweils größeren Grundstück muß ein Transformatorhäuschen, das 2 qm Fläche benötigt, gebaut werden. Der Gesamtpreis für das Grundstück (das Transformatorhäuschen wird nicht berechnet) beläßt sich auf K DM. Der Kaufpreis jedes einzelnen Grundstücks soll sich nach dem Verhältnis der beiden Flächen $X_1 \cdot X_2$ und $X_1 \cdot X_3$ – 2 berechnen. Unabhängig voneinander werden die Strecken X_1 , X_2 und X_3 bei jedem Grundstückpaar gemessen, wobei folgende Verteilungen angenommen werden können:

$$X_1 \sim N(15, 0,015), \quad X_2 \sim N(10, 0,01) \quad \text{und} \quad X_3 \sim N(20, 0,02).$$

Da sich der Kaufpreis der einzelnen Grundstücke aus dem Gesamtpreis und

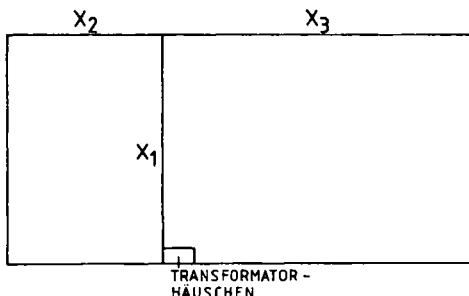


Abb. 7: Lage eines Grundstückpaars im Beispiel

dem Verhältnis der Flächen berechnet, möchte man wissen, wie groß die Standardabweichung der Funktion

$$\frac{Z_1}{Z_2 - 2} = \frac{X_1 \cdot X_2}{X_1 \cdot X_3 - 2} = f(Z_1, Z_2)$$

ist. Nach dem Fehlerfortpflanzungsgesetz bestimmen wir die Varianz der Funktion $f(Z_1, Z_2)$ wie folgt, vgl. (y),

$$\begin{aligned}\text{Var}(X_1 \cdot X_2) &\simeq (\text{EX}_2)^2 \cdot \text{Var } X_1 + (\text{EX}_1)^2 \cdot \text{Var } X_2 \\ &= 10^2 \cdot 0,015 + 15^2 \cdot 0,01 = 3,75 = \text{Var } Z_1,\end{aligned}$$

$$\text{E}(X_1 \cdot X_2) = \text{EX}_1 \cdot \text{EX}_2 = 15 \cdot 10 = 150 = \text{EZ}_1,$$

$$\begin{aligned}\text{Var}(X_1 \cdot X_3) &\simeq (\text{EX}_3)^2 \cdot \text{Var } X_1 + (\text{EX}_1)^2 \cdot \text{Var } X_3 \\ &= 20^2 \cdot 0,015 + 15^2 \cdot 0,02 = 10,5 = \text{Var } Z_2,\end{aligned}$$

$$\text{E}(X_1 \cdot X_3) = \text{EX}_1 \cdot \text{EX}_3 = 15 \cdot 20 = 300 = \text{EZ}_2,$$

$$\begin{aligned}\text{Cov}(Z_1, Z_2) &= \text{Cov}(X_1 \cdot X_2, X_1 \cdot X_3) = \text{EX}_2 \cdot \text{EX}_3 \cdot \text{Var } X_1 \\ &= 10 \cdot 20 \cdot 0,015 = 3,\end{aligned}$$

so daß sich

$$\begin{aligned}\text{Var}(f(Z_1, Z_2)) &= \text{Var} \frac{Z_1}{Z_2 - 2} \\ &= \frac{1}{(\text{EZ}_2 - 2)^2} \cdot \left[\text{Var } Z_1 + \left(\frac{\text{EZ}_1}{\text{EZ}_2 - 2} \right)^2 \cdot \text{Var } Z_2 \right. \\ &\quad \left. - 2 \cdot \frac{\text{EZ}_1}{\text{EZ}_2 - 2} \text{Cov}(Z_1, Z_2) \right] \\ &= \frac{1}{(300 - 2)^2} \cdot \left[3,75 + \left(\frac{150}{300 - 2} \right)^2 \cdot 10,5 - 2 \cdot \frac{150}{300 - 2} \cdot 3 \right] \\ &= 0,0000382\end{aligned}$$

ergibt. Damit ist die gesuchte Standardabweichung (mittlerer Fehler) gerade $\sigma = \sqrt{\text{Var}(f(Z_1, Z_2))} \simeq \sqrt{0,0000382} = 0,0061806$.

- (d) Ein minderwertiges Erz, das aus technischen und ökonomischen Gründen nicht alleine verhüttbar ist, muß, damit es verhüttbar wird, mit einem höherwertigen Erz, das auch alleine verhüttbar ist, gemischt werden. Es interessiert nun, welchen Prozentsatz des vorhandenen Metalls im minderwertigen Erz man mit einem bestimmten Verfahren erschmelzen kann, vgl. Hartung/Schmitz (1980). Bevor in einem Versuch dieser Prozentsatz bestimmt werden kann, muß man durch Analysen zunächst ermitteln, wie hoch die Metallanteile im minderwertigen und im höherwertigen Erz tatsächlich sind. Natürlich ist man gezwungen, diese Analysen auf Stichprobenbasis durchzuführen; vgl. hierzu auch Abschnitt 4 in Kap. XI und Elpelt/Hartung (1982).

Wegen der Analysegenauigkeit und der Ungenauigkeiten, die dadurch entstehen, daß man nur einen Teil des vorhandenen Erzes analysiert, hinterher aber eine Aussage über das gesamte Erz machen will, das als Naturprodukt sowieso schon gewissen Schwankungen unterliegt, kann der Schätzwert für die Metallgehalte in den Erzen nur innerhalb einer Streubreite s ermittelt werden.

Nehmen wir einmal an, daß die Analysen folgende Ergebnisse lieferten:

Metallgehalt beim höherwertigen Erz:

$$M(h)\% \pm s_{M(h)}\% = 50\% \pm 10\%;$$

Metallgehalt beim minderwertigen Erz:

$$M(m)\% \pm s_{M(m)}\% = 20\% \pm 3\%.$$

Nun wird in einer Grundqualitätsprüfung zunächst nur höherwertiges Erz verhüttet; dabei erfaßt man die verbrauchte Erzmenge und die Schlacke in Tonnen:

$$\text{Erzmenge: } E(h) t \pm s_{E(h)} t = 150 t \pm 10 t;$$

$$\text{Schlackemenge: } S(h) t \pm s_{S(h)} t = 120 t \pm 10 t.$$

Nun wird der Metallgehalt der Schlacke analysiert; es ergab sich

$$MS(h)\% \pm s_{MS(h)}\% = 2\% \pm 1\%.$$

Der Prozentsatz an erschmelzbarem Metall im höherwertigen Erz ist dann wegen

Erschmolzenes Metall aus höherwertigem Erz

= Metall in höherwertigem Erz – Metall in Schlacke

= Erzmenge × Metallgehalt/100 – Schlackemenge × Metallgehalt in Schlacke/100

gegeben durch

$$Q(h)\% = 100\% \frac{E(h) \cdot M(h) - S(h) \cdot MS(h)}{E(h) \cdot M(h)} = 100\% \left(1 - \frac{S(h) \cdot MS(h)}{E(h) \cdot M(h)} \right)$$

$$= 100\% \left(1 - \frac{120 \cdot 2}{150 \cdot 50} \right) = 96,8\%.$$

Wie groß ist aber der mittlere Fehler $s_{Q(h)}$?

Hier kommt das Fehlerfortpflanzungsgesetz zur Anwendung:

$$s_{Q(h)}\% \approx \sqrt{\frac{1}{E(h)^2 \cdot M(h)^2} \cdot \left(S(h)^2 \cdot s_{MS(h)}^2 + MS(h)^2 \cdot s_{S(h)}^2 \right.}$$

$$\quad \left. + \frac{S(h)^2 \cdot MS(h)^2}{E(h)^2 \cdot M(h)^2} \cdot (E(h)^2 \cdot s_{M(h)}^2 + M(h)^2 \cdot s_{E(h)}^2) \right)}^{1/2} \cdot 100\%$$

$$= \sqrt{\frac{1}{150^2 \cdot 50^2} \cdot \left(120^2 \cdot 1^2 + 2^2 \cdot 10^2 \right.}$$

$$\quad \left. + \frac{120^2 \cdot 2^2}{150^2 \cdot 50^2} \cdot (150^2 \cdot 10^2 + 50^2 \cdot 10^2) \right)}^{1/2} \cdot 100\%$$

$$= 1,8\%.$$

d.h. $Q(h)\% \pm s_{Q(h)}\% = 96,8\% \pm 1,8\%$, wenn man unabhängige Größen annimmt.

Nun tauschen wir einen Teil des höherwertigen Erzes gegen das zu untersu-

chende minderwertige Erz aus. Bei diesem Zulageversuch, vgl. auch **Abb. 8**, werden wiederum Erzmenge und Schlackemenge bestimmt:

Erzmenge Zulageversuch: $E(m+h) t \pm s_{E(m+h)} t = 200 t \pm 10 t$;

Schlackemenge Zulageversuch: $S(m+h) t \pm s_{S(m+h)} t = 180 t \pm 15 t$.

Der prozentuale Anteil des minderwertigen Erzes am Gesamterz im Zulageversuch (Mischungsverhältnis) sei

$$A(m)\% \pm s_{A(m)}\% = 60\% \pm 3\%$$

Nun wird wiederum der Metallgehalt (beim Zulageversuch)

$$MS(m+h)\% \pm s_{MS(m+h)}\% = 3\% \pm 0,1\%$$

bestimmt. Dann ist gerade

Erschmolzenes Metall aus minderwertigem Erz

= Insgesamt erschmolzenes Metall – erschmolzenes Metall aus höherwertigem Erz,

und der Prozentsatz $Q(m)$ erschmelzbaren Metalls vom Gesamtmetallanteil aus minderwertigem Erz ist gegeben durch

$$\begin{aligned} Q(m)\% &= 100\% \frac{\text{erschmolzenes Metall aus minderwertigem Erz}}{\text{Metallanteil im minderwertigem Erz}} \\ &= 100\% + \frac{M(h)(100 - A(m))(100 - Q(h))}{A(m) \cdot M(m)} \% \\ &\quad - \frac{10000 \cdot MS(m+h) \cdot S(m+h)}{E(m+h) \cdot A(m) \cdot M(m)} \% \\ &= 100\% + \frac{50(100 - 60)(100 - 96,8)}{60 \cdot 20} \% - \frac{10000 \cdot 3 \cdot 180}{200 \cdot 60 \cdot 20} \% \\ &= 100\% + 5,3\% - 22,5\% = 82,8\%. \end{aligned}$$

Der mittlere Fehler bei der Berechnung von $Q(m)$ muß wieder mittels Fehlerfortpflanzungsgesetz bestimmt werden. Dabei setzen wir die Unabhängigkeit aller Größen voraus.

$$\begin{aligned} s_{Q(m)}\% &= \left\{ \left[\frac{(100 - A(m)) \cdot (100 - Q(h))}{A(m) \cdot M(m)} \cdot s_{M(h)} \right]^2 \right. \\ &\quad + \left[\frac{M(h) \cdot (100 - A(m))}{A(m) \cdot M(m)} \cdot s_{Q(h)} \right]^2 \\ &\quad + \left[\frac{10000 \cdot MS(m+h)}{E(m+h) \cdot A(m) \cdot M(m)} \cdot s_{S(m+h)} \right]^2 \\ &\quad + \left[\frac{10000 \cdot S(m+h)}{E(m+h) \cdot A(m) \cdot M(m)} \cdot s_{MS(m+h)} \right]^2 \\ &\quad \left. + \left[\frac{10000 \cdot MS(m+h) \cdot S(m+h)}{E(m+h)^2 \cdot A(m) \cdot M(m)} \cdot s_{E(m+h)} \right]^2 \right\}^{1/2} \end{aligned}$$

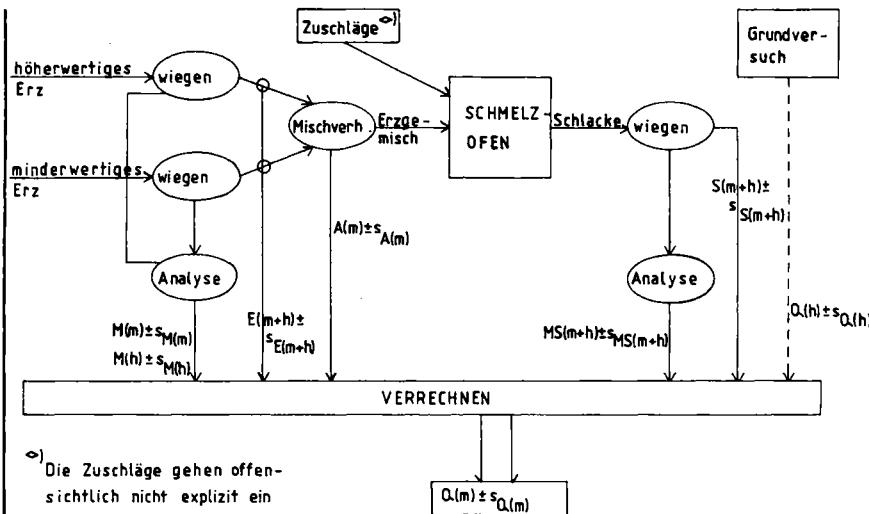


Abb. 8: Schematischer Ablauf des Zulageversuchs bei der Qualitätsbeurteilung eines nicht alleine verhüttbaren Erzes

$$\begin{aligned}
 & + \left[\frac{1}{A(m) \cdot M(m)^2} \cdot \left(\frac{10000 \cdot MS(m+h) \cdot S(m+h)}{E(m+h)} \right. \right. \\
 & \quad \left. \left. - M(h) \cdot (100 - A(m)) \cdot (100 - Q(h)) \right) s_{M(m)} \right]^2 \\
 & + \left[\frac{1}{A(m)^2 \cdot M(m)} \cdot \left(\frac{10000 \cdot MS(m+h) \cdot S(m+h)}{E(m+h)} \right. \right. \\
 & \quad \left. \left. - 100 \cdot M(h) \cdot (100 - Q(h)) \right) s_{A(m)} \right]^{1/2} \% \\
 & = \left\{ \left[\frac{(100 - 60) \cdot (100 - 96,8)}{60 \cdot 20} \cdot 10 \right]^2 + \left[\frac{50 \cdot (100 - 60)}{60 \cdot 20} \cdot 1,8 \right]^2 \right. \\
 & \quad \left. + \left[\frac{10000 \cdot 3}{200 \cdot 60 \cdot 20} \cdot 15 \right]^2 + \left[\frac{10000 \cdot 180}{200 \cdot 60 \cdot 20} \cdot 0,1 \right]^2 + \left[\frac{10000 \cdot 3 \cdot 180}{200^2 \cdot 60 \cdot 20} \cdot 10 \right]^2 \right. \\
 & \quad \left. + \left[\frac{1}{60 \cdot 20^2} \left(\frac{10000 \cdot 3 \cdot 180}{200} - 50 \cdot (100 - 60) \cdot (100 - 96,8) \right) \cdot 3 \right]^2 \right. \\
 & \quad \left. + \left[\frac{1}{60^2 \cdot 20} \left(\frac{10000 \cdot 3 \cdot 180}{200} - 100 \cdot 50 \cdot (100 - 96,8) \right) \cdot 3 \right]^{1/2} \% \right\}^{1/2} \% \\
 & = (1,06667^2 + 3^2 + 1,875^2 + 0,75^2 + 1,125^2 + 2,575^2 + 0,45833^2)^{1/2} \% \\
 & = \sqrt{22,322} \% = 4,7\%
 \end{aligned}$$

Die prozentuale Angabe des erschmelzbaren Metalls im minderwertigen Erz

müßte somit etwa

$$Q(m)\% \pm s_{Q(m)}\% = 82,8\% \pm 4,7\%$$

lauten.

- (e) Ganz ähnliche Problemstellungen wie bei der Bestimmung des prozentualen Anteils von erschmelzbarem Metall in nicht allein verhüttbarem Erz treten bei der **energetischen Futterwertprüfung** von Mischfutter (Kraftfutter) für Wiederkäuer auf. Die Mischfutter werden dabei in Verdauungsversuchen an Hammeln geprüft. Man ermittelt in einem Grundversuch zunächst die Verdaulichkeit der Inhaltsstoffe (z. B. Rohprotein, Rohfaser, organische Substanz, N-freie-Extraktstoffe) im Rauhfutter (Heu), tauscht dann in einem Zulageversuch einen Teil des Rauhfutters gegen Mischfutter aus und berechnet dann aus der Differenz der Verdaulichkeiten von Rauhfutter und Rauhfutter-Mischfutter die speziellen Verdaulichkeiten der Nährstoffe im Mischfutter, vgl. Abb. 9. Diese Verdaulichkeiten der einzelnen Bestandteile des Mischfutters werden benötigt, um nach der Stärkewertformel von Oskar Kellner den Stärkewert des Mischfutters bestimmen zu können; denn aufgrund des Stärkewertes erfolgt die Zuteilung des Mischfutters zu einer qualitativen Futterklasse. Eine ausführliche Darstellung der hier nur kurz angedeuteten Problematik findet man bei Hartung/Schmitz (1979). Eine energetische Wertprüfung von Nahrungsmitteln für Menschen wirft im Prinzip gleiche Probleme auf, ist in ihrer Durchführung aber erheblich komplizierter (Vermeidung von Woyzeck-Symptomen).

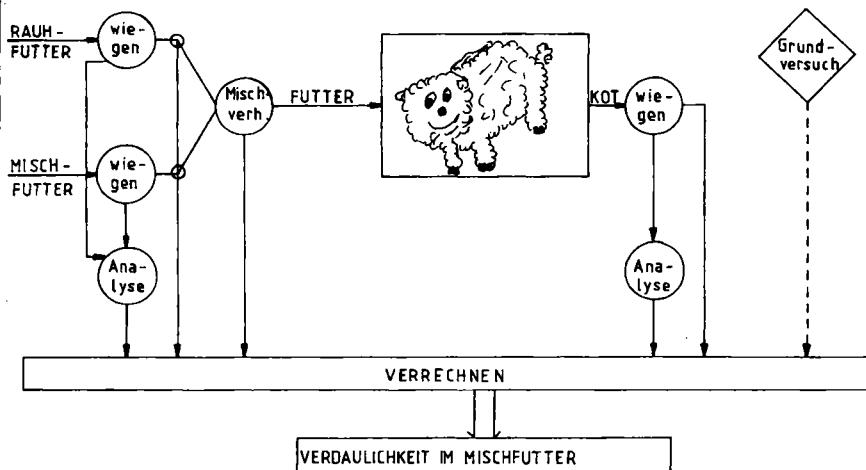


Abb. 9: Schematischer Ablauf des Zulageversuchs bei der Qualitätsbeurteilung von Mischfutter

4.3. Kontrolle und Erfassung von Meßfehlern

4.3.1. Kontrolle und Ermittlung systematischer Fehler

Am Anfang dieses Abschnitts 4 wurden als **systematische Fehler** diejenigen Fehler bezeichnet, die bei gleichen Meßvorgängen die gleiche Struktur haben. Diese Fehlerart entsteht durch Unvollkommenheit der Meßgeräte, der Meß- bzw. Analyse-

verfahren und der Meßgegenstände, aber auch durch Einflüsse der Umwelt, wie z. B. Temperatur, Luftdruck, Feuchte, Spannung und fremde elektrische oder magnetische Felder. Ebenso spielen die persönlichen Einflüsse der Beobachter (Aufmerksamkeit, Übung, Sehschärfe, Schätzungsverfahren) eine große Rolle. Es könnte sein, daß ein Beobachter mehr zum Auf-, ein anderer mehr zum Abrunden von Skalenwerten neigt. Falls die Größenordnung und das Vorzeichen eines systematischen Fehlers bekannt sind, so ist dies in der Auswertung zu berücksichtigen, und es sind die Meßergebnisse eventuell abzuändern.

Es gibt jedoch auch systematische Fehler, die nicht auf einfache Weise erfaßbar sind, die aber in manchen Fällen abgeschätzt und dann natürlich auch berücksichtigt werden können.

Wie könnte ein leicht erfaßbarer systematischer Fehler aussehen? Wenn man weiß, daß ein Meßgerät – aus rein mechanischen Gründen – im Intervall $[0; 1]$ jeweils einen um 2%, im Intervall $[1; 3]$ einen um 1% zu hohen Meßwert anzeigt, so sind die gemessenen Größen in diesen Intervallen um diese Größenordnungen zu vermindern. Oftmals ist diese Information jedoch nicht vorhanden, und es ist meist sehr schwierig, Größenordnung oder Vorzeichen des systematischen Fehlers bei einem Meßvorgang zu bestimmen.

Beispiel: In der Tabelle 8 dieses Kapitels sind 15 verschiedene Ergebnisse für die Messung der Astronomischen Einheit mit zugehörigen Streubereichen angegeben. Obwohl die Astronomische Einheit eine Naturgröße ist, sind diese Meßwerte sehr unterschiedlich; teilweise sind sogar die angegebenen Streubereiche disjunkt, obwohl gleiche Meßverfahren verwandt wurden. Hier müssen also systematische Fehler vorliegen, die die Größenordnung der Meßwerte beeinflußt haben. Eine Möglichkeit das Problem anzugehen, wäre es etwa, mit diesen unterschiedlichen Werten eine neue Größenbestimmung der Naturgröße „Astronomische Einheit“ vorzunehmen.

Etwas anders ist das Problem gelagert, falls man nun den systematischen Fehler eines Meßgerätes aus einem ganzen **Meßpark** bestimmen möchte. Man kann dann alle übrigen Meßgeräte des Meßparkes beibehalten und nur diesen einen Apparat, für den man sich interessiert, gegen ein äquivalentes Gerät austauschen. Natürlich stellt man sich nun die Frage, ob es notwendig ist, wenn man den systematischen Fehler eines gesamten Meßparkes und nicht nur eines einzelnen Apparates bestimmen will, alle möglichen Austauschkombinationen neu durchzumessen. Dies soll an einem Beispiel erläutert werden.

Beispiel: Um einen physikalischen Sachverhalt zu messen, benötigt man drei Meßapparate A, B und C. Um Informationen über den systematischen Fehler dieses Meßparkes zu erhalten, besitzt man von jedem Instrument zwei Ausgaben, also A_1 und A_2 , B_1 und B_2 , C_1 und C_2 . Ist es nun notwendig, von jedem der $2^3 = 8$ Kombinationsmöglichkeiten A_i, B_i, C_i ($i = 1, 2$) eine Meßreihe mit n Wiederholungsmessungen zu erstellen?

Diese Frage kann mit einem klaren „Nein“ beantwortet werden, denn durch eine geschickte Kombination der Geräte kommt man mit weniger Wiederholungen aus. Mit den Kombinationen

- (1) A_1, B_1, C_1
- (2) A_1, B_2, C_2
- (3) A_2, B_1, C_1
- (4) A_2, B_2, C_2

hat man für jeden Meßapparat die benötigte Anzahl von Wiederholungen, denn die Summe der beiden Kombinationen (1) und (2) beinhaltet den systematischen Fehler

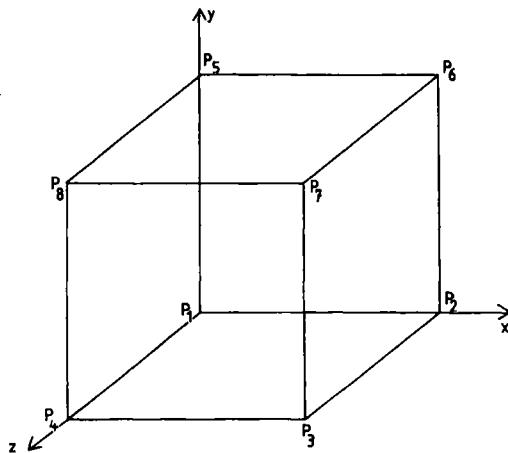


Abb. 10: Bildliche Darstellung des Meßparks im Beispiel

der beiden Apparate B und C, genauso wie die Summe von (3) und (4). Wünschte man sich zur Bestimmung des systematischen Fehlers für jeden Meßapparat 32 Wiederholungen, so benötigt man anstelle der insgesamt $8 \cdot 16 = 128$ dann nur $4 \cdot 16 = 64$ Wiederholungen. Bei einem noch größeren Meßpark und einem größeren Stichprobenumfang n ermöglicht es diese Methode – was schon aus rein wirtschaftlichen Gründen wünschenswert ist –, mit erheblich weniger Messungen auszukommen.

Den Austausch zweier Meßinstrumente kann man sich an der Abb. 10 verdeutlichen. Die Meßinstrumente A_i , B_i bzw. C_i für $i = 1, 2$ messen jeweils die Einflußgrößen X, Y bzw. Z. Im Punkt P_1 hat man die Geräte A_1 , B_1 , C_1 verwandt. Da die Geräte A_1 und A_2 die Größe X messen, gelangt man durch Austausch von A_1 durch A_2 zum Punkt P_2 . Im Punkt P_7 verwendet man die Geräte A_2 , B_2 , C_2 . Es entsprechen also die vorhin verwandten Kombinationen den Punkten P_1 , P_8 , P_6 , P_3 . Man möchte den wahren Wert der gesuchten Größe ermitteln, und der würde sich hier bildlich gesprochen im Mittelpunkt M des Würfels befinden. Da man keine Kombination bzw. keinen Eckpunkt bevorzugen kann, ergibt sich als vernünftige Schätzung für den wahren Wert das arithmetische Mittel der vier ausgewählten Punkte P_1 , P_8 , P_6 und P_3 , also

$$\begin{aligned} M &= \frac{P_1 + \frac{P_3 - P_1}{2} + P_6 + \frac{P_8 - P_6}{2}}{2} \\ &= \frac{2 \cdot P_1 + P_3 - P_1 + 2 \cdot P_6 + P_8 - P_6}{4} \\ &= \frac{P_1 + P_3 + P_6 + P_8}{4}. \end{aligned}$$

Hätte man dagegen konservativ die Punkte P_1 , P_2 , P_4 , P_5 ausgewählt, so hätte sich

$$\begin{aligned}
 M &= P_1 + \left(\frac{P_2 - P_1}{2} \right) + \left(\frac{P_5 - P_1}{2} \right) + \left(\frac{P_4 - P_1}{2} \right) \\
 &= \frac{P_2 + P_4 + P_5 - P_1}{2}
 \end{aligned}$$

ergeben. Je nach Auswahl der Punkte erhält man auch verschiedene Größen für die gemeinsame Varianz, denn einmal wird der Zähler durch vier, das andere Mal nur durch zwei dividiert. Der Preis einer konservativen Auswahl ist also, im Durchschnitt den Fehler zu verdoppeln.

Zusammenfassend kann man sagen: Durch Austausch des Meßparkes oder des Meßverfahrens, bzw. durch Wiederholung des Meßvorganges in anderen Laboratorien kann man sich über die Größenordnung des systematischen Fehlers bewußt werden, um ihn dann in geeigneter Art und Weise bei der Analyse des Versuches zu berücksichtigen.

Bei der **Kalibration von Meß- und Analyseverfahren (Eichen)** versucht man ein neues Verfahren so zu justieren bzw. zu skalieren, daß die mit ihm erzielten Meßergebnisse mit denen eines erprobten Standardverfahrens vergleichbar sind, d. h. man justiert so, daß die systematischen Fehler beider Meßverfahren gleich sind. Das Instrumentarium der Kalibration ist die **inverse Regressionsrechnung**; vgl. Scheffé (1973), Graybill (1976).

4.3.2. Die Verwendung von Kontrollkarten

Um Meß- und Analysevorgänge auch über längere Zeiten hinweg kontrollieren zu können, bedient man sich sogenannter **Kontrollkarten**, vgl. auch Kap. VI, auf denen man charakteristische Größen von Meßvorgängen einträgt. Dazu werden in regelmäßigen Abständen je n Messungen vorgenommen.

Je nach der gewünschten Information oder dem zu untersuchenden Sachverhalt unterscheidet man zwischen \bar{x} -, \bar{x} -s-, \bar{x} -R-, p- und np-Karten. Kann man für die zu untersuchenden Meßgrößen bestimmte Verteilungsannahmen treffen, so existieren auf einer Kontrollkarte eine obere und eine untere **Kontrollgrenze**, sowie eine obere

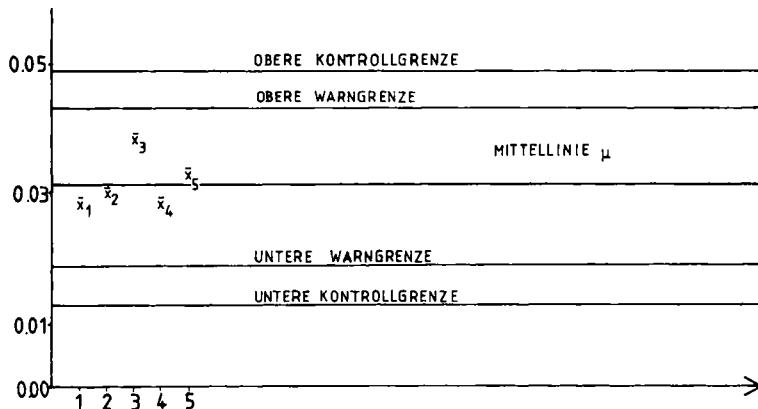


Abb. 11: Beispiel einer \bar{x} -Karte

und eine untere **Warngrenze**. Nimmt man für die Meßgrößen eine $N(\mu, \sigma^2)$ -Verteilung an, so wird man vor allem \bar{x} -, \bar{x} -s- oder \bar{x} -R-Kontrollkarten verwenden.

Auf der \bar{x} -Karte, vgl. Abb. 11, trägt man die Mittellinie μ , die Kontrollgrenzen

$$\mu + 3 \cdot \sigma / \sqrt{n} \quad \text{und} \quad \mu - 3 \cdot \sigma / \sqrt{n}$$

sowie die Warngrenzen

$$\mu + 2 \cdot \sigma / \sqrt{n} \quad \text{und} \quad \mu - 2 \cdot \sigma / \sqrt{n}$$

ein. Auf der \bar{x} -s-Karte werden neben diesen Grenzen auch Grenzen für die Standardabweichung der je n Messungen eingetragen: die Mittellinie σ , die Kontrollgrenzen

$$\sigma + 3 \cdot \sigma \cdot \sqrt{\gamma_n^2 - 1} \quad \text{und} \quad \sigma - 3 \cdot \sigma \cdot \sqrt{\gamma_n^2 - 1}.$$

sowie die Warngrenzen

$$\sigma + 2 \cdot \sigma \cdot \sqrt{\gamma_n^2 - 1} \quad \text{und} \quad \sigma - 2 \cdot \sigma \cdot \sqrt{\gamma_n^2 - 1}.$$

Dabei sind die Werte γ_n in Abschnitt 1.3 des Kap. IV vertafelt.

Die \bar{x} -R-Karte enthält neben den Grenzen für den Mittelwert noch Grenzen für die Spannweite R der Messungen.

Zu jedem Zeitpunkt i werden nun n Messungen vorgenommen und je nach Sachverhalt ihr Mittelwert \bar{x}_i , ihre Standardabweichung s_i und/oder ihre Spannweite R_i berechnet und in die jeweilige Kontrollkarte eingetragen, vgl. Abb. 11.

Besonders bei $B(n, p)$ -verteilten Meßwerten finden die p - und die np -Karte Verwendung. Hier werden dann Grenzen für p bzw. np eingetragen. Zu jedem Zeitpunkt i entnimmt man etwa n Produkte aus der laufenden Produktion, berechnet $\hat{p}_i = m_i/n$ und trägt \hat{p}_i in die p -Karte ein bzw. man trägt in die np -Karte direkt die Anzahl m_i der defekten Produkte unter den n ausgewählten ein, vgl. auch Kap. VI.

Kommt es nun häufig vor, daß Kontroll- und Warngrenze auf einer Kontrollkar-

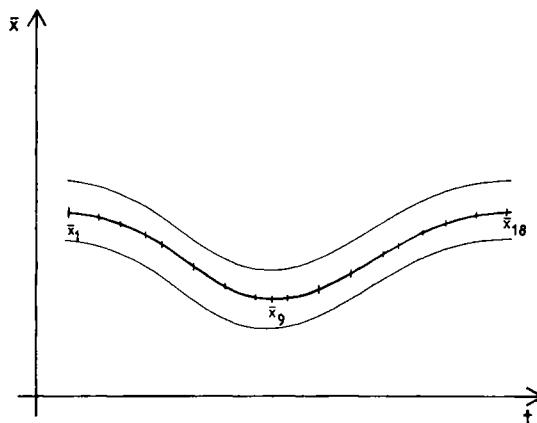


Abb. 12: \bar{x} -Karte für das Körpermengen von Ratten mit eingezeichneter Streubreite (konstant)

te von den beobachteten Werten überschritten werden, so ist zum einen die gemachte Verteilungsannahme, zum anderen das Meß- bzw. Analyseverfahren zu überprüfen. Neben dem häufigen Überschreiten der Kontrollgrenzen können jedoch noch andere Phänomene auftreten.

Beispiel: Mißt man z. B. in einem Labor das Körpergewicht von Ratten, so könnte es durch eine Verschlechterung der Nahrungsqualität vorübergehend zu einem Verlust an Körpergewicht führen, das sich jedoch nach Verbesserung der Nahrung wieder erhöht. Auf einer Kontrollkarte würde sich damit der optische Eindruck der Abb. 12 ergeben. Schon schwieriger ist es, falls nicht alle Ratten gleich auf die Verminderung der Nahrungsqualität reagieren. Es könnte sein, daß Ratten mit einem höheren Körpergewicht einen wesentlich höheren Gewichtsverlust erleiden als Ratten mit einem geringeren Körpergewicht. In diesem Fall würde nicht nur der Mittelwert sinken, sondern sich die Varianz auch noch verringern. Für den Beobachter ergibt sich damit das Bild aus Abb. 13 auf seiner Kontrollkarte.

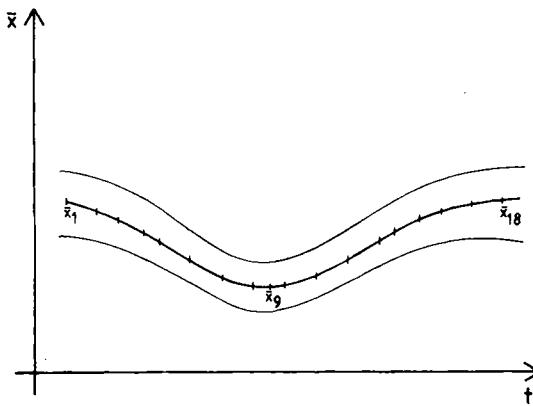


Abb. 13: \bar{x} -Karte für das Körpergewicht von Ratten mit eingezeichneter Streubreite (schwankend)

4.3.3. Ringversuche: Inter- und Intralaboratorielle Vergleiche

Eine große Rolle bei der Untersuchung systematischer und zufälliger Fehler bei Meß- und Analysemethoden spielen sogenannte **Ringversuche**. Läßt man gleichartige Proben etwa von zehn verschiedenen Laboratorien analysieren, so würden, falls keine Meßfehler auftreten, alle Laboratorien zu den gleichen Analyseergebnissen kommen. Sind nun die Ergebnisse in acht oder neun Laboratorien nahezu gleich und weichen sie in ein oder zwei Laboratorien stark ab, so wird man diese Abweichungen genauer untersuchen, denn offensichtlich liegen hier systematische Fehler in der Analyse vor. Einen solchen Ringversuch nennt man auch einen **intralaboratoriellen Vergleich**.

Ringversuche können aber nicht nur zum Vergleich zwischen (inter) verschiedenen Laboratorien sondern auch zum Vergleich innerhalb (intra) der Laboratorien (**intralaboratorielle Vergleiche**) dienen. In jedem Labor könnten die betreffenden Analysen noch von verschiedenen Mitarbeitern und von diesen an verschiedenen Tagen durchgeführt werden. Man spricht dann von einem **hierarchischen Aufbau** des Ringversuches. Die Analyse des zufälligen Fehlers erfolgt dann mittels Methoden der Varianzanalyse (Modelle II), vgl. Kap. XI und Hartung/Elpelt (1984).

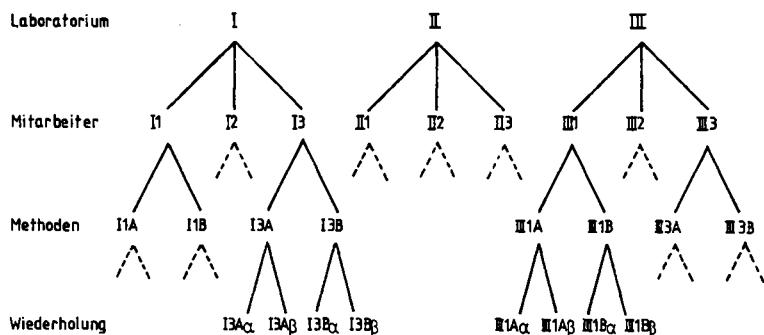


Abb. 14: Beispielhafter hierarchischer Aufbau eines Ringversuchs

Beispiel: In Abb. 14 ist ein solcher hierarchischer Aufbau eines Ringversuches grafisch dargestellt: Drei Laboratorien I, II, III untersuchen identische Proben. In jedem Labor wird die Analyse von je drei verschiedenen Mitarbeitern 1, 2, 3 nach zwei Analysemethoden A, B an jeweils zwei Tagen α , β durchgeführt.

Will man n Laboratorien miteinander vergleichen, so kann man dies auch mit Hilfe eines Streudiagramms tun.

Beispiel: In zehn Laboratorien werden jeweils zwei gleiche Wasserproben auf ihren Härtegrad hin untersucht. In jedem Labor ergibt sich also ein Paar von Analysergebnissen. Stellt man diese Paare nun als Punkte in einem Diagramm dar, so erhält man das Streudiagramm, vgl. Abb. 15. Im Labor 4 ergab sich zum Beispiel der Härtegrad der Probe 1 zu 7 und der der Probe 2 zu 5. Dieses Ergebnis wird im Streudiagramm durch den Punkt 4 mit den Koordinaten (7, 5) dargestellt. Laboratorien, deren Ergebnisse nicht im Zentralbereich des Streudiagramms liegen (hier

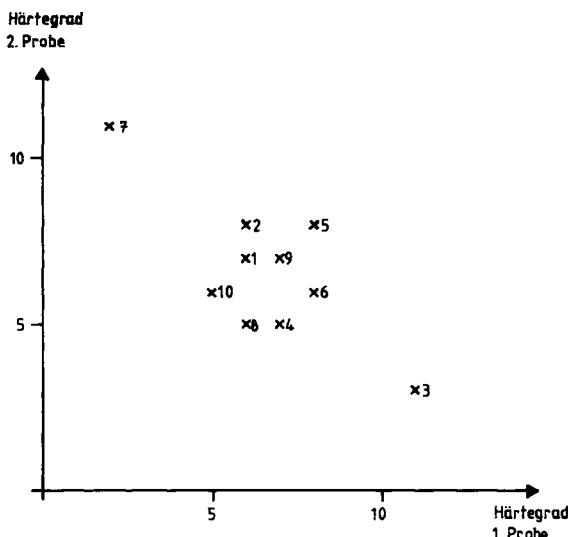


Abb. 15: Streudiagramm für die Analyse von Wasserproben in 10 Laboratorien

Tab. 9: Kritische Werte $R_{0,05}$ (I, J) und wahre Grenzen des Testniveaus ($\alpha_1 \leq \alpha \leq \alpha_2$) für Youden's extremen Rangsummentest

I	J	$R_{0,05}(I,J)$	α_1	α_2	I	J	$R_{0,05}(I,J)$	α_1	α_2
3	4	0	0,069	0,069	9	3	0	0,024	0,025
5	0	0,024	0,024		4	2	0,040	0,041	
6	1	0,053	0,058		5	4	0,037	0,038	
7	1	0,021	0,022		6	7	0,055	0,058	
8	2	0,038	0,041		7	9	0,041	0,043	
9	3	0,059	0,064		8	12	0,050	0,052	
10	3	0,026	0,028		9	15	0,056	0,059	
15	7	0,044	0,048		10	17	0,040	0,041	
					15	32	0,044	0,046	
4	4	0	0,030	0,031	10	3	1	0,076	0,080
5	1	0,045	0,047		4	3	0,067	0,070	
6	2	0,052	0,055		5	5	0,049	0,050	
7	3	0,055	0,059		6	8	0,057	0,060	
8	4	0,056	0,059		7	10	0,037	0,039	
9	5	0,055	0,058		8	13	0,039	0,040	
10	6	0,053	0,056		9	17	0,057	0,060	
15	11	0,037	0,039		10	20	0,054	0,056	
					15	36	0,044	0,045	
5	3	0	0,078	0,080	11	3	1	0,064	0,066
4	1	0,077	0,080		4	3	0,051	0,053	
5	2	0,064	0,067		5	6	0,060	0,063	
6	3	0,052	0,054		10	22	0,050	0,052	
7	4	0,040	0,042		15	40	0,044	0,045	
8	5	0,032	0,033		12	3	1	0,054	0,056
9	7	0,053	0,056		4	3	0,039	0,041	
10	8	0,040	0,042		5	6	0,043	0,045	
15	16	0,059	0,062		10	24	0,046	0,048	
					15	44	0,044	0,045	
6	3	0	0,055	0,056	13	3	1	0,046	0,047
4	1	0,045	0,046		4	4	0,061	0,064	
5	2	0,032	0,032		5	7	0,053	0,056	
6	4	0,052	0,054		10	26	0,044	0,046	
7	5	0,033	0,034		15	48	0,044	0,045	
8	7	0,044	0,046		14	3	1	0,040	0,041
9	9	0,053	0,056		4	4	0,049	0,051	
10	11	0,061	0,064		5	7	0,040	0,041	
15	20	0,053	0,055		10	28	0,042	0,044	
					15	52	0,044	0,046	
7	3	0	0,040	0,041	15	3	1	0,035	0,036
4	1	0,029	0,029		4	4	0,040	0,042	
5	3	0,045	0,047		5	8	0,049	0,051	
6	5	0,052	0,055		10	31	0,054	0,056	
7	7	0,056	0,058		15	56	0,044	0,046	
8	3	0	0,031	0,031					
4	2	0,056	0,059						
5	4	0,058	0,062						
6	6	0,053	0,056						
7	8	0,047	0,049						
8	10	0,040	0,041						
9	13	0,055	0,057						
10	15	0,044	0,046						
15	28	0,046	0,048						

| Labor 3 und Labor 7), arbeiten offenbar mit einem systematischen Fehler im Vergleich zu den restlichen.

Ein Ringversuch kann nun nicht nur dazu dienen, verschiedene Methoden oder Laboratorien miteinander zu vergleichen oder zu untersuchen, sondern bietet vielmehr auch die Möglichkeit, weitere Aussagen über die zu untersuchenden Objekte bzw. Proben zu machen.

Werden etwa I Objekte nach je J verschiedenen Verfahren oder mit je J verschiedenen Meßinstrumenten untersucht, so läßt sich mit Hilfe von **Youden's extremen Rangsummentest** überprüfen, ob alle I Objekte bzgl. der interessierenden Eigenschaft gleich sind oder ob sie sich unterscheiden. Dazu werden zunächst den Meßergebnissen an jedem der J Meßgeräte getrennt Rangzahlen zugeordnet. Das kleinste Ergebnis erhält den Rang 1, das größte den Rang I. Bei gleichen Meßergebnissen an einem Gerät (Bindungen) wird zufällig entschieden (zum Beispiel durch Münz- oder Würfelwurf), welche Rangzahl welchem Meßergebnis zugeordnet wird. Für das i-te Objekt ($i = 1, \dots, I$) bildet man sodann die Rangsummen r_i über alle Meßgeräte und prüft mittels Youden's extremen Rangsummentest die Hypothese

H_0 : Für alle J Geräte ist jede der $I!$ Permutationen der Ränge der Objekte gleichwahrscheinlich,

die im Grunde besagt, daß alle Objekte bzgl. der interessierenden Eigenschaft gleich sind, wobei systematische Fehler der Meßgeräte nicht eingehen. Die Nullhypothese H_0 wird zum Niveau α verworfen, falls für mindestens eine der I Rangsummen r_i gilt:

$$r_i \leq J + R_\alpha(I, J) \quad \text{oder} \quad r_i \geq I \cdot J - R_\alpha(I, J).$$

Für $\alpha \approx 0,05$ und einige Kombinationen I, J sind die kritischen Werte $R_{0,05}(I, J)$ in Tab. 9 angegeben. (Bei den angegebenen kritischen Werten wird das Niveau 0,05 nicht genau eingehalten, das wahre Niveau liegt zwischen den jeweils mit angegebenen Grenzen α_1 und α_2 .) Weitere kritische Werte und Grenzen findet man bei Thompson/Willke (1963).

Beispiel: Mit $J = 4$ verschiedenen Thermometern wurde jeweils die Temperatur von $I = 4$ Objekten gemessen. Die sich ergebenden Meßwerte, vgl. Youden (1960), sind in Tab. 10 angegeben, wo auch schon die zugehörigen Rangzahlen und Rangsummen r_i eingetragen wurden (bei der Bindung am 1. Thermometer wurden die Rangzahlen per Münzwurf vergeben).

Wir wollen die Hypothese

H_0 : Für jedes der vier Thermometer ist jede der $4! = 24$ Permutationen der Rangzahlen gleichwahrscheinlich, d.h. die Temperaturen der Objekte sind gleich,

Tab. 10: Temperaturen x_{ij} von vier Objekten gemessen mit vier Thermometern sowie Rangzahlen (in Klammern) und Rangsummen r_i für den extremen Rangsummentest von Youden

Objekt i		Thermometer j			Rangsummen r_i
	1	2	3	4	
1	36 (4)	38 (2)	36 (2)	30 (2)	10
2	17 (1)	18 (1)	26 (1)	17 (1)	4
3	30 (3)	39 (3)	41 (4)	34 (4)	14
4	30 (2)	45 (4)	38 (3)	33 (3)	12

testen. Mit $I = 4$ und $J = 4$ ergibt sich aus Tab. 9 der Wert

$$R_{0,05}(4, 4) = 0$$

(das wahre Testniveau liegt zwischen $\alpha_1 = 0,030$ und $\alpha_2 = 0,031$). Somit ist

$$J + R_{0,05}(I, J) = 4 + R_{0,05}(4, 4) = 4 + 0 = 4 \quad \text{und}$$

$$I \cdot J - R_{0,05}(I, J) = 4 \cdot 4 - R_{0,05}(4, 4) = 16 - 0 = 16$$

und die Hypothese H_0 wird verworfen, denn

$$r_2 = 4 \leq J + R_{0,05}(I, J) = 4.$$

Die Temperatur des zweiten Objekts hebt sich also signifikant von denen der übrigen Objekte ab.

4.3.4. Präzision, Spezifität, Richtigkeit und Sensibilität von Meß- und Analyseverfahren

In diesem Abschnitt sollen die Begriffe Präzision, Spezifität, Richtigkeit und Sensitivität eines Meß- bzw. Analyseverfahrens vorgestellt werden.

Verwendet man bei Messungen oder Analysen ein bestimmtes Verfahren, so ist man natürlich daran interessiert zu wissen, wie gut, wie präzise dieses ist. Die **Präzision** (**precision**) kann durch den zufälligen Meß- oder Analysefehler beschrieben werden.

Eine hohe Präzision erreicht man natürlich dann, wenn der zufällige Fehler, d. h. die Varianz bzw. die Standardabweichung des Verfahrens als Maß für den zufälligen Fehler, sehr klein ist. Daher wird als Maß für die Präzision oft eine inverse Funktion der Standardabweichung – etwa $1/\sigma$ – verwandt. Je größer der Wert dieser Funktion ist, desto präziser ist das Meßverfahren; je kleiner er ist, desto unpräziser ist das Verfahren.

Der Mangel an Präzision, die **Unpräzision**, wird oft durch die Varianz σ^2 des Meß- oder Analyseverfahrens beschrieben. Wie kann man nun die Unpräzision, d. h. die Varianz σ^2 , eines Meßverfahrens im konkreten Falle schätzen? Hierzu eignen sich besonders die Modelle II der Varianzanalyse, vgl. Kap. XI, Abschnitt 4, die direkt verwendbar sind, wenn Wiederholungsmessungen an einem Objekt möglich sind. Ist dies nicht der Fall, sei es aus Kostengründen, aus Nicht-Reproduzierbarkeitsgründen oder weil es sich um zerstörende Prüfungen handelt, so ist die Präzision eines Meßverfahrens nur im Vergleich zu der eines anderen Verfahrens bestimmbar; vgl. etwa Grubbs (1973), Klösener (1983), Hartung/Heine (1984), Hartung/Elpelt (1984), Hartung/Klösener (1985).

Natürlich wäre es auch sinnvoll, Präzision bzw. Unpräzision eines Meß- oder Analyseverfahrens in Abhängigkeit von den zu messenden Größen anzugeben: bei großen Beobachtungswerten fällt eine Standardabweichung nicht so sehr ins Gewicht, wie es dieselbe Standardabweichung bei kleinen Beobachtungswerten tut. Bei kleinen Werten wird man also von einem Meßverfahren eine größere Präzision verlangen.

Neben der Präzision interessiert man sich oft auch für die **Spezifität** (**specificity**) eines Meß- oder Analyseverfahrens. Interessiert man sich z. B. bei der Analyse einer chemischen Substanz für die Anteile der einzelnen Inhaltsstoffe, so soll die Spezifität des Verfahrens ein Maß für die Güte der Bestimmbarkeit dieser Anteile sein. Die Spezifität eines Verfahrens ist dann gut, wenn bei Änderung des Gehalts eines

Inhaltsstoffes sich auch nur dieses Analyseergebnis ändert und die anderen Ergebnisse gleich bleiben.

Spezifität und Präzision sind Maße, die sich am zufälligen Fehler eines Verfahrens orientieren. Bei der Betrachtung der **Richtigkeit** oder **Genauigkeit (accuracy)** eines Meß- oder Analyseverfahrens liegt das schwerpunktmäßige Interesse dagegen in der Ermittlung des systematischen Fehlers. Während die Präzision also in direktem Zusammenhang zur Streuung von Meßergebnissen steht, wird die Richtigkeit eines Verfahrens danach beurteilt, wie weit der wahre Wert vom Mittelwert aller Meßergebnisse entfernt ist. Der „wahre“ Wert kann dabei auch

- empfohlen bzw. gebräuchlich sein, wie dies etwa bei Naturkonstanten der Fall ist,
- vorgegeben sein oder
- als ein hypothetischer Wert festgelegt werden.

Zur Ermittlung der Richtigkeit eines Meß- oder Analyseverfahrens werden im wesentlichen folgende Vorgehensweisen verwandt:

- Vergleichsversuche durch Parallelbestimmungen mit einer anderen Analysemethode,
- Zusatzversuche (bei der Analyse von Substanzen etwa durch Hinzufügen bekannter Mengen zur Ursprungssubstanz),
- Mischversuche (z. B. Mischen einer zu untersuchenden Substanz in verschiedenen Konzentrationen oder Mengen),
- Ringversuche, vgl. Abschnitt 4.3.3.

Die **Totale Richtigkeit** impliziert nun lediglich, daß der systematische Fehler – nicht aber der Gesamtfehler als Summe von zufälligem und systematischem Fehler – den Wert Null besitzt, d. h. der systematische Fehler (Verzerrung bzw. Bias der Verteilung der Beobachtungsdaten) charakterisiert die **Accuracy** einer Verteilung der Meßwerte in der interessierenden Grundgesamtheit.

Die Präzision eines Meß- bzw. Analyseverfahrens orientiert sich also lediglich am zufälligen, die Richtigkeit lediglich am systematischen Meßfehler. Wie sind Meßverfahren insgesamt nun zu bewerten?

Haben zwei Meßverfahren die gleiche Richtigkeit, so wird man natürlich dasjenige bevorzugen, welches die größere Präzision hat und umgekehrt ist von zwei Meßverfahren gleicher Präzision sicherlich stets dasjenige zu bevorzugen, das die größere Accuracy aufweist.

Schwieriger ist die Situation zu beurteilen, wenn ein Verfahren eine hohe Präzision und eine geringe Richtigkeit, das andere eine geringe Präzision, aber eine hohe Richtigkeit hat. Hier kann man keine generellen Regeln für die Wahl des einen oder des anderen Verfahrens aufstellen; vielmehr muß bei der Entscheidung die konkrete Situation berücksichtigt werden. Dabei ist zu beachten: Ein unpräzises Meßverfahren ist meist sehr schwer zu präzisieren, bei einem Verfahren mit geringer Accuracy dagegen besteht häufig die Möglichkeit, es zu justieren. In solchen Fällen sollte dann der Präzision mehr Gewicht zukommen als der Accuracy, d. h. man sollte sich für das Verfahren mit der höheren Präzision entscheiden.

Ein weiteres Maß für die Güte eines Meß- oder Analyseverfahrens ist seine **Empfindlichkeit, Sensitivität (sensitivity)**, unter der man in der Regel den kleinsten als

signifikant von Null verschieden nachweisbaren Meßwert (erfaßbaren Mindestgehalt bei einer Analyse) versteht.

Um diese Nachweisgrenze zahlenmäßig erfassen zu können, ist es notwendig, zunächst eine hinreichend große Zahl von **Blindversuchen** durchzuführen. Blindversuche müssen dabei stets unter den gleichen Bedingungen stattfinden, wie die späteren „wirklichen“ Analysen, vgl. auch Abschnitt 5; lediglich den interessierenden Inhaltsstoff sollen die bei Blindversuchen verwandten Proben nicht enthalten. Ansonsten sollte ihre Zusammensetzung derjenigen der wirklich zu analysierenden Proben entsprechen, insbesondere was die Verunreinigung der Proben mit Fremdstoffen anbelangt.

Man bestimmt dann den mittleren festgestellten Gehalt \bar{x}_{BL} des interessierenden Stoffes (der eigentlich Null sein sollte) in den Proben der Blindversuche sowie die Standardabweichung s_{BL} dieser Gehaltswerte und berechnet hieraus als Maß für die Empfindlichkeit des Verfahrens die **Kaisersche Nachweisgrenze**

$$\underline{x} = \bar{x}_{BL} + 3s_{BL}.$$

Die Kaisersche Nachweisgrenze \underline{x} wird dann (bei Normalverteilungsannahme und nicht zu kleinem Stichprobenumfang) als das kleinste zum 99% Niveau signifikant verschieden von Null zu betrachtende Analyseergebnis angesehen, vorausgesetzt natürlich, daß es sich hierbei überhaupt um ein „vernünftiges“ Analyseverfahren zum Nachweis des interessierenden Inhaltsstoffes handelt.

4.4. Das Ausreißer-Problem

Beispiel: Ein Körper unterliegt einer gleichförmigen Bewegung; es wurden 12 Messungen x_1, \dots, x_{12} seiner Geschwindigkeit durchgeführt, wobei sich die Meßergebnisse aus Tab. 11 ergaben.

Tab. 11: Geschwindigkeit eines Körpers bei 12 Wiederholungsmessungen

Messung i	1	2	3	4	5	6	7	8	9	10	11	12
Geschwindigkeit x_i (in m/sec)	41	39	42	40	44	40	46	39	37	41	36	41

Sieht man sich die Meßergebnisse genauer an, so gewinnt man den Eindruck, daß der 7. und 11. Meßwert, die stark von den übrigen Werten abweichen, vielleicht gar nicht zur präzisierten Stichprobe gehören. Es besteht sogar die Gefahr, daß bei Auswertung der Meßreihe diese ausreißer-verdächtigen Einzelwerte das Ergebnis verfälschen würden.

Ist das starke Abweichen einzelner Beobachtungswerte von den restlichen Daten nicht durch z. B. Meß-, Rechen-, Schreib-, Datenerfassungsfehler zu erklären, d. h. sind systematische Fehler ausgeschlossen, so kann man mittels sogenannter **Ausreißer-Tests** überprüfen, ob diese Daten überhaupt zur präzisierten Stichprobe zu rechnen sind. Sind sie dies signifikant nicht, so wird man sie bei einer Auswertung der Daten gar nicht berücksichtigen, um so eine Verfälschung der Ergebnisse zu vermeiden.

Im folgenden werden verschiedene Ausreißertests vorgestellt und an obigem Beispiel verdeutlicht. Alle diese Tests verlangen, daß zunächst die interessierende Meßreihe der Größe nach geordnet wird, vgl. Tab. 12.

Tab. 12: Geordnete Meßreihe zu den Daten aus Tab. 11

Nummer der Messung	11	9	2	8	4	6	1	10	12	3	5	7
Bezeichnung des Meßwertes $x_{(i)}$	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$	$x_{(11)}$	$x_{(12)}$
Meßwert (Geschwin- digkeit in m/sec)	36	37	39	39	40	40	41	41	41	42	44	46

A. Der David-Hartley-Pearson-Test

Kann man voraussetzen, daß die n Meßwerte einer Stichprobe Realisationen aus einer Normalverteilung sind, so kann mit Hilfe des **David-Hartley-Pearson-Tests** die Hypothese

H_0 : Der kleinste (oder der größte) Meßwert gehört zur Stichprobe getestet werden. Die Hypothese H_0 wird zum Niveau α verworfen, falls gilt

$$Q = \frac{R}{s} = \frac{x_{(n)} - x_{(1)}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{(i)} - \bar{x})^2}} > Q_{n; 1-\alpha};$$

einige Quantile $Q_{n,y}$ sind in **Tab. 13** angegeben, weitere findet man bei David et al. (1954).

Tab. 13: Quantile $Q_{n,y}$ des David-Hartley-Person-Tests

n	$Q_{n, 0,95}$	$Q_{n, 0,99}$
3	2,00	2,00
4	2,43	2,45
5	2,75	2,80
6	3,01	3,10
7	3,22	3,34
8	3,40	3,54
9	3,55	3,72
10	3,69	3,88
12	3,91	4,13
15	4,17	4,43
20	4,49	4,79
30	4,89	5,25
40	5,15	5,54
50	5,35	5,77
100	5,90	6,36

Wird die Hypothese H_0 verworfen, so wird der kleinste oder der größte Wert in der Meßreihe als Ausreißer gewertet, je nach dem, welcher den größten Abstand vom Mittelwert

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{(i)}$$

hat; sind beide gleich weit vom Mittelwert entfernt, so werden beide Meßwerte ($x_{(1)}$ und $x_{(n)}$) eliminiert.

Beispiel: Aus den Daten der Tab. 12 ergibt sich

$$R = x_{(12)} - x_{(1)} = 46 - 36 = 10 \quad \text{und} \quad s = \sqrt{\frac{1}{11} \sum_{i=1}^{12} (x_{(i)} - \bar{x})^2} = 2,74,$$

so daß sich zum 1% Niveau weder der kleinste noch der größte Wert der Meßreihe als Ausreißer eliminieren läßt, denn

$$Q = \frac{R}{s} = \frac{10}{2,74} = 3,65 < 4,13 = Q_{12; 0,99}.$$

B. Der Grubbs-Test

Auch der **Ausreißertest von Grubbs** basiert auf einer Normalverteilungsannahme für die Grundgesamtheit, aus der eine vorliegende Meßreihe $x_{(1)}, \dots, x_{(n)}$ zu gewinnen war.

Die Hypothese

$$H_{01}: x_{(1)} \text{ ist kein Ausreißer}$$

wird zum Niveau α verworfen, falls

$$T_1 = \frac{\bar{x} - x_{(1)}}{s} > T_{n; 1-\alpha}$$

und die Hypothese

$$H_{0n}: x_{(n)} \text{ ist kein Ausreißer}$$

wird zum Niveau α verworfen, falls

$$T_n = \frac{x_{(n)} - \bar{x}}{s} > T_{n; 1-\alpha}$$

Tab. 14: Kritische Werte $T_{n, \gamma}$ des Grubbs-Tests

n	$T_{n, 0,95}$	$T_{n, 0,99}$
3	1,15	1,16
4	1,46	1,49
5	1,67	1,75
6	1,82	1,94
7	1,94	2,10
8	2,03	2,22
9	2,11	2,32
10	2,18	2,41
12	2,29	2,55
15	2,41	2,71
20	2,56	2,88
30	2,75	3,10
40	2,87	3,24
50	2,96	3,34
100	3,21	3,60

gilt. In Tab. 14 sind einige kritische Werte $T_{n,y}$ angegeben; weitere findet man bei Grubbs/Beck (1972).

Beispiel: Zum 1% Niveau wollen wir die Hypothesen H_{01} und H_{0n} für die Meßreihe aus Tab. 12 testen. Mit $s = 2,74$ und $\bar{x} = 40,5$ ergibt sich

$$T_1 = \frac{\bar{x} - x_{(1)}}{s} = \frac{40,5 - 36}{2,74} = 1,64 < 2,55 = T_{12; 0,99} \quad \text{und}$$

$$T_n = \frac{x_{(12)} - \bar{x}}{s} = \frac{46 - 40,5}{2,74} = 2,01 < 2,55 = T_{12; 0,99}.$$

Beide Hypothesen können zum 1% Niveau also nicht verworfen werden, d. h. weder der kleinste noch der größte Meßwert sind zum 1% Niveau signifikante Ausreißer.

C. Dixon's r-Statistiken

Die **Ausreißertests von Dixon** überprüfen, ob ein einzelner Meßwert – nämlich $x_{(n)}$ bzw. $x_{(1)}$ – unter Normalverteilungsannahme ein Ausreißer ist. In die Teststatistiken gehen neben $x_{(1)}$ bzw. $x_{(n)}$ zwei weitere Meßwerte $x_{(i)}$ ein: Für die Hypothese H_{01} , daß $x_{(1)}$ ein Ausreißer ist, der größte und der zweitkleinste Meßwert und für die Hypothese H_{0n} , daß $x_{(n)}$ ein Ausreißer ist, der kleinste und der zweitgrößte Meßwert. Dabei werden bei einigen Teststatistiken einzelne Meßwerte der Meßreihe vermieden, vgl. Tab. 15. Dadurch kann dann erreicht werden, daß nur ein relativ zentraler Bereich der Meßwerte in die Berechnung der Teststatistik eingeht. Betrachten wir etwa die Teststatistik r_{12}^1 , so wird hier durch das Vermeiden von $x_{(n)}$ und $x_{(n-1)}$ erreicht, daß es sich hier nicht auswirkt, wenn $x_{(n)}$ und $x_{(n-1)}$ sehr weit vom Mittelwert entfernt sind. Die Hypothesen H_{01} und H_{0n} werden zum Niveau α verworfen, falls die benutzte Teststatistik r_{ij}^1 bzw. r_{ij}^n ($i = 1, 2; j = 0, 1, 2$) den kritischen

Tab. 15: r-Statistiken von Dixon

in der Meßreihe werden die Werte ... vermieden	r-Statistik für die Hypothese	
	$H_{01}: x_{(1)}$ ist kein Ausreißer	$H_{0n}: x_{(n)}$ ist kein Ausreißer
—	$r_{10}^1 = \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}$	$r_{10}^n = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$
$x_{(n)}$ bzw. $x_{(1)}$	$r_{11}^1 = \frac{x_{(2)} - x_{(1)}}{x_{(n-1)} - x_{(1)}}$	$r_{11}^n = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(2)}}$
$x_{(n)}, x_{(n-1)}$ bzw. $x_{(1)}, x_{(2)}$	$r_{12}^1 = \frac{x_{(2)} - x_{(1)}}{x_{(n-2)} - x_{(1)}}$	$r_{12}^n = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(3)}}$
$x_{(2)}$ bzw. $x_{(n-1)}$	$r_{20}^1 = \frac{x_{(3)} - x_{(1)}}{x_{(n)} - x_{(1)}}$	$r_{20}^n = \frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(1)}}$
$x_{(2)}, x_{(n)}$ bzw. $x_{(n-1)}, x_{(1)}$	$r_{21}^1 = \frac{x_{(3)} - x_{(1)}}{x_{(n-1)} - x_{(1)}}$	$r_{21}^n = \frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(2)}}$
$x_{(2)}, x_{(n)}, x_{(n-1)}$ bzw. $x_{(n-1)}, x_{(1)}, x_{(2)}$	$r_{22}^1 = \frac{x_{(3)} - x_{(1)}}{x_{(n-2)} - x_{(1)}}$	$r_{22}^n = \frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(3)}}$

Tab. 16: Kritische Werte der Dixon-Tests zum 1% Niveau

n	$r_{10,n;0,99}$	$r_{11,n;0,99}$	$r_{12,n;0,99}$	$r_{20,n;0,99}$	$r_{21,n;0,99}$	$r_{22,n;0,99}$
6	0,70	0,81	0,93	0,84	0,95	1,00
7	0,64	0,74	0,84	0,78	0,89	0,95
8	0,59	0,68	0,76	0,71	0,83	0,89
9	0,56	0,64	0,70	0,67	0,78	0,84
10	0,53	0,60	0,66	0,63	0,73	0,79
11	0,50	0,57	0,62	0,60	0,68	0,75
12	0,48	0,54	0,59	0,58	0,64	0,70
13	0,47	0,52	0,55	0,56	0,62	0,67
14	0,45	0,50	0,54	0,54	0,59	0,64
15	0,44	0,49	0,52	0,52	0,57	0,62
20	0,39	0,43	0,46	0,46	0,51	0,54
25	0,36	0,39	0,42	0,43	0,46	0,49
30	0,34	0,37	0,39	0,40	0,43	0,46

Wert $r_{ij,n;1-\alpha}$ übersteigt. Einige kritische Werte $r_{ij,n;\gamma}$ für $\gamma = 0,99$ sind in **Tab. 16** angegeben, weitere findet man bei Dixon (1951).

Beispiel: Zur Illustration wollen wir mit sämtlichen r-Statistiken die Hypothese

$H_{0n}: x_{(12)}$ ist kein Ausreißer

(für die Meßreihe aus Tab. 12) zum Niveau $\alpha = 0,01$ überprüfen. Es ist

$$r_{10}^n = \frac{46 - 44}{46 - 36} = 0,200 < 0,48 = r_{10,12;0,99},$$

$$r_{11}^n = \frac{46 - 44}{46 - 37} = 0,222 < 0,54 = r_{11,12;0,99},$$

$$r_{12}^n = \frac{46 - 44}{46 - 39} = 0,285 < 0,59 = r_{12,12;0,99},$$

$$r_{20}^n = \frac{46 - 42}{46 - 36} = 0,400 < 0,58 = r_{20,12;0,99},$$

$$r_{21}^n = \frac{46 - 42}{46 - 37} = 0,444 < 0,64 = r_{21,12;0,99} \quad \text{und}$$

$$r_{22}^n = \frac{46 - 42}{46 - 39} = 0,571 < 0,70 = r_{22,12;0,99},$$

so daß sich die Hypothese H_{0n} mit Hilfe keiner der r-Statistiken widerlegen läßt. Auch die Dixon-Tests erlauben also nicht die Annahme, daß $x_{(12)}$ signifikant zum 1% Niveau ein Ausreißer in der Meßreihe unseres Beispiels ist.

D. Test auf ein Ausreißerpaar

Bislang haben wir nur Tests kennengelernt, die überprüfen, ob der kleinste bzw. der größte Wert einer Meßreihe ein Ausreißer ist. Grubbs/Beck (1972) geben nun Tests an zum Prüfen der Hypothese (unter Normalverteilungsannahme)

H_{01} : $x_{(1)}$ und $x_{(2)}$ sind keine Ausreißer und
 H_{0n} : $x_{(n)}$ und $x_{(n-1)}$ sind keine Ausreißer.

Kann man die Hypothese H_{01} zum Niveau α verwerfen, so sind $x_{(1)}$ und $x_{(2)}$ (die beiden kleinsten Werte der Meßreihe) zu diesem Niveau ein Ausreißerpaar, kann man H_{0n} zum Niveau α verwerfen, so bilden $x_{(n)}$ und $x_{(n-1)}$ (die beiden größten Werte in der Meßreihe) ein Ausreißerpaar. Mit

$$S_0^2 = \sum_{i=1}^n (x_{(i)} - \bar{x})^2, \quad S_{n-1,n}^2 = \sum_{i=1}^{n-2} (x_{(i)} - \bar{x}_{n,n-1})^2,$$

$$S_{1,2}^2 = \sum_{i=3}^n (x_{(i)} - \bar{x}_{1,2})^2, \quad \text{wobei } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_{(i)},$$

$$\bar{x}_{n,n-1} = \frac{1}{n-2} \sum_{i=1}^{n-2} x_{(i)} \quad \text{und} \quad \bar{x}_{1,2} = \frac{1}{n-2} \sum_{i=3}^n x_{(i)}$$

ist, wird die Hypothese H_{01} zum Niveau α verworfen, falls

$$S_{1,2}^2 / S_0^2 < s_{n;\alpha}$$

gilt, und die Hypothese H_{0n} zum Niveau α verworfen, falls gilt:

$$S_{n,n-1}^2 / S_0^2 < s_{n;\alpha}.$$

In Tab. 17 sind einige kritische Werte $s_{n;\alpha}$ angegeben; weitere finden sich bei Grubbs/Beck (1972).

Tab. 17: Kritische Werte $s_{n;\alpha}$ für die Tests auf ein Ausreißerpaar

n	$s_{n,0.05}$	$s_{n,0.01}$
5	0,018	0,004
6	0,056	0,019
7	0,102	0,044
8	0,148	0,075
9	0,191	0,108
10	0,231	0,141
12	0,300	0,204
15	0,382	0,286
20	0,480	0,391
30	0,601	0,527
40	0,672	0,610
50	0,720	0,667
100	0,833	0,802

Beispiel: Für die Meßreihe aus Tab. 12 wollen wir zum 1% Niveau sowohl die Hypothese H_{01} als auch die Hypothese H_{0n} überprüfen. Dazu müssen zunächst einige Hilfsgrößen berechnet werden. Mit

$$\bar{x} = 40,5, \quad \bar{x}_{1,2} = 41,3 \quad \text{und} \quad \bar{x}_{12,11} = 39,6$$

ergibt sich

$$S_0^2 = 83, \quad S_{1,2}^2 = 44,1 \quad \text{sowie} \quad S_{12,11}^2 = 32,4.$$

Somit ist

$$S_{1,2}^2/S_0^2 = 44,1/83 = 0,531 > 0,204 = s_{12;0,01} \quad \text{und}$$

$$S_{11,12}^2/S_0^2 = 32,4/83 = 0,390 > 0,204 = s_{12;0,01},$$

so daß weder H_{01} noch H_{0n} verworfen werden können. Es ist zum 1% Niveau also weder signifikant, daß $x_{(1)}$ und $x_{(2)}$ noch daß $x_{(11)}$ und $x_{(12)}$ ein Ausreißerpaar bilden.

4.5. Transformationen

Jeder beobachtete Wert x läßt sich als Realisation einer Zufallsvariablen X auffassen. Die meisten hier vorgestellten Verfahren stellen an die Verteilung solcher Zufallsvariablen gewisse Anforderungen (z. B. Normalität). In manchen Fällen sind diese nur zum Teil erfüllt; durch geeignete **Transformation** $g(X)$ der Zufallsvariablen lassen sich die gewünschten Eigenschaften mindestens annähernd erreichen. Liegen konkrete Daten vor, so müssen diese natürlich der gleichen Transformation $g(x)$ unterzogen werden wie die entsprechende Zufallsvariable. Die gebräuchlichsten Transformationen werden im folgenden vorgestellt, vgl. auch Abschnitt 4.2 sowie die Abschnitte 1.3 und 2.3 in Kap. XIV.

A. Die reziproke Transformation

Entstammen vorliegende Beobachtungswerte einer Verteilung mit starker Schiefe, so erreicht man durch Anwendung der **reziproken Transformation**

$$g(x) = x^{-1}$$

oftmals eine recht gute Anpassung an eine Normalverteilung. Hier arbeitet man dann also nicht mit den beobachteten Werten, sondern mit den Kehrwerten. Natürlich kann und sollte diese Transformation nur dann angewandt werden, wenn die beobachteten Daten den Wert Null bzw. Werte nahe Null nie bzw. nur selten annehmen.

Die reziproke Transformation wird besonders dann verwandt, wenn die Kehrwerte der Beobachtungsdaten selbst eine inhaltliche, z. B. physikalische, Bedeutung haben.

B. Die Wurzel-Transformation

Vor allem dann, wenn eine vorliegende Beobachtungsreihe einer Poissonverteilung entstammt, wird die **Wurzel-Transformation** verwandt. Sie tritt in vielen Varianten auf:

$$g(x) = \sqrt{x},$$

$$g(x) = \sqrt{x + c},$$

$$g(x) = \sqrt{x + \sqrt{x + 1}} \quad \text{oder auch}$$

$$g(x) = \sqrt{c \cdot x},$$

wobei c eine Konstante ist. Zur Transformation von Daten aus einer Poissonverteilung wird sie zumeist speziell in der Form

$$g(x) = \sqrt{x + 3/8}$$

und zur Transformation von Daten aus einer χ^2 -Verteilung in der Form

$$g(x) = \sqrt{2 \cdot x}$$

verwandt. Bei Anwendung der Transformation

$$g(x) = \sqrt{x + 3/8}$$

auf Daten aus einer Poissonverteilung wird approximativ nicht nur eine Normalverteilung erreicht, sondern vielmehr auch eine konstante Varianz von $1/4$, vgl. Abschnitt 4.2.

Daher wird sie, obwohl die Poissonverteilung eine recht schiefe Verteilung ist, der reziproken Transformation zumeist vorgezogen, bei der man hier keine konstante Varianz erreicht.

Beispiel: (die Daten sind entnommen aus Student (1907): "On the error of counting with a haemacytometer", Biometrika 5, S. 351–360)

In 400 Volumeneinheiten einer Suspension wurde jeweils die Anzahl i der Hefezellen ermittelt. Wir gehen zunächst davon aus, daß diese Anzahl poissonverteilt mit Parameter $\lambda = 5$ ist, vgl. auch das Histogramm aus Abb. 16. Die Daten sind in Tab. 18 angegeben und sollen nun mittels der Wurzel-Transformation

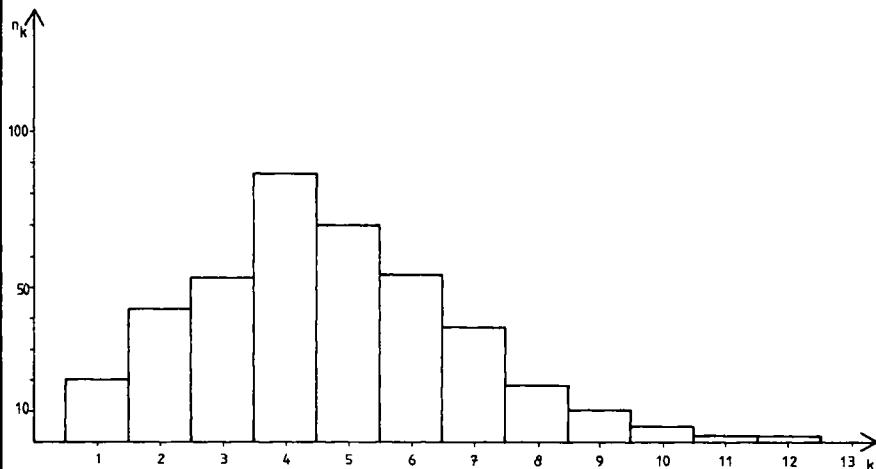


Abb. 16: Histogramm zur Anzahl der Hefezellen in einer Einheit einer Suspension, vgl. Tab. 18

$$g(i) = \sqrt{i + 3/8}$$

in Daten aus einer Normalverteilung mit Varianz $\sigma^2 = 1/4$ überführt werden. Die sich ergebenden Werte sind ebenfalls in Tab. 18 zu finden; es ist z. B. für $i = 4$:

$$g(4) = \sqrt{4 + 3/8} = \sqrt{4,375} = 2,09.$$

Wie berechnet man nun den Mittelwert μ dieser Normalverteilung? Mit

$$\lambda' = \lambda - 1/4 + \frac{8c - 3}{32\lambda} = 5 - \frac{1}{4} + \frac{8 \cdot 3/8 - 3}{32 \cdot 5} = 5 - \frac{1}{4} + 0 = 4,75$$

erhält man

$$\mu = E(g(i)) \approx g(\lambda') = \sqrt{\lambda' + 3/8} = \sqrt{4,75 + 3/8} = 2,26.$$

Tab. 18: Anzahl der Hefezellen in einer Suspension und Arbeitstabelle

Anzahl i der Hefezellen in einer Volumeneinheit	Anzahl n_i der Einheiten mit i Hefezellen	transformierte Daten $g(i) = \sqrt{i + 3/8}$	relative Häufigkeit einer Volumeneinheit mit i Hefezellen $n_i/400$	$b_i = \frac{g(i) + g(i+1)}{2}$	$h_i = \frac{n_i/400}{b_i - b_{i-1}}$
0	0	0,61	0,0000	0,890	-
1	20	1,17	0,0500	1,355	0,11
2	43	1,54	0,1075	1,685	0,33
3	53	1,83	0,1325	1,960	0,48
4	86	2,09	0,2150	2,200	0,90
5	70	2,31	0,1750	2,415	0,81
6	54	2,52	0,1350	2,615	0,675
7	37	2,71	0,0925	2,800	0,50
8	18	2,89	0,0450	2,975	0,26
9	10	3,06	0,0250	3,140	0,15
10	5	3,22	0,0125	3,295	0,08
11	2	3,37	0,0050	3,440	0,0344
12	2	3,51	0,0050	3,580	0,0357
> 12	0	3,65	0,0000	-	-

Wir wollen nun noch die Güte der Transformation überprüfen. Dazu müssen wir ein neues normiertes Histogramm erstellen und dies mit der Dichte der $N(2,26, 0,25)$ -Verteilung vergleichen. Da in den Originaldaten (Anzahl der Hefezellen) nur ganzzahlige Werte auftreten, müssen zunächst die relativen Häufigkeiten $n_i/400$ einer Volumeneinheit mit i Hefezellen berechnet werden, vgl. Tab. 18, und die Anfangs- bzw. Endpunkte

$$b_i = \frac{g(i) + g(i+1)}{2}$$

der Balken des normierten Histogramms berechnet werden, vgl. Tab. 18. Die Höhe dieser Balken ergibt sich dann zu

$$h_i = \frac{n_i/400}{b_i - b_{i-1}}.$$

In Abb. 17 ist das normierte Histogramm sowie die Dichte der $N(2,26, 0,25)$ -Verteilung graphisch dargestellt; man erkennt, λ ist zu groß gewählt.

C. Die Logarithmische Transformation

Die Logarithmische Transformation

$$g(x) = \ln(x + w),$$

wobei w eine Konstante ist, wird vor allem zur Transformation von Daten aus einer Lognormalverteilung in solche aus einer Normalverteilung verwandt. Die Lognormalverteilung spielt z. B. in der Bevölkerungsstatistik, aber auch in der Zuverlässigkeitstheorie als Lebensdauerverteilung eine Rolle.

Außerdem findet diese Transformation ihre Verwendung bei der Stabilisierung der Varianz von Daten aus einer Verteilung mit konstantem Variationskoeffizienten. Ist also die Standardabweichung einer Verteilung proportional zu ihrem Mit-

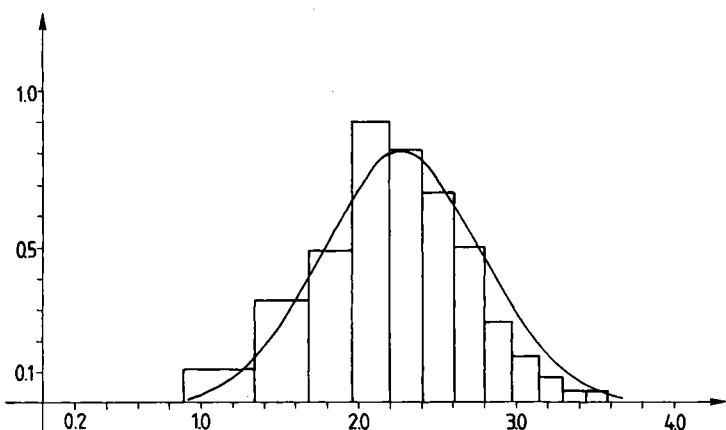


Abb. 17: Normiertes Histogramm und Dichtefunktion der $N(2,26, 0,25)$ -Verteilung

telwert, so wird durch die Logarithmische Transformation eine Varianzstabilisierung erreicht, vgl. auch Formel (ε) in Abschnitt 4.2.

D. Die Box-Cox-Transformation

In manchen Fällen lässt sich auch durch die **Box-Cox-Transformation** mit den Parametern λ und w

$$g(x) = \begin{cases} \frac{(x + w)^\lambda - 1}{\lambda}, & \text{falls } \lambda \neq 0 \\ \ln(x + w), & \text{falls } \lambda = 0 \end{cases}$$

die als Spezialfälle auch einige der schon erwähnten Transformationen enthält, eine Annäherung an eine Normalverteilung sowie eine Stabilisierung der Varianz erreichen, vgl. Box/Cox (1964) sowie Abschnitt 4.2, Kap. XII (2.4), Kap. XIV (1.3).

E. Die Arcus-Sinus-Transformation

So wie die Wurzeltransformation bei Daten aus einer Poissonverteilung angewandt wird, spielt die **Arcus-Sinus-Transformation**

$$g(x) = \sqrt{n + w_1} \cdot \arcsin \sqrt{\frac{x + w_2}{n + w_3}}$$

bei Realisationen von mit Parametern n und p binomialverteilten Zufallsvariablen eine entscheidene Rolle.

Wählt man dabei speziell

$$w_1 = 0,5, \quad w_2 = 0,375 \quad \text{und} \quad w_3 = 0,75,$$

so erreicht man nicht nur eine recht gute Anpassung an eine Normalverteilung, sondern auch eine Varianzstabilisierung: Entstammen die Beobachtungen einer $B(n, p)$ -Verteilung, so entstammen die transformierten Werte approximativ einer $N(\mu, 0,25)$ -Verteilung mit

$$\mu = E(g(x)) \quad g(n \cdot p).$$

Tab. 19: 200 Realisationen aus einer $B(8, 0,5)$ -Verteilung und Arbeitstabelle zur Berechnung des normierten Histogramms aus Abb. 18

Beobachtete Anzahl k	Häufigkeit n_k einer Beobachtung k	Transformierte Daten $g(k) = \sqrt{8,5} \arcsin \sqrt{\frac{k+0,375}{8,75}}$	$b_k = \frac{g(k) + g(k+1)}{2}$	$h_k = \frac{n_k}{200}$ $\frac{b_k - b_{k-1}}{b_k}$
0	0	0,61	0,900	–
1	5	1,19	1,395	0,05
2	24	1,60	1,775	0,32
3	42	1,95	2,120	0,61
4	56	2,29	2,460	0,82
5	46	2,63	2,805	0,67
6	20	2,98	3,185	0,26
7	7	3,39	3,680	0,07
8	0	3,97	–	–

Beispiel: In Tab. 19 sind 200 Realisationen einer $B(8, 0,5)$ -verteilten Zufallsvariablen sowie die zugehörigen transformierten Daten

$$g(k) = \sqrt{8,5} \cdot \arcsin \sqrt{\frac{k+0,375}{8,75}}$$

angegeben.

Um die Güte der erreichten Anpassung an eine Normalverteilung mit Erwartungswert

$$\mu \approx g(n \cdot p) = g(4) = \sqrt{8,5} \cdot \arcsin \sqrt{\frac{4,375}{8,75}} = \sqrt{8,5} \cdot 0,785 = 2,29$$

und Varianz

$$\sigma^2 = 0,25$$

zu überprüfen, ist in Abb. 18 neben der Dichte der $N(2,29, 0,25)$ -Verteilung das normierte Histogramm der transformierten Daten angegeben. Die Anfangs- und

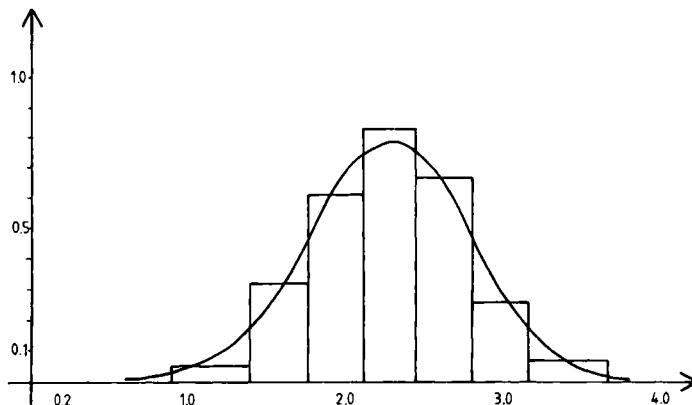


Abb. 18: Dichtefunktion der $N(2,29, 0,25)$ -Verteilung und normiertes Histogramm von 200 transformierten Realisationen aus einer $B(8, 0,5)$ -Verteilung, vgl. Tab. 19

Endpunkte der Histogrammbalken berechnen sich dabei zu

$$b_k = \frac{g(k) + g(k+1)}{2},$$

und die Balkenhöhe zu

$$h_k = \frac{n_k / 200}{b_k - b_{k-1}},$$

wobei n_k die Häufigkeit einer Beobachtung k angibt.

F. Die Fishersche Z-Transformation

Vor allem in der Korrelationsrechnung spielt die **Fishersche Z-Transformation**

$$g(x) = \operatorname{arctanh}(x)$$

eine entscheidende Rolle. Auf sie soll an dieser Stelle nicht näher eingegangen werden, da sie im Abschnitt 1 des Kap. IX ausführlich diskutiert wird.

Es sei noch erwähnt, daß man sich die Anwendung von Transformationen sehr gut überlegen muß, vor allem wenn nur leichte Verletzungen von Voraussetzungen vorliegen, denn es ist schwierig, manchmal unmöglich, die so gewonnenen Aussagen sinnvoll zu interpretieren.

5. Allgemeine Aspekte der Planung von Versuchen

Ein Experiment zerfällt in drei aufeinander folgende Phasen, nämlich die **Planung**, die **Durchführung** und die **Auswertung**. Wesentlich ist zunächst einmal die Formulierung der zu beantwortenden Versuchsfrage, d. h. man muß sich darüber im Klaren sein, was man eigentlich wissen möchte und auf welche Grundgesamtheit sich eine Aussage beziehen soll, denn die Wahl des Versuchstyps ist abhängig von der interessierenden Grundgesamtheit.

Allgemein lassen sich die drei folgenden **Grundsätze der Planung eines Versuches** formulieren:

1. Jede Stufe der Einflußfaktoren bzw. Behandlungen muß stets auf mehrere Versuchsobjekte angewandt werden bzw. von mehreren Versuchsobjekten getragen werden;
2. Die Versuchsobjekte sollten zufällig ausgewählt werden und zwar aus der Grundgesamtheit, auf die man schließen möchte;
3. Die Zuordnung der Versuchsobjekte zu den Behandlungen bzw. die Auswahl der Objekte, die eine Faktorstufe tragen, sollte zufällig erfolgen, um systematische Fehler auszuschalten.

Möchte man den Einfluß eines einzelnen Faktors, der in p Ausprägungen bzw. auf p Stufen interessiert, auf ein bestimmtes Merkmal einer Grundgesamtheit, auf die sich eine Versuchsaussage erstrecken soll, untersuchen, so wählt man zunächst $N = p \cdot n$ Objekte aus der interessierenden Grundgesamtheit zufällig aus (**einfache Zufallsauswahl**, vgl. Abschnitt 1 dieses Kapitels). Sodann ordnet man zufällig jeweils n Objekten eine Stufe bzw. eine Behandlung des interessierenden Faktors zu und führt dann das Experiment durch, d. h. man mißt die Ausprägungen des Merk-

mals, auf das der Faktor einen Einfluß haben könnte, an allen N Versuchsobjekten. Diese Art der Zuordnung stellt einen **vollständig randomisierten Versuchsplan** dar, der mit Methoden der **einfachen Varianzanalyse** ausgewertet wird.

Ebenfalls mit den Verfahren der einfachen Varianzanalyse kann man Versuche auswerten, bei denen der interessierende Einflußfaktor in der Grundgesamtheit bereits vorliegt. Man zieht dann je n Versuchsobjekte mit jeder der p Faktorstufen und vergleicht diese.

Interessiert man sich etwa für den Faktor Blutdruck bei Männern und Frauen, so wählt man z. B. zufällig 10 Männer und 10 Frauen aus der Grundgesamtheit aus und mißt deren Blutdruck. Weitere eventuelle Einflußgrößen – also Faktoren, die einen Einfluß auf das interessierende Merkmal haben – werden hierbei natürlich vernachlässigt. Dies soll am folgenden Beispiel verdeutlicht werden.

Beispiel: Eine pharmazeutische Firma möchte ein neues Medikament einführen und daher zunächst feststellen, ob dieses zu besseren Ergebnissen als das alte führt. Um dieses zu untersuchen, kann man etwa gleich großen Gruppen von Personen je eines dieser Medikamente verabreichen und die Wirkung feststellen. Der Haupteinflußfaktor ist also die Medikamentenbehandlung. Weitere Einflußfaktoren, wie z. B. Alter, Geschlecht und Gewicht, können das Versuchsergebnis beeinflussen, werden aber zunächst vernachlässigt.

Ein offensichtlich schlecht geplanter Versuch würde darin bestehen, Medikament A nur Frauen und Medikament B nur Männern zu verabreichen. Denn dann lassen unterschiedliche Gruppenergebnisse nicht notwendigerweise auf unterschiedliche Wirkungen der Medikamente schließen; sie könnten auch allein aufgrund der so unterschiedlichen Gruppenzusammensetzung entstanden sein. Die **Effekte der Faktoren überlagern sich vollständig** und der vornehmlich interessierende Haupteffekt, die Wirkung der Medikamente, ist nicht mehr **identifizierbar**. Ein besserer Versuchsplan, bei dem die Nebeneinflußfaktoren nicht vernachlässigt werden, besteht darin, den vollständig randomisierten Versuchsplan zu verwenden, d. h. die Versuchspersonen den Behandlungen streng zufällig (mit Hilfe von Würfeln, Losverfahren, Zufallszahlen aus Tabellen (vgl. Anhang) und Rechnern) zuzuweisen (**randomisieren**), vgl. auch Abschnitt 8 in Kap. III. Hierbei soll man sich vor nur scheinbar zufälligen Zuteilungen hüten, z. B. Auswahlverfahren aufgrund des Alphabets oder der Geburtstage, vgl. Abschnitt 3 dieses Kapitels.

Bei diesem Versuchsplan hofft man, daß die zufällige Zuteilung der Versuchsobjekte zu den Behandlungen eventuelle Nebeneinflußfaktoren ausgleicht. Das braucht natürlich nicht der Fall zu sein; es könnte ja bei der zufälligen Zuteilung auch passieren, daß überwiegend Männer in der einen und überwiegend Frauen in der anderen Gruppe sind. Daher sollte man versuchen, den Faktor Geschlecht und auch andere Einflußfaktoren, deren Bedeutung so groß ist, daß man ihre Zuordnung nicht dem Zufall überlassen kann, schon im Modell zu berücksichtigen und nur den Ausgleich von wirklich unwesentlichen und natürlich auch von unbekannten Einflüssen dem Randomisieren zu überlassen. Im obigen Beispiel kommt etwa zum Faktor Medikamentenbehandlung als ein Nebeneffekt das Geschlecht hinzu. Dieser kann durch sogenannte **Blockbildung** berücksichtigt werden.

Die Grundgesamtheit zerfällt aufgrund eines Faktors in r Schichten (Klassen) – bei einer Einteilung in Männer und Frauen etwa ist $r = 2$ –, so daß man jetzt eine **geschichtete Stichprobe** zieht und zwar so, daß man aus jeder Schicht eine Zufallsstichprobe vom Umfang $n \cdot p$ zieht. Diese r Zufallsstichproben bilden jeweils einen

Block. Jedem Versuchsobjekt in den r Blöcken wird nun zufällig eine der p Behandlungen zugeordnet; dann wird der Versuch durchgeführt. Im Beispiel würde man jetzt also den $n \cdot p$ Frauen und den $n \cdot p$ Männern getrennt die p Medikamente zufällig zuordnen.

Einen solchen Versuch nennt man auch ein **Blockexperiment**. Die Auswahl der Versuchsobjekte für ein Blockexperiment ist in Abb. 19 auch graphisch veranschaulicht, wobei die Versuchsergebnisse dann durch Zufallsvariablen Y_{ijk} beschrieben werden.

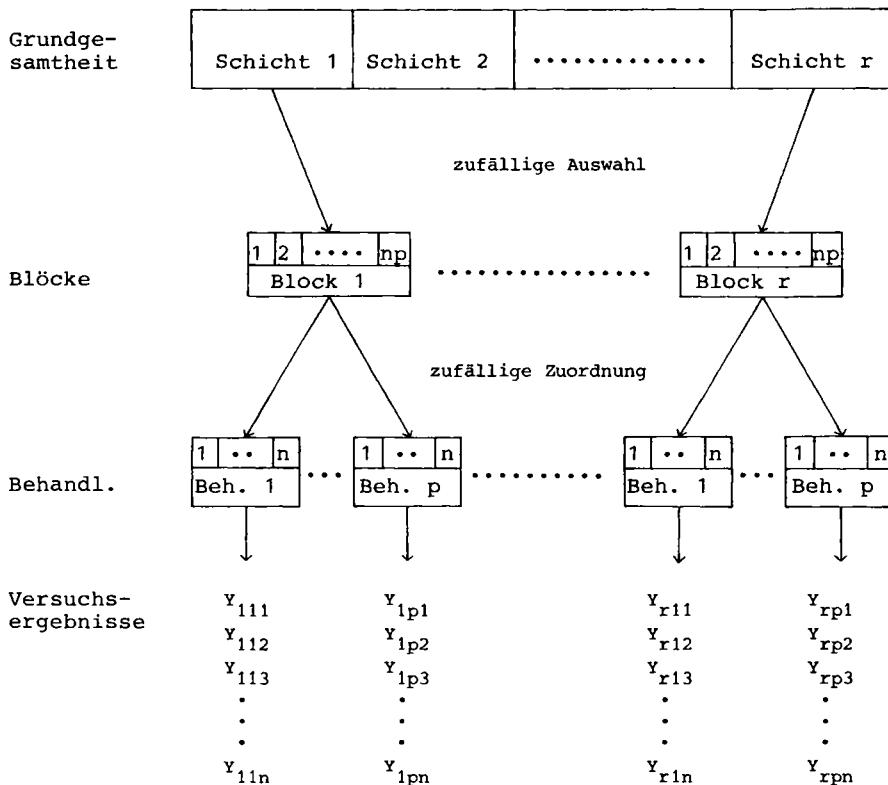


Abb. 19: Auswahl der Versuchsobjekte für ein Blockexperiment

Beispiel:

- (a) Der Einfluß dreier verschiedener Belichtungsdauern (Kurztag, Langtag, Dauerlicht) auf das Nadelgewicht von Föhrensämlingen nach der Aufzucht soll untersucht werden. Als Blockfaktor werden vier verschiedene Herkunftsorte der Samenmaterialien im Experiment berücksichtigt. Die Grundgesamtheit zerfällt also in 4 Klassen. Aus jeder Klasse werden zufällig $2 \cdot 3$ Sämlinge ausgewählt; diese je 6 Sämlinge bilden je einen der 4 Blöcke. In jedem Block wird dann zufällig je 2 Föhrensämlingen einer der 3 Belichtungsdauern zugeordnet. Nach der Aufzucht mißt man dann die Nadelgewichte der insgesamt 24 Föhrensämlinge.

- (b) Man möchte den Einfluß zweier verschiedener Lademethoden auf die Feuergeschwindigkeiten eines Geschützes untersuchen und man weiß, daß diese abhängig ist von der Mannschaft. Als Blockfaktor werden 3 verschiedene Mannschaften betrachtet. Eine besteht aus schweren, eine aus mittelschweren und eine aus leichten Männern. Jede Mannschaft bedient das Geschütz nach jeder der beiden Lademethoden. Gemessen werden dann die Feuergeschwindigkeiten.

Die Voraussetzungen der zufälligen Auswahl lassen sich häufig nicht verwirklichen, z. B. erfolgt die Blockbildung vielmals erst nach Auswahl der Versuchsstoffe. Der Bezug zur Grundgesamtheit ist in solchen Fällen jedoch oft unklar. Bestimmt man etwa die Erträge bei zwei verschiedenen Düngerbehandlungen auf den Parzellen von zehn Feldern, so ist es ein Unterschied, ob die Felder aus der Grundgesamtheit aller Felder ausgewählt und nachträglich nur aufgrund der Tatsache, daß es sich um verschiedene Felder handelt, zu Blöcken erklärt wurden, oder ob die Grundgesamtheit zuerst in 10 Klassen bzgl. der Bodengüte homogener Felder aufgeteilt wird und aus jeder Klasse jeweils ein Feld zufällig ausgewählt wird, auf dessen Parzellen die beiden Düngerbehandlungen angewandt werden. Läßt man weitere Einflußfaktoren wie etwa Bearbeitung und Klima, wie hier geschehen, außer acht, so hat letzteres Vorgehen eine größere Aussagekraft bzgl. der Grundgesamtheit aller Felder, da diese hier besser repräsentiert wird.

Beispiel: Möchte man etwa bei Frauen ein altersspezifisches Merkmal, wie z. B. den Blutdruck, untersuchen, wählt dazu zufällig einige Frauen aus der Grundgesamtheit aller Frauen aus und teilt diese dann in Altersklassen ein, so kann es z. B. passieren, daß eine Altersklasse sehr schwach besetzt bzw. überhaupt nicht besetzt ist, so daß die Versuchspersonen die Grundgesamtheit der Frauen offensichtlich schlecht repräsentieren. Besser wäre es, von vornherein aus jeder Altersklasse eine bestimmte Anzahl von Frauen zufällig auszuwählen.

Häufig tritt das sogenannte **einfache Blockexperiment** auf, bei dem jede Behandlung nur einmal in jedem Block angewandt wird ($n = 1$). Das liegt meist daran, daß die **natürliche Klassengröße** in der Grundgesamtheit nicht größer ist als die Anzahl verschiedener Behandlungen. Man kann also nicht mehr aus jeder Klasse der Grundgesamtheit einen Teil auswählen, da dann zu viele Blöcke im Versuch zu berücksichtigen sind, sondern kann aus der Grundgesamtheit nur aus einigen Klassen, diese dann aber z. T. auch vollständig (**Klumpenstichprobe**) Versuchsstoffe auswählen und als Blöcke im Versuch betrachten.

Als **Beispiel** betrachten wir einen Mastversuch bei Schweinen: 4 verschiedene Futtermischungen sollen im Hinblick auf die Mastleistung (z. B. durchschnittliche tägliche Zunahme der Tiere) untersucht werden. Die Grundgesamtheit der zur Mast in Frage kommenden Schweine des gleichen Geschlechts (um geschlechtspezifische Faktoren hier nicht berücksichtigen zu müssen) zerfällt in die Klassen aller verschiedenen Würfe eines bestimmten Alters bei Mastbeginn, wenn man derart nach genetischen und haltungsbedingten Gesichtspunkten klassifiziert. Für den Mastversuch wählt man nun z. B. 25 Würfe und daraus je 4 Ferkel des festgelegten Geschlechts zufällig aus dieser Grundgesamtheit aus. Jedem dieser 4 Ferkel eines Wurfes wird dann eine der 4 Futtermischungen zufällig zugeordnet. Wollte man jede Futtermischung mindestens 2 mal pro Block verwenden, so wäre es notwendig, 8 Ferkel des festgelegten Geschlechts aus jeweils einem Wurf auszuwählen. Dies würde die Anzahl der zur Auswahl stehenden Würfe stark verringern und somit auch die Grundgesamtheit verkleinern, auf die sich eine Aussage aufgrund des Experimentes beziehen kann.

Ähnliche Probleme treten etwa im medizinischen und psychologischen Bereich auf, wo man bei **differentiellen Behandlungen** als Blöcke Geschwistergruppen (und bevorzugt Zwillingspaare) verwendet. Hierbei muß allerdings sichergestellt sein, daß die so erreichte **Homogenität der Versuchseinheiten** nicht die **Verallgemeinerungsfähigkeit** der Versuchsergebnisse auf eine Grundgesamtheit, die z. B. nicht nur Zwillinge umfaßt, gefährdet.

Oft ist es aber auch so, daß die natürliche Klassengröße Eins ist (z. B. ein Tier, eine Person) und an diesem einen Objekt mehrere Behandlungen durchgeführt werden. In solchen Fällen ist zu beachten, daß sich die verschiedenen Behandlungen in ihrer Wirkung nicht überlagern sollten, was z. B. dadurch sichergestellt werden kann, daß zwischen Versuchen, an denen das selbe Versuchsobjekt beteiligt ist, genügend Zeit verbleibt, in der eventuelle Behandlungsnachwirkungen neutralisiert werden. Dies ist insbesondere kritisch bei Medikamentenversuchen, wo mehrere Behandlungen an der gleichen Person durchgeführt werden müssen, weil anders die benötigte Homogenität bzgl. übriger Störeinflüsse nicht gewährleistet werden kann.

Werden in einem Block alle Behandlungen angewandt und werden die Behandlungen zufällig den Versuchsobjekten zugeordnet, so spricht man von einem **Experiment in vollständig randomisierten Blöcken**. Wenn nicht ausdrücklich anders vermerkt, beinhaltet diese Bezeichnung, daß der Block vollständig ist in dem Sinne, daß jede Behandlung im Block auch angewandt wird. Besteht der Block nur aus einem Versuchsobjekt, so entfällt natürlich das Randomisieren. Allenfalls über die Reihenfolge, in der die Behandlungen angewandt werden, ist zu randomisieren, falls verbleibende Nachwirkungen doch nicht auszuschließen sind; allerdings ist in diesem Fall zu überlegen, ob nicht Methoden der **multivariaten Varianzanalyse** anzuwenden sind, vgl. Kap. X in Hartung/Elpelt (1984).

Von einem **unvollständigen Blockplan** spricht man, wenn die Anzahl der Versuchseinheiten in einem Block kleiner ist als die Anzahl der verschiedenen Behandlungsstufen.

Wenn der Einfluß zweier frei zuzuordnender Faktoren (verschiedene Behandlungsarten), die auf r bzw. p Stufen interessieren, auf ein Merkmal untersucht werden soll, so werden zunächst $n \cdot r \cdot p$ Versuchseinheiten zufällig aus der interessierenden Grundgesamtheit entnommen und dann jeweils n Versuchseinheiten zufällig eine der $r \cdot p$ Stufenkombinationen der beiden Faktoren zugeordnet; eine solche Zuordnung nennt man auch **Kreuzklassifikation**, da jede Stufe des einen mit jeder Stufe des anderen Faktors kombiniert angewandt wird.

Beispiel: Der Einfluß von 5 verschiedenen Diäten und 3 verschiedenen Medikamentenbehandlungen auf das Merkmal Blutdruck soll untersucht werden. Aus der Grundgesamtheit der möglichen Patienten werden dann z. B. $10 \cdot 15$ Versuchspersonen zufällig ausgewählt. Je 10 Versuchspersonen wird dann zufällig eine der 15 möglichen Kombinationen von Diäten und Medikamentenbehandlungen zugeordnet.

Einen solchen Versuch wertet man dann mit Methoden der **zweifachen Varianzanalyse** aus, die gleichfalls zur Anwendung kommen, wenn die Stufen der beiden Faktoren fest vorliegen und deren Einfluß auf ein bestimmtes Merkmal untersucht (ohne weitere Behandlungsmethoden) wird. Hier zerfällt die Grundgesamtheit in $r \cdot p$ Klassen, die jeweils eine der Stufenkombinationen tragen; aus jeder Klasse werden dann zufällig n Versuchsobjekte entnommen und bzgl. des interessierenden Merkmals analysiert (**geplante Erhebung bei zwei Faktoren**).

Beispiel: Der Einfluß des Geschlechts und der Größe des Betriebes einer Branche auf das Anfangseinkommen eines dort Beschäftigten in vergleichbarer Berufsgruppe soll untersucht werden. Die Betriebsgrößen mögen in 7 Kategorien eingeteilt sein; das Geschlecht liegt natürlich auf 2 Stufen vor. Aus den so entstehenden 24 Klassen der Grundgesamtheit werden dann z. B. je 30 Personen zufällig ausgewählt und nach ihrem Einkommen befragt.

Liegen Stufen des einen Faktors nur innerhalb jeweils einer Stufe des zweiten Faktors vor, so kommt man zu **Modellen der hierarchischen Klassifikation**.

Beispiel: Es wird der regionale Einfluß sowie der Effekt verschiedener landwirtschaftlicher Betriebe auf die durchschnittliche Milchmengenleistung der Kühe untersucht. Die Gebiete stellen die Stufen des einen Faktors und jeder Betrieb eine Stufe des anderen Faktors dar; natürlich liegt jeder Betrieb nur innerhalb eines Gebietes und ist nicht beliebig mit den Stufen des Faktors Region kombinierbar. Die beiden Faktoren sind also nicht kreuzklassifizierbar, sie liegen vielmehr nur in hierarchischer Anordnung vor.

Werden nun zwei Faktoren auf p bzw. q Stufen unter zusätzlicher Berücksichtigung eines Blockfaktors auf r Stufen in Bezug auf den Einfluß auf ein interessierendes Merkmal untersucht, so kommen wir zu einem **$p \times q$ faktoriellen Experiment in vollständig randomisierten Blöcken**, wenn wir den r Blöcken mit je $p \cdot q$ Versuchseinheiten, die zufällig aus den entsprechenden Klassen der Grundgesamtheit ausgewählt sind, jeweils jede der $p \cdot q$ Stufenkombinationen der beiden Faktoren rein zufällig zuordnen.

Beispiel: Im bereits erwähnten Versuch, in dem 5 Diäten mit 3 Medikamentenbehandlungen kombiniert werden, empfiehlt es sich, falls der Versuch stationär durchgeführt wird, als Blockfaktor verschiedene Krankenhäuser bzw. Stationen in Krankenhäusern zu berücksichtigen.

Sind hierbei Korrelationen derjenigen Beobachtungen, die eine Stufe eines der beiden Faktoren gemeinsam haben, zu erwarten, so sei auf das **Modell des Split-Plot-in-Strips-Experiments** hingewiesen.

Häufig können aus versuchstechnischen oder sachlogischen Bedingungen heraus die Stufen der Faktoren nicht derart beliebig kombiniert und randomisiert zugeteilt werden. Kann man im Block zunächst nur die Stufen des einen Faktors und erst dann innerhalb der Stufen dieses Faktors die Stufen des zweiten Faktors zufällig zuordnen, so ist ein **Modell der Split-Plot-Experimente in vollständig randomisierten Blöcken** angebracht. Die Methoden der Split-Plot-Experimente kommen auch dann bevorzugt zur Anwendung, wenn die Randomisationsmöglichkeiten anderweitig eingeschränkt sind.

Beispiel: In 5 Branchen werden in 12 verschiedenen Regionen und 7 verschiedenen Kategorien von Betriebsgrößen die prozentualen Steigerungen der Auftragseingänge untersucht. Betrachten wir die 5 Branchen als Stufen des Blockfaktors, so haben wir zunächst innerhalb jeder Branche die gleichen Gebiete. Wird nun innerhalb eines jeden solchen Gebietes je ein Betrieb jeder der 7 Kategorien ausgewählt, so kommt man zum Split-Plot-Modell und zwar zum Modell des Split-Plot in vollständig randomisierten Blöcken, wenn man nur davon ausgeht, daß die Erhebungsergebnisse korreliert sind, wenn sie im gleichen Gebiet zur gleichen Branche gehören. Berücksichtigt man auch eine eventuelle Korrelation der Ergebnisse bei Betrieben der gleichen Größenkategorie innerhalb der gleichen Branche, so kommt man zum Modell des Split-Plot-in-Strips.

Berücksichtigt man **mehrere Faktoren**, so sind im Experiment ähnliche Vorgehensweisen zu beachten. Allgemein bei Einbeziehung von drei Faktoren, unter denen natürlich auch Blockfaktoren sein können, stehen die Methoden der **dreifachen Varianzanalyse** bzw. auch die **höheren Split-Plot-Experimente** zur Verfügung.

Eines der wesentlichen Ziele der Untersuchung des Einflusses mehrerer Faktoren auf ein interessierendes Merkmal ist es, die **Interaktionen** dieser Faktoren – auch **Wechselwirkungen** genannt – zu untersuchen. Unter einer Interaktion versteht man, daß die Faktoren sich im Verhältnis zueinander unterschiedlich auf wechselnden Stufen eines anderen Faktors verhalten. Dies mag am Beispiel einer **zweifachen Wechselwirkung**, d. h. anhand der Interaktion zweier Faktoren in einem Blockexperiment illustriert werden.

Beispiel: Der Einfluß von drei verschiedenen Dosierungen eines blutdrucksenkenden Medikaments auf den Blutdruck soll untersucht werden. Die Grundgesamtheit aller möglichen Versuchspersonen wird, da ein Einfluß des Geschlechts auf die Wirkung nicht ausgeschlossen werden kann, in männliche und weibliche Personen eingeteilt. Aus jeder der beiden Klassen werden dann 15 Versuchspersonen zufällig ausgewählt und je 5 Personen aus jedem der beiden Blöcke wird zufällig eine der 3 Dosierungen des Medikaments zugeordnet. Gemessen wird dann die Anzahl der Einheiten, um die der Blutdruck bei Behandlung mit der jeweiligen Dosierung gesenkt wird. Die durchschnittlichen Ergebnisse bei jeder Stufenkombination der Faktoren Dosierung und Geschlecht sind in **Tab. 20** zusammengestellt.

Tab. 20: Durchschnittliche Blutdrucksenkung bei Behandlung von männlichen und weiblichen Versuchspersonen mit 3 Dosierungen eines blutdrucksenkenden Medikaments

Dosierung \ Geschlecht	männlich	weiblich
Dosierung		
1	7	3
2	4,5	8
3	9	5,5

Diese durchschnittlichen Versuchsergebnisse sind in **Abb. 20** graphisch dargestellt.

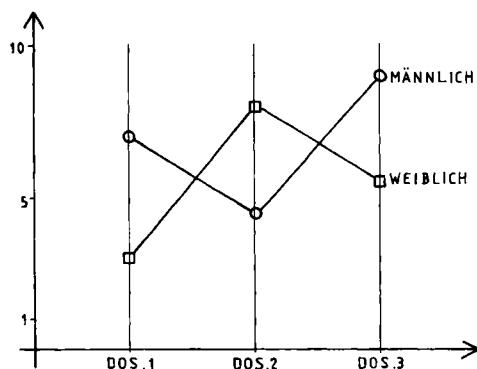


Abb. 20: Durchschnittliche Blutdrucksenkung bei Behandlung von männlichen und weiblichen Versuchspersonen, vgl. Tab. 20

Der Abb. 20 entnimmt man, daß offenbar Wechselwirkungen zwischen den beiden Faktoren bestehen, denn, wenn keine Interaktionen vorliegen, könnte die grafische Darstellung etwa eine Gestalt wie in Abb. 21 haben.

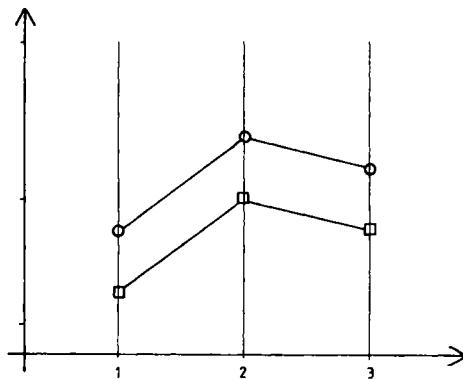


Abb. 21: Graphische Gestalt der durchschnittlichen Versuchsergebnisse, falls keine Wechselwirkungen vorliegen

Liegen zwischen zwei Faktoren keine Wechselwirkungen vor, so verlaufen die Verbindungslien der durchschnittlichen Werte parallel. Leichte Abweichungen von der Parallelität können allerdings zufallsbedingt sein. Daher testet man, ob die Wechselwirkung zwischen Faktoren statistisch signifikant gesichert ist.

Treten zwischen den Einflußfaktoren keine Interaktionen auf, so lassen sich mitunter die **Versuchsanlagen erheblich verkleinern**, was auch finanzielle Vorteile mit sich bringt; die gebräuchlichsten Anlagen sind die sogenannten **Lateinischen Quadrate**, **Griechisch-Lateinischen Quadrate**, **Lateinischen Rechtecke**, **unvollständigen balancierten Blockpläne**, **Zwei- und Dreifachgitter** (**Lattice Designs**).

Wird eine **Störvariable** (Einflußfaktor) nicht kategorisiert, so daß man sie in Form von Blockbildung bei der Durchführung eines Experimentes berücksichtigen kann, sondern wird diese Störgröße als quantitative Variable (**Kovariable**) mitgeführt, wobei sie auch erst im Experiment miterhoben werden kann, so kommt man zu **Modellen der Kovarianzanalyse**.

Beispiel: Der Einfluß von 4 verschiedenen Narkotika auf die Zeit, die bis zur Einschläferung vergeht, soll untersucht werden. Eine qualitative Störvariable, die durch Blockbildung berücksichtigt werden kann, ist etwa die Trinkgewohnheit der Versuchspersonen. Man könnte die Grundgesamtheit etwa in 4 Klassen (kein Alkoholtrinker, schwacher Alkoholtrinker, mittlerer Alkoholtrinker, starker Alkoholtrinker) aufteilen. Eine quantitative Störvariable im Experiment, die als Kovariable mitgeführt werden kann, ist etwa das Gewicht der Versuchspersonen. Durch Gewichtsklassenbildung könnte man auch diese Störgröße kategorisieren und durch Blockbildung berücksichtigen. Eine andere Störvariable ist etwa der Pulsschlag der Versuchsperson bei Beginn der Narkotisierung.

Es wird dann unter Berücksichtigung der oder des qualitativen Einflußfaktors (hier z. B. der 4 verschiedenen Narkotika) bzgl. der quantitativen Störvariablen (hier etwa Gewicht, Pulsschlag) eine Regression durchgeführt.

Oft ist es angebracht, wesentlich kompliziertere Modelle als die hier erwähnten zu konstruieren.

Natürlich werden in den vorgestellten Modellen **weitere Annahmen** gemacht, die in der Praxis allerdings nur annäherungsweise erfüllt sein können. Häufig wird etwa vorausgesetzt, daß die zufälligen Versuchsfehler **normalverteilt** sind. Hier läßt sich jedoch sagen, daß die Verteilungsannahme nicht zum Schätzen der Behandlungseffekte benötigt wird und daß die Testverfahren auf signifikante Unterschiede (F-Test) in Bezug auf Verletzungen der Voraussetzungen über die Fehlerterme in gewisser Hinsicht **robust** sind.

Eine Verzerrung der Meßwerte (systematische Fehler, vgl. Abschnitt 4 dieses Kapitels) kann zum einen durch gutes Experimentieren vermieden und zum anderen durch **Randomisation** in den meisten Fällen ausgeglichen werden. In ausgewogenen (balancierten) Versuchsplänen ist auch die Verletzung der Voraussetzung, daß alle Fehler gleiche Varianz σ^2 haben, nicht besonders schwerwiegend. Eine Möglichkeit besteht in solchen Fällen darin, varianzstabilisierende Transformationen, vgl. Abschnitt 4, anzuwenden und dabei (hoffentlich) auch Normalität zu erreichen.

Die schwerwiegendsten Fehler bei der Verwendung des **F-Testes**, vgl. Kapitel XI, treten dann auf, wenn die Fehlerterme **korreliert** sind. In Laborversuchen tritt dieser Fehler allerdings kaum auf, da solche Experimente gut zu kontrollieren sind; im Bereich der Biologie und der Landwirtschaft etwa wird die Unkorreliertheit durch Randomisation zumindest annähernd erreicht. Zum Beispiel bei ökonomischen und ökologischen Zeitreihen, wo die Fehler wohl oft korreliert sind, sei auch auf Verfahren der **Zeitreihenanalyse**, vgl. Kap. XII, hingewiesen. Eine Verletzung der Normalitätsannahme hat, vgl. Hotelling (1960), wenig Einfluß auf die Ergebnisse des F-Testes, falls die Verteilung der das Experiment beschreibenden Zufallsgrößen nicht allzu schief ist. In der **Varianzanalyse** wird zudem durch Randomisation erreicht, daß der F-Test den entsprechenden verteilungsfreien, jedoch äußerst aufwendigen (Permutations-)Test gut approximiert.

Allgemein kann man in einem Experiment zwischen drei verschiedenen **Arten von Einflußfaktoren** unterscheiden:

- **Haupteinflußfaktoren**,
- **kontrollierte weitere Einflußfaktoren**, die als qualitative Störgrößen durch Blockbildung bzw. als quantitative Störgrößen (Kovariablen) berücksichtigt werden können,
- **nicht-kontrollierbare weitere Einflußfaktoren**, die weitgehend z. B. durch Randomisation ausgeschaltet werden müssen. Andere Möglichkeiten der Ausschaltung solcher Einflußfaktoren sind etwa in klinischen Versuchen die Anlage eines Versuches als Blind- bzw. Doppelblindversuch. Bei einem Blindversuch ist der Versuchsperson die Art des verabreichten Medikamentes nicht bekannt; unter einem Doppelblindversuch versteht man ein Experiment, bei dem sowohl dem Probanden als auch dem Versuchsdurchführenden die Art des verabreichten Medikamentes nicht bekannt ist. Erreicht werden soll durch solche Versuchsanordnungen die Ausschaltung psychologischer Effekte. Dies ist insbesondere dann wichtig, wenn als Kontrollgruppen Versuchspersonen mitgeführt werden, denen ein Placebo (**Scheinbehandlung**) verabreicht wird, vgl. auch Abschnitt 6.4. Darüber hinaus empfiehlt es sich in vielen Fällen, Kontrollgruppen mit „Nichtbehandlung“ mitzuführen. Unbekannte – eventuell erhebliche – nicht kontrollierte Einflußfaktoren können so aufgefangen werden.

lierbare Einflußfaktoren können nämlich nur dann ausgeglichen werden, wenn eine Kontrollgruppe eben diesen Einflußfaktoren mit unterworfen wird, aber den oder die Haupteinflußfaktoren nicht erhält.

Auf einige weitere Probleme der Versuchsplanung, die auch von allgemeinem Interesse sind, wird in Abschnitt 6 im Zusammenhang mit der Planung klinischer Versuche noch eingegangen.

Die Auswertung einiger der hier angesprochenen Modelle der Varianzanalyse findet sich in Kapitel XI. Die Analyse komplizierterer Versuchsanlagen muß, da sie sehr umfangreich ist, einer eigenen Darstellung vorbehalten bleiben, vgl. Hartung/Elpelt (1984), Hartung/Werner (1984); es sei aber auch auf die Bücher von Fisher (1966, 1972), Cochran/Cox (1957), Kempthorne (1952) und Anderson/Bancroft (1952) hingewiesen, die sich ausführlich mit Problemen der Planung von Versuchen und der Varianzanalyse beschäftigen. Für eine Darstellung der mathematischen Theorie der optimalen Versuchsplanung sei auf das Buch von Krafft (1978) verwiesen.

Auf die Analyse rein qualitativer Daten, die man natürlich auch durch Übergang zu Klassen, vgl. Abschnitt 2 in Kapitel I, bei ursprünglich quantitativen Daten erhalten kann, wird ausführlich in Kapitel VII eingegangen.

6. Anlage von klinischen Versuchen

Die pharmakologische und medizinische Erforschung prophylaktischer oder auch therapeutischer Behandlungen von Krankheiten läuft im wesentlichen in vier Stufen ab:

1. In Laborversuchen wird, so weit als möglich, durch Experimente in vitro und auch in vivo, an Tieren Basiswissen gesammelt.
2. Da sich auf diese Art keine direkt auf den Menschen übertragbaren Ergebnisse gewinnen lassen, sind bei günstigem Ausgang der Laborversuche, die Resultate am Menschen zu überprüfen und Strategien zur Anwendung der betreffenden Behandlung zu entwickeln.
3. Erhält man auch hier den Eindruck, daß die neue Behandlung den bisher üblichen Methoden gleichwertig oder gar überlegen ist, so bleibt in einer Vergleichsstudie nunmehr die Güte der neuen Behandlung relativ zu den Standardmethoden festzustellen.
4. Bestätigen sich die positiven Eigenschaften auch in der dritten Stufe, so bleiben noch in einer Langzeitbeobachtung seltene Nebeneffekte und Spätfolgen zu erforschen.

Ein **klinischer Versuch** eignet sich insbesondere für die vergleichende Betrachtung verschiedener Behandlungen. Dabei können zum einen theoretische Interessen im Vordergrund stehen (**non-therapeutic clinical research, explanatory approach**), zum anderen kann beim klinischen Versuch hauptsächlich interessieren, welche Behandlung einer Erkrankung für die Praxis empfehlenswert ist (**therapeutic clinical research, pragmatic approach**).

Beispiel:

- (a) Vergleicht man etwa eine synthetisch gewonnene Droge mit einer von der chemischen Struktur her eng verwandten natürlichen Droge, der man sich bislang

bedient hat, so dient ein solcher klinischer Versuch der nichttherapeutischen klinischen Forschung.

- (b) Untersucht man in einer klinischen Studie, ob bei einer bestimmten Krebserkrankung eine Radiotherapie, eine Chemotherapie, ein operativer Eingriff oder eine Kombination dieser Methoden angewandt werden sollte, so dient dies der therapeutischen klinischen Forschung.

Konkret spricht man nun immer dann von einem klinischen Versuch, wenn unter Berücksichtigung ethischer Aspekte ein **kontrolliertes Experiment** zum Vergleich der Wirkungsweisen mehrerer Behandlungen am Menschen durchgeführt wird. Der klinische Versuch ist somit ein in allen Einzelheiten von vornherein festgelegtes Experiment. Die **Behandlungen** sind eine Vielzahl prophylaktischer oder therapeutischer Maßnahmen, die gemäß den in einem Versuchsplan festgelegten Regeln am Patienten angewandt werden. In speziellen Fällen wird hier auch eine Nicht- oder Scheinbehandlung miteinbezogen.

Die Anzahl der Versuchspersonen spielt eine geringere Rolle als ein guter Versuchsplan, der es erlaubt, aus detaillierten Einzelbeobachtungen Aussagen über die durchschnittliche Wirkung der verschiedenen Behandlungen zu gewinnen. Zumeist wird es dabei vonnöten sein, durch Zusammenarbeit mehrerer Ärzte und Krankenhäuser für eine ausreichende Anzahl geeigneter Patienten zu sorgen, wobei noch einmal darauf hingewiesen sei, daß jede Versuchsperson prinzipiell jeder Behandlung unterzogen werden könnte.

Auch bei noch so sorgfältigen Einzelbeobachtungen werden sich die Therapien nicht vollständig durch klinische Versuche erforschen lassen, da die oft sehr ernsten Spätfolgen erst lange nach Ablauf der meisten Experimente feststellbar sind. Um zumindest im Rahmen des Möglichen weitgehend unverfälschte Ergebnisse, insbesondere auch bei Labor- und Röntgenuntersuchungen, zu erhalten, sollte die Möglichkeit des Blindversuches, vgl. Abschnitt 6.4, in seinen verschiedenen Ausprägungen erwogen werden.

Ein durchdachtes Vorgehen bei klinischen Versuchen ist in besonderem Maße dann erforderlich, wenn nicht nur Unterschiede zwischen den Behandlungen entdeckt, sondern auch Ursachen für diese Unterschiede erkannt werden sollen.

Über den Versuch und den Versuchsplan ist stets ein **Versuchs-Protokoll** anzufertigen, daß etwa wie folgt gegliedert werden kann:

1. Die Krankheit

Genaue Beschreibung der Stadien, die im Versuch berücksichtigt beziehungsweise nicht berücksichtigt werden sollen;

2. Die Versuchspersonen

Ein- und Ausschlußkriterien: Wer kommt als Versuchsperson in Frage, und wann ist jemand aus dem laufenden Versuch auszuschließen?

Welche Charakteristika der Versuchspersonen sind dem Protokoll beizufügen? Zum Beispiel: Alter, Geschlecht, Gewicht, soziales Umfeld usw.;

bereits erfolgte Behandlungen der bewußten Krankheit beziehungsweise Vorbeugungsmaßnahmen, Schwere der Erkrankung beziehungsweise Größe des Risikos, andere Krankheiten (auch in der Familie), mögliche Kontraindikationen;

3. Die Behandlungen

Wesentliche Charakteristika der verschiedenen Strategien; detaillierte Darstel-

lung jeder einzelnen Strategie mit Zeitplan, Dosierung, Kontroll- und Nachuntersuchungen usw.;

4. Der Versuch

Offener Versuch, einfacher oder doppelter Blindversuch?

Welcher Versuchsplan (im engeren Sinne) kommt zur Anwendung?

Wann und wie hat eine Randomisation zu erfolgen?

Welche Variablen werden beobachtet?

Wann und wie finden die dafür benötigten Messungen und Kontrollmessungen statt?

5. Anzahl der Versuchspersonen

Fest vorgegebener Umfang oder sequentielles Vorgehen?

6. Statistische Auswertung der Daten.

Bei jedem einzelnen Schritt der Planung, Durchführung und Datenanalyse des klinischen Versuches ist darauf zu achten, daß die erhaltenen Daten tatsächlich eine Antwort auf die Ausgangsfrage liefern und nicht die zu beobachtenden Behandlungseffekte durch Störeffekte überlagert werden.

Einige Probleme, die im Versuchsplan auftreten, vgl. auch Abschnitt 5, werden in diesem Abschnitt detailliert behandelt.

Daß alle erst den Daten entnommenen Auffälligkeiten einer genaueren Klärung in einem speziell darauf abgestimmten Folgeversuch bedürfen, sollte eigentlich keiner weiteren Erörterung bedürfen.

Unabhängig vom speziellen Versuchsplan und den anzuwendenden Behandlungen lassen sich die am behandelten Patienten vorzunehmenden Beobachtungen nach ihrer Zielsetzung klassifizieren (natürlich müssen im konkreten Experiment nicht alle Variablentypen vorkommen):

Zielvariablen (primary observations) sind von direktem Interesse für die Beurteilung der Behandlungen, wie etwa die Erythrozytenzahl;

Ersatzvariablen (substitute primary observations) werden verwendet, um nicht direkt meßbare, „subjektive“ Größen quantitativen Analysen zugänglich zu machen, ohne daß eine funktionale Beziehung zwischen der Ziel- und der Ersatzvariablen zu bestehen braucht. Zum Beispiel kommt man vielleicht zu einer brauchbaren Messung der individuellen Schmerzempfindung während eines bestimmten Stadiums der Behandlung, indem man einen Konzentrationstest durchführt;

Erklärende Variablen (explanatory observations) können eingeführt werden, um die zuvor definierten beiden Variablentypen näher zu beleuchten. Zum Beispiel können hier bei der Schmerzermittlung insbesondere Messungen der endokrinalen Sekretion durchgeführt werden, speziell die Produktion von Epinephrin der Nebenniere käme vielleicht in Betracht;

Begleitende Variablen (supplementary observations, concomitant variables) sorgen für eine Korrektur der erhobenen Daten durch explizite Berücksichtigung wichtiger Störgrößen. So liefert etwa das Körpergewicht wichtige Informationen über den Blutdruck. Begleitende Variablen können zur Blockbildung, vgl. Abschnitt 5 und 6.3, verwendet werden und so die Präzision des Versuches erhöhen, gegebenenfalls auch gewisse Wechselwirkungen erkennen lassen.

Kontrollvariablen (observations for checking) sollen bei der Überprüfung der Korrektheit des Einhaltens der Grundvoraussetzungen des Versuchsplans helfen. So

kann etwa mit Hilfe der Untersuchungszeiten überprüft werden, ob die hier in der Regel unvermeidlichen Unterschiede zu einem systematischen Fehler geführt haben. Wesentliche Methoden zur Analyse und Auswertung (von Versuchen mit mehreren Einflußfaktoren) liefern die **Regressions- und Varianzanalyse**, vgl. Kap. X und XI und bei qualitativen Daten die Kontingenztafelanalyse, vgl. Kap. VII.

6.1. Ethische Probleme bei klinischen Versuchen

Die meisten Ärzte und auch Patienten geben die Notwendigkeit zu, neue Behandlungen zu entwickeln. Sicher steht es auch außer Frage, daß die nun einmal benötigten Versuchspersonen keinem unnötigen Risiko und zusätzlichem Schmerz ausgesetzt werden sollten.

Wie diese Risiken für einige Menschen gegen mögliche Vorteile für viele Menschen gegeneinander ausgewogen werden können, ist Gegenstand dieses Abschnittes. Hier sollen vornehmlich statistische Aspekte solcher ethischen Fragen angedeutet werden.

Wenn ein Arzt eine Behandlung für besser erkennt, aus welchen mehr oder minder subjektiven Beweggründen auch immer, wird er sie bevorzugt oder gar ausschließlich anwenden wollen. Das Ansinnen, streng zufällig einigen seiner Patienten eine als schlechter empfundene Behandlung zuzumuten, mag ihm nicht mit seinem Berufsethos vereinbar erscheinen.

Wenn ein anderer (von ihm respektierter!) Arzt eine andere Behandlung besser findet, mag der Arzt schon eher einem kontrollierten Experiment beipflichten.

Dabei steht es natürlich außer Frage, daß man bei der Durchführung eines Versuches auch stets das beste anwendbare Modell zu wählen hat, wobei dann wieder versuchsplanerische Vorteile mit zusätzlichen Risiken für die Probanden verbunden sein könnten.

Die ethischen Probleme werden durch sehr praktische Schwierigkeiten noch verstärkt. So findet es weithin Beifall, daß den Patienten mitgeteilt werden soll, um welche Behandlung es sich im einzelnen handelt – was nicht impliziert, daß kein Blindversuch mehr durchgeführt werden könnte. Allerdings wird man wohl kaum sehr viele Patienten finden können, denen es völlig gleichgültig ist, ob sie zum Beispiel einer Operation oder einer Chemotherapie unterzogen werden. Ebenso wird man schlecht sagen können, daß man gegebenenfalls mit einem Placebo impfen wird.

Auch wenn sich für den letzten angesprochenen Fall genügend viele Probanden fänden, kann die Autosuggestion, vgl. Abschnitt 6.4, weiteres Unheil anrichten: Sie kann hier dann sogar dazu führen, daß ein Wirkmittel als Placebo empfunden wird und psychosomatische Störungen tatsächlich den Wirkstoff blockieren.

Bei sehr schweren und bei schmerzhaften Erkrankungen ist es besonders problematisch, als hochwirksam erachtete Behandlungen vorzuhalten.

Gab es vorher überhaupt keine brauchbare Behandlung, so ist das Problem noch lösbar, da man einfach mittels einer Gruppe kontrolliert, ob die Behandlung Verbesserungen bewirkt.

Völlig neue Behandlungen (zumal bei Medikamenten) sind sehr selten, meist handelt es sich um geringfügige Modifikationen und Kombinationen gut eingeführter Mittel, so daß man in der Regel davon ausgehen kann, daß das neue Mittel wohl nicht wesentlich schlechter als die Stammform ist.

6.2. Auswahl und Zuordnung von Versuchspersonen

Natürlich kommen in einem klinischen Experiment nur solche Menschen als Versuchspersonen in Betracht, die von der interessierenden Krankheit befallen beziehungsweise bedroht sind.

Geht es im Versuch nun vor allem darum, den „Wert“ einer Behandlung festzustellen, so wird man zudem insbesondere daran interessiert sein, daß Art und Schwere der Erkrankung für alle Probanden in etwa gleich sind; steht dagegen die Entscheidung, welche Behandlung in Zukunft zu empfehlen ist, im Vordergrund, so muß die **Repräsentativität** der Versuchspersonen gewahrt werden.

Um eine streng zufällige Zuordnung der Probanden zu den Behandlungen vornehmen zu können, muß bei der Auswahl eines Probanden für den Versuch sichergestellt sein, daß keine **Kontraindikationen** für jede der interessierenden Behandlungen bei ihm vorhanden sind. Oft ist es jedoch so, daß Kontraindikationen erst während der Behandlung sichtbar werden; daher sollte jeder Versuchsplan feste Regeln für die Handhabung solcher Fälle beinhalten, und zudem sollte die Auswahl der Probanden natürlich sehr sorgfältig erfolgen. Dazu sollte der über die Aufnahme in das Experiment entscheidende Arzt nicht nur medizinische Gesichtspunkte berücksichtigen, sondern gleich die verschiedenen Behandlungsstrategien den aus medizinisch-versuchsplanerischer Sicht akzeptablen Probanden verständlich darlegen, so daß jene schon gleich ausgeschlossen werden können, die gegen eine der Methoden voreingenommen sind.

Das Problem zu spät entdeckter Kontraindikationen läßt sich zumindest teilweise beseitigen, wenn man bereits bei der Voruntersuchung alle Bewerber ablehnt, bei denen auch nur ein Verdacht auf mögliche Komplikationen besteht. Die ausgewählten Versuchspersonen können dann (zumindest formal) stets nach Maßgabe der ihnen zuteil werdenden Behandlungen verschiedenen Gruppen zugewiesen werden.

Im Interesse der **Verallgemeinerbarkeit** der Ergebnisse wird man jede einzelne Gruppe möglichst **heterogen** wünschen und zur Gewährleistung der **Vergleichbarkeit** darauf achten, daß die Gruppen untereinander möglichst **homogen** sind.

Homogenität der Gruppen untereinander impliziert, daß zu Beginn des Versuches alle Gruppen möglichst gleich strukturiert sind. So ist darauf zu achten, daß in allen als relevant erachteten Aspekten weitgehende Übereinstimmung besteht, wobei im einzelnen etwa an die Altersstruktur, das soziale Umfeld, den allgemeinen Gesundheitszustand und die Schwere der Erkrankung beziehungsweise die Risikoträchtigkeit zu denken ist. Insbesondere ist auch auf eine Ausgewogenheit der Merkmalskombinationen zu achten, damit nicht zum Beispiel in der einen Gruppe alle Frauen jung und alle Männer alt sind, in einer anderen Gruppe hingegen der umgekehrte Fall vorliegt.

Daß Strukturgleichheit in der Tat erzielt worden ist, sollte vor Beginn des Versuches mit geeigneten Methoden verifiziert werden.

Allerdings hat man darüber hinaus zu berücksichtigen, daß hier Behandlungen als komplexe Strategien verstanden werden, welche alle Einzelheiten für die gesamte Dauer des Versuches festlegen. Unterschiede in den Behandlungseffekten möchte man jedoch auf die wesentlichen Besonderheiten der betreffenden Behandlungen zurückführen können. Aus diesem Grund tut es not, alle anderen Details, soweit irgend möglich, für alle Behandlungen und alle Versuchspersonen konstant zu halten.

Wenn dies aus technischen Gründen nicht möglich ist, wird man (schon bei der Versuchsplanung) durch Randomisation diesen Mangel auszugleichen trachten, so daß zum Beispiel nicht die Visiten oder Laboruntersuchungen stets in der gleichen Reihenfolge ablaufen.

Nicht weniger wichtig ist es jedoch, den am Versuch beteiligten Ärzten und Probanden unabhängig von der im einzelnen angewandten Behandlung die Motivation zu vermitteln, nicht vom Versuchsplan abzuweichen. Auf dieses spezielle Problem werden wir später noch näher eingehen, vgl. Abschnitt 6.4.

6.2.1. Die retrospektive Zuordnung

Daß prinzipiell die für die verschiedenen Therapien beziehungsweise Prophylaxen zu bildenden Gruppen untereinander möglichst homogen sein sollten, ist schon seit einiger Zeit bekannt und auch weithin anerkannt. Die konsequente Verwirklichung dieses Grundsatzes hingegen wurde und wird in praxi oft einer bequemeren Art der Zuteilung der Versuchspersonen auf die einzelnen Gruppen geopfert.

Dies ist im besonderen Maße dann der Fall, wenn zum Beispiel aus Kostengründen weitgehend auf vorliegendes Material zurückgegriffen werden soll und nur, wenn etwa die Zahl der Beobachtungen insbesondere für die neuen Behandlungen nicht ausreicht, Zusatzuntersuchungen eigens angesetzt werden.

In der täglichen Praxis werden den Patienten mit einem bestimmten Leiden oder gegen ein bestimmtes Leiden zwar nicht nach einem vorgegebenen (zufälligen) Schema, aber doch, in Abhängigkeit von der jeweiligen Entscheidung des behandelnden Arztes, jeweils andere Behandlungen zugewiesen, so daß sich durch Sammeln der Einzeldaten ein Überblick über den Erfolg der Therapie oder der Prophylaxe gewinnen läßt (**retrospektive Zuordnung**).

Varianten dieses Verfahrens ziehen in Betracht, daß unterschiedliche Kliniken und Arztpraxen durch ihre differenzierte Ausstattung jeweils andere Behandlungen bevorzugt anwenden, so daß man auch äußerlich eine „natürliche“ Gruppeneinteilung vorgegeben hat. Bei einem solchen Vorgehen wird aber übersehen, daß Ärzte die Behandlungen nach dem status praesens des Patienten festlegen, so daß zum Beispiel als besonders riskant empfundene Therapien mitunter entweder nur bei sehr leichten oder aber sehr schweren Stadien der Krankheit angewandt werden. Vielleicht wird auch eine kostspielige Behandlung nur an finanziell kräftigen Privatpatienten (mit entsprechend günstigen allgemeinen Lebensbedingungen) vorgenommen, da die Therapie von den Krankenkassen noch nicht anerkannt und somit auch nicht bezahlt wird.

Solchen Problemen will man bisweilen dadurch begegnen, daß alle Patienten zusätzlich noch von einigen wenigen Spezialisten untersucht werden. Da diese Experten ihre Diagnose auf Einzelheiten aufbauen, welche nur zum Teil protokolliert (und auch protokollierbar) sind, dürfte eine Nachprüfbarkeit ihrer Befunde in der Regel nicht in ausreichendem Maße gegeben sein.

Bei der „historischen Kontrolle“ werden hingegen für die bisher üblichen Verfahren bereits abgeschlossene Krankenberichte beziehungsweise epidemiologische Studien herangezogen, die dann durch Fachleute ausgewertet werden. Da sich hier die Analyse durch die Experten vorwiegend auf schriftlich vorliegendes Datenmaterial stützt, ist eine leichtere Nachprüfbarkeit der Befunde gegeben. Allerdings hat sich bei derartigen Verfahren in praxi oftmals gezeigt, daß dem Gebot der Homogenität zwischen den Gruppen bei weitem nicht in ausreichendem Maße genüge getan

worden ist, da die hier unvermeidlichen großen Unterschiede in Zeit und Raum in der Regel zu starken Verzerrungen der Behandlungseffekte führen.

Beispiel: Das erratische Verhalten infektiöser Erkrankungen ist wohlbekannt, aber auch bei anderen Krankheiten lassen sich regionale (etwa durch die Auswirkungen eines Chemiewerkes auf seine Umgebung) und auch temporale (zum Beispiel durch verbesserte Diagnose- und Früherkennungsmethoden) Veränderungen feststellen, so daß die verschiedenen Behandlungen bei unterschiedlichen Stadien der Erkrankung angewandt werden, wodurch die Vergleichbarkeit gefährdet wird.

Bereits durch den Umstand, daß für die neuen Behandlungen Versuchspersonen gesucht werden, treten Verzerrungen auf, da durch das Bekanntwerden des Versuches das Bewußtsein der von ihm Betroffenen wächst und schon leichtere Fälle sich um eine Aufnahme in das Experiment bemühen, die sich sonst überhaupt noch keiner Behandlung unterzogen hätten.

Dieser letzte Effekt läßt sich an einem **Beispiel** demonstrieren: Bei einem kontrollierten Experiment von R. Paterson und M. H. Russel (vgl. Fac. Radiol. 10, 1959) wurde postoperativ einer Gruppe von an Brustkrebs Erkrankten systematisch eine Radiotherapie verabfolgt, der anderen Gruppe blieb es freigestellt, an der Radiotherapie teilzunehmen oder nicht. Es zeigte sich während der sechs Jahre der Beobachtung, daß die Sterblichkeitsraten sich für beide Gruppen fast nicht unterschieden, hingegen mit der Zeit deutlich sanken. Wäre die eine Behandlung in den ersten drei, die andere in den letzten drei Jahren erfolgt, so hätte sich ein Unterschied in den Behandlungseffekten leicht „beweisen“ lassen.

6.2.2. Zuordnung auf freiwilliger Basis

Bei der retrospektiven Zuordnung, vgl. Abschnitt 6.2.1, erweist es sich als äußerst schwierig, für die Vergleichbarkeit der Struktur der Gruppen zu sorgen; daß sich die durchgeführten Behandlungen in der Tat nur in den eigentlich interessierenden Punkten unterscheiden, dürfte gleichfalls die Ausnahme sein.

Letzterem Problem kann man durch ein kontrolliertes Experiment begegnen, in dem man die Versuchspersonen vor die Wahl stellt, einer neuen oder einer Standardbehandlung unterzogen zu werden. Eine Strukturgleichheit allerdings wird sich auch hier nur im Ausnahmefall ergeben.

Beispiel: Bei einer der ersten Schutzimpfungen gegen Keuchhusten, für die nach diesem Verfahren kategorisiert worden ist, erwiesen sich 47% der geimpften, aber nur 20% der nicht geimpften Kinder als Einzelkinder, so daß sich die geimpfte Gruppe alleine durch die geringere Ansteckungsgefahr vorteilhaft von der Kontrollgruppe abheben mußte.

Ganz allgemein gibt es (in direkter Abhängigkeit von der betreffenden Krankheit) eine ganze Reihe medizinischer, psychologischer und soziologischer Gründe, die dazu führen, daß bei einem gegebenen Versuch die Freiwilligen überwiegend einem mehr oder minder eng umgrenzten Personenkreis angehören. Speziell das Bewußtsein, persönlich besonders gefährdet zu sein, spielt dabei eine entscheidende Rolle.

Lediglich in einem sehr frühen Stadium der Untersuchung einer Behandlung wird man auf Freiwillige zurückgreifen dürfen (die man dann aber streng zufällig den verschiedenen Gruppen zuzuteilen hat), um erste Ergebnisse zu erhalten, auch wenn deren externe Validität mehr als zweifelhaft ist.

6.2.3. Pseudaleatorische und aleatorische Zuordnung

Beim **alternierenden (pseudaleatorischen) Verfahren** werden die Versuchspersonen in der Reihenfolge, in der sie für den Versuch zugelassen werden, in zyklischer Ordnung den verschiedenen Behandlungen zugewiesen. Hierbei besteht jedoch die Möglichkeit, daß der die Vorauswahl treffende Arzt die Zuordnung bewußt oder auch unbewußt verfälscht – vielleicht sogar im Bemühen, einer vermeintlichen Verzerrung entgegen zu wirken.

Bei der **aleatorischen Zuordnungsvorschrift** wird die Reihenfolge, in der die in den Versuch aufgenommenen Probanden den verschiedenen Behandlungen zugewiesen werden, im vorab rein zufällig festgelegt und in durchnumerierte, verschlossene Umschlägen dem die Vorauswahl treffenden Arzt überreicht. Dieser übergibt dann dem k-ten Probanden den k-ten Umschlag. – Dabei kann das Verfahren natürlich problemlos auf mehrere Schichten bzw. Blöcke ausgeweitet werden, indem man einfach für jede der Schichten, Blöcke eine getrennte Numerierung der Umschläge vornimmt.

6.2.4. Einige weitere Zuordnungsverfahren

In bestimmten Fällen lassen sich an jeder einzelnen Versuchsperson nacheinander alle Behandlungen erproben, wobei jeweils die Reihenfolge der Verfahren streng zufällig festzulegen ist, um über die Effekte der gewählten Reihenfolge zu randomisieren. Dieses Zuordnungsverfahren (**cross-over trial** oder **within patient trial** genannt) bleibt den Krankheiten oder Erkrankungsrisiken vorbehalten, die zumindest während der Dauer des Versuches ein konstantes Niveau halten und die nach jeder einzelnen Behandlung „schnell“ wieder auf diesen status quo ante zurückkehren, wie dies etwa bei rheumatoider Arthritis der Fall ist. Fiebererkrankungen wie zum Beispiel Malaria sind hierfür sehr viel weniger geeignet, da bei ihnen rasch entweder der Genesungsprozeß beginnt oder der Exitus eintritt.

Zwei andere Zuordnungsverfahren sind als **Namensstichprobe** beziehungsweise als **Geburtstagsstichprobe** bekannt. Bei ihnen wird nach dem Anfangsbuchstaben des Nachnamens beziehungsweise nach der Jahreszahl (oder etwa dem Geburtsmonat) die Kategorisierung vorgenommen. Da die Verteilung der Anfangsbuchstaben des Nachnamens für Angehörige verschiedener Völkerstaaten stark schwankt, ist für Namensstichproben bei gemischten Populationen mit Verzerrungen zu rechnen, die allerdings auch bei Geburtstagsstichproben auftreten und die nur durch kompliziertere Verfahren ausgeglichen werden können, vgl. auch Abschnitt 3 dieses Kapitels und Schach/Schach (1978, 1979).

Insgesamt ist zu sagen, daß es, wenn immer die Möglichkeit besteht, ratsam ist, von vornherein eine streng zufällige Zuordnung der Probanden zu den interessierenden Behandlungen vorzunehmen.

6.3. Die Vergleichbarkeit der Versuchsergebnisse

In den 30er Jahren stellt Sir Ronald Aylmer Fisher drei Forderungen auf, deren Erfüllsein erst einen sinnvollen Vergleich verschiedener Behandlungen möglich machen :

Die **Wiederholbarkeit (replication)**, die **Mischung bzw. Randomisation (randomization)**, die **Blockbildung bzw. Schichtung (blocking, planned grouping)**, vgl. auch Abschnitt 5.

Beispiel: Gehen wir einmal davon aus, daß zwei Menschen gleich alt, gleich groß, gleich schwer usw. sind, dem gleichen Geschlecht angehören, unter der gleichen Krankheit leiden und gleiche Krankheitssymptome zeigen, so wäre zu erwarten, daß eine gleiche Behandlung der beiden Personen zu gleichen Ergebnissen führt. In der Praxis ist dies jedoch nicht der Fall, denn so ähnlich sich zwei Menschen auch sind, wird es immer Unterschiede (z. B. genetischer Art) zwischen ihnen geben. Diese Unterschiede (Versuchsfehler) lassen sich statistisch nur dann erfassen, wenn eine Therapie auf mehrere möglichst homogene Schichten (Personen) angewandt werden kann, d.h. wenn Wiederholungen möglich sind.

Randomisieren, d.h. Zuordnung der Behandlungen – gegebenenfalls innerhalb der Blöcke – mittels Zufallsmechanismen wie z. B. Würfeln oder Zufallszahlen aus Rechnern oder Tafeln (s. Anhang), soll gegen systematische Verzerrungen schützen, die sich wegen der Vielzahl der (teilweise gänzlich unbekannten) Einflußfaktoren durch kein noch so ausgeklügeltes systematisches Vorgehen vermeiden lassen. Auch durch die Randomisation gleichen sich natürlich nicht alle Störgrößen völlig aus – lediglich bei vielfacher Wiederholung des Versuches kann man auf ein „Aus-mitteln“ hoffen – aber die zu verwendenden Versuchspläne (im engeren Sinne) wie etwa die Blockpläne berücksichtigen diesen Umstand teilweise.

Schichtung bzw. Blockbildung wird wie die Randomisation durchgeführt, um die durch Störgrößen verursachte Variation der Versuchsergebnisse zu verkleinern, wobei allerdings nunmehr die Störgröße nicht „vergessen“, sondern als begleitende Variable zur Blockbildung herangezogen wird.

6.4. Auto- und Heterosuggestion, Blindversuche

Daß das Vertrauen in eine Behandlung bisweilen schon für einen Behandlungserfolg ausreicht, ist wohlbekannt. Etwa bei 35% der Betroffenen ließ sich so eine Heilung erreichen bei Angina, Kopfschmerzen, Seekrankheit, Übelkeit, Husten, Erkältung, postoperativen Schmerzen und anderem mehr.

Über ein besonders schönes **Beispiel** wird von Wolf (Proc. Ass. Res. nerv. ment. Dis., 37, 1959) berichtet: Ein Patient, der seit Jahren an Asthma gelitten hat, konnte bislang durch kein Medikament Linderung erfahren. Eine vielfach gepriesene neue Droge wurde ihm von seinem Arzt verschrieben, woraufhin sofort Erleichterung auftrat. Nach mehreren solchen Erfolgen erprobte der Arzt ein **Placebo (Scheinmedikament)**, worauf der Patient aber nicht ansprach. Nach mehrfacher Wiederholung dieses kontrollierten Experimentes teilte der begeisterte Arzt der Herstellerfirma das Ergebnis mit – und bekam zur Antwort, daß er stets nur mit einem Placebo beliefert worden sei.

Auf der anderen Seite können Placebos aber nicht nur Heilerfolge erzielen, sondern auch negative Effekte wie z. B. Schweißausbrüche oder Nesselfieber hervorrufen.

Bis heute ist es noch nicht gelungen, für diese **Placebo-Effekte** besonders anfällige Personengruppen zu entdecken; keine der äußerlichen Merkmalskombinationen erwies sich als brauchbarer Indikator, und auch psychologische Untersuchungen liefern keine eindeutige Antwort.

Es ist etwa umstritten, was der Placebo-Effekt nun eigentlich zu messen erlaubt: Ist es die Wirkung des Umstandes, überhaupt behandelt (und in einen Versuch mit einbezogen) zu sein oder ist es der natürliche Verlauf der Dinge (ohne jede Behandlung)?

Beim einfachen **Blindversuch** weiß der Proband nicht, welcher Behandlung er unterzogen wird, so daß die Möglichkeit der Autosuggestion entfällt. Aus diesem Grund wird er glaubwürdigere Aussagen über subjektive Empfindungen („Schmerzen“, „Verbesserung“) machen können; gleichzeitig werden aber auch die objektiven Meßwerte zuverlässiger, da psychosomatische Störungen durch den Glauben oder Nicht-Glauben an die Wirkung weniger häufig auftreten.

Ferner kann sich die Zahl der den Versuch aus eigenem Antrieb – ohne äußeren Anlaß – abbrechenden Versuchspersonen auf diese Art vermindern, falls das Auscheiden durch die eigene Einschätzung der Wirkung der Behandlung und des persönlichen Nutzens motiviert ist.

Um nun auch die Heterosuggestion zu beseitigen, wird im **Doppel-Blindversuch** auch dem betreuenden Arzt nicht mitgeteilt, welche Medikation er vornimmt. Durch den Doppel-Blindversuch prägen sich die angesprochenen Aspekte für den Probanden noch stärker aus, und analoge Wirkungen lassen sich auch für den Arzt feststellen. Letzterer wird, falls er Probanden verschiedener Gruppen betreut und keine der vertrauten Behandlungen deutliche Nebenwirkungen aufweist, nicht so leicht während des Versuchs zu einer spekulativen Beurteilung der relativen Qualität der einzelnen Behandlungen kommen, so daß eine weitere Quelle der Verfälschung beseitigt wird.

Bei pragmatischer Betrachtung des Problems kann man den Placebo-Effekt ohne weiteres auch positiv bewerten: Wenn auf diese preiswerte Art und ohne das Risiko schädlicher Nebenwirkungen und Spätfolgen eine Verbesserung des Gesundheitszustandes des Patienten erzielt werden kann, so ist dies doch erfreulich.

Auf der anderen Seite ist die Anwendung des Blindversuches auf bestimmte Situationen beschränkt, da man zum Beispiel kaum Scheinoperationen durchführen wird. (Von einem Doppel-Blindversuch wird man hier wohl aus technischen Gründen ganz absehen müssen.) Allerdings sind Blindversuche ethisch problematisch, wenn man dabei etwa vor die Entscheidung gestellt wird, Menschen erst für so gefährdet zu erklären, daß sie sich der Behandlung unterziehen sollten, und dann nur eine Scheinbehandlung durchzuführen. Psychosomatische Komplikationen können hier sehr leicht ein Abbrechen des Blindversuches erforderlich machen.

Gleichfalls wird man den Versuch nicht in der wünschenswerten Form fortsetzen können, wenn sich wider Erwarten bei dem wirksamen Präparat zum Beispiel toxische Effekte zeigen.

Diese Möglichkeiten sollten, falls ein Blindversuch geplant ist, bei den Ein- und Ausschlußkriterien für die Versuchspersonen berücksichtigt werden.

6.5. Sequentielle Studien

Bei Tierversuchen im Laboratorium bereitet es in der Regel keinerlei Schwierigkeiten, zu Beginn einer Studie ausreichend viele Versuchstiere zu erhalten und den Versuch dann mit diesen Tieren durchzuführen.

Bei Menschen erweist es sich bisweilen als sehr viel schwieriger, ausreichend viele Versuchspersonen zu finden, die allen Zugangskriterien genügen und die man dann auch nicht beliebig lange warten lassen kann, ehe man mit der Behandlung beginnt. Aus diesem Grund wird man um zeitlich versetzte Startzeiten für den Versuch in vielen Fällen nicht herumkommen. Nun könnte man natürlich dann so lange weiterforschen, bis man eine fest vorgegebene Anzahl von Probanden aufgenommen

hat, aber es liegt doch nahe, nicht zuletzt aus Kosten- und Zeitgründen, aufzuhören, sobald das Ergebnis „klar ist“.

Die statistischen Verfahren für dieses Vorgehen sind unter dem Begriff der **sequentiellen Analyse (sequential analysis)** bekannt. Die sequentielle Analyse geht weitgehend auf Dodge/Romig (1929) sowie Wald (1945, 1947) zurück. In Kap. IV, Abschnitt 6, und Kap. VI werden solche sequentiellen Verfahren bei bestimmten Verteilungsannahmen für die interessierende Grundgesamtheit vorgestellt.

Die Vorteile der sequentiellen Analyse sind zum einen der im Mittel kleinere Stichprobenumfang zur Absicherung vorgegebener Fehlergrenzen und zum anderen der deutlich sichtbare Versuchsverlauf, der sich sicherlich motivierend auf die am Versuch beteiligten Ärzte auswirkt.

Diese Vorteile können allerdings nur dann voll ausgenutzt werden, wenn die Ergebnisse auch wirklich „schnell“ greifbar sind, so daß langwierige Behandlungen nicht in Betracht kommen. In praxi wird man den Schwierigkeiten, die durch den zeitlich versetzten Beginn der Untersuchungen im Kleinen gegeben sind, dadurch begegnen, daß man eine Blockbildung mit der begleitenden Variablen „Start der Behandlung“ durchführt.

Ein Nachteil der sequentiellen Analyse besteht darin, daß sie es nicht gestattet, Untergruppen zu berücksichtigen, und auch darin, daß nur zwei Behandlungen mit einer einzigen Variablen berücksichtigt werden können. Ein weiterer Nachteil sequentieller Versuche ist es, daß, wenn wirklich die behandelnden Ärzte ständigen Überblick über den Verlauf der Untersuchung haben, wieder Auto- und Heterosuggestion für eine Verzerrung des Ergebnisses sorgen können. Auch hier tritt die Frage auf, ob man eine schon „nahezu“ als schlecht erwiesene Behandlung weiter verabreichen darf.

6.6. Ein Beispiel

Je nach Problematik und Ziel einer klinischen Studie, müssen recht unterschiedliche statistische Verfahren zur Auswertung des Versuches herangezogen werden. Daher kann hier kein allgemeingültiges Auswertungskonzept vorgestellt werden und das Vorgehen soll lediglich an einem Beispiel exemplarisch vorgestellt werden.

Die „National Foundation against Infantile Paralysis“ hat im Jahre 1954 einen Großversuch durchgeführt, um die Wirkung des Salk-Vakzines bei „Poliomyelitis anterior acuta“ zu überprüfen, vgl. Schwartz et al. (1980).

In den USA betrug die Erkrankungsrate langfristig 0,35% pro Jahr, mit starken Schwankungen von einem Jahr zum anderen. In Voruntersuchungen wurden einhundert Gebiete ausgewählt, in denen die Rate sogar 0,5% betrug und zudem ein großer Anteil sechz- bis neunjähriger Kinder und geeignete medizinische Einrichtungen zu finden waren. In den Versuch wurden Schulkinder der Klassen 1 bis 3 im Alter von 6 bis 8 Jahren aufgenommen; die Eltern der in Frage kommenden Kinder wurden um schriftliches Einverständnis gebeten.

Die eine (streng zufällig ausgewählte) Hälfte der Kinder wurde mit den abgetöteten Viren nach Jonas Edward Salk geimpft, die anderen mit einem Placebo (einer Salzlösung). Der Impfstoff ließ sich nur durch eine Codenummer identifizieren, so daß ein Doppel-Blindversuch zur Anwendung kam.

Alle Kinder der angegebenen Grundgesamtheit, bei denen Poliomyelitis festge-

stellt wurde, ob in den Versuch aufgenommen oder nicht, wurden genau notiert und untersucht.

An einer Stichprobe von 2% wurde vor und nach der Impfung eine serologische Untersuchung angestellt, um die Veränderung des Bestandes an Antikörpern festzustellen.

749 236 Kinder aus 84 Gebieten in elf Staaten der USA erwiesen sich bei der Vorauswahl als geeignet, von denen 401974 Kinder tatsächlich in den Versuch aufgenommen werden konnten. Bei den restlichen fehlte meistens die Genehmigung der Eltern für die Einbeziehung. In den fünfzehn Monaten des Versuchs ergab sich für die mit Salk-Vakzine geimpften Kinder eine Erkrankungsrate von 0,28% und für die Kontrollgruppe eine von 0,70%. Für den entsprechenden Versuch in der Bundesrepublik Deutschland sei verwiesen auf S. Koller (1954, Medizinische Wochenschrift, vol. 62, p. 273–276 und 307–310).

Will man nun untersuchen, ob die (theoretische) Erkrankungsrate bei geimpften p_1 und mit Placebo behandelten Kindern p_2 gleich ist, d.h. ob die im Versuch erhaltenen empirischen Erkrankungsraten nur zufällig verschieden sind, oder ob sie als signifikant verschieden angesehen werden können, so wird man von einer Binomialverteilung $B(1, p_1)$ bzw. $B(1, p_2)$ der Erkrankung eines Kindes ausgehen und die Hypothese

$$H_0: p_1 = p_2$$

testen, vgl. auch Kap. VII, Abschnitt 1.3; dort ist auch die Bestimmung notwendiger Stichprobenumfänge zur Absicherung eines Fehlers 1. Art α und 2. Art β vorgeführt worden.

Bei vielen klinischen Versuchen – insbesondere bei Blockbildung bzw. Berücksichtigung mehrerer Einflußfaktoren – erfolgt die Anlage und Auswertung nach Modellen und Methoden der Varianzanalyse, vgl. Abschnitt 5 dieses Kapitels und Kapitel XI, oder bei qualitativen Daten mit Methoden der Kontingenztafelanalyse, die ausführlich in Kap. VII behandelt werden.

Allgemein läßt sich sagen, daß nahezu alle statistischen Verfahren eine ihrer wesentlichen Anwendungsmöglichkeiten in diesem Bereich haben, was zum Teil durch unsere Auswahl der Beispiele zu den jeweiligen Verfahren verdeutlicht wird.

7. Skalierung von Merkmalsausprägungen und Testergebnissen

Beobachtet man an n Objekten aus einer Grundgesamtheit verschiedene Merkmale, so möchte man in der Statistik stets deren Ausprägungen untersuchen, auf die Merkmalsausprägungen in der Grundgesamtheit zurückzuschließen und Zusammenhänge zwischen den Merkmalen aufzudecken.

Sind die Ausprägungen der beobachteten Merkmale **metrisch** skaliert, d.h. sind die Abstände zwischen den Ausprägungen interpretierbar (z.B. Körpergröße, Stromverbrauch in Kilowattstunden, Entfernung, Stromstärken), so sind die in diesem Rahmen vorgestellten Methoden wie z.B. die **Regressionsanalyse**, die **Varianzanalyse** oder die **Korrelationsrechnung** ohne Schwierigkeiten direkt anwendbar. Diese Verfahren setzen zwar fast alle normalverteilte Merkmale voraus, was sich jedoch mittels Datentransformation, vgl. Abschnitt 4.5, oft leicht erreichen läßt. Auch die multivariaten Verfahren, wie z.B. **multivariate Regressions- und Varianzanalyse**, **Diskriminanzanalyse**, **Clusteranalyse** lassen sich zur gemeinsamen

Analyse aller erhobenen Merkmale verwenden. Ist die Zahl der interessierenden Merkmale sehr groß, so kann man auch wenige Faktoren, die sie beschreiben, mittels **Faktorenanalyse** extrahieren.

Bei **ordinalen Merkmalsausprägungen**, d. h. Ausprägungen, die einer Reihenfolge unterliegen, deren Abstände aber nicht interpretierbar sind (Punktzahlen, Noten usw.), lassen sich die vorgestellten Verfahren für **Rangzahlen** oder auch die Methoden der **Kontingenztafelanalyse** (vgl. Kap. VII) anwenden.

Auch ist es möglich, ordinale Merkmale mit endlich vielen möglichen Ausprägungen zu quantifizieren. Zwei Verfahren, die dies möglich machen, sollen hier vorgestellt werden.

Bei der **marginalen Normalisierung** nach Fechner (1860) werden die Ausprägungen eines ordinalen Merkmals aufgrund einer Stichprobe von n Objekten, an denen das Merkmal beobachtet wird, so skaliert, daß die skalierten Ausprägungen möglichst gut an eine Standardnormalverteilung angepaßt sind. Das Verfahren wird im folgenden kurz beschrieben.

Wir bezeichnen die ordinalen Merkmalsausprägungen mit $1, \dots, k$ und zählen, wieviele der n beobachteten Objekte die i -te Ausprägung ($i = 1, \dots, k$) tragen. Diese Anzahlen seien n_1, \dots, n_k .

Die relative Häufigkeit der i -ten Ausprägung ist dann

$$p_i = \frac{n_i}{n} \quad \text{für } i = 1, \dots, k.$$

Nun werden die Summenhäufigkeiten

$$\alpha_j = \sum_{i=1}^j p_i \quad \text{für } j = 1, \dots, k-1,$$

sowie die zugehörigen Quantile

$$u_{\alpha_j} = \Phi^{-1}(\alpha_j) \quad \text{für } j = 1, \dots, k-1$$

der Standardnormalverteilung bestimmt, vgl. Tabelle im Anhang. Mit

$$z_j = \varphi(u_{\alpha_j}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} u_{\alpha_j}^2} \quad \text{für } j = 1, \dots, k-1$$

ergeben sich die gesuchten Skalierungspunkte für die Ausprägungen $1, \dots, k$ zu

$$\begin{aligned} s_1 &= -z_1/p_1, & s_k &= z_{k-1}/p_k \quad \text{und} \\ s_j &= (z_{j-1} - z_j)/p_j \quad \text{für } j = 2, \dots, k-1. \end{aligned}$$

Beispiel: Bei $n = 136$ Schülern wird die Note in Physik ermittelt. $n_1 = 11$ Schüler haben eine 1, $n_2 = 26$ eine 2, $n_3 = 48$ eine 3, $n_4 = 41$ eine 4 und $n_5 = 10$ Schüler eine 5. Diese Noten (Ausprägungen des Merkmals Physiknote) sollen nun mittels marginaler Normalisierung skaliert werden. Es ist

$$\begin{aligned} p_1 &= \frac{11}{136} = 0,0809, & p_2 &= 0,1912, & p_3 &= 0,3529, \\ p_4 &= 0,3015 \quad \text{und} & p_5 &= 0,0735, \end{aligned}$$

so daß sich mit

$$\begin{aligned} \alpha_1 &= p_1 = 0,0809, & \alpha_2 &= p_1 + p_2 = 0,2721, & \alpha_3 &= 0,6250 \quad \text{und} \\ \alpha_4 &= 0,9265 \end{aligned}$$

gerade

$$u_{\alpha_1} = u_{0,0809} = \Phi^{-1}(0,0809) = -1,40, \quad u_{\alpha_2} = -0,61, \\ u_{\alpha_3} = 0,32 \quad \text{und} \quad u_{\alpha_4} = 1,45$$

ergibt. Somit ist

$$z_1 = \varphi(u_{\alpha_1}) = \varphi(-1,40) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot 1,40^2} = 0,150, \quad z_2 = 0,331, \\ z_3 = 0,379 \quad \text{und} \quad z_4 = 0,139,$$

d.h. die gesuchten Skalierungspunkte sind

- für die Note 1: $s_1 = -z_1/p_1 = -1,854$,
 für die Note 2: $s_2 = (z_1 - z_2)/p_2 = -0,947$,
 für die Note 3: $s_3 = -0,136$,
 für die Note 4: $s_4 = 0,796$ und
 für die Note 5: $s_5 = z_4/p_5 = 1,891$.

Ein häufig verwandtes Verfahren bei der Skalierung (Konstruktion) von Noten, Bewertungen, „Intelligenzquotienten“ usw. ist das **Prozentrangverfahren**, vgl. Lienert (1969), Gutjahr (1971). n Personen werden einem Test bestehend aus verschiedenen Fragen und Aufgaben unterzogen. Jede Antwort wird mit einer Punktzahl bewertet, die Gesamtpunktzahl wird berechnet und so skaliert, daß sie in ein vorgegebenes **Bewertungssystem** eingegordnet werden kann.

Zunächst bestimmt man die Häufigkeit n_i der Personen mit Gesamtpunktzahl i , für $i = 1, \dots, k$, wenn k die maximale Punktzahl ist und berechnet die Summenhäufigkeiten

$$N_i = \sum_{j=1}^i n_j \quad \text{für } i = 1, \dots, k.$$

Dann lassen sich die sogenannten **Prozentränge** zur Gesamtpunktzahl i bestimmen:

$$P_i \% = 100 P_i = \frac{100}{n} \left(N_i - \frac{n_i}{2} \right) = \frac{50}{n} (2N_i - n_i) \quad \text{für } i = 1, \dots, k.$$

Zu den Prozenträngen werden die zugehörigen Quantile der Standardnormalverteilung

$$u_{P_i} = \Phi^{-1}(P_i) \quad \text{für } i = 1, \dots, k$$

bestimmt und zur Bewertung z.B. gemäß **Tab. 21** verwandt, vgl. auch Rasch et al. (1978).

Tab. 21: Einige Bewertungsskalen

Bezeichnung der Skala	Transformation	Wertebereich	Mittelwert	$\sqrt{\text{Varianz}}$
T-Note (nach McCall)	$T_i = 50 + 10 u_{p_i}$	20 ... 80	50	10
C-Note (nach Guilford)	$C_i = 5 + 2 u_{p_i}$	-1 ... 11	5	2
Schulnote (nach Lienert)	$S_i = 3 - u_{p_i}$	1 ... 5	3	1
L-Note	$L_i = 10 + 5 u_{p_i}$	-5 ... 25	10	5
Z-Note	$Z_i = 100 + 10 u_{p_i}$	70 ... 130	100	10
„Intelligenzquotient“ (nach Wechsler)	$IQ = 100 + 15 u_{p_i}$	55 ... 145	100	15

Beispiel: Eine Stichprobe von $n = 150$ Personen unterzieht sich einem Test, bei dem maximal 20 Punkte erreichbar sind. Die Testergebnisse mit Summenhäufigkeiten, Prozenträngen und zugehörigen Quantilen der Standardnormalverteilung sind in **Tab. 22** zusammengefaßt; zudem sind dort zu jeder Punktzahl die zugehörigen Schulnoten (nach Lienert) und „Intelligenzquotienten“ (nach Wechsler) angegeben.

Tab. 22: Ergebnisse und Arbeitstabelle zur Bewertungsfindung

Punktzahl i	Häufig- keit n_i	Summen- häufigkeit N_i	Prozent- rang $P_i \%$	Quantil u_{p_i}	Schul- note $S_i = 3 - u_{p_i}$	IQ $= 100 + 15 u_{p_i}$
0	1	1	0,333	-2,71	5,71	59,35
1	2	3	1,333	-2,22	5,22	66,70
2	2	5	2,667	-1,93	4,93	71,05
3	1	6	3,667	-1,79	4,79	73,15
4	5	11	5,667	-1,58	4,58	76,30
5	3	14	8,333	-1,38	4,38	79,30
6	4	18	10,667	-1,25	4,25	81,25
7	9	27	15,000	-1,04	4,04	84,40
8	7	34	20,333	-0,83	3,83	87,55
9	15	49	27,667	-0,60	3,60	91,00
10	12	61	36,667	-0,34	3,34	94,90
11	14	75	45,333	-0,12	3,12	98,20
12	25	100	58,333	0,21	2,79	103,15
13	13	113	71,000	0,56	2,44	108,40
14	10	123	78,667	0,80	2,20	112,00
15	5	128	83,667	0,98	2,02	114,70
16	7	135	87,667	1,16	1,84	117,40
17	9	144	93,000	1,48	1,52	122,20
18	3	147	97,000	1,89	1,11	128,35
19	1	148	98,333	2,13	0,87	131,95
20	2	150	99,333	2,48	0,52	137,20

Nach diesen Ergebnissen würden also die Schulnoten etwa so vergeben:

0– 4 Punkte: Note 5

5– 9 Punkte: Note 4

10–12 Punkte: Note 3

13–17 Punkte: Note 2

18–20 Punkte: Note 1.

Beim „Intelligenztest“ würde man eine durchschnittliche „Intelligenz“ (IQ = 100) bestätigen, wenn eine Person zwischen 11 und 12 Punkten hat.

Bei diesem Verfahren ist zum einen kritisch anzumerken, daß nur die Gesamtpunktzahl zur Bewertung (z. B. Zuordnen eines „Intelligenzquotienten“) herangezogen wird. Es kann so passieren, daß Personen, die ganz unterschiedliche Ergebnisse bei einzelnen Aufgaben haben, die gleiche Bewertung erhalten. Es wäre hier wohl günstiger, jede Aufgabe einzeln zu bewerten und ein mehrdimensionales Bewertungssystem zu verwenden.

Zum anderen wird ein Bewertungssystem nur aufgrund einer **Lernstichprobe** von n Objekten (Personen) aufgestellt. Die Ergebnisse der Skalierung sind natürlich

abhängig von dieser Lernstichprobe, so daß das Bewertungssystem eigentlich nur für weitere Objekte eine Gültigkeit hat, wenn diese aus der Grundgesamtheit stammen, die durch die Lernstichprobe repräsentiert wird. Durch Nichtberücksichtigung dieses Sachverhaltes kann es zu erheblichen Fehlbeurteilungen kommen. Konkret heißt dies z. B., daß ein „Intelligenztest“, der in den USA entwickelt wurde, nicht blind in einem anderen Kulturreis angewandt werden darf, da er auf amerikanische, nicht aber z. B. auf europäische Verhältnisse zugeschnitten ist.

Die quasi einzige Möglichkeit **nominalskalierte Merkmale**, d. h. Merkmale, deren Ausprägungen keiner Ordnung unterliegen, direkt auszuwerten, bietet die **Kontingenztafelanalyse**, vgl. Kap. VII. Eine (indirekte) Auswertungsmöglichkeit besteht darin, die nominalen Merkmalsausprägungen zunächst zu skalieren: Bei Vorliegen qualitativen (insbesondere bei z. T. nominalen Ausprägungen) Datenmaterials kann man die Daten mittels Darstellung in Form von Kontingenztafeln über **Verfahren der mehrdimensionalen Skalierung in Kontingenztafeln** einer quantitativen Datenanalyse zugänglich machen; man vgl. hierzu Fisher (1940), Lancaster (1957), Kündert/Bargmann (1972), Bargmann (1979), Hartung/Elpelt (1984, Kap. V).

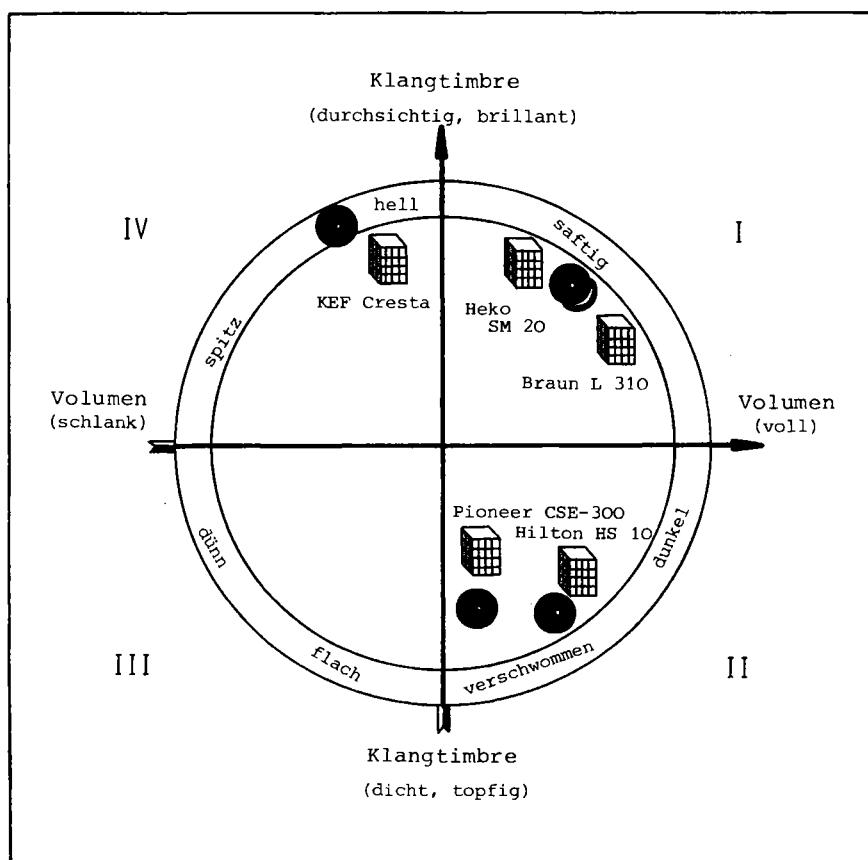


Abb. 22: Lautsprecherhimmel für 5 Boxen einer Preisklasse

Eine **Ähnlichkeitsskala** ist eine relative Skala zur Beurteilung der Lage der Objekte zueinander. Die Ähnlichkeit zweier Objekte i und j nennt man auch einen **Distanzindex** $d(i, j)$. Ein Distanzindex zur Beurteilung der Ähnlichkeit von Objekten muß dabei mindestens die folgenden Eigenschaften besitzen:

Es muß gelten

$$d(i, i) = 0 \quad \text{und} \quad d(i, j) = d(j, i) \quad \text{für alle Paare } (i, j).$$

Weiterhin soll der Distanzindex ein monotones Maß sein, d. h. ist das Objektpaar (i', j') verschiedener als das Paar (i, j) , so soll stets

$$d(i, j) < d(i', j')$$

gelten. Eine ausführliche Darstellung solcher Ähnlichkeitsmaße findet man z. B. bei Pfanzagl (1971) und Krauth (1980).

Die **multidimensionale Skalierung (MDS)** dient nun, ausgehend von einer Ähnlichkeitsskala, zur Bestimmung metrischer Faktoren, die zwar selbst nicht beobachtbar und i. d. R. nicht direkt interpretierbar sind, die Objekte aber beschreiben, vgl. z. B. Kruskal/Wish (1978) und Hartung/Elpelt (1984, Kap. VI).

Die **Faktorenanalyse** dient zum Erkennen von Strukturen in komplexen Datensituationen und kann so auch zu einer (ebenfalls mehrdimensionalen) Skalierung von Objekten bzgl. dieser Strukturfaktoren verwandt werden; man vgl. hierzu etwa Harman (1967), Überla (1971), Weber (1974) und Hartung/Elpelt (1984, Kap. VIII).

Beispiel: Anhand einer Vielzahl von Kriterien zur Beurteilung der Qualität von Lautsprecherboxen wurden zwei charakteristische künstliche Größen bestimmt, mit denen sich die Qualität der Boxen beschreiben lässt, und die somit als (zweidimensionale) Beurteilungsskala der untersuchten Lautsprecherboxen dienen, vgl. Harden (1972): Lautsprecherboxen, in: Hifi-Report 72/73, fono forum. Die beispielhafte **Abb. 22** ist dieser Arbeit entnommen.

Kapitel VI: Qualitätskontrolle

Ein Ziel einer jeden Produktion ist es, brauchbare Waren, d.h. Produkte, deren Eigenschaften bestimmten Anforderungen des Produzenten bzw. des Verbrauchers genügen, herzustellen. Um die geforderte Qualität der Erzeugnisse zu garantieren, sind Kontrollen der zugelieferten Teile, laufende Kontrollen in der Produktion sowie der fertigen Waren unerlässlich (zur Durchführung vgl. auch Kap. IX, 2.1 und Kap. XIII, 4). Sicher wird man zunächst meinen, eine Totalkontrolle sei, zumindest wenn man die dadurch entstehenden Kosten unberücksichtigt läßt, in jedem Fall einer Stichprobenkontrolle vorzuziehen. Denkt man jedoch einmal an zerstörende Kontrollen, wie sie etwa bei Lebensdauerprüfungen, Belastungsprüfungen etc. vorkommen können, so sieht man, daß eine Totalkontrolle sogar sinnlos sein kann. Es gibt aber auch gute Gründe dafür, eine Stichprobenkontrolle bei nicht zerstörenden Prüfungen durchzuführen. So ist bei der verhältnismäßig kleinen Stichprobe viel eher eine sorgfältige, durch qualifiziertes Personal durchgeführte Untersuchung der einzelnen Stücke möglich als bei einer Totalkontrolle, bei der man etwa durch nicht genügend ausgebildetes Personal oder Ermüdungserscheinungen der Kontrolleure trotz der höheren Kosten schlechtere Ergebnisse erhalten kann als bei einer geeigneten Stichprobenkontrolle. Aber auch allein wegen der weitaus höheren Kosten wird man oft auf eine Totalkontrolle verzichten. Es sollen hier nun zunächst statistische Verfahren in der Eingangs- und Endkontrolle besprochen werden. Später gehen wir dann noch auf die Kontrolle der laufenden Produktion sowie kurz auf kontinuierliche Stichprobenpläne ein.

Die Darstellung erfolgt hier in Anlehnung an das Standardlehrbuch der Qualitätskontrolle von Uhlmann (1966; 2. Auflage 1982), auf das der interessierte Leser an dieser Stelle – auch für weitergehende Ausführungen – hingewiesen sei.

1. Stichprobenpläne in der Eingangs- und Endkontrolle

Aufgrund des Untersuchungsergebnisses einer aus einer Partie entnommenen zufälligen Stichprobe soll entschieden werden, ob die gesamte Partie gewissen Qualitätsansprüchen genügt und damit angenommen werden kann. Diese Entscheidungen werden mit Hilfe statistischer Tests gefällt, die im wesentlichen schon in Kapitel IV vorgestellt wurden, wie z.B. Tests für die Parameter der Binomial-, hypergeometrischen, Poisson- und Normalverteilung. Es ist klar, daß hierbei Fehlentscheidungen – Annahme einer Partie, obwohl die Spezifikationen nicht eingehalten wurden und umgekehrt – vorkommen können. Es läßt sich jedoch in Abhängigkeit vom jeweiligen Ausschußanteil die Wahrscheinlichkeit angeben, mit der solch eine Partie angenommen bzw. abgelehnt wird, die **Annahme- bzw. Ablehnungswahrscheinlichkeit**. Die dadurch bestimmte Funktion, die man als **Operationscharakteristik (OC)** oder **Annahmekennlinie** bezeichnet, berechnet sich als $OC = 1 - \text{Gütfunktion}$ des verwendeten statistischen Tests (vgl. Abschnitt 4 in Kapitel III).

Die Annahmekennlinie eines idealen Prüfverfahrens, wie etwa einer fehlerfreien **Totalkontrolle**, das mit Sicherheit feststellt, ob in einer Partie der Ausschußanteil einen vorgegebenen Wert p_0 übersteigt oder nicht, ist in **Abb. 1** für $p_0 = 2\%$ dargestellt und im Vergleich dazu ist in **Abb. 2** die typische OC eines realen **Stichproben-**

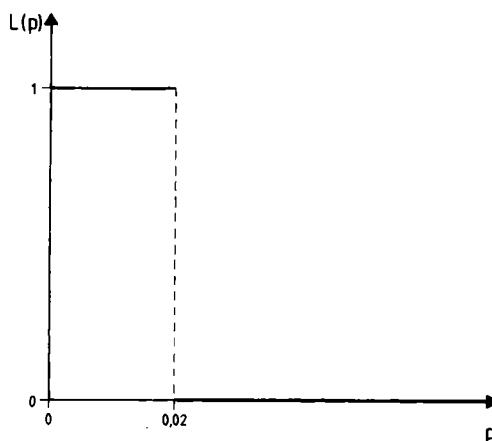


Abb. 1: OC eines idealen Prüfverfahrens

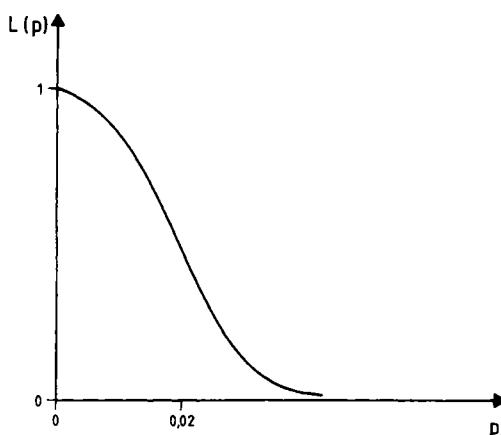


Abb. 2: OC eines realen Stichprobenverfahrens

verfahrens zu sehen. Haben z. B. die angelieferten Partien einen Ausschußprozentsatz von 1%, so läßt sich aus Abb. 1 eine Annahmewahrscheinlichkeit von 100% ablesen; die Partien werden also mit Sicherheit angenommen. Dagegen ergibt sich aus Abb. 2 für das reale Prüfverfahren eine Annahmewahrscheinlichkeit von etwa 90%, d.h. von den Partien mit 1% Ausschuß werden durchschnittlich 10% abgelehnt, obwohl der Anteil der schlechten Stücke den zulässigen Wert von 2% nicht erreicht.

Bei einem Ausschußprozentsatz von 4% in der zu prüfenden Ware liest man aus Abb. 1 für das ideale Verfahren die Annahmewahrscheinlichkeit von 0% ab, d.h. es wird stets die richtige Entscheidung getroffen, solche nicht den Anforderungen entsprechenden Partien abzulehnen. Aus Abb. 2 ergibt sich für den Ausschußanteil 4% ein Wert der OC von etwa 10%, so daß das reale Verfahren in der gleichen Situation durchschnittlich 10% der Partien fälschlicherweise annimmt. Die Fähig-

keit eines Verfahrens, zwischen „guten“ und „schlechten“ Partien zu unterscheiden, läßt sich also sehr gut an Hand der OC beurteilen.

Wie sehen nun solche Verfahren und deren OC konkret aus?

1.1. Einfache Stichprobenpläne für qualitative Merkmale

Stellen wir uns vor, in einer Partie vom Umfang N wären insgesamt M Teile unbrauchbar. Der Ausschußanteil ist also $100 \cdot \frac{M}{N} \%$. Aufgrund der Anzahl der defekten Stücke m in einer relativ kleinen Stichprobe der Größe n soll entschieden werden, ob die Partie angenommen oder abgelehnt wird. Dabei soll eine Annahme möglichst nur dann erfolgen, wenn der unbekannte Anteil $p = \frac{M}{N}$ defekter Teile in der gesamten Partie nicht zu groß ist. Damit ist klar, daß man das Los annimmt, wenn m klein genug ausfällt – etwa $m \leq c$ – und sonst ablehnt. In der Regel wird man eine Stichprobenentnahme ohne Zurücklegen durchführen und deshalb, wegen der dann hypergeometrisch verteilten Testgröße m , die folgende Operationscharakteristik erhalten:

$$L_{N,n,c}\left(\frac{M}{N}\right) = \sum_{m=0}^c \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}.$$

Berechnungen lassen sich zum Teil wesentlich vereinfachen, wenn man folgende Näherungen benutzt:

$$L_{n,c}(p) = \sum_{m=0}^c \binom{n}{m} p^m (1-p)^{n-m} \quad (\text{Binomialverteilung})$$

ist im allgemeinen hinreichend genau, falls $n \leq \frac{N}{10}$ gilt;

$$L_{n,c}^*(p) = \sum_{m=0}^c \frac{n^m p^m}{m!} e^{-np} \quad (\text{Poissonverteilung})$$

ist in der Regel anwendbar, falls $n \leq \frac{N}{10}$ und zusätzlich n groß und p klein ist; bei

den hier auftretenden Problemen ist die Annäherung durch $L_{n,c}^*(p)$ meist schon hinreichend genau für $p < 0,1$; die Normalverteilungsapproximation

$$L_{n,c}(p) \simeq \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \quad \text{mit} \quad \xi = \frac{c + \frac{1}{2} - np}{\sqrt{np(1-p)}}$$

ist meist nur für $np(1-p) > 9$ brauchbar und liefert daher bei den in der Praxis relevanten Problemen wegen der zu kleinen Stichprobenumfänge meist nur eine recht grobe Abschätzung.

Oft sind die folgenden Beziehungen recht nützlich: $L_{n,c}(p) = \beta$ genau dann, wenn

$$(1) \quad p = \frac{(c+1)F_{2(c+1), 2(n-c); 1-\beta}}{n-c + (c+1)F_{2(c+1), 2(n-c); 1-\beta}} = \frac{c+1}{c+1 + (n-c)F_{2(n-c), 2(c+1); \beta}}$$

(vgl. Abschnitt 3.1 in Kap. IV) und $L_{n,c}^*(p) = \beta$ genau dann, wenn

$$(2) \quad p = \frac{1}{2n} \chi_{2(c+1); 1-\beta}^2 \quad (\text{vgl. Abschnitt 3.4 in Kap. IV}).$$

Das zentrale Problem besteht nun darin, **Stichprobenpläne**, also **Stichprobenumfang n** und **Annahmezahl c** zu finden, deren Operationscharakteristik möglichst gut vorgegebene Spezifikationen einhält. Dabei ist im Hinblick auf die entstehenden **Prüfkosten** die Stichprobengröße n so gering wie möglich zu halten. Zwei oft in der Praxis auftauchende Vorgehensweisen sollen hier jetzt ausführlich besprochen werden.

A. Vorgabe zweier Punkte der Operationscharakteristik

Gesucht ist ein Stichprobenplan mit minimalem Stichprobenumfang, dessen OC an zwei Stellen vorgegebene Werte ungefähr annimmt. Dabei ist es üblich, an der einen Stelle, die dann als **Annahmegrenze (Acceptable Quality Level = A.Q.L.)** bezeichnet wird, einen relativ hohen Wert der OC, etwa 90 oder 95%, zu fordern, während die andere Stelle, die man **Ablehngrenze (Limiting Quality = L.Q.)** nennt, einen relativ kleinen Wert, etwa 5 oder 10%, haben soll. Man bezeichnet die Größe 1-(Wert der OC an der Annahmegrenze) als **Produzentenrisiko**, da sie die Wahrscheinlichkeit angibt, mit der eine ausreichend gute Partie abgelehnt wird. Der Wert der OC an der Ablehngrenze gibt dagegen die Wahrscheinlichkeit an, mit der eine relativ schlechte Partie noch angenommen wird und erhält deshalb die Bezeichnung **Konsumentenrisiko**. Wie man unter derartigen Nebenbedingungen einen dazu passenden Stichprobenplan findet, soll an einem Beispiel demonstriert werden.

Beispiel: Gesucht sind ein möglichst kleiner Stichprobenumfang n und eine Annahmezahl c so, daß zu der Annahmegrenze $p_{1-\alpha} = 0,5\%$ das Produzentenrisiko $\alpha = 5\%$ und zur Ablehngrenze $p_\beta = 2,5\%$ das Konsumentenrisiko $\beta = 5\%$ gehört. Die Partiegröße sei $N = 6000$. Der gesuchte Stichprobenplan soll also eine Operationscharakteristik haben, für die (zumindest annähernd) gilt

$$L_{N,n,c}(p_{1-\alpha}) = 1 - \alpha \quad \text{und} \quad L_{N,n,c}(p_\beta) = \beta.$$

$L_{N,n,c}(p_{1-\alpha})$ und $L_{N,n,c}(p_\beta)$ können bei den endlich vielen möglichen n und c auch nur endlich viele Werte annehmen, so daß es meist nicht möglich ist, die obigen Gleichungen exakt zu erfüllen. Deshalb fordert man nur, daß weder das Produzenten- noch das Konsumentenrisiko die angegebenen Werte übersteigt, also

$$L_{N,n,c}(p_{1-\alpha}) \geq 1 - \alpha \quad \text{und} \quad L_{N,n,c}(p_\beta) \leq \beta.$$

Eine Annahmezahl c, mit der diese beiden Ungleichungen erfüllt werden, läßt sich nur für hinreichend großes n finden. Das kleinste n, mit dem das möglich ist, löst zusammen mit der zugehörigen Annahmezahl c unser Problem. Eine Lösung läßt sich prinzipiell mit Hilfe von Tafeln der hypergeometrischen oder, meist in sehr guter Näherung, mit denen der Binomialverteilung finden. Einfacher ist es, wenn man die Approximation durch die Poissonverteilung und die Beziehung (2) benutzt. Den oben aufgeführten Ungleichungen entspricht dann

$$\frac{\chi_{2(c+1); 1-\beta}^2}{2p_\beta} \leq n \leq \frac{\chi_{2(c+1); \alpha}^2}{2p_{1-\alpha}}.$$

Da in unserem Beispiel

$$\frac{\chi^2_{2(c+1); 0,95}}{0,05} > \frac{\chi^2_{2(c+1); 0,05}}{0,01} \quad \text{für } c = 0, 1, 2, 3 \text{ und}$$

$$\frac{\chi^2_{2(c+1); 0,95}}{0,05} = 366,2 \leq 394,0 = \frac{\chi^2_{2(c+1); 0,05}}{0,01} \quad \text{für } c = 4,$$

ist der gesuchte Stichprobenplan durch die Annahmezahl $c = 4$ und den Stichprobenumfang $n = 367$ gegeben. Mit der Beziehung (1) können wir uns nun schnell überzeugen, ob die Benutzung der Poissonverteilung berechtigt war und haben zugleich eine wirksame Kontrolle gegen Rechenfehler. Wir erhalten hier

$$p_{1-\alpha} = \frac{c+1}{c+1 + (n-c)F_{2(n-c), 2(c+1); 1-\alpha}} = \frac{5}{5 + 363F_{726, 10; 0,95}} = 0,0054,$$

$$p_\beta = \frac{(c+1)F_{2(c+1), 2(n-c); 1-\beta}}{n-c+(c+1)F_{2(c+1), 2(n-c); 1-\beta}} = \frac{5F_{10,726; 0,95}}{363+5F_{10,726; 0,95}} = 0,0246.$$

Die OC nimmt also an der Stelle 0,0054 den Wert 95% an. An der von uns festgelegten Annahmegrenze 0,005 ergibt sich daher wegen des fallenden Verlaufs der Annahmekennlinie ein etwas höherer Wert als die geforderten 95%. Dagegen ist die OC auf den Wert 5% schon etwas früher als vorgegeben gesunken, nämlich an der Stelle 0,0246, so daß bei einem Ausschubanteil der angelieferten Partien von 0,025 die Annahmewahrscheinlichkeit unter 5% liegt. Da der Stichprobenumfang

$n = 367 < 600 = \frac{N}{10}$ ist, haben wir eine hinreichend genaue Approximation

der hypergeometrischen Verteilung durch die Binomialverteilung und damit die Aufgabe, einen Stichprobenplan zu finden, für den $L_{N, n, c}(0,005) \geq 0,95$ und $L_{N, n, c}(0,025) \leq 0,05$ gilt, gelöst, vgl. Abb. 3. Wie gut war nun die Approximation durch die Poissonverteilung? Läßt sich eventuell noch ein Plan mit besserer Anpassung an die vorgegebenen Daten und/oder kleinerem Stichprobenumfang finden? Dazu berechnen wir für einige weitere Pläne mit Beziehung (1) wie oben die Annahmegrenze $p_{1-\alpha}$ und die Ablehngrenze p_β , vgl. Tab. 1, und tragen zur besseren Veranschaulichung die Werte in Abhängigkeit vom Stichprobenumfang n in ein

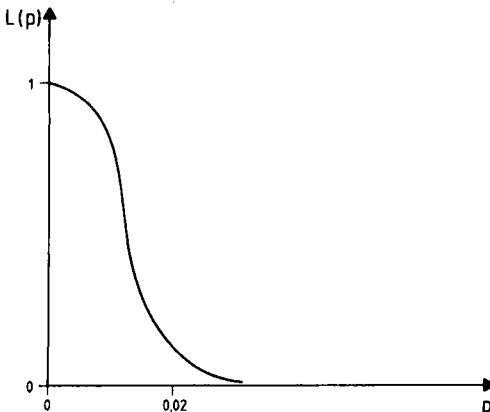


Abb. 3: OC eines einfachen Stichprobenplans mit Annahmezahl $c = 4$ und Stichprobenumfang $n = 367$

Tab. 1: Annahme- und Ablehngrenze zu vorgegebener Annahmezah c und vorgegebenem Stichprobenumfang n

$c = 3$			$c = 4$			$c = 5$		
n	$p_{1-\alpha}$	p_β	n	$p_{1-\alpha}$	p_β	n	$p_{1-\alpha}$	p_β
270	0,00509	0,0282	350	0,00567	0,0258	400	0,00657	0,0259
280	0,00491	0,0272	360	0,00550	0,0251	410	0,00641	0,0253
290	0,00474	0,0263	367	0,00540	0,0246	420	0,00626	0,0247
300	0,00450	0,0254	370	0,00535	0,0244	430	0,00611	0,0241
310	0,00443	0,0246	380	0,00521	0,0238	440	0,00597	0,0236
320	0,00429	0,0238	390	0,00508	0,0232	450	0,00580	0,0231
			400	0,00495	0,0226	460	0,00571	0,0226
						470	0,00559	0,0221
						500	0,00525	0,0208

Diagramm ein. Die von uns gewünschten Annahme- und Ablehngrenzen markieren wir durch waagerechte Linien, vgl. Abb. 4.

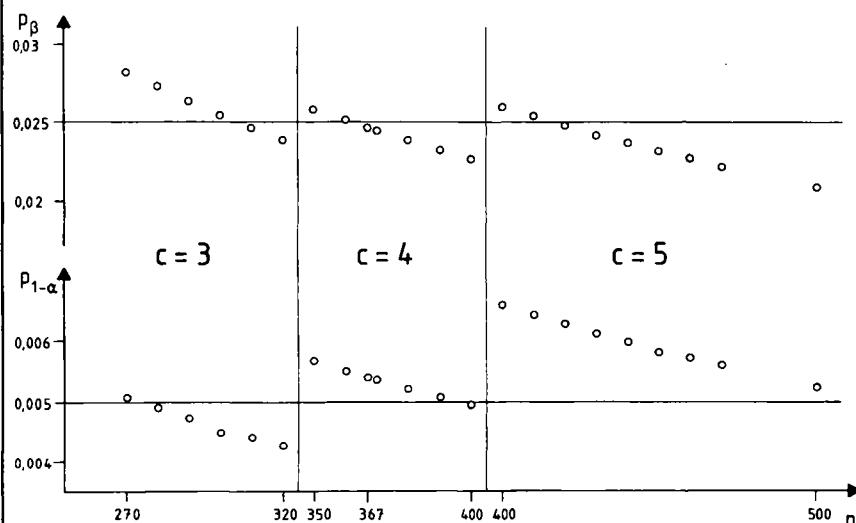


Abb. 4: Annahme- und Ablehngrenzen für verschiedene Annahmezahlen c und Stichprobenumfänge n

Die Forderung $L_{N,n,c}(0,005) \geq 0,95$ und $L_{N,n,c}(0,025) \leq 0,05$ wird dann, wie wir oben gesehen haben, von solchen Plänen erfüllt, für die sowohl $p_{1-\alpha}$ als auch p_β im Diagramm zwischen diesen beiden Linien eingezeichnet sind. Mit einer Annahmezahl $c = 3$ lässt sich das nicht erreichen; entweder $p_{1-\alpha}$ oder/und p_β liegt dann außerhalb des markierten Bereichs. Bei einer Annahmezahl $c = 4$ liefern Stichprobenumfänge zwischen 360 und 395 Pläne, deren Grenzen $p_{1-\alpha}$ und p_β zwischen den beiden Linien liegen und damit den obigen Bedingungen genügen. Dabei wird die Annahmegrenze $p_{1-\alpha} = 0,005$ genau eingehalten, wenn der Stichprobenumfang $n = 395$ beträgt. Die Ablehngrenze hat dann aber nicht den gewünschten Wert $p_\beta = 0,025$, während es bei $n = 361$ genau umgekehrt ist. Der zunächst gefundene

Stichprobenumfang $n = 367$ ist also ein annehmbarer Kompromiß. Läßt man zu, daß sowohl Produzenten- als auch Konsumentenrisiko die gewünschten 5% geringfügig überschreiten, so kann einiges an Prüfarbeit eingespart werden, wenn man den Stichprobenplan mit dem Stichprobenumfang $n = 295$ und der Annahmezahl $c = 3$ benutzt.

B. Vorgabe des Indifferenzpunktes und der Steilheit

Anstatt zwei Punkte der OC vorzugeben, läßt sich z.B. auch ein Punkt und die Steigung der OC in diesem Punkt vorgeben. Üblicherweise arbeitet man dann mit dem Indifferenzpunkt $p_{50\%}$, d.h. demjenigen Punkt, an dem die OC den Wert 50% annimmt. Die Steigung der OC wird an dieser Stelle annähernd maximal und eignet sich daher vorzüglich zur Beschreibung der **Trennschärfe** des korrespondierenden Stichprobenverfahrens. Aus Gründen der guten Tabellierbarkeit so beschriebener Stichprobenpläne benutzt man jedoch nicht die Steigung selbst, sondern die Steilheit

$$h = - \left. \frac{p}{L_{n,n,c}(p)} \frac{dL_{n,n,c}(p)}{dp} \right|_{p=p_{50\%}}$$

Mit der Näherung der OC durch die Poissonverteilung erhält man

$$h \approx \frac{2(np_{50\%})^{c+1}}{c!} e^{-np_{50\%}}.$$

Zusammen mit der Beziehung (2) für den **Indifferenzpunkt**

$$p_{50\%} = \frac{1}{2n} \chi^2_{2(c+1), 0,5}$$

läßt sich hiermit aus vorgegebenen Daten für h und $p_{50\%}$ der zugehörige Stichprobenplan bestimmen. Zur weiteren Vereinfachung können diese beiden Gleichungen in guter Näherung ersetzt werden durch

$$\frac{\pi}{2} h^2 \approx c + 0,73 \quad \text{und} \quad np_{50\%} \approx c + 0,67.$$

Es ist auch hier wieder ratsam, die OC des damit bestimmten Stichprobenplans daraufhin zu untersuchen, ob die Vorgabewerte hinreichend genau eingehalten werden. Ist $n < N/10$, so reicht normalerweise die Überprüfung mit Hilfe der Binomialapproximation aus, die gegeben ist durch

$$h = - \left. \frac{p}{L_{n,c}(p)} \frac{dL_{n,c}(p)}{dp} \right|_{p=p_{50\%}} = 2p_{50\%} \frac{n!}{c!(n-c-1)!} p_{50\%}^c (1-p_{50\%})^{n-c-1}$$

mit $p_{50\%} = \frac{c+1}{c+1+(n-c)F_{2(n-c), 2(c+1), 0,5}}$.

Auf der Vorgabe des Indifferenzpunktes und der Steilheit beruht das Philips-Standard-Stichprobensystem. Ausführliche Tafeln findet man in dem Buch von Schaafsma/Willemse (1970).

Beispiel: Zur Prüfung von Partien mit einer Stückzahl $N = 2000$ ist ein Stichprobenplan gesucht, dessen OC eine Steilheit $h = 1,3$ aufweist. Der auch **Prüfpunkt** genannte Indifferenzpunkt $p_{50\%}$ soll den Wert 0,02 annehmen, so daß Partien

mit einem Ausschußanteil unter 0,02 im Schnitt in mehr als 50% der Fälle angenommen und bei einem Ausschußprozentsatz über 0,02 überwiegend abgelehnt werden.

Wir wollen zunächst mit Hilfe der Beziehung

$$\frac{\pi}{2} h^2 \simeq c + 0,73$$

die Annahmezahl c bestimmen. Da c ganzzahlig ist, läßt sich natürlich nicht jede gewünschte Steilheit realisieren. Für $c = 1$ ergibt sich aus der obigen Gleichung eine Steilheit $h = 1,05$ und für $c = 2$ ergibt sich $h = 1,32$. Soll der vorgegebene Wert nicht unterschritten werden, so hat man $c = 2$ zu wählen, wie man direkt ersieht aus

$$c \geq \frac{\pi}{2} h^2 - 0,73 = 1,92.$$

Der Stichprobenumfang kann nun mit Hilfe von

$$np_{50\%} \simeq c + 0,67$$

bestimmt werden. Hier kann wegen der Ganzzahligkeit von n der vorgegebene Wert für $p_{50\%}$ in der Regel nicht genau eingehalten werden. Normalerweise läßt man nur zu, daß dieser Vorgabewert für den Indifferenzpunkt unterschritten wird, was zu einer Verschärfung der Abnahmeprüfung führt, so daß nur Stichprobenumfänge mit

$$n \geq \frac{c + 0,67}{p_{50\%}} = 133,5$$

in Frage kommen. Das gesuchte Prüfverfahren hat also einen Stichprobenumfang $n = 134$ und eine Annahmezahl $c = 2$. Da $n = 134 < 200 = \frac{N}{10}$, liefert die Binomialverteilung hinreichend genaue Werte für

$$\begin{aligned} p_{50\%} &= \frac{c + 1}{c + 1 + (n - c) F_{2(n-c), 2(c+1); 0,5}} \\ &= \frac{3}{3 + 132 F_{264, 6; 0,5}} = 0,0199 \quad \text{und} \\ h &= 2 p_{50\%} \frac{n!}{c!(n-c-1)!} p_{50\%}^c (1 - p_{50\%})^{n-c-1} \\ &= 2 \cdot 0,0199 \cdot \frac{134 \cdot 133 \cdot 132}{2} \cdot 0,0199^2 \cdot 0,9801^{131} \\ &= 1,332. \end{aligned}$$

Damit ist unsere Aufgabe gelöst. Um besser mit dem im vorigen Abschnitt behandelten Problem vergleichen zu können, berechnen wir noch die Annahme- und Ablehngrenzen ($\alpha = \beta = 0,05$)

$$p_{1-\alpha} = \frac{3}{3 + 132 F_{264, 6; 0,95}} = 0,0062 \quad \text{und}$$

$$p_\beta = \frac{3 F_{6, 264; 0,95}}{132 + 3 F_{6, 264; 0,95}} = 0,0455.$$

Die OC für diesen Stichprobenplan ist in **Abb. 5** dargestellt.

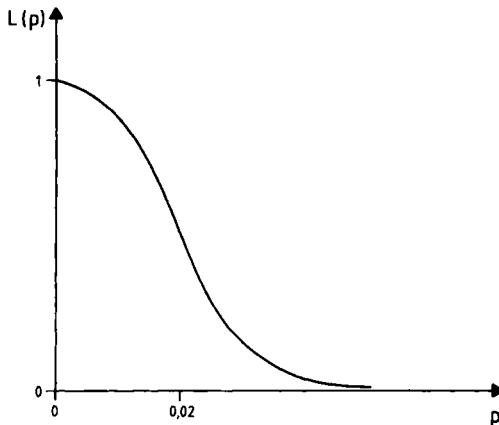


Abb. 5: OC eines einfachen Stichprobenplans mit Annahmezahl $c = 4$ und Stichprobenumfang $n = 134$

Bei nicht zerstörenden Kontrollen kann man abgelehnte Partien einer Totalkontrolle unterziehen und dabei gleichzeitig die schlechten Stücke durch gute ersetzen. Durch diese Vorgehensweise wird der erwartete Anteil schlechter Stücke, der sogenannte **mittlere Durchschlupf** (*average outgoing quality = AOQ*) von p auf $p \cdot L_{N, n, c}(p)$ reduziert. Der **Höchstwert des mittleren Durchschlupfes** (*average outgoing quality limit = AOQL*), also das Maximum der Funktion $p \cdot L_{N, n, c}(p)$ über p , kann ebenso wie etwa oben die Steilheit der OC zur Charakterisierung des verwendeten Stichprobenplans dienen. Durch die Totalkontrolle der abgelehnten Partien verbunden mit dem Aussortieren der schlechten Stücke wird zwar der Ausschuß verringert, aber gleichzeitig die Anzahl der zu prüfenden Stücke von n auf $n \cdot L_{N, n, c}(p) + N \cdot (1 - L_{N, n, c}(p))$ erhöht. Stichprobenpläne, die diese Anzahl für einen bestimmten Ausschußprozentsatz (z. B. den, mit dem die Partien erfahrungs-gemäß angeliefert werden) minimieren unter Vorgabe der AOQL, sind bei Dodge/Romig (1959) tabelliert (vgl. auch Schaafsma/Willemze (1970)). Anstelle der AOQL kann man aber auch die Annahmegrenze, die Ablehngrenze oder ähnliches vorgeben und dann den Prüfaufwand unter dieser Nebenbedingung minimieren.

1.2. Mehrfache und sequentielle Stichprobenpläne für qualitative Merkmale

Während im vorherigen Abschnitt aufgrund einer einzigen Stichprobe entschieden wurde, ob eine Partie angenommen oder abgelehnt wird, wollen wir hier zunächst eine verhältnismäßig kleine Stichprobe entnehmen, deren Untersuchungsergebnis entweder unmittelbar zu einer Entscheidung führt oder in Zweifelsfällen die Entnahme einer zweiten Stichprobe veranlaßt. Bei einem 2-fachen Stichprobenplan

muß dann endgültig über die Annahme der Partie entschieden werden. Dagegen kann bei q-fachen Stichprobenplänen die Entscheidung bis nach der q-ten Stichprobe aufgeschoben werden. Die sogenannten fortschreitenden oder sequentiellen Methoden entscheiden nach der Kontrolle eines jeden Exemplars, ob angenommen, zurückgewiesen oder weiter kontrolliert wird, so daß hier die Anzahl der zu kontrollierenden Exemplare nicht im Vorhinein begrenzt wird.

Durch eine geeignete Wahl solcher komplizierterer Verfahren läßt sich im Vergleich zu einstufigen Stichprobenplänen mit etwa der gleichen OC der **durchschnittliche Prüfaufwand (average sample number = ASN)** reduzieren. Ein gutes doppeltes Stichprobenschema wird in der Regel etwa mit 70–80% der Kontrollen auskommen, die ein einfacher Stichprobenplan erfordert. Eine weitere Reduzierung der Prüfarbeit kann durch **mehrfache** und schließlich **sequentielle Verfahren** erreicht werden. Die Einsparungen sind dann aber nicht mehr so groß wie beim Übergang vom einfachen zum doppelten System. Da außerdem die Anzahl der zu kontrollierenden Stücke immer stärkeren Schwankungen unterliegt, was erhebliche organisatorische Schwierigkeiten mit sich bringen kann, kommen neben den einfachen fast nur noch **doppelte Stichprobenpläne** zur Anwendung. Mit diesen wollen wir uns nun näher beschäftigen.

A. Doppelte Stichprobenpläne

Ein doppelter Plan wird durch die beiden Stichprobenumfänge n_1 und n_2 und drei natürliche Zahlen c_1 , c_2 und c_3 festgelegt. Sind dann in der ersten Stichprobe der Größe n_1 etwa m_1 schlechte Stücke, so wird die Partie angenommen, falls $m_1 \leq c_1$; abgelehnt, falls $m_1 > c_2$ und für $c_1 < m_1 \leq c_2$ die Entscheidung bis nach der zweiten Stichprobe aufgeschoben. Hat man in der zweiten Stichprobe vom Umfang n_2 noch m_2 Fehler gefunden, lehnt man die Partie ab, wenn die Anzahl der Fehler $m_1 + m_2$ in beiden Stichproben c_3 übersteigt; ist dagegen $m_1 + m_2 \leq c_3$, so wird die Partie angenommen. Solch ein Verfahren hat die OC

$$\sum_{m_1=0}^{c_1} \frac{\binom{M}{m_1} \binom{N-M}{n_1-m_1}}{\binom{N}{n_1}} + \\ + \sum_{m_1=c_1+1}^{c_2} \sum_{m_2=0}^{c_3-m_1} \frac{\binom{M}{m_1} \binom{M-m_1}{m_2} \binom{N-M}{n_1-m_1} \binom{N-M-n_1+m_1}{n_2-m_2}}{\binom{N}{n_1} \binom{N-n_1}{n_2}}$$

Mit der Binomialapproximation erhält man $\left(p = \frac{M}{N} \right)$

$$\sum_{m_1=0}^{c_1} \binom{n_1}{m_1} p^{m_1} (1-p)^{n_1-m_1} + \sum_{m_1=c_1+1}^{c_2} \sum_{m_2=0}^{c_3-m_1} \binom{n_1}{m_1} \binom{n_2}{m_2} p^{m_1+m_2} (1-p)^{n_1+n_2-m_1-m_2}$$

für die OC und mit der Poissonapproximation

$$\sum_{m_1=0}^{c_1} \frac{(n_1 \cdot p)^{m_1}}{m_1!} e^{-n_1 p} + \sum_{m_1=c_1+1}^{c_2} \sum_{m_2=0}^{c_3-m_1} \frac{n_1^{m_1} \cdot n_2^{m_2}}{m_1! \cdot m_2!} p^{m_1+m_2} e^{-(n_1+n_2)p}.$$

Der tatsächliche Stichprobenumfang ist

n_1 falls $m_1 \leq c_1$ oder $m_1 > c_2$ und $n_1 + n_2$ falls $c_1 < m_1 \leq c_2$,
so daß man als durchschnittlichen Stichprobenumfang (ASN) erhält:

$$n_1 + n_2 \cdot \sum_{m=c_1+1}^{c_2} \frac{\binom{M}{m} \binom{N-M}{n_1-m}}{\binom{N}{n_1}}.$$

Die Binomialapproximation liefert hierbei

$$n_1 + n_2 \cdot \sum_{m=c_1+1}^{c_2} \binom{n_1}{m} p^m (1-p)^{n_1-m}.$$

Für einen doppelten Stichprobenplan mit $n_1 = n_2 = 50$, $c_1 = 0$ und $c_2 = c_3 = 2$ ist der ASN in Abb. 6 dargestellt (vgl. auch Uhlmann (1966, S. 125)). Im Vergleich dazu hat ein einfacher Stichprobenplan etwa der gleichen OC einen von dem Ausschußanteil p unabhängigen Stichprobenumfang $n = 94$ und eine Annahmezahl $c = 2$. Es sei an dieser Stelle darauf hingewiesen, daß nicht jeder zweistufige Plan günstiger als ein entsprechender einstufiger ist (vgl. Hamaker/v. Strik (1955)). Genauso wie bei den einstufigen Verfahren läßt sich auch hier der Stichprobenplan durch 2 Punkte der OC bzw. den Indifferenzpunkt und die Steilheit usw. kennzeichnen, doch wegen der vielen Parameter nicht mehr eindeutig dadurch festlegen. Man gibt deshalb noch Verhältnisse zwischen den Kenngrößen fest vor. Gebräuchlich sind

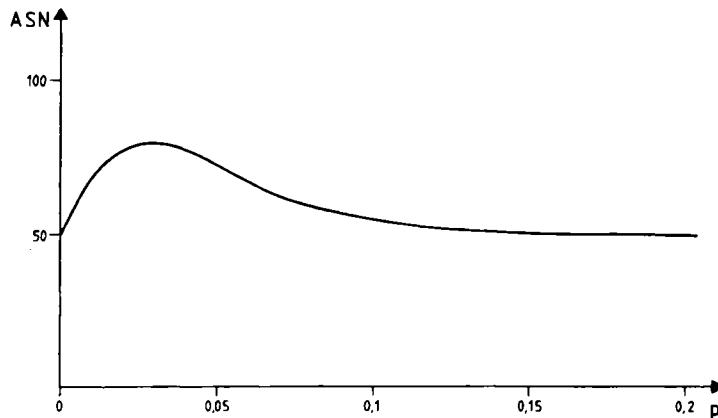


Abb. 6: Durchschnittlicher Stichprobenumfang bei einem zweistufigen Stichprobenplan

$$\begin{aligned} c_2 &= c_3, \\ c_2 &= 3 \cdot c_1 \quad \text{oder} \quad c_2 = 5 \cdot c_1 \quad \text{und} \\ n_2 &= n_1 \quad \text{oder} \quad n_2 = 2 \cdot n_1. \end{aligned}$$

Damit erhält man einerseits möglichst einfache Pläne und andererseits bieten diese zweistufigen Pläne gegenüber den einstufigen wirklich einen Vorteil.

Für die Bestimmung zweistufiger Stichprobenpläne sei auf die Tabellen in Wagner (1959) (Vorgabe der Annahme- und Ablehngrenzen) und Schaafsma/Willemze (1970) (Vorgabe des Indifferenzpunktes und der Steilheit) hingewiesen. Die Stichprobenprüfung kann aber auch hier wie im einstufigen Fall mit einer Totalkontrolle bei Ablehnung der Partie verbunden werden. Dazu sei wieder auf Dodge/Romig (1959) verwiesen.

B. Sequentielle Stichprobenpläne

Wenden wir uns nun noch den **sequentiellen Tests** zu. Nach der Prüfung eines jeden Stichprobenelements wird entschieden, ob die Partie anzunehmen, abzulehnen oder ob ein weiteres Element zu prüfen ist. Konkret gibt man drei Konstanten a , b und c vor mit $0 < c < 1$ und $-a < 0 < b$ und nimmt die Partie an, wenn zum ersten Mal für die Anzahl m_k der schlechten Stücke bei den ersten k Stichprobenelementen gilt

$$m_k \leq c \cdot k - a.$$

Wenn zum ersten Mal

$$m_k \geq c \cdot k + b$$

ist, wird die Partie abgelehnt, und solange

$$c \cdot k - a < m_k < c \cdot k + b$$

ist, wird ein weiteres Stichprobenelement geprüft.

Soll die OC dieses sequentiellen Verfahrens an der Annahmegrenze $p_{1-\alpha}$ näherungsweise den Wert $1 - \alpha$ und an der Ablehngrenze p_β näherungsweise den Wert β annehmen, so muß gelten (Verwendung der Binomialapproximation)

$$a = \frac{-\ln \frac{\beta}{1-\alpha}}{\ln \frac{p_\beta(1-p_{1-\alpha})}{p_{1-\alpha}(1-p_\beta)}}, \quad b = \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{p_\beta(1-p_{1-\alpha})}{p_{1-\alpha}(1-p_\beta)}}, \quad c = \frac{\ln \frac{1-p_{1-\alpha}}{1-p_\beta}}{\ln \frac{p_\beta(1-p_{1-\alpha})}{p_{1-\alpha}(1-p_\beta)}}.$$

Für $\alpha = \beta$ und damit $a = b$ lassen sich auch einfache Beziehungen zum Indifferenzpunkt $p_{50\%}$ und zur Steilheit h angeben:

$$a = b = h(1 - p_{50\%}) \quad \text{und} \quad c = p_{50\%}.$$

Für den durchschnittlichen Stichprobenumfang (ASN) des sequentiellen Verfahrens ergibt sich ungefähr

$$\begin{aligned} \frac{b - (a + b)L(p)}{p - c} &\quad \text{für } p \neq c \quad \text{und} \\ \frac{a \cdot b}{c(1 - c)} &\quad \text{für } p = c. \end{aligned}$$

wobei mit L die OC des Verfahrens bezeichnet wird.

Beispiel: Gesucht ist ein sequentielles Stichprobenverfahren, zu dessen Annahmegrenze $p_{1-\alpha} = 0,02$ das Produzentenrisiko $\alpha = 0,1$ gehört und das eine Ablehn-

grenze $p_\beta = 0,05$ mit dem Konsumentenrisiko $\beta = 0,1$ besitzt. Wir erhalten wegen $\alpha = \beta$ nun

$$a = b = \frac{-\ln \frac{\beta}{1-\alpha}}{\ln \frac{p_\beta(1-p_{1-\alpha})}{p_{1-\alpha}(1-p_\beta)}} = \frac{-\ln \frac{0,1}{0,9}}{\ln \frac{0,05(1-0,02)}{0,02(1-0,05)}} = 2,32 \quad \text{und}$$

$$c = \frac{\ln \frac{1-p_{1-\alpha}}{1-p_\beta}}{\ln \frac{p_\beta(1-p_{1-\alpha})}{p_{1-\alpha}(1-p_\beta)}} = \frac{\ln \frac{1-0,02}{1-0,05}}{\ln \frac{0,05(1-0,02)}{0,02(1-0,05)}} = 0,033.$$

Das Prüfverfahren muß also wie folgt durchgeführt werden. Nach jeder Prüfung stellt man fest, welche von den drei folgenden Ungleichungen erfüllt ist. Gilt für die Anzahl m_k der bis dahin gefundenen schlechten Stücke

$$m_k \leq c \cdot k - a = 0,033 \cdot k - 2,32,$$

wobei k die Anzahl aller bisher durchgeföhrten Untersuchungen ist, so nimmt man die Partie an; ist

$$m_k \geq c \cdot k + b = 0,033 \cdot k + 2,32,$$

so lehnt man die Partie ab; und bei

$$c \cdot k - a = 0,033 \cdot k - 2,32 < m_k < 0,033 \cdot k + 2,32 = c \cdot k + b$$

muß ein weiteres Stichprobenelement geprüft werden.

Die Entscheidung, ob man die Partie annehmen, ablehnen oder weiter kontrollieren soll, läßt sich auch mit Hilfe eines graphischen Verfahrens herbeiführen.

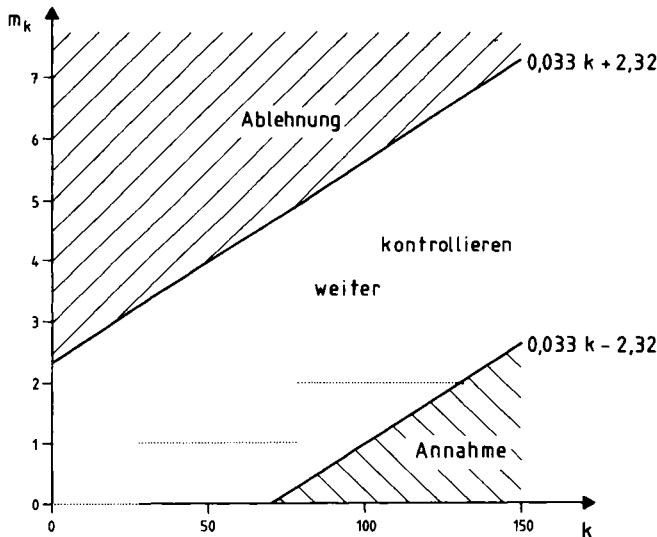


Abb. 7: Stichprobenergebnis bei einem sequentiellen Prüfplan für qualitative Merkmale mit $c = 0,033$ und $a = b = 2,32$

In einem Diagramm werden die beiden durch $c \cdot k - a = 0,033 \cdot k - 2,32$ und $c \cdot k + b = 0,033 \cdot k + 2,32$ festgelegten Geraden eingezeichnet. Nun trägt man nach jeder Prüfung eines Partielements einen Punkt in das **Diagramm** ein, dessen Koordinaten durch die Anzahl der bis dahin durchgeföhrten Untersuchungen k und die dabei gefundene Anzahl schlechter Stücke m_k gegeben sind. Liegt der eingezeichnete Punkt zwischen den beiden Geraden, so kontrolliert man weiter, doch sobald eine Eintragung außerhalb dieses Bereichs vorgenommen wird, kommt es zur endgültigen Entscheidung; dabei führt ein Überschreiten der oberen Grenze zur Ablehnung und ein Eindringen in den Bereich unterhalb der anderen Geraden zur Annahme der Partie. Aufgrund der in Abb. 7 eingetragenen Stichprobenergebnisse wird man also die Partie annehmen. Wie groß ist nun der Prüfaufwand bei diesem sequentiellen Verfahren? Für einige spezielle Ausschußanteile p in den angelieferten Partien lässt sich der ungefähre durchschnittliche Stichprobenumfang (ASN) einfach berechnen. Man erhält (vgl. auch Abb. 8)

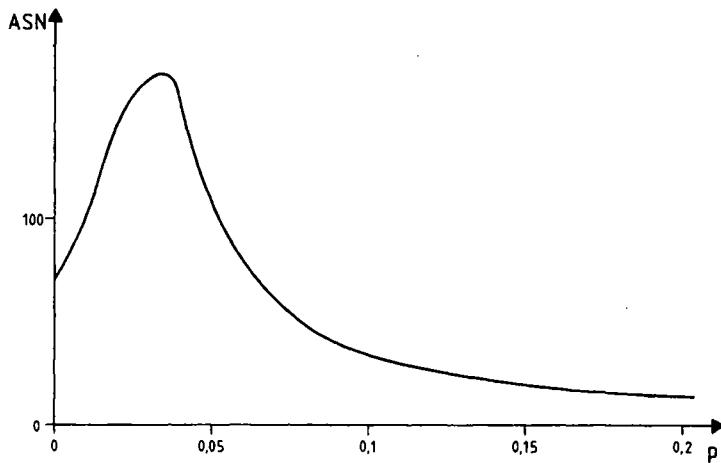


Abb. 8: Durchschnittlicher Prüfaufwand bei dem sequentiellen Prüfplan aus Abb. 7

$$\frac{b = (a+b)L(p)}{p - c} = \begin{cases} \frac{b - (a+b) \cdot 1}{0 - c} = \frac{a}{c} = 70,3, & \text{für } p = 0 \\ \frac{b - (a+b) \cdot (1-\alpha)}{p_{1-\alpha} - c} = 142,8 & \text{für } p = p_{1-\alpha} = 0,02 \\ \frac{b - (a+b) \cdot \beta}{p_\beta - c} = 109,2, & \text{für } p = p_\beta = 0,05 \\ \frac{b - (a+b) \cdot 0}{1 - c} = \frac{b}{1 - c} = 2,4, & \text{für } p = 1 \\ \frac{a^2}{c \cdot (1 - c)} = 168,7, & \text{für } p = c = 0,033 \end{cases}$$

Im Vergleich dazu hat ein einfacher Plan mit etwa den gleichen Güteeigenschaften einen vom Ausschußanteil p unabhängigen Stichprobenumfang $n = 295$ bei

einer Annahmezahl von 9. Die Einsparung von Prüfarbeit ist also erheblich. Das ist auch der alleinige Grund, sequentielle Methoden anzuwenden, denn die gleiche Güte läßt sich ja auch mit den einfachen Verfahren erreichen, allerdings mit durchschnittlich größeren Stichprobenumfängen. Bei den **einstufigen Verfahren** läßt sich jedoch auf recht einfache Weise der Prüfaufwand etwas reduzieren. Man kann nämlich die **Kontrolle abbrechen**, sobald die Anzahl der schlechten Stücke die Annahmezahl überschreitet, denn dann ist sicher, daß die Partie abgelehnt wird, unabhängig vom weiteren Untersuchungsergebnis. Entsprechend läßt sich auch bei **zwei- und mehrstufigen Verfahren** Prüfarbeit einsparen, vgl. Abb. 9. Die geringen durchschnittlichen Stichprobenumfänge der sequentiellen Verfahren lassen sich allerdings in der Regel, zumindest wenn die Ausschußanteile der angelieferten Partien selten zwischen den Werten $p_{1-\alpha}$ und p_β liegen, nicht erreichen. Man wird deshalb immer dann zu sequentiellen Verfahren greifen, wenn die Einzelprüfungen aufwendig oder/und teuer sind und dafür in Kauf nehmen, mit relativ komplizierten Stichprobenplänen zu arbeiten, bei denen zudem der Stichprobenumfang in der Regel starken Schwankungen unterliegt, wodurch sich z. B. organisatorische Probleme ergeben können.

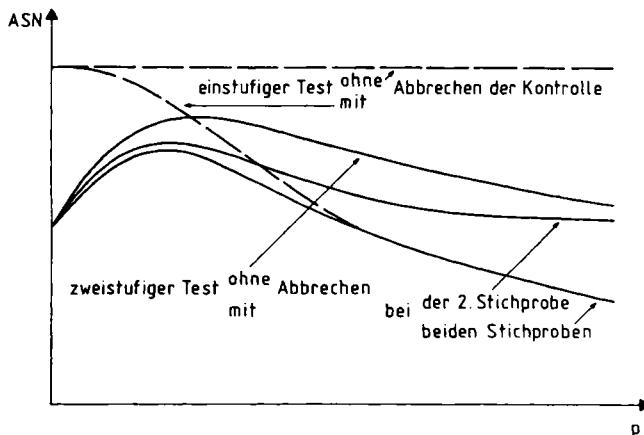


Abb. 9: Charakteristischer Verlauf des durchschnittlichen Prüfaufwandes mit und ohne Abbrechen der Kontrolle

Obwohl sequentielle Verfahren zunächst im Rahmen der Qualitätskontrolle entwickelt wurden, werden sie inzwischen auch in vielen anderen Gebieten der **Technik**, der **Medizin**, der **Naturwissenschaften** usw. angewendet, wo es darauf ankommt, den durchschnittlichen Stichprobenumfang möglichst gering zu halten und dabei gleichzeitig sowohl den Fehler erster Art α als auch den Fehler zweiter Art β unter Kontrolle zu halten, vgl. Abschnitt 6 in Kap. IV.

1.3. Stichprobenpläne für quantitative Merkmale

Stellt man bei den zu überprüfenden Stücken nur fest, ob sie brauchbar oder unbrauchbar sind, so kommen die in 1.1 und 1.2 besprochenen Stichprobenpläne für qualitative Merkmale zum Einsatz. Trifft man dagegen die Entscheidung „gut“ oder „schlecht“ aufgrund eines quantitativen Merkmals wie etwa Länge, Gewicht, Lebensdauer usw., so lassen sich auch die Stichprobenpläne für quantitative Merk-

male verwenden, die gegenüber denjenigen für qualitative Merkmale den Vorteil haben, mit wesentlich geringeren Stichprobengrößen auszukommen. Da sich gleichzeitig der Aufwand für die einzelnen Prüfungen erhöht, wird allerdings der Gesamtprüfaufwand bzw. werden die Kosten dafür nicht in jedem Fall sinken.

Wir nehmen nun an, daß die zur Charakterisierung der Partieelemente benutzte Größe x normalverteilt ist mit Mittelwert $w + \mu$ und Varianz σ^2 . Ein Teil wird als brauchbar angesehen, wenn für den zugehörigen Wert x gilt

$$\begin{aligned} x &\geq w - d \text{ bzw. } x \leq w + d \text{ (einseitige Kriterien) oder} \\ w - d &\leq x \leq w + d \text{ (zweiseitiges Kriterium).} \end{aligned}$$

Für den erwarteten Anteil schlechter Stücke in der Partie erhält man dann einseitig

$$\begin{aligned} p(\mu) &= \Phi\left(\frac{-\mu - d}{\sigma}\right) \text{ bzw.} \\ p(\mu) &= \Phi\left(\frac{\mu - d}{\sigma}\right) \text{ oder zweiseitig} \\ p(\mu) &= \Phi\left(\frac{-\mu - d}{\sigma}\right) + \Phi\left(\frac{\mu - d}{\sigma}\right), \end{aligned}$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet. Eine Stichprobe vom Unfang n liefert die Meßwerte x_1, \dots, x_n (unabhängige Realisierungen einer $N(w + \mu, \sigma^2)$ -verteilten Zufallsvariablen), mit Hilfe derer über die Annahme der Partie entschieden wird. Im folgenden soll nur noch der zweiseitige Fall besprochen und zudem, damit sich der Rechenaufwand in Grenzen hält, $d \geq 3\sigma$ angenommen werden, wodurch viele Anwendungsfälle abgedeckt sind.

Bezeichnet \bar{x} den Stichprobenmittelwert, so ist bei bekanntem σ die Partie abzulehnen, wenn für eine vorgegebene Konstante c gilt

$$|\bar{x} - w| > d + \sigma c.$$

Ist dagegen

$$|\bar{x} - w| \leq d + \sigma c,$$

so nimmt man die Partie an.

Als OC ergibt sich in Abhängigkeit vom Ausschußanteil p in den zu prüfenden Partien näherungsweise (wegen $d \geq 3\sigma$)

$$L(p) \simeq \Phi((c - u_p) \cdot \sqrt{n}),$$

wobei u_p das p -Quantil der Standardnormalverteilung bezeichnet. Bei unbekannter Varianz wird in der Prüfvorschrift σ durch die empirische Standardabweichung

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

ersetzt, so daß die Partie abgelehnt wird, falls

$$|\bar{x} - w| > d + s \cdot c$$

gilt und angenommen, falls gilt

$$|\bar{x} - w| \leq d + s \cdot c.$$

Die OC erhält man hinreichend genau (wegen $d \geq 3\sigma$) als Wert der Verteilungsfunktion der nichtzentralen t-Verteilung mit $n - 1$ Freiheitsgraden und Nichtzentralitätsparameter $\delta = u_p \cdot \sqrt{n}$ an der Stelle $c \cdot \sqrt{n}$.

Näherungsweise gilt die einfachere Beziehung

$$L(p) \simeq \Phi \left(\frac{(c - u_p) \cdot \sqrt{n}}{\sqrt{1 + \frac{c^2}{2}}} \right),$$

wobei u_p wiederum das p -Quantil der $N(0, 1)$ -Verteilung bezeichnet. Wie bei den qualitativen Verfahren konstruiert man auch hier Stichprobenpläne, deren OC bestimmte vorgegebene Eigenschaften hat. Fordert man einen Wert $1 - \alpha$ der OC an der **Annahmegrenze** $p_{1-\alpha}$ und einen Wert β an der **Ablehngrenze** p_β , so erhält man für die Konstante c , sowohl bei bekannter als auch bei unbekannter Varianz, näherungsweise

$$\begin{aligned} c &\simeq \frac{u_{1-\alpha} u_{p_\beta} - u_\beta u_{p_{1-\alpha}}}{u_{1-\alpha} - u_\beta} \\ &= \frac{u_\alpha u_{p_\beta} + u_\beta u_{p_{1-\alpha}}}{u_\alpha + u_\beta} \left(= \frac{u_{p_\beta} + u_{p_{1-\alpha}}}{2} \text{ für } \alpha = \beta \right). \end{aligned}$$

Dabei ist bei bekannter Varianz ein Stichprobenumfang von

$$n \simeq \left(\frac{u_{1-\alpha} - u_\beta}{u_{p_\beta} - u_{p_{1-\alpha}}} \right)^2 = \left(\frac{u_\alpha + u_\beta}{u_{p_\beta} - u_{p_{1-\alpha}}} \right)^2$$

und bei unbekannter Varianz von

$$\begin{aligned} n &\simeq \left(\frac{u_{1-\alpha} - u_\beta}{u_{p_\beta} - u_{p_{1-\alpha}}} \right)^2 \cdot \left(1 + \frac{c^2}{2} \right) \\ &= \left(\frac{u_\alpha + u_\beta}{u_{p_\beta} - u_{p_{1-\alpha}}} \right)^2 \cdot \left(1 + \frac{c^2}{2} \right) \\ &= \frac{2(u_\alpha + u_\beta)^2 + (u_\alpha u_{p_\beta} + u_\beta u_{p_{1-\alpha}})^2}{2(u_{p_\beta} - u_{p_{1-\alpha}})^2} \\ &= \left(\frac{2u_\alpha}{u_{p_\beta} - u_{p_{1-\alpha}}} \right)^2 \left(1 + \frac{1}{2} \left(\frac{u_{p_\beta} + u_{p_{1-\alpha}}}{2} \right)^2 \right) \text{ für } \alpha = \beta \end{aligned}$$

zu wählen.

Beispiel: Gesucht ist ein Stichprobenplan für ein quantitatives normalverteiltes Merkmal mit einem Produzentenrisiko $\alpha = 10\%$ an der Annahmegrenze $p_{1-\alpha} = 1\%$ und einem Konsumentenrisiko $\beta = 10\%$ an der Ablehngrenze $p_\beta = 2\%$. Wir erhalten

$$c \simeq \frac{u_{p_{1-\alpha}} + u_{p_\beta}}{2} = \frac{-2,32635 - 2,05375}{2} = -2,19 \text{ und}$$

$$n \simeq \left(\frac{-1,28155 - 1,28155}{-2,05375 - (-2,32635)} \right)^2 = 88,41$$

bei bekannter Varianz;

$$\begin{aligned} n &\simeq \left(\frac{u_{\alpha} + u_{\beta}}{u_{p_{\beta}} - u_{p_{1-\alpha}}} \right)^2 \cdot \left(1 + \frac{c^2}{2} \right) \\ &= \left(\frac{-1,28155 - 1,28155}{-2,05375 - (-2,32635)} \right)^2 \cdot \left(1 + \frac{(-2,19)^2}{2} \right) = 300,42 \end{aligned}$$

bei unbekannter Varianz.

Im Vergleich dazu benötigt ein etwa gleich guter Stichprobenplan für qualitative Merkmale einen Stichprobenumfang von $n = 978$. Mit dem oben berechneten Stichprobenplan sollen nun Partien geprüft werden, die Drehteile mit einem Soll-durchmesser von 17 cm enthalten. Die einzelnen Teile können als brauchbar angesehen werden, solange ihr Durchmesser vom Sollwert um höchstens 0,1 cm abweicht. Bei einer bekannten Standardabweichung $\sigma = 0,03$ cm führt eine Stichprobe vom Umfang $n = 89$ zu einer Ablehnung der Partie, wenn für den Mittelwert der Stichprobe \bar{x} gilt

$$|\bar{x} - w| = |\bar{x} - 17| > d + \sigma \cdot c = 0,1 - 0,03 \cdot 2,19 = 0,0343.$$

Ist dagegen

$$|\bar{x} - 17| \leq 0,0343,$$

so wird die Partie angenommen.

Bei unbekannter Varianz ist ein Stichprobenumfang $n = 301$ erforderlich, und die Partie wird aufgrund des Stichprobenmittelwertes \bar{x} abgelehnt, wenn gilt

$$|\bar{x} - 17| > 0,1 - 2,19 \cdot s,$$

wobei s die aus der Stichprobe berechnete empirische Standardabweichung ist.

Hat man aber

$$|\bar{x} - 17| \leq 0,1 - 2,19 \cdot s,$$

so führt das zur Annahme der Partie. An dieser Stelle sei noch einmal daran erinnert, daß wir die Stichprobenpläne unter der Voraussetzung $d \geq 3\sigma$ berechnet haben.

Einen Stichprobenplan, der im **Indifferenzpunkt** $p_{50\%}$ die **Steilheit** h besitzt, läßt sich im Fall bekannter Varianz näherungsweise noch recht einfach angeben durch

$$n \simeq \left(\frac{h}{2p_{50\%}} \right)^2 \cdot e^{-(u_{p_{50\%}})^2} \quad \text{und} \quad c \simeq u_{p_{50\%}}.$$

Im Fall unbekannter Varianz sei auf die Tabellen in Schaafsma/Willemze (1970) hingewiesen.

Beispiel: Es soll ein Stichprobenplan mit der Steilheit $h = 1,5$ und dem Prüfungspunkt $p_{50\%} = 2,5\%$ konstruiert werden. Die Varianz σ^2 sei dabei als bekannt vorausgesetzt und es gelte $d \geq 3\sigma$. Wir erhalten dann

$$c \simeq u_{p_{50\%}} = u_{0,025} = -1,96$$

und benötigen einen Stichprobenumfang

$$n \approx \left(\frac{h}{2p_{50\%}} \right)^2 \cdot e^{-(u_{p50\%})^2} = \left(\frac{1,5}{2 \cdot 0,025} \right)^2 \cdot e^{-(1,96)^2} = 19,3.$$

Die Kontrollen müssen also wie folgt durchgeführt werden: Man entnimmt der zu prüfenden Partie eine Stichprobe vom Umfang $n = 20$ und berechnet deren Mittelwert \bar{x} . Ist die absolute Abweichung vom Sollwert w größer als $d - 1,96 \cdot \sigma$, so wird die Partie abgelehnt, und sonst angenommen.

Auch für quantitative Merkmale lassen sich mehrfache und **sequentielle Prüfpläne** entwickeln. Hier wollen wir uns auf die Vorstellung eines sequentiellen Verfahrens beschränken. Wie zuvor sei $d \geq 3\sigma$ vorausgesetzt. Sind x_1, \dots, x_k die Prüfergebnisse der ersten k Stichprobenelemente, so werden die einzelnen Abweichungen $x_i - w$ vom Sollwert w berechnet und summiert. Ist nun

$$(d + \sigma c)k + \sigma b \leq \left| \sum_{i=1}^k (x_i - w) \right|,$$

so lehnt man die Partie ab, ist

$$\left| \sum_{i=1}^k (x_i - w) \right| \leq (d + \sigma c)k - \sigma a,$$

so wird die Partie angenommen und bei

$$(d + \sigma c)k - \sigma a < \left| \sum_{i=1}^k (x_i - w) \right| < (d + \sigma c)k + \sigma b$$

muß ein weiteres Stück geprüft werden. Die Konstanten a , b und c müssen so gewählt sein, daß das Prüfverfahren die geforderten Güteeigenschaften besitzt. Soll die OC an der Annahmegrenze $p_{1-\alpha}$ den Wert $1 - \alpha$ und an der Ablehngrenze p_β den Wert β annehmen, so werden die gestellten Anforderungen ungefähr erfüllt durch (wegen $d \geq 3\sigma$)

$$c = \frac{u_{p_{1-\alpha}} + u_{p_\beta}}{2}, \quad a = \frac{-\ln \frac{\beta}{1-\alpha}}{u_{p_\beta} - u_{p_{1-\alpha}}} \quad \text{und} \quad b = \frac{\ln \frac{1-\beta}{\alpha}}{u_{p_\beta} - u_{p_{1-\alpha}}} \quad (= a \text{ für } \alpha = \beta).$$

Die OC ergibt sich in Abhängigkeit des Ausschubanteils p in der zu prüfenden Partie als

$$L(p) = \frac{1 - \left(\frac{1-\beta}{\alpha} \right)^h}{\left(\frac{\beta}{1-\alpha} \right)^h - \left(\frac{1-\beta}{\alpha} \right)^h} \quad \text{mit} \quad h = \frac{u_{p_{1-\alpha}} + u_{p_\beta} - 2 \cdot u_p}{u_{p_\beta} - u_{p_{1-\alpha}}}.$$

Für den zu erwartenden Stichprobenumfang erhält man (falls $u_{p_{1-\alpha}} + u_{p_\beta} \neq 2u_p$):

$$n(p) = \frac{L(p) \ln \left(\frac{\beta}{1-\alpha} \right) + (1-L(p)) \ln \left(\frac{1-\beta}{\alpha} \right)}{(u_{p_\beta} - u_{p_{1-\alpha}})(u_{p_{1-\alpha}} + u_{p_\beta} - 2u_p)/2}.$$

Beispiel: Es soll ein sequentieller Prüfplan erstellt werden, dessen OC an der Annahmegrenze $p_{1-\alpha} = 1\%$ den Wert $1 - \alpha = 90\%$ und an der Ablehngrenze $p_\beta = 2\%$ den Wert $\beta = 10\%$ annimmt. Man berechnet

$$c = \frac{u_{p_{1-\alpha}} + u_{p_\beta}}{2} = \frac{-2,326 - 2,054}{2} = -2,19 \text{ und}$$

$$a = \frac{-\ln \frac{\beta}{1-\alpha}}{u_{p_\beta} - u_{p_{1-\alpha}}} = \frac{-\ln \frac{0,1}{0,9}}{-2,054 - (-2,326)} = 8,08 = b.$$

Die Prüfvorschrift lautet also:

Bestimme nach jeder Entnahme eines Stichprobenelements die Abweichung des Meßwertes x_k vom Sollwert w und addiere sie zur Summe der Abweichungen aus den vorangegangen $k - 1$ Messungen. Der Betrag dieser Summe

$$\left| \sum_{i=1}^k (x_i - w) \right|$$

wird nun verglichen mit

$$(d + \sigma c)k - \sigma a = (d - 2,19 \cdot \sigma)k - 8,08 \cdot \sigma$$

$$\text{und } (d + \sigma c)k + \sigma b = (d - 2,19 \cdot \sigma)k + 8,08 \cdot \sigma.$$

Liegt er unterhalb des ersten, kleineren Wertes oder ist gleich diesem, so nimmt man die Partie an, ist er größer oder gleich dem zweiten Wert, so wird die Partie abgelehnt und sonst, d. h. falls

$$(d - 2,19 \cdot \sigma)k - 8,08 \cdot \sigma < \left| \sum_{i=1}^k (x_i - w) \right| < (d - 2,19 \cdot \sigma)k + 8,08 \cdot \sigma$$

gilt, wird weitergeprüft.

Diese Entscheidungen lassen sich auch mit Hilfe eines graphischen Verfahrens herbeiführen. Dazu werden in ein Diagramm die beiden durch

$$(d - 2,19 \cdot \sigma)k - 8,08 \cdot \sigma \text{ und } (d - 2,19 \cdot \sigma)k + 8,08 \cdot \sigma$$

festgelegten Geraden eingezeichnet. Man trägt nun nach jeder Prüfung eines Stichprobenelements einen Punkt in das Diagramm ein, dessen Koordinaten durch die Anzahl k der bis dahin durchgeföhrten Messungen und den dazugehörigen Wert von $\left| \sum_{i=1}^k (x_i - w) \right|$ gegeben sind. Solange man Punkte zwischen den beiden Geraden einzeichnet, wird weiter kontrolliert, doch sobald eine der beiden Geraden überschritten wird, kommt es zur Entscheidung. Ist $\sigma = 0,0207$ und $d = 0,081$, so erhält man die beiden Geraden

$$(0,081 - 2,19 \cdot 0,0207)k - 8,08 \cdot 0,0207 = 0,036k - 0,167$$

$$\text{und } (0,081 - 2,19 \cdot 0,0207)k + 8,08 \cdot 0,0207 = 0,036k + 0,167,$$

die in Abb. 10 eingetragen sind. Die dort zusätzlich eingezeichneten Datenpunkte führen zur Ablehnung der Partie, denn sie überschreiten die durch $0,036k + 0,167$ gegebene Linie.

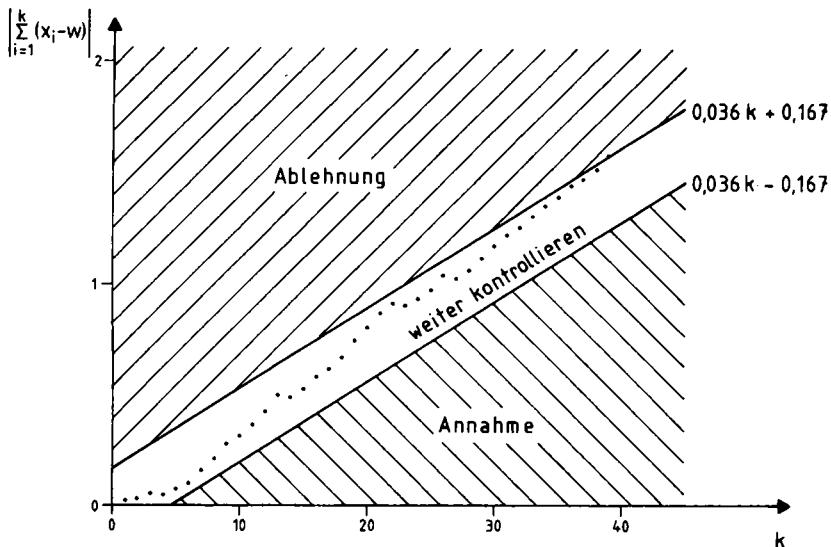


Abb. 10: Stichprobenergebnis bei einem sequentiellen Prüfplan für quantitative Merkmale mit $c = 0,036$ und $\sigma_a = \sigma_b = 0,167$

2. Laufende Kontrolle der Produktion (Kontrollkarten)

Der Produktion werden laufend kleine Stichproben vom Umfang n entnommen, aufgrund derer entschieden wird, ob die Fertigung noch wie gewünscht verläuft oder ob Störungen aufgetreten sind, die ein Eingreifen in den Produktionsprozeß erforderlich machen.

2.1 Laufende Kontrolle bei quantitativen Merkmalen

Wir wollen zunächst den Fall betrachten, daß die einzelnen Produkte durch eine Größe charakterisiert werden können, die normalverteilt ist mit dem Mittelwert μ und der Varianz σ^2 . Aus dem Stichprobenergebnis x_1, \dots, x_n (in der Praxis meist n zwischen 4 und 12) soll nun mit Hilfe statistischer Tests auf etwaige Veränderungen der Parameter μ und/oder σ^2 geschlossen werden. Die Sollwerte μ_0 und σ_0 für μ und σ werden entweder vorgegeben oder aus einem vorherigen Produktionsablauf mittels m Stichproben vom Umfang n geschätzt. Um hinreichend genaue Schätzungen zu bekommen, sollten mindestens $n \cdot m \geq 200$ Meßwerte vorliegen. Diese Schätzwerte sollten gelegentlich überprüft und gegebenenfalls korrigiert werden. \bar{x}_j bezeichne den Mittelwert, s_j die (empirische) Standardabweichung und R_j die Spannweite der j -ten dieser m Stichproben vom konstanten Umfang n . Als Schätzwert für μ_0 benutzen wir das Gesamtmittel

$$\bar{x} = \frac{1}{m} \sum_{j=1}^m \bar{x}_j$$

und für σ_0 entweder

$$\bar{s}' = \gamma_n \bar{s} = \frac{\gamma_n}{m} \sum_{j=1}^m s_j \quad \text{oder} \quad \bar{s}'' = \bar{R}/\alpha_n = \frac{1}{\alpha_n \cdot m} \sum_{j=1}^m R_j,$$

vgl. Kapitel IV, Abschnitt 1.3, mit den Korrekturfaktoren γ_n und $1/\alpha_n$, aber in der Regel – wegen des größeren Rechenaufwandes – nicht die bei kleinem n nur geringfügig bessere Ergebnisse liefernde, mit γ_{mn} korrigierte empirische Streuung der gesamten $n \cdot m$ großen Stichprobe.

Zur Überprüfung des Mittelwertes μ bei einer laufenden Produktion benutzt man z. B. das Mittel \bar{x} des Stichprobenergebnisses x_1, \dots, x_n als Testgröße und lehnt die Hypothese $\mu = \mu_0$ ab, wenn

$$\bar{x} \leq \mu_0 - c \frac{\sigma}{\sqrt{n}} \quad \text{oder} \quad \bar{x} \geq \mu_0 + c \frac{\sigma}{\sqrt{n}}.$$

Dabei werden die Testschranken von zwei unterschiedlich scharfen Tests benutzt. Liegt das Stichprobenmittel im Ablehnbereich des schärferen Tests, so führt das unmittelbar zum Eingreifen in die Produktion, während nur Anlaß zu „erhöhter Aufmerksamkeit“ besteht bzw. eine zusätzliche Stichprobe entnommen wird, wenn \bar{x} allein im Ablehnbereich des schwächeren Tests liegt. Die Testschranken des schwächeren Tests nennt man deshalb **Warngrenzen** und die des schärferen **Kontroll- oder Regelgrenzen**. Es ist üblich, die Niveaus der beiden Tests mit 5% und 1% festzulegen, was gleichbedeutend ist mit Werten der Operationscharakteristiken an der Stelle μ_0 von 95% und 99%;

$$\mu_0 - c \frac{\sigma}{\sqrt{n}} \quad \text{und} \quad \mu_0 + c \frac{\sigma}{\sqrt{n}}$$

liefern dann mit $c = 1,96$ die Warn- und mit $c = 2,58$ die Regelgrenzen. (Oft wird auch $c = 2$ für die Warn- und $c = 3$ für die Regelgrenze benutzt.) Für die normalerweise in der Praxis auftretenden Stichprobenumfänge können μ_0 durch \bar{x} und/oder σ durch s' oder s'' ersetzt werden, ohne daß sich die Testniveaus wesentlich verändern.

Beispiel: Bei einer Produktion von Widerständen soll mit Hilfe von Kontrollkarten überwacht werden, ob der Mittelwert die gewünschten $10 \text{ k}\Omega$ einhält. Es ist bekannt, daß der Fertigungsprozeß mit einer Standardabweichung $\sigma = 0,5 \text{ k}\Omega$ läuft. Der Produktion werden stündlich Stichproben vom Umfang $n = 5$ entnommen und deren Mittelwerte in eine Kontrollkarte eingetragen, auf der die Warn- und Regelgrenzen durch waagerechte Linien angedeutet sind. Wie müssen diese Grenzen hier nun gewählt sein? Die Warngrenzen zur Sicherheitswahrscheinlichkeit 95% ergeben sich zu

$$\mu_0 - 1,96 \frac{\sigma}{\sqrt{n}} = 10 - 1,96 \frac{0,5}{\sqrt{5}} = 9,56$$

$$\text{und} \quad \mu_0 + 1,96 \frac{\sigma}{\sqrt{n}} = 10 + 1,96 \frac{0,5}{\sqrt{5}} = 10,44,$$

während man als Kontrollgrenzen zur Sicherheitswahrscheinlichkeit 99% erhält:

$$\mu_0 - 2,58 \frac{\sigma}{\sqrt{n}} = 10 - 2,58 \frac{0,5}{\sqrt{5}} = 9,43$$

$$\text{und} \quad \mu_0 + 2,58 \frac{\sigma}{\sqrt{n}} = 10 + 2,58 \frac{0,5}{\sqrt{5}} = 10,57.$$

Solange die Mittelwerte der Stichproben innerhalb der Warngrenzen liegen, also für

$$9,56 < \bar{x} < 10,44,$$

kann man annehmen, daß die Fertigung normal, d. h. mit dem geforderten Mittelwert verläuft. Liegt \bar{x} zwar noch innerhalb der Regel-, aber nicht mehr innerhalb der Warngrenzen

$$9,43 < \bar{x} \leq 9,56 \quad \text{oder} \quad 10,44 \leq \bar{x} < 10,57,$$

so kann das ein erstes Anzeichen für eine Störung im Produktionsverlauf, also eine Veränderung des Mittelwertes, sein. Man wird daher sofort eine zusätzliche Stichprobe untersuchen und die Fertigung stoppen, wenn auch der Mittelwert dieser Stichprobe außerhalb der Warngrenzen liegt. Ist er dagegen innerhalb dieser Grenzen, so kann der Verdacht einer Störung als nicht ausreichend begründet abgewiesen werden. Sobald ein Stichprobenmittelwert sogar außerhalb der Regelgrenzen liegt, ist ein unmittelbares Eingreifen in den Produktionsprozeß erforderlich. In Abbildung 11 tritt dieser Fall bei der 7. Stichprobe auf. Bei der 11. Stichprobe sind nur die Warngrenzen überschritten worden, so daß eine zusätzliche Stichprobe gezogen wurde, deren Mittelwert durch einen Kreis angekennzeichnet ist. Da dieser innerhalb der Warngrenzen liegt, kann die Fertigung ungestört weiterlaufen.

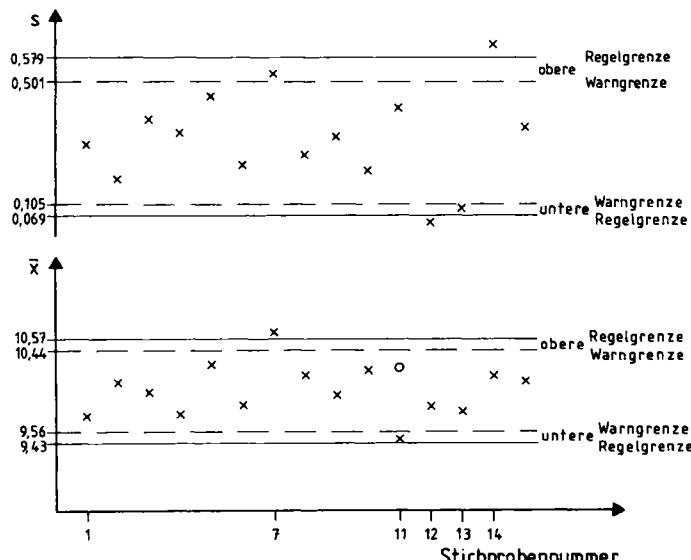


Abb. 11: Kontrollkarte für den Mittelwert \bar{x} und die Streuung s

Zur Überprüfung der Streuung kann man als Testgröße die empirische Streuung s des Stichprobenergebnisses x_1, \dots, x_n verwenden. Als Warn- bzw. Kontrollgrenzen ergeben sich dann jeweils

$$\sigma_0 \sqrt{\frac{1}{n-1} \chi^2_{n-1; \alpha/2}} \quad \text{und} \quad \sigma_0 \sqrt{\frac{1}{n-1} \chi^2_{n-1; 1-\alpha/2}},$$

wobei für α die jeweiligen Testniveaus einzusetzen sind. σ_0 kann auch hier in der Regel durch geschätzte Werte ersetzt werden.

Man wird zunächst meinen, die unteren Warn- und Regelgrenzen für die Streuung wären überflüssig, da ein Unterschreiten dieser Grenzen auf einen sehr gleichmäßig verlaufenden Produktionsprozeß schließen läßt, der Erzeugnisse liefert, die die gestellten Qualitätsansprüche in hohem Maße befriedigen. Diese unnötig hohe Qualität kann nun durch ein zu langsames und damit teures Arbeiten entstanden sein, was schließlich zu einer Verteuerung der Produktion und damit geringerer Konkurrenzfähigkeit führen kann. Die kleinen Werte für die empirische Streuung können andererseits durch mangelnde Aufmerksamkeit oder bewußte Manipulation der Kontrolleure zustandegekommen sein. Deshalb kann es sinnvoll sein, auch mit unteren Warn- und Regelgrenzen zu arbeiten.

Beispiel: Zur Überprüfung der Streuung soll eine Kontrollkarte für einen Stichprobenumfang $n = 5$ entworfen werden. Es ist ein Wert $\sigma_0 = 0,3$ gewünscht. Man erhält für die Warngrenzen (Sicherheitswahrscheinlichkeit 95%)

$$\sigma_0 \sqrt{\frac{1}{n-1} \chi^2_{n-1; 0,05/2}} = 0,3 \cdot 0,35 = 0,105$$

$$\text{und } \sigma_0 \sqrt{\frac{1}{n-1} \chi^2_{n-1; 1 - 0,05/2}} = 0,3 \cdot 1,67 = 0,501,$$

während die Regelgrenzen (Sicherheitswahrscheinlichkeit 99%) die Werte

$$\sigma_0 \sqrt{\frac{1}{n-1} \chi^2_{n-1; 0,01/2}} = 0,3 \cdot 0,23 = 0,069$$

$$\text{und } \sigma_0 \sqrt{\frac{1}{n-1} \chi^2_{n-1; 1 - 0,01/2}} = 0,3 \cdot 1,93 = 0,579$$

annehmen; vgl. auch Abb. 11.

Bei der oft wünschenswerten gleichzeitigen Kontrolle von Mittelwert und Streuung muß man sich darüber im klaren sein, daß die Wahrscheinlichkeit, daß sowohl \bar{x} als auch s innerhalb der Warn- bzw. Regelgrenzen liegen, nur $0,95^2 = 90,25\%$ bzw. $0,99^2 = 98,01\%$ beträgt.

Andere, teilweise leichter zu bestimmende Testgrößen wie etwa Spannweite, Median, Extremwert usw. können auch zur Kontrolle eines Produktionsprozesses herangezogen werden, sollen hier aber nicht mehr besprochen werden.

2.2 Laufende Kontrolle bei qualitativen Merkmalen

Wir wollen uns nun noch mit der Kontrolle von Erzeugnissen beschäftigen, die durch qualitative Merkmale wie etwa Farbe, Oberflächengüte usw. charakterisiert werden und dort die Anzahl m der schlechten Stücke in einer Stichprobe vom Umfang n (meist n zwischen 20 und 60) als Testgröße benutzen (vgl. auch Abschnitt 1.1). Soll der vorgegebene bzw. geschätzte bisherige Ausschußanteil p_0 nicht überschritten werden, so erhält man die passende Warn- bzw. Kontrollgrenze als kleinstmögliche natürliche Zahl c mit

$$\sum_{m=0}^c \binom{n}{m} p_0^m (1-p_0)^{n-m} \geq 1 - \alpha,$$

wobei $1 - \alpha$ der Wert der OC an der Stelle p_0 ist, oft 95% bzw. 99%. Hierbei kann es durchaus passieren, daß bei sehr kleinem n Warn- und Kontrollgrenze zusammenfallen. Es sei noch darauf hingewiesen, daß sich bei größerem n die Testschranke c näherungsweise durch die Poisson- oder Normalapproximation finden läßt (vgl. Abschnitt 1.1). Die Kontrollergebnisse werden in Kontrollkarten eingetragen, in die vorher schon die Warn- und Kontrollgrenzen eingezeichnet wurden. Verwenden kann man dabei etwa die von der Deutschen Gesellschaft für Qualität e. V. herausgegebenen DGQ-Qualitätsregelkarten (DGQ 178), die vom Beuth-Verlag unter der Bestellnummer 32165 vertrieben werden. In Abbildung 12 ist eine solche ausgefüllte Karte wiedergegeben. Dort ist $n = 30$, $p_0 = 0,02$ und $\alpha = 0,05$ bzw. $\alpha = 0,01$ gewählt worden. Da die Binomialverteilung diskret ist, sind die exakten Niveaus von Warngrenze ($c = 2$) und Kontrollgrenze ($c = 3$) hier nicht 5% und 1%, sondern 2,171% und 0,03%. Die Ergebnisse zusätzlicher – wegen Überschreitung der Warmlinie gezogener Stichproben – sind in Abb. 12 mit einem Kreis ange deutet. Wird dem Verantwortlichen derart das Qualitätsergebnis sichtbar gemacht, so führt dies meist zu einem sorgfältigeren Arbeiten und damit zu einer deutlichen Qualitätsverbesserung.

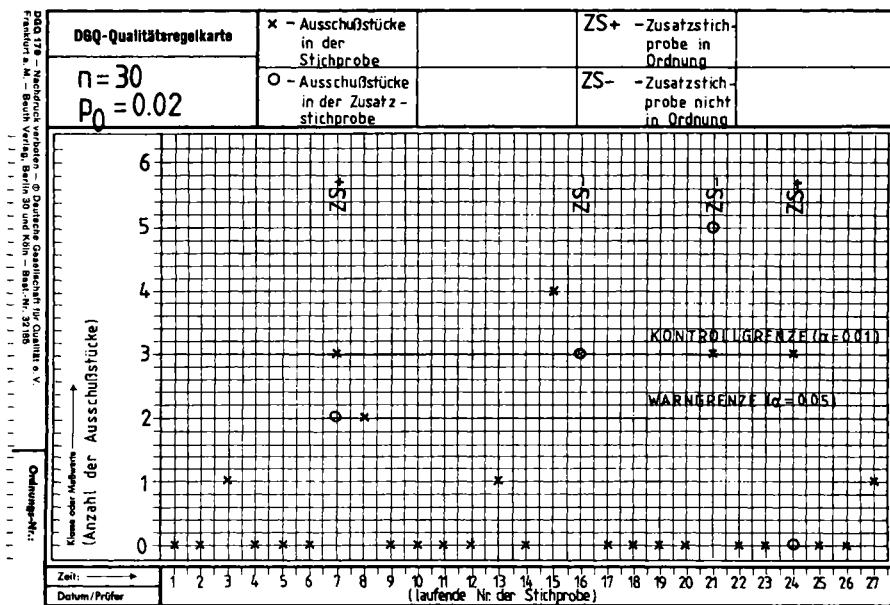


Abb. 12: Kontrollkarte für die Wahrscheinlichkeit p mit Warn- und Kontrollgrenzen

3. Kontinuierliche Stichprobenpläne

Mit den sowohl bei der laufenden Kontrolle der Produktion als auch in der Eingangs- und Endkontrolle anwendbaren **kontinuierlichen Stichprobenplänen (continuous sampling plan = CSP)** soll durch eine in der Intensität wechselnde, in der Fertigungsreihenfolge ablaufende Kontrolle – verbunden mit einem Reparieren bzw. Austauschen der schlechten Stücke – sichergestellt werden, daß der durchschnittliche Ausschußprozentsatz nach der Kontrolle einen vorgegebenen Wert nicht übersteigt. Der erste und wohl einfachste Plan (**CSP-1**) wurde von Dodge (1943) veröffentlicht. Man beginnt mit einer Totalkontrolle, prüft also in der Reihenfolge der Fertigung jedes Stück daraufhin, ob es gut oder schlecht ist. Hat man i -mal hintereinander gute Stücke gefunden, so geht man zu einer Stichprobenkontrolle über, bei der von jeweils k Stücken nur noch eines geprüft wird. Sobald jedoch ein Fehler gefunden wird, kommt wieder die Totalkontrolle zur Anwendung und zwar solange, bis man i -mal hintereinander gute Stücke gefunden hat usw. Ist der durchschnittliche Ausschußanteil konstant gleich einem meist unbekannten Wert p – man sagt, der Produktionsprozeß ist unter statistischer Kontrolle –, so ergibt sich als **AOQ**, dem **durchschnittlich ausgelieferten Ausschußanteil**:

$$\frac{(k-1)p(1-p)^i}{1+(k-1)(1-p)^i}.$$

Dieser Ausdruck wird am größten und ist dann gleich dem **AOQL**, dem **Höchstwert des mittleren Durchschlupfes**, wenn p die Gleichung

$$(1+i) \cdot p - 1 = (k-1)(1-p)^{i+1}$$

löst. Gibt man sich nun einen Wert für den AOQL vor, sagen wir l , so muß bei vorgegebenem i

$$k \approx 1 + \frac{il}{\left(1 - \frac{1+il}{1+i}\right)^{i+1}}$$

gewählt werden. Der ausgelieferte Ausschußanteil l ergibt sich dann gerade für

$$p = \frac{1+il}{1+i}.$$

Je nach Vorgabe von i erhält man so verschiedene Prüfpläne mit gleichem AOQL, die sich jedoch im durchschnittlichen Prüfaufwand unterscheiden. In Abhängigkeit vom angelieferten Ausschußanteil p erhält man für den Prüfanteil:

$$\frac{1}{1+(k-1)(1-p)^i}.$$

Es sei noch erwähnt, daß mit diesem Prüfverfahren der durchschnittlich ausgelieferte Ausschußanteil selbst dann durch $\frac{k-1}{k+i}$ begrenzt bleibt, wenn der Produktionsprozeß nicht unter statistischer Kontrolle ist.

Zahlreiche weitere kontinuierliche Stichprobenpläne sind inzwischen entwickelt und untersucht worden. Dabei werden zum Teil andere Kriterien für den Wechsel der Kontrollintensität herangezogen (**sampling plan A = SPA** (vgl. Wald/Wolfowitz (1945)); **sampling plan B = SPB; SPC, ...**) oder man benutzt noch weitere Stichprobenanteile (**multi-level continuous sampling plans = MLP**).

Kapitel VII: Analyse diskreten Datenmaterials in Form von Kontingenztafeln

Wir haben uns in den übrigen Kapiteln hauptsächlich mit der Analyse von quantitativen Datenmaterial beschäftigt. Dabei sind insbesondere Zusammenhänge zwischen verschiedenen Größen untersucht worden. Die Vielfalt der verwendeten Methoden reicht von der Bestimmung eines Korrelationskoeffizienten über die einfache Regressionsanalyse bis hin zu den komplizierteren Verfahren der Varianzanalyse.

In diesem Kapitel sollen nun Methoden bereitgestellt werden, die es uns gestatten, auch bei **nominal skalierten Größen** Zusammenhänge zu erfassen. Solche Größen sind z. B. Farben, Hersteller von Autos oder Waschmaschinen, Geschmacksbeschreibungen von Speisen und Getränken, Meinungen und Ansichten von Bürgern und Politikern, Oberflächenbeschaffenheit von Metallteilen oder auch nicht-quantifizierbare Angaben über den Gesundheitszustand.

Die hier vorgestellten Verfahren lassen sich auch bei quantitativem Datenmaterial verwenden, indem Klasseneinteilungen vorgenommen werden und dann mit den Besetzungszahlen dieser Klassen weitergearbeitet wird.

Wir wollen nun die unterschiedlichen Fragestellungen, die bei qualitativen und quantitativen Größen auftreten können, an zwei **Beispielen** erläutern.

Tab. 1: Gewicht und Größe von 9 männlichen Personen

Person	1	2	3	4	5	6	7	8	9
Gewicht (in kg)	87	79	62	71	70	62	84	80	81
Größe (in cm)	180	179	165	167	176	169	177	183	174

In **Tab. 1** sind Gewicht und Größe von 9 männlichen Personen angegeben. Zwischen dem Gewicht g bzw. dessen Erwartungswert und der Größe h wird man eine lineare Beziehung der Form

$$g = ah + b$$

vermuten. Hat man nun die Gültigkeit dieser Beziehung überprüft und die Konstanten a und b aufgrund der vorliegenden Daten geschätzt, vgl. Kap. X, so lässt sich zu jeder Größe h das Gewicht g schätzen. Das gilt auch für solche Größen h , die in Tab. 1 überhaupt nicht auftauchen.

Solche Prognosen sind bei nominalen Daten nicht möglich. Betrachten wir dazu ein Beispiel. In **Tab. 2** sind für einen bestimmten Zeitabschnitt die Verkaufszahlen eines Autoherstellers angegeben. Dabei wird unterschieden, ob es sich um eine Limousine oder ein Coupé handelt, und es wird zusätzlich als zweites Merkmal die Farbe angegeben.

Konnten wir uns beim obigen Beispiel fragen, welche funktionale Beziehung zwischen den beiden Merkmalen bzw. deren Erwartungswerten besteht, so kann man hier zunächst nur die Frage stellen, ob überhaupt ein Zusammenhang zwischen den Merkmalen vorhanden ist. Bei der Beantwortung dieser Frage gehen wir

Tab. 2: Verkaufszahlen eines Autoherstellers

Art \ Farbe	rot	grün	gelb	blau
Limousine	358	783	523	872
Coupé	147	62	126	281

von der folgenden Überlegung aus: Bezeichnen wir in unserem Beispiel mit p_l die Wahrscheinlichkeit, daß eine Limousine verkauft wird, und mit p_b die Wahrscheinlichkeit des Verkaufs eines blauen Fahrzeugs, so müßte bei Unabhängigkeit der beiden Merkmale Art und Farbe des Autos die Wahrscheinlichkeit p_{lb} für den Verkauf einer blauen Limousine gegeben sein durch

$$p_{lb} = p_l \cdot p_b.$$

Für die anderen Merkmalsausprägungen könnte man analoge Beziehungen aufstellen. Abhängigkeiten der beiden Merkmale sind vorhanden, wenn einige dieser Beziehungen nicht mehr erfüllt sind. Dies äußert sich dann auch bei den beobachteten Häufigkeiten. So scheint in unserem Beispiel bei der Limousine im Gegensatz zum Coupé die Farbe grün sehr beliebt zu sein; dagegen entscheiden sich beim Coupé relativ viele Kunden für die Farbe rot. Die Beliebtheit der einzelnen Farben scheint also davon abhängig zu sein, ob es sich um eine Limousine oder um ein Coupé handelt.

Eine wichtige Aufgabe dieses Kapitels besteht nun darin zu testen, ob tatsächlich solche **Zusammenhänge** bestehen oder ob es sich dabei nur um zufällige Schwankungen handelt. In Abschnitt 3 werden wir außerdem **Maßzahlen** für die Stärke solcher Zusammenhänge kennenlernen.

Kehren wir nun noch einmal zu unserem ersten Beispiel zurück. Teilen wir das Gewicht und die Größe der männlichen Personen in mehrere Klassen ein, so können wir die Ergebnisse in einer **Kontingenztafel** darstellen. Für eine recht grobe Einteilung, nämlich nur zwei Klassen für jedes Merkmal, ist dies in **Tab. 3** geschehen.

Tab. 3: Gewicht und Größe von 9 männlichen Personen in je zwei Klassen eingeteilt, vgl. auch Tab. 1

Größe \ Gewicht	≤ 75	> 75
≤ 175	3	1
> 175	1	4

Auch hier ist noch zu erkennen, daß zwischen Gewicht und Größe ein Zusammenhang besteht: Männliche Personen mit einer Körpergröße über 175 cm scheinen zumeist auch schwerer als 75 kg zu sein, während kleinere Personen in der Regel wohl leichter sind. Ein linearer Zusammenhang der beiden Merkmale ist hier (bei Klassenbildung) jedoch nicht mehr zu erkennen.

Die im Beispiel durchgeführte **Senkung des Skalenniveaus** kann vor allem dann sinnvoll sein, wenn wir Zusammenhänge unterschiedlich skalierten Merkmale un-

tersuchen wollen. Mit dieser Prozedur können wir nämlich ein einheitliches Skalenniveau herstellen. Das wird in den meisten Fällen die Nominalskala sein. Nach einer solchen **Niveaualgleichung** können dann die Methoden dieses Kapitels zur Auswertung herangezogen werden.

In den Beispielen haben wir es bisher immer nur mit zwei Merkmalen zu tun gehabt. Dadurch konnten wir uns im wesentlichen auf die Erörterung der Frage beschränken, ob **Abhängigkeit** vorliegt oder nicht.

Bei mehreren Merkmalen stellt sich sogleich die Frage, welche **Abhängigkeiten** denn vorliegen. Dies führt zu einer großen Anzahl von möglichen Abhängigkeits-hypothesen, die in Abschnitt 2 mit Hilfe des **loglinearen Modells** formuliert und dann auch getestet werden.

Es wird auffallen, daß wir bei den bisherigen Ausführungen alle Merkmale gleich behandelt haben. Wir haben nicht, wie z. B. bei der **Varianzanalyse**, vgl. Kap. XI, nur die Abhängigkeit eines Beobachtungswertes von mehreren Einflußgrößen untersucht. Das soll nun geschehen. Dazu betrachten wir das **Beispiel aus Tab. 4**.

Tab. 4: Erfolge und Mißerfolge von Operationstechniken

Erfolg				Mißerfolg			
Klinik				Klinik			
Operations-technik	1	2	3	Operations-technik	1	2	3
1	37	128	97	1	12	15	8
2	21	82	157	2	3	9	17

Dort sind die Anzahlen der Erfolge und Mißerfolge von zwei Operationstechniken in drei verschiedenen Kliniken angegeben. Wir könnten jetzt ganz allgemein untersuchen, ob zwischen den drei Merkmalen Klinik, Operationstechnik, Erfolg irgendwelche Zusammenhänge bestehen und herausfinden, welche das im einzelnen sind. Da wohl im wesentlichen interessiert, ob und welchen Einfluß die beiden Faktoren Klinik und Operationstechnik auf den Ausgang der Behandlung haben, läßt sich hier auch eine spezielle, gezieltere Vorgehensweise anbringen, nämlich die Verwendung eines **Logit-Modells**. Dabei hat ein Merkmal, in unserem Beispiel der Erfolg, eine besondere Stellung: Alle anderen Merkmale werden nur zur Erklärung dieses einen herangezogen.

An dieser Stelle können Parallelen zu Modellen der Varianzanalyse gezogen werden. Betrachten wir dazu folgendes

Beispiel: Es soll mittels eines Modells der Varianzanalyse untersucht werden, wie der Ertrag einer Getreidesorte von den Einflußfaktoren Düngung und Bodenbeschaffenheit abhängt. Während wir hierbei den Erfolg in einer metrischen Skala (Angabe des Ertrages) messen können, haben wir in dem obigen Beispiel nur die Merkmalsausprägungen Erfolg und Mißerfolg betrachtet. Aber auch dabei läßt sich eine quantitative Aussage machen, indem man die Wahrscheinlichkeit für den Erfolg angibt. Diese Wahrscheinlichkeit hat ihre Entsprechung in den erwarteten Erträgen, während die beobachteten relativen Häufigkeiten ihre Entsprechung in den tatsächlich aufgetretenen Erträgen haben.

Wie schon die Bezeichnung Logit-Modell andeutet, werden die Wahrscheinlichkeiten und die relativen Häufigkeiten in einer logarithmischen Skala gemessen.

Das Logit-Modell für die Kontingenztafelanalyse unterscheidet sich vom Modell I der Varianzanalyse, vgl. Kap. XI, im wesentlichen durch die anderen Verteilungsannahmen. Hier handelt es sich um diskrete Verteilungen, während dort stetige Verteilungen, meist irgendwelche Normalverteilungen, angenommen werden. Das Ziel ist jedoch immer das gleiche: Es sollen die zufälligen Schwankungen bzw. Fehler von den wirklichen Einflüssen getrennt werden. Diese Einflüsse werden bei den Kontingenztafeln durch theoretische Wahrscheinlichkeiten und beim Modell I der Varianzanalyse durch Erwartungswerte von in der Regel normalverteilten Zufallsgrößen beschrieben.

Zum Schluß wollen wir noch an einem **Beispiel** andeuten, wie Kontingenztafeln bei **Prognosen** helfen können. Dazu muß zunächst die Tafel der theoretischen Wahrscheinlichkeiten, vgl. **Tab. 5**, mit Hilfe der beobachteten Häufigkeiten geschätzt werden. Nehmen wir einmal an, dabei hätten sich die Daten aus **Tab. 6** ergeben.

Tab. 5: 3×3 Tafel von theoretischen Wahrscheinlichkeiten

		Merkmal 2	y_1	y_2	y_3
		Merkmal 1			
	x_1		p_{11}	p_{12}	p_{13}
	x_2		p_{21}	p_{22}	p_{23}
	x_3		p_{31}	p_{32}	p_{33}

Tab. 6: Beispiel einer 3×3 Tafel von relativen Häufigkeiten

		Merkmal 2	y_1	y_2	y_3
		Merkmal 1			
	x_1		0,027	0,250	0,051
	x_2		0,127	0,040	0,170
	x_3		0,180	0,045	0,110

Weiß man noch gar nichts, so wird man die in **Tab. 7** wiedergegebenen Kombinationen von Merkmalsausprägungen als wahrscheinlichste Ausgänge in Betracht ziehen.

Tab. 7: Wahrscheinlichste Ausgänge eines Experimentes nach Tab. 6

Merkmal 1	x_1	x_2	x_2	x_3	x_3
Merkmal 2	y_2	y_1	y_3	y_1	y_3

Mit einer Wahrscheinlichkeit von

$$0,250 + 0,127 + 0,170 + 0,180 + 0,110 = 0,837$$

wird man eine dieser fünf Kombinationen erwarten können. Kennt man schon die Ausprägung eines Merkmals, hat man etwa x_1 beobachtet, so erwartet man für das Merkmal 2 die Ausprägung y_2 . Die unter x_1 bedingte Wahrscheinlichkeit für das Auftreten von y_2 ist nämlich

$$\frac{0,250}{0,027 + 0,250 + 0,051} = 0,762.$$

Somit ist bei dieser Vorgehensweise die Wahrscheinlichkeit für eine richtige Prognose gerade 0,762.

(Auch in diesem Kapitel werden wir aus Notationsgründen bezeichnungsmäßig nicht immer zwischen Zufallsvariablen und ihren Realisationen unterscheiden.)

1. Die 2×2 -Felder-Tafel

In einer Studie wurde neben den Rauchgewohnheiten auch das Geschlecht der befragten $n = 250$ Personen erhoben. Es waren 45 weibliche Raucher und 80 weibliche Nichtraucher unter den Befragten. Die Anzahl der männlichen Raucher betrug 55 und die der männlichen Nichtraucher 70. Dieses Ergebnis lässt sich übersichtlich in einer **2×2 -Kontingenztafel** (auch **Vierfeldertafel** genannt) anordnen, wie in **Tab. 8** geschehen; steht ein Punkt anstelle eines Laufindexes, so bedeutet dies, daß bzgl. dieses Indexes summiert wird.

Tab. 8: Beziehung zwischen Geschlecht und Rauchgewohnheit

Geschlecht \ Rauchgewohnheit			Σ
	Raucher	Nichtraucher	
weiblich	$n_{11} = 45$	$n_{12} = 80$	$n_{1 \cdot} = 125$
männlich	$n_{21} = 55$	$n_{22} = 70$	$n_{2 \cdot} = 125$
Σ	$n_{\cdot 1} = 100$	$n_{\cdot 2} = 150$	$n = 250$

Zu dieser Kontingenztafel der empirischen Häufigkeiten n_{ij} korrespondiert eine Tafel der theoretischen Wahrscheinlichkeiten p_{ij} bzw. der erwarteten Häufigkeiten $m_{ij} = np_{ij}$, siehe **Tab. 9**.

Tab. 9: Zu Tab. 8 korrespondierende Tafeln der theoretischen Wahrscheinlichkeiten p_{ij} und der erwarteten Häufigkeiten m_{ij}

RG \ G	Σ			Σ	Σ			
	R	N	Σ		G	R	N	
W	p_{11}	p_{12}	$p_{1 \cdot}$		W	m_{11}	m_{12}	$m_{1 \cdot}$
M	p_{21}	p_{22}	$p_{2 \cdot}$		M	m_{21}	m_{22}	$m_{2 \cdot}$
Σ	$p_{\cdot 1}$	$p_{\cdot 2}$	1		Σ	$m_{\cdot 1}$	$m_{\cdot 2}$	n

Dabei steht bei der Angabe des Geschlechts G ein W für weiblich und ein M für männlich. Bei der Rauchgewohnheit RG steht R für Raucher und N für Nichtraucher.

Für die uns unbekannten Werte p_{ij} bzw. np_{ij} sollen nun Hypothesen aufgestellt und getestet werden.

1.1. Hypothesen für die 2×2 -Felder-Tafel

A. Die Unabhängigkeitshypothese

Nehmen wir einmal an, bei der Untersuchung wurden die $n = 250$ Personen mit einer einzigen Zufallsstichprobe ausgewählt und nicht etwa mit zwei Stichproben von jeweils 125 männlichen und weiblichen Bewohnern der Gemeinde. Wir können dann sowohl RG als auch G als Zufallsvariable betrachten. Eine naheliegende Frage ist die nach der Abhängigkeit zwischen diesen beiden Variablen. Wir formulieren dazu das Testproblem

$$H_0^u: p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \quad \text{gegen} \quad H_1^u: \text{Es gibt } (i, j) \text{ mit } p_{ij} \neq p_{i\cdot} \cdot p_{\cdot j},$$

was man einfacher schreiben kann:

$$H_0^u: p_{11} = p_{1\cdot} \cdot p_{\cdot 1} \quad \text{gegen} \quad H_1^u: p_{11} \neq p_{1\cdot} \cdot p_{\cdot 1}.$$

H_0^u bedeutet gerade, daß Geschlecht und Rauchgewohnheit unabhängig voneinander sind. Deshalb spricht man hier auch von einer **Unabhängigkeitshypothese**.

Führen wir die Studie wie oben beschrieben durch, so ist keine der Randsummen n_i , n_j vorher festgelegt.

B. Die Homogenitätshypothese

Wir hätten allerdings auch von vornherein festlegen können, daß 125 Männer und 125 Frauen befragt werden sollen. Dann wären n_1 und n_2 vorher determiniert gewesen. Die beiden Randsummen könnten dann z. B. auch so gewählt werden, daß sie die Anteile von Frauen und Männern in der Gesamtbevölkerung widerspiegeln. Wird die Studie derart angelegt, so hat man es mit 2 Zufallsstichproben zu tun: eine vom Umfang n_1 aus der Grundgesamtheit der Frauen und eine weitere vom Umfang n_2 aus der Grundgesamtheit der Männer. Hier bietet es sich an, die Frage zu untersuchen, ob die Rauchgewohnheiten in beiden Populationen die gleichen sind, d. h. ob der Anteil p_w der Raucher unter den Frauen der gleiche ist wie der Anteil p_M der Raucher unter den Männern. Das führt uns nun zur Aufstellung der **Homogenitätshypothese**

$$H_0^h: p_w = p_M \quad \text{gegen} \quad H_1^h: p_w \neq p_M.$$

C. Beziehungen zwischen den Hypothesen

Wir haben also nur aufgrund unterschiedlicher Auffassungen über die Art und Weise, wie die Daten erhoben wurden, zwei verschiedene Hypothesen, nämlich die Unabhängigkeitshypothese und die Homogenitätshypothese, formulieren können. Bei vielen Problemstellungen ist es jedoch nicht evident, welche Auffassung eher angemessen ist.

Glücklicherweise ist das auch gar nicht nötig, denn wir können für die beiden Hypothesen die selbe Teststatistik und den selben kritischen Wert verwenden.

Wir wollen nun auch eine Verbindung zwischen den beiden Hypothesen herstellen. Denken wir uns dazu die Kontingenztafel durch das Ziehen einer **bivariaten Stichprobe** erzeugt. Wir beobachten also die Realisationen (x_{1i}, x_{2i}) eines Zufallsvektors (X_1, X_2) , dessen Verteilung in **Tab. 10** wiedergegeben ist.

Tab. 10: Verteilung eines 0-1 wertigen (dichotomen) Zufallsvektors (X_1, X_2)

$X_1 \backslash X_2$	0	1	Σ
0	p_{11}	p_{12}	$p_{1\cdot}$
1	p_{21}	p_{22}	$p_{2\cdot}$
Σ	$p_{\cdot 1}$	$p_{\cdot 2}$	1

Auf das obige Beispiel bezogen bedeutet $X_1 = 1$ gerade, daß die befragte Person ein Mann war, $X_2 = 1$ bedeutet, daß es sich um einen Nichtraucher handelt, usw. Die Größen W, M, R und N sind also **numerisch codiert**.

Setzen wir

$$p_1 = P(X_2 = 0/X_1 = 0) = \frac{P(X_1 = 0, X_2 = 0)}{P(X_1 = 0)} = \frac{p_{11}}{p_{1\cdot}},$$

so entspricht das dem oben eingeführten p_w , wenn man von dem anderen wahrscheinlichkeitstheoretischen Modell einmal absieht;

$$p_2 = P(X_2 = 0/X_1 = 1) = \frac{P(X_1 = 1, X_2 = 0)}{P(X_1 = 1)} = \frac{p_{21}}{p_{2\cdot}}.$$

entspricht dann dem p_m .

Ist $p_1 = p_2 = p$, so gilt

$$\begin{aligned} P(X_2 = 0) &= P(X_2 = 0/X_1 = 0) \cdot P(X_1 = 0) \\ &\quad + P(X_2 = 0/X_1 = 1) \cdot P(X_1 = 1) \\ &= p_1 \cdot P(X_1 = 0) + p_2 \cdot P(X_1 = 1) \\ &= p(P(X_1 = 0) + P(X_1 = 1)) = p. \end{aligned}$$

Da nun gilt

$$P(X_2 = 0) = p_{\cdot 1} \text{ und } p = \frac{p_{11}}{p_{1\cdot}},$$

ergibt sich

$$p_{11} = p_{1\cdot} \cdot p_{\cdot 1}.$$

Aus $p_1 = p_2$ folgt also die (stochastische) Unabhängigkeit und umgekehrt.

1.2. Tests auf Unabhängigkeit in der 2×2 -Tafel

1.2.1. Der χ^2 -Test

Ein Test auf Unabhängigkeit, also ein Test für

$$H_0^u: p_{11} = p_{1\cdot} \cdot p_{\cdot 1} \text{ gegen } H_1^u: p_{11} \neq p_{1\cdot} \cdot p_{\cdot 1},$$

läßt sich bei genügend großem Stichprobenumfang n , etwa $n \geq 60$, mit der χ^2 -Teststatistik von Pearson

$$\chi^2 = \frac{n \cdot (n_{11} \cdot n_{22} - n_{12} \cdot n_{21})^2}{n_{1\cdot} n_{\cdot 1} n_{2\cdot} n_{\cdot 2}}$$

durchführen. Unter H_0^u ist diese Prüfgröße approximativ χ^2_1 -verteilt. Wollen wir zum Niveau α testen, so ist demnach H_0^u zu verwerfen, wenn $\chi^2 > \chi^2_{1;1-\alpha}$ ist.

Liegt der Stichprobenumfang n ungefähr im Bereich zwischen 20 und 60, so sollte man mit der **nach Yates korrigierten Teststatistik**

$$\chi^2_{\text{corr}} = \frac{n(|n_{11} \cdot n_{22} - n_{12} \cdot n_{21}| - n/2)^2}{n_1 \cdot n_2 \cdot n_1 \cdot n_2}$$

arbeiten. H_0^u ist dann bei $\chi^2_{\text{corr}} > \chi^2_{1;1-\alpha}$ zu verwerfen.

1.2.2. Exakte Tests

Bei noch kleineren Stichprobenumfängen und auch bei starker Asymmetrie der Vierfeldertafel ist die Verwendung des **exakten Tests von Fisher** angezeigt. Dieser Test kann aber ebenso gut bei großen Stichprobenumfängen angewandt werden. Da die Bestimmung des kritischen Bereichs recht aufwendig ist – man benutzt z. B. Quantile der hypergeometrischen Verteilung – macht man trotz der hervorragenden Güteeigenschaften jedoch selten Gebrauch von dieser Möglichkeit. Wir wollen nun an einem Beispiel erläutern, wie man zu diesem Test kommt. Dabei soll zunächst noch ein **weiterer exakter Test** vorgestellt werden, der sich im Gegensatz zum exakten Test von Fisher auch auf größere Kontingenztafeln übertragen lässt. Bei der Vierfeldertafel wird dieser Test jedoch kaum benutzt, da hier der exakte Test von Fisher bei etwa gleichem Rechenaufwand die besseren Güteeigenschaften hat und zudem die Möglichkeit bietet, einseitig, d. h. auf positive oder negative (stochastische) Abhängigkeit, zu testen. Bezüglich einer Vertafelung kritischer Werte sei auf Krüger et al. (1981) hingewiesen.

A. Ein exakter Test, der auf größere Kontingenztafeln übertragbar ist

Nehmen wir einmal an, wir hätten die in **Tab. 11** wiedergegebene Kontingenztafel erhalten.

Tab. 11: Beispiel für eine 2×2 -Tafel

$X_1 \backslash X_2$	0	1	Σ
0	1	2	3
1	3	4	7
Σ	4	6	10

Wir berechnen nun die Wahrscheinlichkeit, mit der wir bei Gültigkeit von H_0^u (und vorgegebenen Randsummen) diese Tafel beobachten können. Erhalten wir einen genügend kleinen Wert, ist es also relativ unwahrscheinlich, daß diese Tafel bei Vorliegen von Unabhängigkeit auftritt, so verwerfen wir die Hypothese H_0^u . In **Tab. 12** sind alle möglichen Tafeln wiedergegeben, die die gleichen Randsummen haben, wie die Vierfeldertafel in Tab. 11.

Es liegen nun insgesamt vier Tafeln mit den gleichen Randsummen vor. Zur Identifizierung dieser Tafeln genügt die Kenntnis von n_{11} , die übrigen Werte lassen sich dann nämlich mit Hilfe der Randsummen berechnen. Unter der Hypothese der Unabhängigkeit gilt

Tab. 12: 2×2 -Tafeln mit gleichen Randsummen wie in Tabelle 11

0	3	3		2	1	3		3	0	3
4	3	7		2	5	7		1	6	7
4	6	10		4	6	10		4	6	10

$$P(n_{11}) = \frac{\binom{n_{1.}}{n_{11}} \binom{n_{2.}}{n_{21}}}{\binom{n}{n_{.1}}} = \frac{n_{1.}! n_{2.}! n_{11}! n_{21}!}{n! n_{11}! n_{12}! n_{21}! n_{22}!}.$$

Das sind natürlich auch die Wahrscheinlichkeiten, mit denen die entsprechenden Kontingenztafeln auftreten können, wenn Unabhängigkeit vorliegt. Im einzelnen erhalten wir hier

$$P(n_{11} = 0) = \frac{3! 7! 4! 6!}{10! 0! 3! 4! 3!} = \frac{1}{6},$$

$$P(n_{11} = 1) = \frac{3! 7! 4! 6!}{10! 1! 2! 3! 4!} = \frac{1}{2},$$

$$P(n_{11} = 2) = \frac{3! 7! 4! 6!}{10! 2! 1! 2! 5!} = \frac{3}{10},$$

$$P(n_{11} = 3) = \frac{3! 7! 4! 6!}{10! 3! 0! 1! 6!} = \frac{1}{30}.$$

Die in Tab. 11 wiedergegebene Kontingenztafel tritt also bei Unabhängigkeit mit der Wahrscheinlichkeit $1/2$ auf.

Wir können nun folgendermaßen einen exakten Test für H_0^u gegen H_1^u durchführen. Die Hypothese H_0^u ist zu verwerfen, wenn die (unter den Randsummen bedingte) Wahrscheinlichkeit, mit der die beobachtete Kontingenztafel bei Gültigkeit von H_0^u auftritt, kleiner als eine vorgegebene Schranke c ist. Als Teststatistik benutzt man hier also die Wahrscheinlichkeit, mit der die tatsächlich beobachtete Kontingenztafel auftritt. Man spricht von einem exakten Test, weil bei der Festlegung des kritischen Bereichs die exakte Verteilung der Teststatistik benutzt wird und nicht etwa eine Approximation wie bei der χ^2 -Statistik.

Wie berechnet man nun das Testniveau α ? Dazu sucht man sich zunächst alle Kontingenztafeln heraus, die die gleichen Randsummen haben wie die tatsächlich beobachtete Tafel, und die unter der Unabhängigkeitshypothese nur mit einer Wahrscheinlichkeit kleiner als c auftreten können. Die Summe dieser Auftrittswahrscheinlichkeiten ist dann das gesuchte Testniveau α . Setzen wir in unserem Beispiel $c = 0,3$, so liegen die Auftrittswahrscheinlichkeiten von zwei Tafeln unterhalb dieses Wertes, nämlich diejenigen mit $n_{11} = 0$ und $n_{11} = 3$. Das Testniveau ergibt sich daraus zu $\alpha = 1/6 + 1/30 = 0,2$. Aufgrund der in Tab. 11 angegebenen Tafel kann demnach die Hypothese der Unabhängigkeit zum Niveau $\alpha = 0,2$ nicht verworfen werden.

Gibt man das Niveau α vor, so findet man folgendermaßen den zugehörigen kritischen Wert c . Die (bedingten) Wahrscheinlichkeiten, mit denen bei Gültigkeit

der Unabhängigkeitshypothese diejenigen Kontingenztafeln auftreten, deren Randsummen mit den tatsächlich beobachteten übereinstimmen, werden nacheinander – beginnend mit dem kleinsten Wert – in der Reihenfolge ihrer Größe aufaddiert. Das macht man solange, bis der Wert α gerade überschritten ist. Die dabei zuletzt addierte Wahrscheinlichkeit ist gleich dem gesuchten kritischen Wert c . Haben wir in unserem Beispiel etwa $\alpha = 0,3$ vorgegeben, so erhalten wir den kritischen Wert $c = 3/10$, denn es gilt

$$1/30 + 1/6 = 0,2 \leq 0,3 \text{ und } 1/30 + 1/6 + 3/10 = 0,5 > 0,3.$$

Man sieht, daß sich für alle Niveaus α zwischen 0,2 und 0,5 der gleiche kritische Wert $c = 0,3$ ergibt.

B. Der exakte Test von Fisher

Wie wir oben schon einmal erwähnt haben, läßt sich der eben besprochene Test auch bei größeren Kontingenztafeln verwenden. Das ist nicht so bei dem **exakten Test von Fisher**, der die Zellhäufigkeit n_{11} als Teststatistik benutzt. Mit der Größe n_{11} kann man zusammen mit den Randsummen eine Vierfeldertafel eindeutig festlegen, was bei größeren Kontingenztafeln nicht mehr der Fall ist.

Der exakte Test von Fisher wird wie folgt durchgeführt: Man bestimmt zwei Konstanten c_1 und c_2 und lehnt H_0^u ab, wenn $n_{11} \leq c_1$ oder $n_{11} > c_2$ ist. c_1 erhält man näherungsweise als $\alpha/2$ -Quantil und c_2 als $1 - \alpha/2$ -Quantil der hypergeometrischen Verteilung $H(n, n_{1.}, n_{.1})$, vgl. Abschnitt 3.2 in Kap. IV. Ist die Verteilung symmetrisch, ist also $n_{1.} = n_{.1}$, so gilt das sogar exakt. Dann stimmt auch der kritische Bereich des exakten Tests von Fisher überein mit dem des weiter oben besprochenen exakten Tests. Aber auch sonst stimmen die kritischen Bereiche dieser beiden Tests ziemlich gut überein. Der exakte Test von Fisher ist dabei der zuerst entwickelte und hat unter allen Tests, die die wünschenswerte Eigenschaft der Unverfälschtheit besitzen, die besten Gütteneigenschaften. Den hier zuerst vorgestellten exakten Test kann man ansehen als eine Nachempfindung des exakten Tests von Fisher für größere Kontingenztafeln.

1.2.3. Einseitige Hypothesen in 2×2 -Tafeln

Wir wollen nun auch **einseitige Hypothesen** testen. Dabei wird die Frage gestellt, ob **positive stochastische Abhängigkeit** vorliegt ($p_{11} \geq p_{1.} \cdot p_{.1}$) oder **negative stochastische Abhängigkeit** ($p_{11} \leq p_{1.} \cdot p_{.1}$). Wir können hierbei den einseitigen exakten Test von Fisher einsetzen. Liegt das Testproblem

$$H_0^1: p_{11} \leq p_{1.} \cdot p_{.1} \text{ gegen } H_1^1: p_{11} > p_{1.} \cdot p_{.1}$$

vor, so lehnt man H_0^1 ab, wenn n_{11} größer ist als das $1 - \alpha$ -Quantil der hypergeometrischen Verteilung $H(n, n_{1.}, n_{.1})$. Bei dem Testproblem

$$H_0^2: p_{11} \geq p_{1.} \cdot p_{.1} \text{ gegen } H_1^2: p_{11} < p_{1.} \cdot p_{.1}$$

wird H_0^2 verworfen, wenn n_{11} kleiner oder gleich dem α -Quantil von $H(n, n_{1.}, n_{.1})$ ist.

Beispiel: Betrachten wir noch einmal das vorangehende Beispiel, vgl. Tab. 11, und testen H_0^1 gegen H_1^1 zum Niveau $\alpha = 1/30$. Da für die beobachtete Tafel gilt $n = 10$, $n_{1.} = 3$ und $n_{.1} = 4$, benötigen wir das $1 - \alpha$ -Quantil von $H(10, 3, 4)$. Das ist gerade gleich 2. Aufgrund der Tafel aus Tab. 11 kann deshalb H_0^1 nicht verworfen werden, denn $n_{11} = 1 \leq 2$.

Ausführliche Tabellen für den Vierfeldertest bei kleinen Stichproben finden sich in Finney et al. (1963), vgl. auch die Dokumenta Geigy Tabellen (1968).

Liegen genügend viele Beobachtungen vor, so kann man für die eben beschriebenen einseitigen Testprobleme auch die χ^2 -Teststatistik mit dem Vorzeichen der Größe $n_{11} \cdot n_{22} - n_{12} \cdot n_{21}$ einsetzen. Ist dabei H_0^1 gegen H_1^1 zu testen, so wird H_0^1 verworfen, falls $\chi^2 > \chi^2_{1-\alpha}$ und $n_{11} \cdot n_{22} > n_{12} \cdot n_{21}$. Beim Testproblem H_0^2 gegen H_1^2 lehnt man H_0^2 ab, wenn $\chi^2 > \chi^2_{1-\alpha}$ und $n_{11} \cdot n_{22} < n_{12} \cdot n_{21}$.

Man kommt zu den gleichen Entscheidungen, wenn man die Teststatistik

$$\sqrt{n} \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1\cdot}n_{2\cdot}n_{\cdot1}n_{\cdot2}}}$$

verwendet und wie folgt vorgeht. Beim Testen von H_0^1 gegen H_1^1 verwirft man die Nullhypothese, wenn diese Teststatistik einen Wert liefert, der größer ist als das $1 - \alpha$ -Quantil der Standardnormalverteilung. Bei dem Testproblem H_0^2 gegen H_1^2 vergleicht man mit dem α -Quantil der Standardnormalverteilung. Wird dieser Wert unterschritten, so führt das zur Ablehnung von H_0^2 .

Wir haben hier stillschweigend vorausgesetzt, daß wir es mit ordinal skalierten Daten zu tun haben, denn nur dann erscheint es überhaupt sinnvoll, von positiver oder negativer stochastischer Abhängigkeit zu sprechen. Haben wir es dagegen, wie im folgenden Beispiel, mit nominal skalierten Daten zu tun, so müssen wir uns um eine andere Interpretation der Hypothese bemühen. Wir werden dabei auf die am Beginn von Abschnitt 1 gemachten Überlegungen zurückgreifen.

Beispiel: Bei einem Schulreifetest an 5005 ungefähr gleichaltrigen Kindern wurden die Ergebnisse nach Geschlecht getrennt aufgeführt. Es ergaben sich die Daten der **Tab. 13**.

Tab. 13: Resultat eines Schulreifetests zum Eintritt in die Volksschule

G \ S	0 (nicht schulreif)	1 (schulreif)	Σ
0 (Mädchen)	121	2350	2471
1 (Jungen)	197	2337	2534
Σ	318	4687	5005

Wie wir schon oben erwähnt haben, ist bei diesen nominal skalierten Daten eine Überprüfung auf positive bzw. negative stochastische Abhängigkeit wenig sinnvoll. Eine sinnvolle Fragestellung ist jedoch, ob Mädchen eher schulreif sind als Jungen. Das können wir auch so formulieren: Ist $P(S = 1/G = 0) > P(S = 1/G = 1)$?

Über eine gleichartige Fragestellung haben wir uns zu Beginn dieses Abschnitts über die Vierfeldertafel schon einmal Gedanken gemacht. Mit den gleichen Überlegungen wie dort sieht man hier, daß $P(S = 1/G = 0) > P(S = 1/G = 1)$ gerade gleichbedeutend ist mit $p_{11} < p_{1\cdot}p_{\cdot1}$. Die oben gestellte Frage läßt sich also mit einem Test für H_0^2 gegen H_1^2 statistisch beantworten. Man stellt dabei fest, daß die Aussage „Mädchen sind eher schulreif als Jungen“ zum Niveau 5% gesichert ist, denn

$$\chi^2 = \frac{5005(121 \cdot 2337 - 2350 \cdot 197)^2}{2471 \cdot 318 \cdot 2534 \cdot 4687} = 17,41 > 2,706 = \chi^2_{1;1-2 \cdot 0,05}$$

und $n_{11}n_{22} - n_{12}n_{21} = -180173 < 0$.

Anstatt die Prüfgröße χ^2 zusammen mit dem Vorzeichen von $n_{11}n_{22} - n_{12}n_{21}$ zu verwenden, kann man genauso gut mit dem Wert der Prüfgröße

$$\sqrt{n} \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_1 n_2 n_{11} n_{22}}} = \sqrt{5005} \frac{121 \cdot 2337 - 2350 \cdot 197}{\sqrt{2471 \cdot 318 \cdot 2534 \cdot 4687}} = -4,17$$

arbeiten. Dieser Wert muß dann mit dem α -Quantil der Standardnormalverteilung $u_{0,05} = -1,645$ verglichen werden. Es ist $-4,17 < -1,645$, so daß wir natürlich zur gleichen Entscheidung gelangen wie oben, nämlich zur Ablehnung der Nullhypothese H_0^2 .

1.3. Tests auf Homogenität in der 2×2 -Tafel

1.3.1. Die Durchführung der Tests

Im Abschnitt 1.2 sind wir davon ausgegangen, daß wir die Daten in unserer Kontingenztafel aufgrund einer einzigen Stichprobe vom Umfang n erhalten haben. Hier werden wir eine andere Annahme machen, die zwar die gleichen Tests liefert, uns aber neue Interpretationsmöglichkeiten bietet. Sind x_{11}, \dots, x_{1n_1} und x_{21}, \dots, x_{2n_2} zwei unabhängige Meßreihen, die einer $B(1, p_1)$ - bzw. $B(1, p_2)$ -Verteilung, vgl. Abschnitt 3.1. in Kap. IV, genügen, so können wir die Daten in einer Kontingenztafel anordnen, wie in Tab. 14 geschehen.

Tab. 14: Anordnung der Ergebnisse zweier unabhängiger binomialverteilter Meßreihen in einer 2×2 -Tafel

		Merkmalswert	0	1	\sum_j
		Population			
1		$n_{11} = n_1 - \sum_{j=1}^{n_1} x_{1j}$	$n_{12} = \sum_{j=1}^{n_1} x_{1j}$		$n_{1.} = n_1$
2		$n_{21} = n_2 - \sum_{j=1}^{n_2} x_{2j}$	$n_{22} = \sum_{j=1}^{n_2} x_{2j}$		$n_{2.} = n_2$
\sum_i		$n_{.1} = n - \sum_{i=1}^2 \sum_{j=1}^{n_i} x_{ij}$	$n_{.2} = \sum_{i=1}^2 \sum_{j=1}^{n_i} x_{ij}$		$n = n_1 + n_2$

Die Zeilensummen $n_{1.}$ und $n_{2.}$ liegen hier schon vor Durchführung des Experiments (der Befragung usw.) fest.

Für die Homogenitätshypothese

$$H_0^h: p_1 = p_2 \text{ gegen } H_1^h: p_1 \neq p_2$$

verwenden wir die gleichen Tests zum Niveau α wie bei der Unabhängigkeitshypothese, also die Teststatistik

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_1 \cdot n_1 \cdot n_2 \cdot n_2}$$

mit dem kritischen Wert $\chi^2_{1;1-\alpha}$ etwa für $n \geq 60$,

$$\chi^2_{\text{corr}} = \frac{n(|n_{11}n_{22} - n_{12}n_{21}| - n/2)^2}{n_1 \cdot n_1 \cdot n_2 \cdot n_2}$$

mit dem gleichen kritischen Wert $\chi^2_{1;1-\alpha}$ etwa für $20 \leq n \leq 60$ und – vorwiegend bei kleinen Stichprobenumfängen – den exakten Test von Fisher.

Die einseitigen Testprobleme

$$H_0^1: p_1 \leq p_2 \text{ gegen } H_1^1: p_1 > p_2$$

$$\text{und } H_0^2: p_1 \geq p_2 \text{ gegen } H_1^2: p_1 < p_2$$

werden ebenfalls wie dort behandelt.

Beispiel: Wir wollen mit den Daten der Tab. 9 testen, ob der Anteil p_M der Raucher bei den Männern signifikant höher ist als der Anteil p_W bei den Frauen. Dabei gehen wir davon aus, daß bei der Befragung zwei Stichproben von je 125 Personen (125 Männer und 125 Frauen) gezogen wurden (was für den anzuwendenden Test – wie wir wissen – keine Bedeutung hat). Hier liegt das einseitige Testproblem

$$H_0^1: p_M \leq p_W \text{ gegen } H_1^1: p_M > p_W$$

vor. H_0^1 ist zu verwerfen, wenn gilt $\chi^2 > \chi^2_{1;1-2\alpha}$ und $n_{11}n_{22} - n_{12}n_{21} < 0$. Welches Vorzeichen von $n_{11}n_{22} - n_{12}n_{21}$ zur Ablehnung von H_0^1 führt, kann man sich leicht selbst an einer extremen Situation klarmachen. Wäre hier zum Beispiel $n_{12} = n_{21} = 125$ und $n_{11} = n_{22} = 0$, so müßte das zur Ablehnung von H_0^1 führen. Dabei ist $n_{11}n_{22} - n_{12}n_{21} = 0 - 125 \cdot 125 < 0$. Mit unseren Daten können wir zum Niveau $\alpha = 0,05$ die Hypothese H_0^1 nicht verwerfen. Es gilt zwar

$$n_{11}n_{22} - n_{12}n_{21} = 45 \cdot 70 - 80 \cdot 55 = -1250 < 0,$$

$$\text{aber } \chi^2 = \frac{250(45 \cdot 70 - 80 \cdot 55)^2}{125 \cdot 100 \cdot 125 \cdot 150} = \frac{5}{3} < \chi^2_{1;1-2 \cdot 0,05} = 2,7.$$

1.3.2. Der erforderliche Stichprobenumfang

Wir haben uns bisher nur damit beschäftigt, eine gegebene Kontingenztafel auszuwerten. Dabei kann es passieren, daß die Frage, die mit der Studie geklärt werden sollte, aufgrund der vorliegenden Daten nicht beantwortet werden kann. Das läßt sich jedoch durch eine richtige Planung der Studie weitgehend verhindern. Dazu müssen wir die folgende Frage untersuchen: Wie groß müssen die Stichprobenumfänge n_1 und n_2 mindestens sein, damit eine gegebene Differenz $p_1 - p_2 = \delta$ mit einer Wahrscheinlichkeit von mindestens $1 - \beta$, d. h. mit einem Fehler 2. Art von höchstens β , durch einen Test zum Niveau α entdeckt wird?

A. Die Formel mittels Approximation der Gütfunktion des χ^2 -Tests

Um diese Frage beantworten zu können, beschaffen wir uns zunächst eine Approximation der Gütfunktion des χ^2 -Tests. Dafür schreiben wir die χ^2 -Statistik in einer anderen Form auf. Setzen wir

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p}) \frac{n_1 + n_2}{n_1 \cdot n_2}}}$$

mit $\hat{p}_1 = \frac{n_{12}}{n_1}$ und $\hat{p}_2 = \frac{n_{22}}{n_2}$ (also den intuitiv naheliegenden Schätzern für p_1 und p_2) und $\hat{p} = (n_{12} + n_{22})/n$, dann ist $\chi^2 = T^2$. Ist $p_1 = p_2$, so ist eine Schätzung der Varianz von $\hat{p}_1 - \hat{p}_2$ gerade gegeben durch

$$\hat{p}(1-\hat{p})(n_{12} + n_{22})/(n_1 \cdot n_2),$$

so daß T unter der Hypothese $p_1 = p_2$ approximativ $N(0, 1)$ -verteilt ist. Ist dagegen $p_1 - p_2 = \delta$, so ist T approximativ $N(\delta, \varepsilon^2)$ -verteilt mit

$$\gamma = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \cdot \frac{\delta}{\sqrt{\frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} \cdot \left(1 - \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}\right)}}$$

und $\varepsilon^2 = \frac{n_2 p_1 (1-p_1) + n_1 p_2 (1-p_2)}{(p_1 n_1 + p_2 n_2)(n_1 + n_2 - p_1 n_1 - p_2 n_2)} (n_1 + n_2).$

Da wegen $\chi^2 = T^2$ ein auf T basierender Test äquivalent ist mit dem χ^2 -Test, haben die beiden Tests auch die gleiche Gütfunktion, nämlich

$$1 - \Phi\left(\frac{u_{1-\alpha} - \gamma}{\varepsilon}\right).$$

Setzt man $n_1 = n_2$ und fordert $\Phi\left(\frac{u_{1-\alpha} - \gamma}{\varepsilon}\right) = \beta$, so ergibt sich

$$n_1 = \frac{\left[u_{1-\alpha} \sqrt{2 \frac{p_1 + p_2}{2} \left(1 - \frac{p_1 + p_2}{2}\right)} - u_\beta \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{\delta^2}$$

als notwendiger Mindeststichprobenumfang.

B. Die Arcus-Sinus-Formel

Eine andere Approximation für n_1 liefert die sogenannte **Arcus-Sinus-Formel**:

$$n_1 = (u_{1-\alpha} - u_\beta)^2 / (2(\arcsin(\sqrt{p_1}) - \arcsin(\sqrt{p_2}))^2).$$

Mit beiden Formeln erhält man ähnliche Ergebnisse, jedoch ist aufgrund der Approximation der Wert von n_1 in der Regel zu klein angesetzt.

C. Die Formel nach Casagrande/Pike/Smith

Daher schlagen Casagrande et al. (1978) eine Modifikation der hier hergeleiteten Funktionen vor, nämlich

$$n_1 = \frac{A(1 + \sqrt{1 + 4\delta/A})^2}{4\delta^2},$$

$$\text{wobei } A = \left(u_{1-\alpha} \sqrt{2 \frac{p_1 + p_2}{2} \left(1 - \frac{p_1 + p_2}{2} \right)} - u_\beta \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right)^2$$

ist. Diese Formel liefert (nach den Autoren) Werte, die in guter Übereinstimmung mit den exakten Werten stehen. Die exakten Werte sind dabei diejenigen, die sich bei der Verwendung des exakten Tests von Fisher ergeben. Da dieser Test optimale Güteeigenschaften besitzt und die exakte Verteilung der Prüfgröße benutzt wird, erhält man damit die minimalen Stichprobenumfänge. Unter Umständen können diese Werte bei der Verwendung anderer Tests nicht mehr die Güte $1 - \beta$ garantieren. Es sind wirklich nur Minimalforderungen, die keinesfalls unterschritten werden sollten.

D. Exakte Stichprobenumfänge

Ausführliche Tafeln für die exakten Werte von n_1 finden sich in oben zitierter Arbeit von Casagrande et al. (1978) und in Hasemann (1978). Hinweise für die graphische Ermittlung von n_1 finden sich in Feigl (1978). **Tab. 15** gibt für einige ausgewählte Werte von δ und p_2 erforderliche Stichprobenumfänge $n_1 = n_2$ an.

Tab. 15: Erforderliche Stichprobenumfänge $n_1 = n_2$ (exakt) für die 2×2 -Tafel (einseitig)

$\delta = p_1 - p_2$	α	Güte $1 - \beta$	p_2				
			0,1	0,2	0,3	0,4	0,5
0,1	0,05	0,8	173	249	302	321	321
		0,9	232	338	408	445	445
	0,01	0,8	269	393	475	519	519
		0,9	344	503	609	667	667
0,2	0,05	0,8	56	73	84	85	84
		0,9	74	97	111	116	111
	0,01	0,8	86	113	129	137	129
		0,9	108	143	163	171	163
0,3	0,05	0,8	30	36	41	41	36
		0,9	39	47	53	53	47
	0,01	0,8	46	55	60	60	55
		0,9	56	69	77	77	69
0,4	0,05	0,8	19	23	23	23	19
		0,9	25	30	31	30	25
	0,01	0,8	29	34	35	34	29
		0,9	35	42	43	42	35
0,5	0,05	0,8	13	15	15	13	—
		0,9	17	18	18	17	—
	0,01	0,8	20	22	22	20	—
		0,9	24	28	28	24	—
0,6	0,05	0,8	10	10	10	—	—
		0,9	12	12	12	—	—
	0,01	0,8	15	16	15	—	—
		0,9	18	18	18	—	—

In Tab. 15 sind die Stichprobenumfänge nur für Werte von p_2 kleiner oder gleich 0,5 tabelliert. Für größere Werte von p_2 ist der erforderliche Stichprobenumfang ebenfalls aus dieser Tabelle zu entnehmen. Dabei muß nur p_2 durch $1 - p_2$ und $p_1 - p_2$ durch $p_2 - p_1$ ersetzt werden.

Beispiel: Zwei Medikamente A und B mit den unbekannten Erfolgswahrscheinlichkeiten p_1 und p_2 sollen miteinander verglichen werden. Bei dem Medikament A rechnet man mit einer Erfolgswahrscheinlichkeit von etwa 70%. Falls das Medikament B eine um mindestens 10% größere Erfolgswahrscheinlichkeit hat, so möchte man das mit einer Wahrscheinlichkeit von 0,9 entdecken. Wir haben hier $p_1 = 0,7$, $p_2 = 0,8$ und $\beta = 0,1$. Der erforderliche Stichprobenumfang kann also nicht direkt aus Tab. 15 abgelesen werden. Wir bilden deshalb $1 - p_2 = 0,2$. Soll zum Niveau $\alpha = 0,01$ getestet werden, so erhält man wegen $p_2 - p_1 = 0,1$ aus der Tabelle einen Mindeststichprobenumfang von $n_1 = n_2 = 503$.

1.4. Tests auf Symmetrie in der 2×2 -Tafel

Das in diesem Abschnitt zu besprechende Testproblem wollen wir an einem Beispiel erläutern.

Beispiel: Um eine ausgeschriebene Mitarbeiterstelle einer Firma bewerben sich insgesamt 20 Interessenten. Die Bewerber werden einem psychologischen Test unterworfen und außerdem vom Personalchef aufgrund eines Vorstellungsgesprächs beurteilt. Fassen wir das Bewerbungsgespräch ebenfalls als einen Test auf, so wird jede der 20 Personen zwei Tests unterzogen. Gehen wir davon aus, daß die Tests nur nach „geeignet“ (Test hat den Wert 1) und „nicht geeignet“ (Test hat den Wert 0) unterscheiden, so können wir das Ergebnis dieses Versuchs in einer Vierfeldertafel wiedergeben, s. **Tab. 16**.

Tab. 16: Ergebnisse zweier Bewerbungstests

		psychologischer Test		Σ
		0	1	
Bewertung durch den Personalchef	0	3	6	9
	1	3	8	11
Σ		6	14	20

Eine Abhängigkeit der beiden Testergebnisse ist von vornherein klar. Ein Bewerber, der den psychologischen Test als geeignet passiert, wird nämlich (sofern der Test und der Personalchef die selben Fähigkeiten beurteilen) mit größerer Wahrscheinlichkeit auch beim Bewerbungsgespräch positiv abschneiden als ein Bewerber, der beim psychologischen Test versagt hat. Es interessiert nun die Frage, ob der psychologische Test und der Test durch den Personalchef gleich schwer sind. Sind beide Tests gleich schwer, so ist (in der üblichen Terminologie) $p_{11} + p_{12} = p_{11} + p_{21}$ und damit $p_{12} = p_{21}$.

Wir können also als Hypothese formulieren:

$$H_0: p_{12} = p_{21} \text{ gegen } H_1: p_{12} \neq p_{21}.$$

Bei gegebenem $k = n_{12} + n_{21}$ ist n_{12} unter H_0 gerade $B(k, 1/2)$ -verteilt.

Wir können also folgendermaßen einen Test zum Niveau α für H_0 gegen H_1 durchführen: Wir bestimmen c_1 als $\alpha/2$ -Quantil und c_2 als $1 - \alpha/2$ -Quantil der $B(k, 1/2)$ -Verteilung. Ist nun $n_{12} \leq c_1$ oder $n_{12} > c_2$, so muß die Hypothese H_0 verworfen werden.

Beispiel: In unserem Beispiel ist $k = 6 + 3 = 9$. c_1 und c_2 erhalten wir demnach als $\alpha/2$ - und $1 - \alpha/2$ -Quantil der $B(9, 1/2)$ -Verteilung. Bei einem Testniveau $\alpha = 10\%$ ergibt sich $c_1 = 0$ und $c_2 = 8$. Die Hypothese H_0 , daß die beiden Einstellungstests gleich schwer sind, kann somit nicht verworfen werden, denn $c_1 < n_{12} \leq c_2$.

1.4.1. Der McNemar-Test

Liegen genügend viele Beobachtungen vor, so läßt sich mit der standardisierten Testgröße

$$T = \frac{n_{12} - k/2}{\sqrt{k/4}} = \frac{n_{12} - (n_{12} + n_{21})/2}{\sqrt{(n_{12} + n_{21})/4}}$$

ein approximativer Test durchführen. Dieser Test, der unter dem Namen **McNemar-Test** bekannt ist, lehnt die Nullhypothese ab, wenn gilt $|T| > u_{1-\alpha/2}$. Der McNemar-Test kann auch wie folgt durchgeführt werden: Man berechnet die Teststatistik

$$\chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} = T^2$$

und vergleicht diesen Wert dann mit dem $1 - \alpha$ -Quantil der χ^2_1 -Verteilung. Ist $\chi^2 > \chi^2_{1; 1-\alpha}$, so wird die Hypothese H_0 verworfen. Bei kleineren Stichprobenumfängen ist es angebracht, mit der korrigierten Statistik

$$\chi^2_{\text{corr}} = \frac{(|n_{12} - n_{21}| - 1)^2}{n_{12} + n_{21}}$$

zu arbeiten.

Beispiel: Das obige Beispiel liefert für die Werte der einzelnen Teststatistiken

$$T = \frac{6 - \frac{6+3}{2}}{\sqrt{\frac{6+3}{4}}} = 1,0,$$

$$\chi^2 = \frac{(6 - 3)^2}{6 + 3} = 1,0,$$

$$\chi^2_{\text{corr}} = \frac{(|6 - 3| - 1)^2}{6 + 3} = 0,44.$$

$|T|$ wird mit dem $1 - \alpha/2$ -Quantil der Standardnormalverteilung $u_{1-0,05} = u_{0,95} = 1,645$ verglichen und die anderen beiden Statistiken mit dem $1 - \alpha$ -Quantil der χ^2 -Verteilung mit einem Freiheitsgrad $\chi^2_{1; 1-0,1} = \chi^2_{1; 0,9} = 2,7$.

1.4.2. Cochran's Q

Wir haben uns bisher in diesem Abschnitt mit dem Vergleich zweier Einstellungstests beschäftigt. Nun wollen wir noch eine Verallgemeinerung auf den Vergleich

mehrerer Tests vornehmen. Dazu betrachten wir das in Tab. 17 angegebene Beispiel.

Tab. 17: Testergebnisse von 9 Personen

Test	Testpersonen	1	2	3	4	5	6	7	8	9	Σ
1		0	0	1	0	1	0	0	1	1	4
2		1	0	0	0	1	1	1	0	1	5
3		1	1	1	1	1	0	0	1	1	7
Σ		2	1	2	1	3	1	1	2	3	16

Ist p_{ij} die Wahrscheinlichkeit dafür, daß der Test i bei der j -ten Person erfolgreich ist, so stellen wir mit der Nullhypothese

$$H_0: p_{1j} = p_{2j} = \dots = p_{rj}$$

die Frage, ob die $r = 3$ Tests gleich schwer sind. Die Entscheidung wird anhand der Ergebnisse von den $s = 9$ Testpersonen mit Hilfe der als Cochran's Q bekannten Teststatistik

$$C_Q = \frac{\sum_{i=1}^r r(r-1)(n_{i\cdot} - \frac{n}{r})^2}{\sum_{j=1}^s n_{\cdot j}(r - n_{\cdot j})} = \frac{r(r-1) \sum_{i=1}^r n_{i\cdot}^2 - (r-1)n^2}{rn - \sum_{j=1}^s n_{\cdot j}^2}$$

gefällt. Ist der berechnete Wert größer als das $1 - \alpha$ -Quantil der χ^2 -Verteilung mit $r - 1$ Freiheitsgraden, so wird die Nullhypothese abgelehnt.

Beispiel: Für unser Beispiel ergibt sich für die Teststatistik

$$\begin{aligned} C_Q &= \frac{3(3-1)(4^2 + 5^2 + 7^2) - (3-1)16^2}{3 \cdot 16 - (2^2 + 1^2 + 2^2 + 1^2 + 3^2 + 1^2 + 1^2 + 2^2 + 3^2)} \\ &= \frac{3 \cdot 2 \cdot 90 - 2 \cdot 256}{48 - 34} \\ &= \frac{28}{14} = 2. \end{aligned}$$

Es ist $\chi^2_{2,0,95} = 5,99$, so daß die Nullhypothese zum Niveau $\alpha = 0,05$ nicht abgelehnt werden kann.

Werden nur zwei Tests miteinander verglichen, ist also $r = 2$, so ergibt sich wieder der oben beschriebene McNemar-Test. Zum Schluß möchten wir noch darauf hinweisen, daß sich die hier vorgestellten Testverfahren auch für die Behandlung anderer praktischer Probleme eignen als den hier vorgestellten Vergleichen von Testmethoden. So können z. B. Medikamente oder Therapiemethoden genauso gut verglichen werden wie etwa Werbemaßnahmen oder Prognoseverfahren.

2. Loglineare Modelle und Tests für $r \times s$ -Tafeln

In Abschnitt 1 haben wir uns schon eingehend mit der 2×2 -Felder-Tafel auseinandergesetzt. Wir wollen uns hier zunächst noch einmal mit dieser Tafel beschäftigen. Daran soll eine Vorgehensweise demonstriert werden, die auch bei beliebigen $r \times s$ -Tafeln anwendbar ist und die man zudem leicht auf mehrdimensionale Kontingenztafeln verallgemeinern kann. Gemeint ist dabei die Darstellung in einem **loglinearen Modell**.

2.1. Das loglineare Modell für die $r \times s$ -Tafel

2.1.1. Entwicklung des Modells am Beispiel der 2×2 -Tafel

Wir wollen uns nun gleich ein konkretes **Beispiel** vorgeben. In einem Naherholungsgebiet sind noch zwei freie Flächen zu gestalten. Dabei soll auf der einen ein Spielplatz entstehen, während die andere entweder eine Minigolfanlage oder eine Rollschuhbahn erhalten soll. Als Planungshilfe wurde auch eine Umfrage unter den Bürgern durchgeführt, deren Ergebnisse in **Tab. 18** zusammengestellt sind.

Tab. 18: Planung von Freizeitanlagen

Standort Freizeitanlage			Σ
	1	2	
Minigolfanlage	587	392	979
Rollschuhbahn	257	307	564
Σ	844	699	1543

Die zu dieser Tafel der empirischen Häufigkeiten n_{ij} korrespondierende Tafel der uns unbekannten, theoretischen Wahrscheinlichkeiten ist in **Tab. 19** dargestellt.

Tab. 19: Planung von Freizeitanlagen: theoretische Wahrscheinlichkeiten zu Tab. 18

Standort Freizeitanlage			Σ
	1	2	
Minigolfanlage	p_{11}	p_{12}	$p_{1.}$
Rollschuhbahn	p_{21}	p_{22}	$p_{2.}$
Σ	$p_{.1}$	$p_{.2}$	1

A. Aufstellung des Modells

Diese theoretischen Wahrscheinlichkeiten p_{ij} lassen sich nun mit den (unbekannten) Parametern u' , $u_{1(i)}$, $u_{2(j)}$ und $u_{12(ij)}$ in der folgenden Form darstellen:

$$p_{ij} = e^{u' + u_{1(i)} + u_{2(j)} + u_{12(ij)}}.$$

Dabei sollen, um eine eindeutige Darstellung zu gewährleisten, die Nebenbedingungen

$$\sum_{i=1}^2 u_{1(i)} = \sum_{j=1}^2 u_{2(j)} = 0$$

$$\text{und } \sum_{i=1}^2 u_{12(ij)} = \sum_{j=1}^2 u_{12(ij)} = 0$$

erfüllt sein.

Bezeichnen wir die erwarteten Häufigkeiten mit m_{ij} , so können wir schreiben

$$\begin{aligned} m_{ij} &= n \cdot p_{ij} = e^{ln n + u' + u_{1(i)} + u_{2(j)} + u_{12(ij)}} \\ &= e^{u + u_{1(i)} + u_{2(j)} + u_{12(ij)}}, \end{aligned}$$

wobei $u = \ln n + u'$ gesetzt ist. Es werden also bis auf eine Ausnahme die gleichen Parameter wie oben verwendet.

Das führt auch bei der Aufstellung zu testender Hypothesen zu gleichartigen Formulierungen für p_{ij} und m_{ij} . Deshalb genügt es, eine dieser beiden Größen zu untersuchen. Das werden meistens die zu erwartenden Häufigkeiten m_{ij} sein. Schreibt man nun

$$\ln m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)},$$

so wird deutlich, warum man hier von einem **loglinearen Modell** spricht. Bevor wir uns der Frage zuwenden, was uns diese Darstellung eigentlich nützt, wollen wir noch zeigen, wie man die hier eingeführten Parameter aus den Häufigkeiten m_{ij} erhält. u ergibt sich als Mittelwert der Logarithmen der m_{ij} zu

$$\begin{aligned} u &= \frac{1}{4} \sum_{i=1}^2 \sum_{j=1}^2 \ln m_{ij} = \frac{1}{4} (\ln m_{11} + \ln m_{12} + \ln m_{21} + \ln m_{22}) \\ &= \ln n + \frac{1}{4} (\ln p_{11} + \ln p_{12} + \ln p_{21} + \ln p_{22}) \\ &= \ln n + u'. \end{aligned}$$

Weiter ergibt sich

$$\begin{aligned} u_{1(i)} &= \frac{1}{2} \sum_{j=1}^2 \ln m_{ij} - u \\ &= \frac{1}{2} (\ln m_{i1} + \ln m_{i2}) - \frac{1}{4} (\ln m_{11} + \ln m_{12} + \ln m_{21} + \ln m_{22}), \\ u_{2(j)} &= \frac{1}{2} \sum_{i=1}^2 \ln m_{ij} - u \\ &= \frac{1}{2} (\ln m_{1j} + \ln m_{2j}) - \frac{1}{4} (\ln m_{11} + \ln m_{12} + \ln m_{21} + \ln m_{22}) \end{aligned}$$

$$\begin{aligned} \text{und } u_{12(ij)} &= \ln m_{ij} - (u + u_{1(i)}) - (u + u_{2(j)}) + u \\ &= \ln m_{ij} - \frac{1}{2} \sum_{j=1}^2 \ln m_{ij} - \frac{1}{2} \sum_{i=1}^2 \ln m_{ij} + \frac{1}{4} \left(\sum_{i=1}^2 \sum_{j=1}^2 \ln m_{ij} \right). \end{aligned}$$

Es reicht hierbei die Angabe der Parameter

$$u, u_{1(1)}, u_{2(1)} \text{ und } u_{12(11)},$$

denn es gelten die folgenden einfachen Beziehungen:

$$u_{1(2)} = -u_{1(1)}, u_{2(2)} = -u_{2(1)} \text{ und}$$

$u_{12(12)} = -u_{12(11)} = u_{12(21)} = -u_{12(22)}$ (dies erhält man direkt aus den oben angegebenen Nebenbedingungen).

Nun ist

$$u = \frac{1}{4} \cdot \ln(m_{11} \cdot m_{12} \cdot m_{21} \cdot m_{22}) = \ln n + \frac{1}{4} \cdot \ln(p_{11} \cdot p_{12} \cdot p_{21} \cdot p_{22}),$$

$$u_{1(1)} = \frac{1}{2}(\ln m_{11} + \ln m_{12}) - u = \frac{1}{4} \cdot \ln \frac{m_{11} \cdot m_{12}}{m_{21} \cdot m_{22}} = \frac{1}{4} \cdot \ln \frac{p_{11} \cdot p_{12}}{p_{21} \cdot p_{22}},$$

$$u_{2(1)} = \frac{1}{2}(\ln m_{11} + \ln m_{21}) - u = \frac{1}{4} \cdot \ln \frac{m_{11} \cdot m_{21}}{m_{12} \cdot m_{22}} = \frac{1}{4} \cdot \ln \frac{p_{11} \cdot p_{21}}{p_{12} \cdot p_{21}}$$

und $u_{12(11)} = \ln m_{11} - (u + u_{1(1)}) - (u + u_{2(1)}) + u$

$$= \frac{1}{4} \cdot \ln \frac{m_{11} \cdot m_{22}}{m_{12} \cdot m_{21}} = \frac{1}{4} \cdot \ln \frac{p_{11} \cdot p_{22}}{p_{12} \cdot p_{21}}.$$

Damit erhält man

$$\begin{aligned} \ln m_{11} &= \frac{1}{4} \ln(m_{11} \cdot m_{12} \cdot m_{21} \cdot m_{22}) + \frac{1}{4} \cdot \ln \frac{m_{11} \cdot m_{12}}{m_{21} \cdot m_{22}} \\ &\quad + \frac{1}{4} \cdot \ln \frac{m_{11} \cdot m_{21}}{m_{12} \cdot m_{22}} + \frac{1}{4} \cdot \ln \frac{m_{11} \cdot m_{22}}{m_{12} \cdot m_{21}} \\ &= \ln n + \frac{1}{4} \cdot \ln(p_{11} \cdot p_{12} \cdot p_{21} \cdot p_{22}) \\ &\quad + \frac{1}{4} \cdot \ln \frac{p_{11} \cdot p_{12}}{p_{21} \cdot p_{22}} + \frac{1}{4} \cdot \ln \frac{p_{11} \cdot p_{21}}{p_{12} \cdot p_{22}} + \frac{1}{4} \cdot \ln \frac{p_{11} \cdot p_{22}}{p_{12} \cdot p_{21}}. \end{aligned}$$

B. Schätzen der Parameter

Wie sieht das nun konkret in unserem **Beispiel** aus? Dort sind die erwarteten Häufigkeiten m_{ij} zwar unbekannt, können aber geschätzt werden durch die beobachteten Häufigkeiten n_{ij} , vgl. Tab. 18. Wir erhalten somit für die Parameter $u, u_{1(1)}, u_{2(1)}$ und $u_{12(11)}$ die Schätzungen

$$\hat{u} = \frac{1}{4} \cdot \ln(n_{11} \cdot n_{12} \cdot n_{21} \cdot n_{22}) = \frac{1}{4} \cdot \ln(587 \cdot 392 \cdot 257 \cdot 307) = 5,906,$$

$$\hat{u}_{1(1)} = \frac{1}{4} \cdot \ln \frac{n_{11} \cdot n_{12}}{n_{21} \cdot n_{22}} = \frac{1}{4} \cdot \ln \frac{587 \cdot 392}{257 \cdot 307} = 0,268,$$

$$\hat{u}_{2(1)} = \frac{1}{4} \cdot \ln \frac{n_{11} \cdot n_{21}}{n_{12} \cdot n_{22}} = \frac{1}{4} \cdot \ln \frac{587 \cdot 257}{392 \cdot 307} = 0,0565,$$

$$\hat{u}_{12(11)} = \frac{1}{4} \cdot \ln \frac{n_{11} \cdot n_{22}}{n_{12} \cdot n_{21}} = \frac{1}{4} \cdot \ln \frac{587 \cdot 307}{392 \cdot 257} = 0,145.$$

Die Schätzungen für die übrigen Parameter erhalten wir direkt aus den oben angegebenen Nebenbedingungen.

C. Die approximative Varianz der Schätzungen

Als Varianz von \hat{u} erhalten wir approximativ

$$\left(\frac{1}{4}\right)^2 \left(\frac{1}{m_{11}} + \frac{1}{m_{12}} + \frac{1}{m_{21}} + \frac{1}{m_{22}} \right) - \frac{1}{n}$$

und für alle anderen Größen

$$\left(\frac{1}{4}\right)^2 \left(\frac{1}{m_{11}} + \frac{1}{m_{12}} + \frac{1}{m_{21}} + \frac{1}{m_{22}} \right).$$

Wir können die Varianz von \hat{u} also schätzen durch

$$\begin{aligned} & \left(\frac{1}{4}\right)^2 \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right) - \frac{1}{n} \\ &= \left(\frac{1}{4}\right)^2 \left(\frac{1}{587} + \frac{1}{392} + \frac{1}{257} + \frac{1}{307} \right) - \frac{1}{1543} = 0,0000646 \end{aligned}$$

und die von $\hat{u}_{1(i)}$, $\hat{u}_{2(j)}$ und $\hat{u}_{12(ij)}$ durch

$$\begin{aligned} & \left(\frac{1}{4}\right)^2 \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right) = \\ &= \left(\frac{1}{4}\right)^2 \left(\frac{1}{587} + \frac{1}{392} + \frac{1}{257} + \frac{1}{307} \right) = 0,0007127. \end{aligned}$$

Damit hat man für \hat{u} die geschätzte Standardabweichung $\sqrt{0,0000646} = 0,008$ und für alle anderen Größen $\sqrt{0,0007127} = 0,027$.

Da die Schätzungen \hat{u} , $\hat{u}_{1(i)}$ usw. approximativ normalverteilt sind, ergeben sich nun näherungsweise die folgenden 95% Konfidenzbereiche (vgl. Abschnitt 1.3 in Kap. IV):

- [5,890; 5,922] für u ,
- [0,215; 0,321] für $u_{1(1)}$,
- [0,004; 0,109] für $u_{2(1)}$ und
- [0,092; 0,198] für $u_{12(11)}$.

D. Die Interpretation der Parameter

Wir wollen uns nun noch um eine Interpretation der Parameter bemühen. In dem hohen **positiven Wert** von $\hat{u}_{1(1)}$ spiegelt sich wider, daß deutlich **mehr als die Hälfte** der Befragten die Minigolfanlage bevorzugen, nämlich 979. Entsprechend ist der große **negative Wert** von $\hat{u}_{1(2)}$ ($= -0,268$) darauf zurückzuführen, daß sich nur 564 Befragte, also weit **weniger als die Hälfte** für die Rollschuhbahn entschieden

haben. Bei der Entscheidung für einen der beiden Standorte ergeben sich weitaus geringere Unterschiede. Das findet dann seinen Niederschlag darin, daß die Werte $\hat{u}_{2(j)}$ recht nahe bei Null liegen. Schließlich weist uns der große positive Wert von $\hat{u}_{12(11)}$ darauf hin, daß die Wahl der Minigolfanlage eine stärkere Bevorzugung des Standortes 1 zur Folge hat, als die Gesamtzahl der 844 Stimmen für diesen Standort erwarten läßt. Die $u_{12(ij)}$ beschreiben also **Wechselwirkungen** zwischen der Wahl der Freizeitanlage und der Wahl des Standortes. In der Tab. 20 sind nun die beobachteten Häufigkeiten n_{ij} in der Form

Tab. 20: Struktur der 2×2 -Tafel aus Tab. 18

		Standort	
		1	2
Freizeitanlage	Minigolfanlage	$367 \cdot 1,31 \cdot 1,06 \cdot 1,16$	$367 \cdot 1,31 \cdot 0,945 \cdot 0,865$
	Rollschuhbahn	$367 \cdot 0,765 \cdot 1,06 \cdot 0,865$	$367 \cdot 0,765 \cdot 0,945 \cdot 1,16$

$$n_{ij} = \hat{m}_{ij} = e^u e^{u_{1(i)}} e^{u_{2(j)}} e^{u_{12(ij)}}$$

aufgeschrieben. Die Einflüsse der einzelnen Faktoren sind so direkt sichtbar, und man erkennt sehr schön die Struktur dieser Kontingenztafel. Es wird aber auch deutlich, daß den konkreten Werten der Parameter des loglinearen Modells nicht unmittelbar eine inhaltliche Bedeutung zukommt.

2.1.2. Das Modell für die allgemeine rxs-Tafel

Die bisher für die 2×2 -Tafel gemachten Überlegungen lassen sich leicht auf die allgemeine $r \times s$ -Tafel übertragen, wobei **r** die Anzahl der Zeilen- und **s** die Anzahl der Spalten-Ausprägungen angibt. Die erwarteten Häufigkeiten m_{ij} lassen sich ebenso wie dort in der Form

$$\ln m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

darstellen. Um eine eindeutige Darstellung zu gewährleisten, wird die Gültigkeit der folgenden Nebenbedingungen gefordert

$$\sum_{i=1}^r u_{1(i)} = \sum_{j=1}^s u_{2(j)} = 0$$

und

$$\sum_{i=1}^r u_{12(ij)} = \sum_{j=1}^s u_{12(ij)} = 0.$$

Analog zum linearen Modell der Varianzanalyse, vgl. Kap. XI, erhalten wir nun für die einzelnen Parameter die Ausdrücke

$$u = \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s \ln m_{ij},$$

$$u_{1(i)} = \frac{1}{s} \sum_{j=1}^s \ln m_{ij} - \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s \ln m_{ij},$$

$$u_{2(j)} = \frac{1}{r} \sum_{i=1}^r \ln m_{ij} - \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s \ln m_{ij}$$

$$\text{und } u_{12(ij)} = \ln m_{ij} - \frac{1}{s} \sum_{j=1}^s \ln m_{ij} - \frac{1}{r} \sum_{i=1}^r \ln m_{ij} + \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s \ln m_{ij}.$$

Wir können diese Werte schätzen, indem wir die uns unbekannten erwarteten Häufigkeiten m_{ij} durch die beobachteten Häufigkeiten n_{ij} ersetzen. Für die Angabe von Konfidenzintervallen benötigen wir noch die Varianzen dieser approximativ normalverteilten Zufallsgrößen. Dafür lassen sich wiederum nur Schätzwerte angeben. Wir werden nun hier – wie auch im folgenden – darauf verzichten, die Schätzwerte in der verwendeten Symbolik als solche zu kennzeichnen, so daß wir schreiben können

$$\begin{aligned} \text{Var}(u) &= \left(\frac{1}{r \cdot s} \right)^2 \sum_{i=1}^r \sum_{j=1}^s \frac{1}{n_{ij}} - \frac{1}{n}, \\ \text{Var}(u_{1(i)}) &= \left(\frac{1}{r \cdot s} \right)^2 \sum_{i=1}^r \sum_{j=1}^s \frac{1}{n_{ij}} + \frac{r-2}{rs^2} \sum_{j=1}^s \frac{1}{n_{ij}}, \\ \text{Var}(u_{2(j)}) &= \left(\frac{1}{r \cdot s} \right)^2 \sum_{i=1}^r \sum_{j=1}^s \frac{1}{n_{ij}} + \frac{s-2}{r^2 s} \sum_{i=1}^r \frac{1}{n_{ij}} \\ \text{und } \text{Var}(u_{12(ij)}) &= \left(\frac{1}{r \cdot s} \right)^2 \sum_{i=1}^r \sum_{j=1}^s \frac{1}{n_{ij}} + \frac{r-2}{rs^2} \sum_{j=1}^s \frac{1}{n_{ij}} \\ &\quad + \frac{s-2}{r^2 s} \sum_{i=1}^r \frac{1}{n_{ij}} + \frac{(r-2)(s-2)}{rs} \cdot \frac{1}{n_{ij}}. \end{aligned}$$

Beispiel: In einem Karosseriebetrieb werden zwei verschiedene Verfahren zur Rostschutzbehandlung der Metallteile verwendet. Zur Untersuchung der Wirksamkeit der beiden Verfahren wurde unter den Kunden der Firma eine Umfrage gemacht, deren Ergebnisse in Form einer 2×3 -Kontingenztafel in **Tabelle 21** zusammenge stellt sind.

Tab. 21: Wirksamkeit zweier Rostschutzverfahren (2×3 -Tafel)

Rostanfälligkeit (j) Rostschutzverfahren (i)	gering	mittel	stark	Σ
I	65	103	106	274
II	74	85	47	206
Σ	139	188	153	480

Aufgrund dieser Daten wollen wir Schätzungen für die Parameter des zugehörigen loglinearen Modells angeben. Dazu sind die Logarithmen der n_{ij} sowie deren Summen über i bzw. j in **Tab. 22** angegeben.

Damit erhalten wir

$$u = \frac{1}{2 \cdot 3} \cdot 26,06919 = 4,345,$$

Tab. 22: $\ln n_{ij}$ mit den Daten aus Tab. 21

$\ln n_{ij}$	j	1	2	3	$\sum_{j=1}^3 \ln n_{ij}$
i					
1		4,1744	4,6347	4,66344	13,47254
2		4,304	4,44265	3,85	12,59665
$\sum_{i=1}^2 \ln n_{ij}$		8,4784	9,07735	8,51344	26,06919

$$u_{1(1)} = \frac{1}{3} \cdot 13,47254 - \frac{1}{2 \cdot 3} \cdot 26,06919 = 0,146 = - u_{1(2)},$$

$$u_{2(1)} = \frac{1}{2} \cdot 8,4784 - \frac{1}{2 \cdot 3} \cdot 26,06919 = - 0,106,$$

$$u_{2(2)} = \frac{1}{2} \cdot 9,07735 - \frac{1}{2 \cdot 3} \cdot 26,06919 = 0,194,$$

$$u_{2(3)} = \frac{1}{2} \cdot 8,51344 - \frac{1}{2 \cdot 3} \cdot 26,06919 = - 0,088,$$

$$u_{12(11)} = 4,1744 - \frac{1}{3} \cdot 13,47254 - \frac{1}{2} \cdot 8,4784 + \frac{1}{2 \cdot 3} \cdot 26,06919 \\ = - 0,211 = - u_{12(21)},$$

$$u_{12(12)} = 4,6347 - \frac{1}{3} \cdot 13,47254 - \frac{1}{2} \cdot 9,07735 + \frac{1}{2 \cdot 3} \cdot 26,06919 \\ = - 0,050 = - u_{12(22)},$$

$$u_{12(13)} = 4,66344 - \frac{1}{3} \cdot 13,47254 - \frac{1}{2} \cdot 8,51344 + \frac{1}{2 \cdot 3} \cdot 26,06919 \\ = 0,261 = - u_{12(23)}.$$

Die Ergebnisse sind noch einmal übersichtlich in **Tabelle 23** zusammengestellt.
Es sollen nun auch die Varianzen geschätzt werden. Zu diesem Zweck berechnen

Tab. 23: Parameter des loglinearen Modells zu Tab. 21

$u_{12(ij)}$	j	1	2	3	$u_{1(i)}$
i					
1		-0,211	-0,050	0,261	0,146
2		0,211	0,050	-0,261	-0,146
$u_{2(j)}$		-0,106	0,194	-0,088	$u = 4,345$

wir zunächst die Kehrwerte der beobachteten Häufigkeiten n_{ij} . Diese sind, zusammen mit deren Summen über i und j , in **Tabelle 24** angegeben.

Tab. 24: $\frac{1}{n_{ij}}$ mit den Daten aus Tab. 21

$\frac{1}{n_{ij}}$	j	1	2	3	$\sum_{j=1}^3 \frac{1}{n_{ij}}$
i					
1		0,015385	0,0097088	0,009434	0,0345278
2		0,0135135	0,0117647	0,0212766	0,0465548
	$\sum_{i=1}^2 \frac{1}{n_{ij}}$	0,0288985	0,0214735	0,0307106	0,0810826

Es ist nun

$$\text{Var } u = \left(\frac{1}{2 \cdot 3} \right)^2 \cdot 0,0810826 - \frac{1}{480} = 0,00017,$$

$$\text{Var } u_{1(i)} = \left(\frac{1}{2 \cdot 3} \right)^2 \cdot 0,0810826 + 0 = 0,00225,$$

$$\text{Var } u_{2(1)} = \left(\frac{1}{2 \cdot 3} \right)^2 \cdot 0,0810826 + \frac{1}{2^2 \cdot 3} \cdot 0,0288985 = 0,00467,$$

$$\text{Var } u_{2(2)} = \left(\frac{1}{2 \cdot 3} \right)^2 \cdot 0,0810826 + \frac{1}{2^2 \cdot 3} \cdot 0,0214735 = 0,00404$$

und $\text{Var } u_{2(3)} = \left(\frac{1}{2 \cdot 3} \right)^2 \cdot 0,0810826 + \frac{1}{2^2 \cdot 3} \cdot 0,0307106 = 0,0048.$

In diesem Beispiel gilt für die Varianzen der Wechselwirkungsterme

$$\text{Var } u_{12(ij)} = \text{Var } u_{2(j)};$$

da die Kontingenztafel nur zwei Zeilen hat, fallen nämlich bei deren Berechnung die beiden Summanden weg, die den Faktor $r - 2$ enthalten.

Wir erhalten nun als Ergänzung zu Tabelle 23 die **Tabelle 25** der zugehörigen Standardabweichungen.

2.2. Hypothesen und Tests in $r \times s$ -Tafeln

In diesem Abschnitt kommen wir zu der wohl wichtigsten Aufgabe, die das loglineare Modell zu erfüllen hat, nämlich die problemlose Formulierung zu testender Hypothesen. Bis jetzt sind nur **saturierte Modelle**, d. h. Modelle, die alle möglichen Effekte enthalten, betrachtet worden. Man fragt sich jedoch, ob überhaupt alle Effekte einen Einfluß haben. Dazu wollen wir nun einige Hypothesen aufstellen.

Tab. 25: Standardabweichungen der Parameter aus Tab. 23 für das loglineare Modell der Kontingenztafel aus Tab. 21.

$\sqrt{\text{Var } u_{12(ij)}}$		j	1	2	3	$\sqrt{\text{Var } u_{1(i)}}$
i						
1			0,068	0,064	0,069	0,047
2			0,068	0,064	0,069	0,047
$\sqrt{\text{Var } u_{2(j)}}$			0,068	0,064	0,069	$\sqrt{\text{Var } u} = 0,013$

2.2.1. Einige Hypothesen für $r \times s$ -Tafeln

A. Die Unabhängigkeits- bzw. Homogenitätshypothese

Die Frage, ob Wechselwirkungen vorliegen, führt zur Aufstellung der Hypothese

$$H_0^1: u_{12(ij)} = 0;$$

diese Hypothese ist uns bei der 2×2 -Tafel schon einmal als **Unabhängigkeits-** bzw. **Homogenitätshypothese** begegnet. Trifft H_0^1 zu, so muß auch hier gelten

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \text{ bzw. } m_{ij} = \frac{m_{i\cdot} \cdot m_{\cdot j}}{n}.$$

Diese Beziehung läßt sich leicht überprüfen. Man benutzt dabei, daß sich die erwarteten Häufigkeiten m_{ij} unter H_0^1 in folgender Form darstellen lassen:

$$m_{ij} = e^{u + u_{1(i)} + u_{2(j)}}.$$

Damit hat man

$$m_{i\cdot} = e^{u + u_{1(i)}} \cdot \sum_{j=1}^s e^{u_{2(j)}} \text{ und } m_{\cdot j} = e^{u + u_{2(j)}} \cdot \sum_{i=1}^r e^{u_{1(i)}};$$

wir erhalten somit

$$\begin{aligned} m_{i\cdot} \cdot m_{\cdot j} &= e^{u + u_{1(i)}} \cdot e^{u + u_{2(j)}} \cdot \left(\sum_{j=1}^s e^{u_{2(j)}} \right) \left(\sum_{i=1}^r e^{u_{1(i)}} \right) \\ &= e^{u + u_{1(i)} + u_{2(j)}} \cdot e^u \cdot \sum_{j=1}^s \sum_{i=1}^r e^{u_{2(j)} + u_{1(i)}} \\ &= e^{u + u_{1(i)} + u_{2(j)}} \cdot \sum_{i=1}^r \sum_{j=1}^s e^{u + u_{1(i)} + u_{2(j)}} \\ &= e^{u + u_{1(i)} + u_{2(j)}} \cdot \sum_{i=1}^r \sum_{j=1}^s m_{ij} \\ &= m_{ij} \cdot n. \end{aligned}$$

Wir erinnern uns, daß bei der Formulierung der Hypothesen für die 2×2 -Tafel mit eingegangen ist, welches (wahrscheinlichkeitstheoretische) Modell zugunregelegt wurde. Je nach Modell kamen wir entweder zur Unabhängigkeits- oder zur Homogenitätshypothese. Das Testverfahren war jedoch immer das gleiche. Durch die Einführung des loglinearen Modells haben wir nicht nur ein einheitliches Test-

verfahren, sondern auch eine einheitliche Formulierung der Hypothese. Wir müssen uns noch nicht einmal darüber Gedanken machen, welche Wahrscheinlichkeitsmodelle überhaupt in Frage kommen. Das kann – wenn überhaupt – nur noch bei der Interpretation der Testergebnisse nötig sein.

B. Die bedingte Gleichverteilungshypothese

Bevor Testverfahren vorgestellt werden, sollen noch einige andere Hypothesen formuliert werden. Da ist zunächst die Hypothese

$$H_0^2: u_{1(i)} = u_{12(ij)} = 0.$$

Man spricht hier auch von einer **bedingten Gleichverteilungshypothese**, denn es gilt für die bedingten Wahrscheinlichkeiten unter H_0^2

$$\frac{p_{ij}}{p_{.j}} = \frac{m_{ij}}{m_{.j}} = \frac{\frac{e^{u+u_{2(j)}}}{\sum_{i=1}^r e^{u+u_{2(j)}}}}{\frac{e^{u+u_{2(j)}}}{r \cdot e^{u+u_{2(j)}}}} = \frac{1}{r}.$$

Entsprechend lässt sich für die bedingte Gleichverteilung des anderen Merkmals die Hypothese

$$H_0^3: u_{2(j)} = u_{12(ij)} = 0$$

aufstellen.

C. Die totale Gleichverteilungshypothese

Schließlich wollen wir noch die **Hypothese der totalen Gleichverteilung** vorstellen:

$$H_0^4: u_{1(i)} = u_{2(j)} = u_{12(ij)} = 0.$$

Unter dieser Hypothese haben alle Wahrscheinlichkeiten p_{ij} den gleichen Wert; das gilt natürlich genauso für die erwarteten Häufigkeiten

$$m_{ij} = e^u = \frac{1}{r \cdot s} \cdot n.$$

Wie man sieht, haben wir uns bei der Auswahl der Hypothesen auf sogenannte **hierarchische Modelle** beschränkt. Bei diesen Modellen impliziert das Auftreten von Effekten höherer Ordnung das Vorhandensein aller zugehörigen Effekte niedrigerer Ordnung. So werden hier etwa durch die Aufnahme der Wechselwirkungsterme $u_{12(ij)}$ auch die Parameter u , $u_{1(i)}$ und $u_{2(j)}$ der entsprechenden einfachen Effekte mit in das Modell einzbezogen. Das Modell, in dem z. B. die Darstellung

$$\ln m_{ij} = u + u_{1(i)} + u_{12(ij)}$$

gelten soll, fällt deshalb nicht in den hier abgesteckten Rahmen.

Es gibt mehrere Gründe, die es ratsam erscheinen lassen, sich auf hierarchische Modelle zu beschränken. Zum einen sind nichthierarchische Modelle, wenn überhaupt, nur äußerst schwer zu interpretieren, und zum anderen ergeben sich auch noch Schwierigkeiten beim Durchführen der Testprozeduren. Die Schätzungen der erwarteten Häufigkeiten m_{ij} lassen sich oft nur auf Umwegen berechnen, etwa über Beziehungen zu geeigneten Modellen unvollständiger Kontingenztafeln. Hierarchische Modelle sind in dieser Hinsicht recht problemlos und gestatten in den meisten Fällen auch eine angemessene Beschreibung der vorliegenden Situation.

2.2.2. Einige Testverfahren für $r \times s$ -Tafeln

Kommen wir nun zu den **Testverfahren** für die oben angegebenen Hypothesen. Dabei werden zunächst unter der Modellannahme der jeweiligen Hypothese die erwarteten Häufigkeiten m_{ij} geschätzt. Das macht man hier in der Regel mit dem **Maximum-Likelihood-Verfahren (ML-Verfahren)**. Dazu müssen jedoch Verteilungsannahmen gemacht werden. Wie wir am Beispiel der Hypothese H_0^1 für die 2×2 -Tafel gesehen haben, gibt es dafür mehrere Möglichkeiten.

Es ist nun zu befürchten, daß wir je nach Verteilungsannahme verschiedene ML-Schätzwerte erhalten. Glücklicherweise stellt sich aber heraus, daß die unter den verschiedenen möglichen Verteilungsannahmen berechneten ML-Schätzungen identisch sind. Für die Hypothese H_0^1 ergibt sich z. B. als Schätzung der erwarteten Häufigkeiten immer

$$m_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}.$$

A. χ^2 - und Likelihood-Quotienten-Test

Das Testverfahren besteht nun darin, diese geschätzten erwarteten Häufigkeiten mit den tatsächlich beobachteten Häufigkeiten mit Hilfe einer Teststatistik zu vergleichen. Dabei stehen zwei Statistiken zur Auswahl, die χ^2 -Teststatistik

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

und die Likelihood-Quotienten-Statistik

$$G^2 = 2 \cdot \sum_{i=1}^r \sum_{j=1}^s n_{ij} \cdot \ln \frac{n_{ij}}{m_{ij}}.$$

Beide Statistiken sind unter der zu testenden Nullhypothese asymptotisch χ^2 -verteilt mit den gleichen Freiheitsgraden; sie sind sogar asymptotisch äquivalent. Es können also folgendermaßen Tests durchgeführt werden. Für die zu testende Nullhypothese werden ML-Schätzungen der erwarteten Häufigkeiten m_{ij} bestimmt und damit die Teststatistik χ^2 bzw. G^2 berechnet. Ist der berechnete Wert größer als das $(1 - \alpha)$ -Quantil der zugehörigen χ^2 -Verteilung, so verwirft man die Hypothese zum Niveau α . Die hier beschriebenen Hypothesen sind zusammen mit den zugehörigen ML-Schätzwerten für die unter der Nullhypothese erwarteten Häufigkeiten m_{ij} und den Freiheitsgraden für die Verteilung der zugehörigen Teststatistiken χ^2 und G^2 in Tab. 26 aufgeführt.

Bevor wir dazu einige Beispiele rechnen, wollen wir noch die dort angegebenen verschiedenen möglichen Beschreibungen der einzelnen Hypothesen erläutern. Bei Gültigkeit der Hypothese

$$H_0^1: u_{12(ij)} = 0$$

hat man die Beziehung

$$m_{ij} = \frac{m_{i\cdot} \cdot m_{\cdot j}}{n},$$

was gleichwertig ist mit

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}.$$

Tab. 26: Hypothesen für die $r \times s$ -Kontingenztafel

Nullhypothese	$\ln m_{ij}$	$\frac{m_{ij}}{m_{i\cdot}}$	$\frac{m_{ij}}{m_{\cdot j}}$	ML-Schätzer für m_{ij}	Freiheits- grade
$H_0^1: u_{12(ij)} = 0$	$u + u_{1(i)} + u_{2(j)}$	$\frac{m_{i\cdot}}{n}$	$\frac{m_{\cdot j}}{n}$	$\frac{n_{i\cdot} n_{\cdot j}}{n}$	$rs - r - s + 1$ $= (r-1)(s-1)$
$H_0^2: u_{1(i)} = u_{12(ij)} = 0$	$u + u_{2(j)}$	$\frac{1}{r}$	$\frac{m_{\cdot j}}{n}$	$\frac{n_{\cdot j}}{r}$	$rs - s$ $= (r-1)s$
$H_0^3: u_{2(j)} = u_{12(ij)} = 0$	$u + u_{1(i)}$	$\frac{m_{i\cdot}}{n}$	$\frac{1}{s}$	$\frac{n_{i\cdot}}{s}$	$rs - r$ $= r(s-1)$
$H_0^4: u_{1(i)} = u_{2(j)} = u_{12(ij)} = 0$	u	$\frac{1}{r}$	$\frac{1}{s}$	$\frac{n}{rs}$	$rs - 1$

Das bedeutet gerade die **Unabhängigkeit** der beiden beobachteten Merkmale. Schreibt man

$$\frac{p_{ij}}{p_{\cdot j}} = p_{i\cdot},$$

so sieht man, daß alle Spalten der Kontingenzttafel die gleiche (bedingte) Wahrscheinlichkeitsverteilung besitzen. Das gleiche gilt auch für die Zeilen der Kontingenzttafel, denn die obige Beziehung läßt sich ebenso gut umformen zu

$$\frac{p_{ij}}{p_{i\cdot}} = p_{\cdot j}.$$

Entsprechende Beziehungen lassen sich auch für die erwarteten Häufigkeiten m_{ij} aufstellen, wie in Tab. 26 geschehen. Ist die Hypothese der Unabhängigkeit nicht erfüllt, so äußert sich das darin, daß sowohl die Spalten als auch die Zeilen nicht mehr die gleiche Verteilung besitzen. Der Grad der Verschiedenheit kann bei den Spalten größer sein als bei den Zeilen und umgekehrt. Diese Asymmetrie bei Abhängigkeit kommt ebenfalls bei Verwendung der später noch einzuführenden (auch als Assoziationsmaße verwendbaren) Teststatistiken τ_A und τ_B zum Ausdruck.

Dagegen sorgen die Statistiken χ^2 und G^2 für eine symmetrische Behandlung des Problems. Nur diese beiden Statistiken sind geeignet, mit den entsprechenden ML-Schätzern für die m_{ij} , die drei anderen Hypothesen zu testen. Im Unterschied zur Hypothese H_0^1 sind hier die bedingten Verteilungen der Zeilen und/oder Spalten nicht nur gleich, sondern es handelt sich hierbei auch noch um Gleichverteilungen.

Beispiel: Wir wollen uns noch einmal mit der Untersuchung über die Wirksamkeit zweier Rostschutzverfahren befassen, deren Ergebnisse in Tab. 21 wiedergegeben wurden. Uns interessiert, ob ein Zusammenhang besteht zwischen der Rostanfälligkeit und dem verwendeten Rostschutzverfahren. Es ist also die Hypothese H_0^1 zu testen. Zur Berechnung der Teststatistiken müssen zunächst die ML-Schätzer $n_{i\cdot} \cdot n_{\cdot j}/n$ für m_{ij} berechnet werden. Man erhält

$$\frac{n_{1 \cdot} \cdot n_{\cdot 1}}{n} = 79,3, \quad \frac{n_{1 \cdot} \cdot n_{\cdot 2}}{n} = 107,3, \quad \frac{n_{1 \cdot} \cdot n_{\cdot 3}}{n} = 87,3,$$

$$\frac{n_{2 \cdot} \cdot n_{\cdot 1}}{n} = 59,7, \quad \frac{n_{2 \cdot} \cdot n_{\cdot 2}}{n} = 80,7, \quad \frac{n_{2 \cdot} \cdot n_{\cdot 3}}{n} = 65,7.$$

Damit ergibt sich

$$\begin{aligned}\chi^2 &= \frac{(65 - 79,3)^2}{79,3} + \frac{(103 - 107,3)^2}{107,3} + \frac{(106 - 87,3)^2}{87,3} \\ &\quad + \frac{(74 - 59,7)^2}{59,7} + \frac{(85 - 80,7)^2}{80,7} + \frac{(47 - 65,7)^2}{65,7} \\ &= 2,6 + 0,17 + 4 + 3,4 + 0,23 + 5,3 = 15,7\end{aligned}$$

$$\begin{aligned}\text{und } G^2 &= 2(65 \cdot \ln \frac{65}{79,3} + 103 \cdot \ln \frac{103}{107,3} + 106 \cdot \ln \frac{106}{87,3} \\ &\quad + 74 \cdot \ln \frac{74}{59,7} + 85 \cdot \ln \frac{85}{80,7} + 47 \cdot \ln \frac{47}{65,7}) \\ &= 2(-12,9 - 4,2 + 20,6 + 15,9 + 4,4 - 15,7) = 16,2.\end{aligned}$$

Beide Teststatistiken sind asymptotisch χ^2 -verteilt mit $r \cdot s - r - s + 1 = 2 \cdot 3 - 2 - 3 + 1 = 2$ Freiheitsgraden. Testen wir zum Niveau $\alpha = 0,01$, so erhalten wir den kritischen Wert als $1 - \alpha = 0,99$ -Quantil der χ^2 -Verteilung mit 2 Freiheitsgraden. Man sieht dafür den Wert 9,21 nach, vgl. die Tafel im Anhang. Beide Teststatistiken verwerfen also die Hypothese. Das hätte man auch aufgrund der früher schon berechneten Wechselwirkungsterme $u_{12(ij)}$ vermutet.

Beispiel: Ein Hersteller von Büroartikeln fertigt Aktenordner in den Farben gelb, rot und blau an. Zur Analyse des Kaufverhaltens wurde für die drei wichtigsten Absatzmärkte A, B und C jeweils die Anzahl der in einem bestimmten Zeitraum geordneten Bestelleinheiten ermittelt. Es ergaben sich die Werte aus **Tabelle 27**.

Tab. 27: Absatzmengen verschiedenfarbiger Ordner

Absatz- markt \ Ordnerfarbe	gelb	rot	blau	Σ
A	546	607	609	1762
B	316	298	317	931
C	458	399	425	1282
Σ	1320	1304	1351	3975

Wir wollen hier testen, ob in den drei Märkten A, B und C die drei Ordnerfarben unterschiedlich beliebt sind; H_0^3 ist also unsere Nullhypothese. Unter dieser Hypothese erhalten wir für die ML-Schätzungen der erwarteten Häufigkeiten m_{ij} nun

$$\frac{n_1}{s} = 587,3 \quad \text{für } m_{11}, m_{12} \text{ und } m_{13},$$

$$\frac{n_2}{s} = 310,3 \quad \text{für } m_{21}, m_{22} \text{ und } m_{23} \text{ und}$$

$$\frac{n_3}{s} = 427,3 \quad \text{für } m_{31}, m_{32} \text{ und } m_{33}.$$

Damit ergibt sich für die Teststatistik χ^2 gerade

$$\begin{aligned}\chi^2 &= \frac{(546 - 587,3)^2}{587,3} + \frac{(607 - 587,3)^2}{587,3} + \frac{(609 - 587,3)^2}{587,3} \\ &\quad + \frac{(316 - 310,3)^2}{310,3} + \frac{(298 - 310,3)^2}{310,3} + \frac{(317 - 310,3)^2}{310,3} \\ &\quad + \frac{(458 - 427,3)^2}{427,3} + \frac{(399 - 427,3)^2}{427,3} + \frac{(425 - 427,3)^2}{427,3} \\ &= 2,90 + 0,66 + 0,8 + 0,1 + 0,49 + 0,14 + 2,21 + 1,87 + 0,012 \\ &= 9,2\end{aligned}$$

$$\begin{aligned}\text{und für } G^2 &= 2 \cdot \left(546 \cdot \ln \frac{546}{587,3} + 607 \cdot \ln \frac{607}{587,3} + 609 \cdot \ln \frac{609}{587,3} \right. \\ &\quad + 316 \cdot \ln \frac{316}{310,3} + 298 \cdot \ln \frac{298}{310,3} + 317 \cdot \ln \frac{317}{310,3} \\ &\quad \left. + 458 \cdot \ln \frac{458}{427,3} + 399 \cdot \ln \frac{399}{427,3} + 425 \cdot \ln \frac{425}{427,3} \right) \\ &= 2(-39,8 + 20,03 + 22,1 + 5,75 - 12,05 + 6,77 + 31,8 - 27,3 - 2,29) \\ &= 10,0.\end{aligned}$$

Beide Teststatistiken sind (bei Gültigkeit der Nullhypothese) asymptotisch χ^2 -verteilt mit $r \cdot s - r = 3 \cdot 3 - 3 = 6$ Freiheitsgraden. Das 0,95-Quantil der χ^2 -Verteilung mit 6 Freiheitsgraden ist 12,6, so daß die Hypothese zum 5%-Niveau nicht verworfen werden kann, egal mit welcher der beiden Statistiken man arbeitet.

In den obigen Beispielen liegen die Werte der beiden Statistiken recht nah beieinander. Da der kritische Wert jeweils relativ stark über- bzw. unterschritten wird, ergeben sich dadurch mit beiden Statistiken die gleichen Entscheidungen. Es können jedoch Situationen auftreten, in denen das nicht so ist. Man muß sich also die Frage stellen, welches die geeigneteren der beiden Statistiken ist. Bei genügend großen Stichprobenumfängen ergeben sich wegen der asymptotischen Äquivalenz der Teststatistiken praktisch die gleichen Werte. Dann ist es gleichgültig, welche man verwendet. Ansonsten werden unterschiedliche Auffassungen vertreten. Lange Zeit wurde fast ausschließlich die χ^2 -Statistik verwendet. Dabei hat sicherlich mit einer Rolle gespielt, daß ohne den Einsatz elektronischer Rechner die Berechnung von G^2 erheblich aufwendiger ist als die von χ^2 .

Es gibt nun gewichtige Gründe, die für die Statistik G^2 sprechen. Diese Statistik beruht auf einem anerkannten, auch sonst häufig benutzten Testprinzip, nämlich

der Verwendung eines ML-Quotienten. Ein zweiter Zugang ist informationstheoretisch begründet. Der Verlust an Information, wenn wir vom saturierten Modell mit allen Parametern übergehen zu dem in der Nullhypothese beschriebenen Modell, läßt sich durch $G^2/2$ schätzen. Die χ^2 -Teststatistik ist dagegen mehr durch heuristische Überlegungen entstanden. So scheinen die Empfehlungen, die Statistik G^2 zu bevorzugen, nicht einer gewissen Grundlage zu entbehren.

Eine gewisse Unsicherheit entsteht nun dadurch, daß zur Durchführung der Tests nicht die exakte Verteilung der Prüfgröße, sondern eine geeignete Approximation durch die χ^2 -Verteilung benutzt wird. Eine grobe Faustregel besagt, daß man bei der χ^2 -Teststatistik mit einer zufriedenstellenden Annäherung rechnen kann, wenn in jeder Zelle mindestens fünf Beobachtungen vorliegen bzw. unter der Hypothese erwartet werden können. Auf Yarnold (1970) geht die folgende Regel zurück: Ist n_0 die Anzahl der Zellen in der Kontingenztafel mit Häufigkeiten kleiner oder gleich fünf, so sollten alle Häufigkeiten mindestens $5 \cdot n_0/n$ sein, damit die Verteilung der χ^2 -Teststatistik ausreichend genau durch die χ^2 -Verteilung beschrieben wird. Ein Vergleich der Statistiken χ^2 und G^2 wird von Margolin/Light (1974) durchgeführt. Es wird gezeigt, daß bei $r \times 2$ -Tafeln mit $n_{11} = n_{12} = n/2$ beim Testen der Hypothese H_0^1 immer gilt

$$\chi^2 < G^2.$$

Das gilt dann natürlich auch bei entsprechenden $2 \times s$ -Tafeln. Bei der Benutzung von G^2 kommt es also, da bei beiden Tests der gleiche kritische Wert benutzt wird, eher zur Ablehnung als bei χ^2 . Die Güte der Anpassung an die χ^2 -Verteilung bei Gültigkeit der Nullhypothese wird an einer sehr kleinen Tafel, nämlich der 3×2 -Tafel, untersucht. Dort sind am ehesten größere Unterschiede zu erwarten. Hierbei zeigt sich (bei gleichen Spaltensummen in der 3×2 -Tafel), daß die Verteilung der χ^2 -Teststatistik meist besser durch die χ^2 -Verteilung beschrieben wird als die von G^2 .

B. Die Statistiken τ_A und τ_B

Eine noch bessere Anpassung an die χ^2 -Verteilung mit $rs - r - s + 1$ Freiheitsgraden scheint man unter der Nullhypothese H_0^1 mit der folgenden Statistik (zum Prüfen von H_0^1) erreichen zu können:

$$(n - 1)(r - 1)\tau_A = (n - 1)(r - 1) \frac{\sum_{j=1}^s \frac{1}{n_{\cdot j}} \sum_{i=1}^r n_{ij}^2 - \frac{1}{n} \sum_{i=1}^r n_{\cdot i}^2}{n - \frac{1}{n} \sum_{i=1}^r n_{\cdot i}^2}.$$

Eine Interpretation des „Assoziationsmaßes“ τ_A findet man in Abschnitt 3. Dort wird gezeigt, daß man diese Statistik durch den Vergleich der „Variabilität“ innerhalb der Spalten der Kontingenztafel mit der „Gesamtvariabilität“ erhalten kann.

Beispiel: Für die Daten aus Tabelle 21 (Vergleich von zwei Rostschutzverfahren) ergibt sich

$$(n - 1)(r - 1)\tau_A = (480 - 1)(2 - 1) \cdot \frac{252,53 - 244,82}{480 - 244,82} = 15,7.$$

Für das 0,99-Quantil der χ^2 -Verteilung mit $r \cdot s - r - s + 1 = 2 \cdot 3 - 2 - 3 + 1 = 2$

Freiheitsgraden ergibt sich 9,21, so daß die Hypothese H_0^1 zum Niveau $\alpha = 0,01$ abgelehnt wird.

Man kann sagen, daß die Gesamtvariabilität signifikant größer ist als die Variabilität innerhalb der drei Gruppen, die durch die wenig, mittel und stark rostanfälligen Metallteile gebildet werden. Es liegt also ein Gruppeneffekt vor, d. h. die drei Gruppen sind unterschiedlich strukturiert hinsichtlich der beiden Arten rotschutzbehandelter Metallteile. Daraus kann auch geschlossen werden, daß die beiden Rostschutzverfahren unterschiedlich wirken.

Wie man sieht, ist die Statistik $(n - 1)(r - 1)\tau_A$ nicht symmetrisch aufgebaut. Werden die Rollen von Zeilen und Spalten vertauscht, so erhält man die Statistik

$$(n - 1)(s - 1)\tau_B = (n - 1)(s - 1) \cdot \frac{\sum_{i=1}^r \frac{1}{n_{i\cdot}} \sum_{j=1}^s n_{ij}^2 - \frac{1}{n} \sum_{j=1}^s n_{\cdot j}^2}{n - \frac{1}{n} \sum_{j=1}^s n_{\cdot j}^2}.$$

Diese ist unter der Nullhypothese H_0^1 ebenfalls χ^2 -verteilt mit $r \cdot s - r - s + 1$ Freiheitsgraden.

Beispiel: Für das obige Beispiel erhält man

$$(n - 1)(s - 1)\tau_B = (480 - 1)(3 - 1) \frac{167,52 - 162,65}{480 - 162,65} = 14,7.$$

Man kann hier sagen, daß die Gesamtvariabilität signifikant größer ist als die Variabilität innerhalb der beiden Gruppen, die durch die verschieden behandelten Metallteile gebildet werden. Es liegt also ein Gruppeneffekt vor, d. h. die beiden Rostschutzverfahren wirken unterschiedlich.

C. Der Test auf Symmetrie nach Bowker

Bei der 2×2 -Tafel haben wir einen Test auf Symmetrie kennengelernt, nämlich den Test von McNemar. Hier soll nun ein entsprechender Test für beliebige $r \times r$ -Kontingenztafeln vorgestellt werden. Die zu testende Hypothese lautet

$$H_0: m_{ij} = m_{ji},$$

bzw. in der Schreibweise des loglinearen Modells

$$H_0: u_{12(ij)} = u_{12(ji)} \quad \text{und} \quad u_{1(i)} = u_{2(i)}.$$

Unter dieser Hypothese erhalten wir für die ML-Schätzungen der erwarteten Häufigkeiten m_{ij} die Werte

$$\frac{n_{ij} + n_{ji}}{2}.$$

Damit ergeben sich die Statistiken

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^r \frac{\left(n_{ij} - \frac{n_{ij} + n_{ji}}{2} \right)^2}{\frac{n_{ij} + n_{ji}}{2}} = \sum_{i>j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}$$

$$\text{und } G^2 = 2 \cdot \sum_{i \neq j} n_{ij} \cdot \ln \frac{2 \cdot n_{ij}}{n_{ij} + n_{ji}},$$

die beide unter der Hypothese H_0 asymptotisch χ^2 -verteilt sind mit $r(r - 1)/2$ Freiheitsgraden.

Die χ^2 -Statistik für den Test auf Symmetrie bei der allgemeinen $r \times r$ -Tafel wurde zuerst von Bowker (1948) angegeben und ist im Fall $r = 2$ identisch mit der McNemar-Teststatistik.

Beispiel: Beim Vergleich zweier Waschmittel hat man die in Tab. 28 aufgeführten Häufigkeiten für die Einstufung in die drei Qualitätsklassen gut, mittel, schlecht erhalten.

Tab. 28: Vergleich zweier Waschmittel

Waschmittel II		schlecht	mittel	gut	Σ
Waschmittel I	schlecht	23	27	15	65
	mittel	45	86	92	223
Σ		77	223	222	522

Wir wollen die Hypothese testen, daß Symmetrie vorliegt und damit beide Waschmittel als gleich gut angesehen werden können. Dazu berechnen wir die Statistiken

$$\begin{aligned} \chi^2 &= \sum_{i>j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \\ &= \frac{(45 - 27)^2}{45 + 27} + \frac{(9 - 15)^2}{9 + 15} + \frac{(110 - 92)^2}{110 + 92} \\ &= 4,5 + 1,5 + 1,6 = 7,6 \end{aligned}$$

$$\begin{aligned} \text{und } G^2 &= 2 \cdot \sum_{i \neq j} n_{ij} \cdot \ln \frac{2 \cdot n_{ij}}{n_{ij} + n_{ji}} \\ &= 2 \left(27 \cdot \ln \frac{2 \cdot 27}{27 + 45} + 45 \cdot \ln \frac{2 \cdot 45}{27 + 45} + 15 \cdot \ln \frac{2 \cdot 15}{15 + 9} \right. \\ &\quad \left. + 9 \cdot \ln \frac{2 \cdot 9}{15 + 9} + 92 \cdot \ln \frac{2 \cdot 92}{92 + 110} + 110 \cdot \ln \frac{2 \cdot 110}{92 + 110} \right) \\ &= 2(-7,77 + 10,04 + 3,35 - 2,59 - 8,59 + 9,39) = 7,66. \end{aligned}$$

Das 0,95-Quantil der χ^2 -Verteilung mit $r(r - 1)/2 = 3 \cdot 2/2 = 3$ Freiheitsgraden hat den Wert 7,81. Die Hypothese der Symmetrie kann also zum Niveau $\alpha = 5\%$ nicht abgelehnt werden.

3. Assoziationsmaße für 2×2 - und $r \times s$ -Tafeln

Oft möchte man die Abhängigkeitsstruktur einer Kontingenztafel durch eine einzige reelle Zahl charakterisieren. Ein solcher Wert wird jedoch, außer vielleicht im Fall der 2×2 -Tafel, nicht alle Aspekte des Zusammenhangs (der Assoziation) der beiden Merkmale beschreiben können. Daher ist es nicht verwunderlich, daß im Laufe der Zeit eine Vielzahl solcher **Assoziationsmaße** vorgeschlagen wurde. Einige davon sind speziell nur für die 2×2 -Tafel verwendbar. Das ist auch der Grund dafür, der Vierfeldertafel einen eigenen Abschnitt zu widmen.

3.1. Assoziationsmaße in der 2×2 -Kontingenztafel

Alle hier vorgestellten Maße lassen sich entweder auf den oft als „**cross-product ratio**“ oder „**odds ratio**“ bezeichneten Quotienten

$$q = \frac{p_{11}p_{22}}{p_{12}p_{21}}$$

oder den „**Korrelationskoeffizienten**“

$$\varrho = \frac{p_{11}p_{22} - p_{21}p_{12}}{\sqrt{p_{11}p_{22}p_{12}p_{21}}}$$

zurückführen. Es sei schon jetzt darauf hingewiesen, daß sich die meisten Assoziationsmaße, die für beliebige $r \times s$ -Tafeln vorgeschlagen werden, im Spezialfall der Vierfeldertafel als Funktion von ϱ schreiben lassen.

3.1.1. Assoziationsmaße, die in Beziehung zum cross-product ratio q stehen

A. Die Eigenschaften des cross-product ratio

Beschreibt man die Vierfeldertafel mit einem loglinearen Modell, so erhält man als Wechselwirkung den Term

$$\ln \frac{p_{11}p_{22}}{p_{12}p_{21}} = \ln q.$$

Dieser Wert oder der Quotient q selbst erscheint daher als geeignet, den Zusammenhang der beiden Merkmale in einer 2×2 -Tafel zu beschreiben; q kann die Null und sämtliche positiven Zahlen annehmen – bei Unabhängigkeit ist $q = 1$ (und damit $\ln q = 0$). Vertauscht man in der Kontingenztafel entweder die Zeilen oder die Spalten, so wird q in $1/q$ überführt und entsprechend $\ln q$ in $-\ln q$. Im Gegensatz zu ϱ und allen später noch zu besprechenden Assoziationsmaßen für allgemeine $r \times s$ -Tafeln, haben die auf q basierenden Maße die folgende **Invarianzeigenschaft**: Multipliziert man Zeilen und/oder Spalten der Kontingenztafel mit beliebigen positiven Zahlen α_i bzw. β_j (haben wir es mit einer Tafel von Wahrscheinlichkeiten zu tun, so muß anschließend natürlich wieder normiert werden), so bleibt das Assoziationsmaß unverändert. Wenn wir also die Tafel aus Tab. 29 überführen in die Tafel aus Tab. 30, so bleibt der Wert von q erhalten und damit auch jedes darauf basierende Maß. Wir haben demnach eine weitgehende Unabhängigkeit von den speziellen Randwahrscheinlichkeiten.

Als Assoziationsmaße werden meist Funktionen von q benutzt, die wie ein Korrelationskoeffizient nur Werte im Bereich von -1 bis $+1$ liefern, vgl. Kap. IX. Die Null soll sich bei Unabhängigkeit der beiden Merkmale ergeben.

Tab. 29

		b_1	b_2
		p_{11}	p_{12}
X_1	a_1	p_{21}	p_{22}
	a_2		

Tab. 30

		b_1	b_2
		$(\sum_{i,j} \alpha_i \beta_j p_{ij})^{-1} \alpha_1 \beta_1 p_{11}$	$(\sum_{i,j} \alpha_i \beta_j p_{ij})^{-1} \alpha_1 \beta_2 p_{12}$
X_1	a_1	$(\sum_{i,j} \alpha_i \beta_j p_{ij})^{-1} \alpha_2 \beta_1 p_{21}$	$(\sum_{i,j} \alpha_i \beta_j p_{ij})^{-1} \alpha_2 \beta_2 p_{22}$
	a_2		

B. Der Q-Koeffizient von Yule

Um 1900 wurde von Yule der **Q-Koeffizient**

$$Q = \frac{p_{11}p_{22} - p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}} = \frac{q - 1}{q + 1}$$

vorgeschlagen. Schreibt man

$$Q = \frac{p_{11}p_{22}}{p_{11}p_{22} + p_{12}p_{21}} - \frac{p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}},$$

so sieht man, daß sich Q als Differenz zweier bedingter Wahrscheinlichkeiten auffassen läßt. Zur Interpretation nun ein Beispiel.

Beispiel: Die unterschiedlichen Rauchgewohnheiten von Ehepartnern lassen sich mit Hilfe theoretischer Wahrscheinlichkeiten p_{ij} in einer Kontingenztafel beschreiben, s. **Tab. 31**.

Tab. 31: Rauchgewohnheiten von Ehepartnern

		Mann	Raucher	Nichtraucher
			p_{11}	p_{12}
Frau	Raucher	p_{21}		p_{22}
	Nichtraucher			

Bei zwei Ehepaaren ist die Wahrscheinlichkeit dafür, daß beide Partner des einen Paars Raucher und die des anderen Nichtraucher sind $p_{11} \cdot p_{22}$. Konträre Verhaltensweisen bei beiden Paaren kommen mit der Wahrscheinlichkeit $p_{12} \cdot p_{21}$ vor. Man sieht nun, daß der Yulesche Q-Koeffizient auf dem Vergleich dieser beiden Wahrscheinlichkeiten beruht. Ein positiver Wert von Q läßt auf gleichartige Rauchgewohnheiten schließen, während negative Werte konträre Verhaltensweisen erwarten lassen.

C. Der Verbundenheitskoeffizient von Yule

Der **Verbundenheitskoeffizient**, den Yule 1912 eingeführt hat, ist gegeben durch

$$Y = \frac{\sqrt{p_{11}p_{22}} - \sqrt{p_{12}p_{21}}}{\sqrt{p_{11}p_{22}} + \sqrt{p_{12}p_{21}}} = \frac{\sqrt{q} - 1}{\sqrt{q} + 1}.$$

Ihm liegt folgende Überlegung zugrunde: Jede 2×2 -Tafel lässt sich durch Multiplikation der Zeilen und Spalten mit geeigneten Werten (dabei bleibt, wie oben erwähnt, q erhalten) so transformieren, daß alle Randwahrscheinlichkeiten gleich $1/2$ sind. Das sieht dann aus wie in Tab. 32 dargestellt.

Tab. 32: Darstellung der p_{ij} als Funktion von $q = \frac{p_{11}p_{22}}{p_{12}p_{21}}$ bei einer 2×2 -Tafel mit Randwahrscheinlichkeiten $1/2$

		b_1	b_2	Σ
		X_1		
a_1	X_2	$\frac{1}{2} \cdot \frac{\sqrt{q}}{\sqrt{q} + 1}$	$\frac{1}{2} \cdot \frac{1}{\sqrt{q} + 1}$	$\frac{1}{2}$
		$\frac{1}{2} \cdot \frac{1}{\sqrt{q} + 1}$	$\frac{1}{2} \cdot \frac{\sqrt{q}}{\sqrt{q} + 1}$	$\frac{1}{2}$
Σ		$\frac{1}{2}$	$\frac{1}{2}$	1

Da nun alle Informationen, die in den Randwahrscheinlichkeiten stecken, eliminiert sind, ist es naheliegend, als Assoziationsmaß

$$2 \cdot \left(\frac{1}{2} \cdot \frac{\sqrt{q}}{\sqrt{q} + 1} - \frac{1}{2} \cdot \frac{1}{\sqrt{q} + 1} \right) = Y$$

zu verwenden.

D. Punkt- und Intervallschätzungen für die Yuleschen Assoziationsmaße

Sowohl Q als auch Y hängen von den in der Regel unbekannten Wahrscheinlichkeiten p_{ij} ab. Wir ersetzen deshalb die p_{ij} durch Schätzwerte n_{ij}/n und erhalten die Zufallsgrößen

$$\hat{Q} = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}} \quad (\text{Q-Koeffizient von Yule})$$

$$\text{und} \quad \hat{Y} = \frac{\sqrt{n_{11}n_{22}} - \sqrt{n_{12}n_{21}}}{\sqrt{n_{11}n_{22}} + \sqrt{n_{12}n_{21}}} \quad (\text{Verbundenheitskoeffizient von Yule})$$

als Schätzer für Q bzw. Y .

Diese sind **asymptotisch normalverteilt** mit Mittelwerten Q bzw. Y und Standardabweichungen

$$\sigma_Q = \frac{1}{2}(1 - Q^2) \sqrt{\frac{1}{n} \cdot \left(\frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}} \right)}$$

und $\sigma_Y = \frac{1}{4}(1 - Y^2) \sqrt{\frac{1}{n} \cdot \left(\frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}} \right)}.$

σ_Q und σ_Y können geschätzt werden durch

$$\hat{\sigma}_Q = \frac{1}{2}(1 - \hat{Q}^2) \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

und $\hat{\sigma}_Y = \frac{1}{4}(1 - \hat{Y}^2) \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}.$

Damit sind wir in der Lage, **approximative Konfidenzintervalle** zum Niveau $1 - \alpha$ für Q und Y anzugeben, nämlich

$$[\hat{Q} - u_{1-\alpha/2} \cdot \hat{\sigma}_Q; \hat{Q} + u_{1-\alpha/2} \cdot \hat{\sigma}_Q] \quad \text{für } Q \text{ und}$$

$$[\hat{Y} - u_{1-\alpha/2} \cdot \hat{\sigma}_Y; \hat{Y} + u_{1-\alpha/2} \cdot \hat{\sigma}_Y] \quad \text{für } Y.$$

Beispiel: Um die Wirksamkeit einer Werbekampagne für einen Kräuterlikör zu testen, wird sowohl vor als auch nach der Durchführung der Kampagne eine Umfrage gemacht. Die Ergebnisse sind in der 2×2 -Tafel der Tab. 33 zusammengestellt.

Tab. 33: Wirksamkeit einer Werbekampagne

Zeitpunkt	Bekanntheitsgrad	nicht bekannt	bekannt	Σ
		vorher	nachher	
		276	163	439
		438	512	950
	Σ	714	675	1389

Der Einfluß der Werbemaßnahme auf den Bekanntheitsgrad des Likörs kann durch ein Assoziationsmaß beschrieben werden. Es ergibt sich hier

$$\hat{q} = \frac{276 \cdot 512}{163 \cdot 438} = 1,98,$$

$$\hat{Q} = \frac{276 \cdot 512 - 163 \cdot 438}{276 \cdot 512 + 163 \cdot 438} = 0,329$$

und $\hat{Y} = \frac{\sqrt{276 \cdot 512} - \sqrt{163 \cdot 438}}{\sqrt{276 \cdot 512} + \sqrt{163 \cdot 438}} = 0,169.$

Für die Standardabweichungen von \hat{Q} und \hat{Y} erhält man die Schätzungen

$$\begin{aligned}\hat{\sigma}_Q &= \frac{1}{2}(1 - 0,329^2) \cdot \sqrt{\frac{1}{276} + \frac{1}{163} + \frac{1}{438} + \frac{1}{512}} \\ &= \frac{1}{2} \cdot 0,8918 \cdot \sqrt{0,01399} = 0,0527\end{aligned}$$

und $\hat{\sigma}_Y = \frac{1}{4}(1 - 0,169^2) \cdot \sqrt{\frac{1}{276} + \frac{1}{163} + \frac{1}{438} + \frac{1}{512}}$

$$= \frac{1}{4} \cdot 0,9714 \cdot \sqrt{0,01399} = 0,0287.$$

Damit ergeben sich die Grenzen für das 95% Konfidenzintervall für Q zu

$$0,329 - 1,96 \cdot 0,0527 = 0,226 \quad \text{und} \quad 0,329 + 1,96 \cdot 0,0527 = 0,432$$

und für Y zu

$$0,169 - 1,96 \cdot 0,0287 = 0,113 \quad \text{und} \quad 0,169 + 1,96 \cdot 0,0287 = 0,225.$$

Es kann also gesichert davon ausgegangen werden, daß die Werbeaktion den Bekanntheitsgrad der Kräuterlikörs gesteigert hat.

Im folgenden werden wir der Einfachheit halber nicht mehr streng zwischen dem Assoziationsmaß und dessen Schätzung unterscheiden. So bezeichnen wir fortan sowohl Q als auch \hat{Q} mit Q, σ_Q und $\hat{\sigma}_Q$ mit σ_Q usw.

E. Eigenschaften der Yuleschen Assoziationsmaße

Wir fassen nun die wichtigsten Eigenschaften von Q und Y zusammen:

Q und Y liegen immer zwischen -1 und +1. Ist wenigstens eine Zelle in der Kontingenztafel gleich Null, so wird entweder der Wert -1 oder +1 angenommen. Bei Unabhängigkeit ergibt sich der Wert Null.

Beide Maße bleiben unverändert, wenn Zeilen und/oder Spalten der Kontingenztafel mit beliebigen positiven Zahlen multipliziert werden. Ein Vertauschen der Zeilen oder Spalten ändert nur das Vorzeichen.

Q und Y spiegeln die gleiche Ordnung wider. Ist also $Q_1 < Q_2$, so gilt auch $Y_1 < Y_2$ und umgekehrt.

Es ist immer $|Y| \leq |Q|$.

3.1.2. Assoziationsmaße, die in Beziehung zum Korrelationskoeffizienten ϱ stehen

A. Der Korrelationskoeffizient ϱ (Phi-Koeffizient) und seine Eigenschaften

Wir wenden uns nun der zweiten Gruppe von Zusammenhangsmaßen für die 2×2 -Tafel zu, nämlich derjenigen, die sich auf dem Korrelationskoeffizienten

$$\varrho = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_1 \cdot n_2 \cdot n_1 \cdot n_2}}$$

begründet. Die Zahl ϱ nennt man auch Phi-Koeffizient. Man erhält ihn durch die folgende Überlegung. Betrachten wir eine 2×2 -Kontingenztafel, etwa die aus Tab. 34.

Tab. 34: Rauchgewohnheiten nach Geschlecht

		Rauchgewohnheit		Σ
		Raucher	Nichtraucher	
Geschlecht	weiblich	p_{11}	p_{12}	$p_{1\cdot}$
	männlich	p_{21}	p_{22}	$p_{2\cdot}$
Σ		$p_{\cdot 1}$	$p_{\cdot 2}$	1

Codiert man die Ausprägungen der beiden Merkmale Rauchgewohnheit und Geschlecht jeweils mit Null und Eins, so ist die Darstellung aus **Tab. 35** gegeben.

Tab. 35: Codierte Tafel zu Tab. 34

		Rauchgewohnheit		Σ
		0	1	
Geschlecht	0	p_{11}	p_{12}	$p_{1\cdot}$
	1	p_{21}	p_{22}	$p_{2\cdot}$
Σ		$p_{\cdot 1}$	$p_{\cdot 2}$	1

Das ist die gemeinsame Verteilung von zwei abhängigen binomialverteilten Zufallsgrößen. Dabei hat das Merkmal Geschlecht eine $B(1, p_{\cdot \cdot})$ und das Merkmal Rauchgewohnheit eine $B(1, p_{\cdot 2})$ -Verteilung. Für Mittelwert und Varianz hat man

$$p_{\cdot 2} \text{ und } p_{\cdot 2}(1 - p_{\cdot 2}) = p_{\cdot 2}p_{\cdot 1}$$

für das Geschlecht bzw.

$$p_{\cdot 1} \text{ und } p_{\cdot 1}(1 - p_{\cdot 1}) = p_{\cdot 1}p_{\cdot 2}$$

für die Rauchgewohnheit.

Als Kovarianz erhält man $p_{22} - p_{\cdot 2} \cdot p_{\cdot 1}$ und damit den Korrelationskoeffizienten

$$\frac{p_{22} - p_{\cdot 2}p_{\cdot 1}}{\sqrt{p_{\cdot 2}p_{\cdot 1}p_{\cdot 1}p_{\cdot 2}}} = \frac{p_{11}p_{22} - p_{21}p_{12}}{\sqrt{p_{\cdot 1}p_{\cdot 2}p_{\cdot 1}p_{\cdot 2}}} = \varrho.$$

Den gleichen Korrelationskoeffizienten erhält man, wenn den Merkmalsausprägungen statt Null und Eins irgendwelche anderen Zahlen zugeordnet werden. Dabei müssen allerdings die zweiten Zeilen und Spalten der Kontingenztafel die größeren Werte erhalten, sonst kann sich das Vorzeichen ändern. Demnach hat ϱ wie Q und Y die Eigenschaft, beim Vertauschen von Zeilen oder Spalten das Vorzeichen zu wechseln.

Die Invarianzeigenschaft, die Q und Y auszeichnet (keine Änderung bei Multiplikation der Zeilen und Spalten mit beliebigen positiven Werten, also Unabhängigkeit von den Randwahrscheinlichkeiten) hat ϱ nicht. Es gilt $|\varrho| = |Y|$, wenn alle Randwahrscheinlichkeiten gleich 0,5 sind, und immer $|\varrho| \leq |Y|$. Dabei wird $|\varrho|$

umso kleiner, je unterschiedlicher die Randwahrscheinlichkeiten sind. Dazu ein Beispiel.

Tab. 36: 2×2 -Tafel mit identischen Randhäufigkeiten

X ₁	X ₂	b ₁	b ₂	Σ
a ₁		20	10	30
a ₂		10	20	30
Σ		30	30	60

Tab. 37: 2×2 -Tafel mit stark unterschiedlichen Randhäufigkeiten

X ₁	X ₂	b ₁	b ₂	Σ
a ₁		20000	10000	30000
a ₂		10	20	30
Σ		20010	10020	30030

Beispiel: Die Tafeln aus **Tab. 36** und **Tab. 37** ergeben für Y aufgrund der Invarianzegenschaft den gleichen Wert. Die Tafel aus Tab. 37 ist nämlich aus der in Tab. 36 wiedergegebenen Tafel durch Multiplikation der ersten Zeile mit 1000 entstanden. Für ρ erhält man bei der ersten Tafel

$$\rho = \frac{20 \cdot 20 - 10 \cdot 10}{\sqrt{30 \cdot 30 \cdot 30 \cdot 30}} = \frac{300}{900} = \frac{1}{3} = Y$$

und bei der zweiten

$$\rho = \frac{20000 \cdot 20 - 10000 \cdot 10}{\sqrt{30000 \cdot 30 \cdot 20010 \cdot 10020}} = \frac{300000}{13433175} = 0,022$$

(also fast Null). Das läßt sich dadurch erklären, daß für Merkmal X₁ bei der Tafel aus Tab. 37 meist die Ausprägung a₁ vorliegt, unabhängig vom Merkmal X₂. Möchte man dagegen aufgrund der Kenntnis von X₁ eine Vorhersage für X₂ machen, so hat man bei beiden Tafeln prinzipiell die gleiche Situation vorliegen. Bei dieser Betrachtungsweise erscheint also Y (bzw. Q) als das geeignete Zusammenspielmaß.

Die Abhängigkeit des Korrelationskoeffizienten ρ von den Randwahrscheinlichkeiten äußert sich noch darin, daß $\rho = -1$ oder $\rho = +1$ nur gelten kann, wenn die Randwahrscheinlichkeiten für das Merkmal X₁ gleich denen für Merkmal X₂ sind ($p_{11} = p_{12}$ und $p_{21} = p_{22}$). Andernfalls ist der Maximalwert von ρ kleiner als Eins. Nur wenn beide Wahrscheinlichkeiten in einer Diagonalen gleich Null sind ($p_{11} = p_{22} = 0$ oder $p_{12} = p_{21} = 0$), ist $|\rho| = 1$. Im Gegensatz zu Q und Y reicht es nicht, wenn nur eine Zelle die Wahrscheinlichkeit Null hat. ρ ist asymptotisch normalverteilt. Die Varianz kann man schätzen durch

$$\sigma_e^2 = \frac{1}{n} \left[1 - \varrho^2 + \left(\varrho + \frac{1}{2} \cdot \varrho^3 \right) \cdot \frac{(n_{1.} - n_{2.})(n_{.1} - n_{.2})}{\sqrt{n_{1.} n_{2.} n_{.1} n_{.2}}} \right. \\ \left. - \frac{3}{4} \cdot \varrho^2 \left(\frac{(n_{1.} - n_{2.})^2}{n_{1.} n_{2.}} + \frac{(n_{.1} - n_{.2})^2}{n_{.1} n_{.2}} \right) \right].$$

Sind alle Randwahrscheinlichkeiten gleich 0,5, so erhält man

$$\sigma_e^2 = \frac{1 - \varrho^2}{n}.$$

Bei Unabhängigkeit (und beliebigen Randwahrscheinlichkeiten) ist immer

$$\sigma_e^2 = \frac{1}{n}.$$

Beispiel: (Bekanntheitsgrad eines Likörs vor und nach einer Werbekampagne). Mit den Daten aus Tab. 33 erhält man:

$$\varrho = \frac{276 \cdot 512 - 163 \cdot 438}{\sqrt{439 \cdot 950 \cdot 714 \cdot 675}} = \frac{69918}{448327} = 0,156, \\ \sigma_e^2 = \frac{1}{1389} \cdot \left[1 - 0,156^2 \right. \\ \left. + \left[0,156 + \frac{1}{2} \cdot 0,156^3 \right] \cdot \frac{(439 - 950)(714 - 675)}{\sqrt{439 \cdot 950 \cdot 714 \cdot 675}} \right. \\ \left. - \frac{3}{4} \cdot 0,156^2 \cdot \left[\frac{(439 - 950)^2}{439 \cdot 950} + \frac{(714 - 675)^2}{714 \cdot 675} \right] \right] \\ = \frac{1}{1389} \cdot \left[0,9757 + 0,1579 \cdot \frac{-19929}{448327} \right. \\ \left. - 0,01825 \cdot (0,626 + 0,00315) \right] \\ = \frac{0,9572}{1389} = 0,00069.$$

Damit ist $\sigma_e = 0,0263$. Fast den gleichen Wert liefert hier

$$\sqrt{1/n} = \sqrt{1/1389} = 0,0268.$$

Als Grenzen für ein approximatives 95%-Konfidenzintervall ergeben sich

$$0,156 - 1,96 \cdot \sigma_e = 0,156 - 1,96 \cdot 0,0263 = 0,104$$

$$\text{und } 0,156 + 1,96 \cdot \sigma_e = 0,156 + 1,96 \cdot 0,0263 = 0,208.$$

B. Der Pearsonsche Kontingenzkoeffizient

Eine naheliegende Möglichkeit zur Definition von Assoziationsmaßen ist noch nicht angesprochen worden, nämlich die Verwendung der Pearsonschen χ^2 -Teststatistik für den Test auf Unabhängigkeit. Für den Fall der 2×2 -Tafel bringt

dieser Ansatz jedoch nichts Neues, denn es gilt

$$\varrho^2 = \frac{(n_{11} \cdot n_{22} - n_{21} \cdot n_{12})^2}{n_{1.} \cdot n_{2.} \cdot n_{.1} \cdot n_{.2}} = \sum_{i=1}^2 \sum_{j=1}^2 \frac{1}{n} \cdot \frac{(n_{ij} - n_{i.} \cdot n_{.j})^2}{n_{i.} \cdot n_{.j}} = \frac{1}{n} \cdot \chi^2.$$

Da der Maximalwert von $\frac{1}{n} \cdot \chi^2$ mit der Tafelgröße wächst, wird als Assoziationsmaß oft der **Pearson'sche Kontingenzkoeffizient**

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

benutzt. Dieser liefert unabhängig von der Tafelgröße immer Werte zwischen Null und Eins. Bei der 2×2 -Tafel ist

$$C = \sqrt{\frac{\varrho^2}{\varrho^2 + 1}}.$$

Daran sieht man sofort, daß hier C als höchsten Wert $\sqrt{1/(1+1)} = 0,707$ annehmen kann. Es sei schon an dieser Stelle erwähnt, daß sich der Maximalwert von C mit wachsender Tafelgröße der Eins nähert.

Zur Konstruktion von Konfidenzintervallen geben wir noch die **Varianz von C** bei der **Vierfeldertafel** an. Es ist näherungsweise

$$\sigma_C^2 = \frac{C^6}{\varrho^6} \cdot \sigma_\varrho^2.$$

3.2. Assoziationsmaße in allgemeinen 2-dimensionalen Kontingenztafeln

Wie schon bei der Vierfeldertafel erwähnt, ist es naheliegend, die χ^2 -Teststatistik für den Test auf Unabhängigkeit zur Konstruktion von Assoziationsmaßen zu verwenden. Dagegen kann der dort eingeführte Korrelationskoeffizient ϱ nicht in kanonischer Weise auf beliebige zweidimensionale Kontingenztafeln verallgemeinert werden.

Zur Bestimmung eines solchen Korrelationskoeffizienten ist es nämlich notwendig, die beiden Merkmale zu skalieren: Jeder Merkmalsausprägung muß eine reelle Zahl zugeordnet werden. Im Gegensatz zur Vierfeldertafel hängt bei größeren Tafeln der Wert des Korrelationskoeffizienten ϱ von dieser Skalierung ab. Glücklicherweise läßt sich ϱ nicht beliebig verändern, sondern es gibt einen Wert ϱ_{\max} , den man nicht überschreiten kann. Die Skalierung, die dieses Maximum ϱ_{\max} liefert, ist derart, daß die Kontingenztafel möglichst gut das Histogramm einer bivariaten Normalverteilung mit dem Korrelationskoeffizienten ϱ_{\max} widerspiegelt. Näheres zu dieser sogenannten **Fisher-Lancaster-Skalierung** findet man etwa bei Lancaster (1957), Hartung/Elpelt (1984, Kap. V); vgl. auch Kap. V, Abschnitt 7. Der Wert ϱ_{\max} erscheint daher als geeignet, den Zusammenhang zwischen den beiden Merkmalen zu beschreiben. In Kap. V wird die **Fechtersche Skalierung** (marginale Normalisierung) ordinaler Merkmale besprochen.

3.2.1. Assoziationsmaße, die in Beziehung zur χ^2 -Statistik stehen

A. Der Pearsonsche Kontingenzkoeffizient für die $r \times s$ -Tafel

Wegen der aufwendigen Berechnung wird dieses Maß ϱ_{\max} allerdings nur selten benutzt. Statt dessen verwendet man oft den **Pearsonischen Kontingenzkoeffizienten**

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad \text{mit} \quad \chi^2 = \sum_i \sum_j \frac{\left(n_{ij} - \frac{n_{i \cdot} n_{\cdot j}}{n} \right)^2}{\frac{n_{i \cdot} n_{\cdot j}}{n}}.$$

Teilen wir nämlich die Ausprägungen der beiden Merkmale einer bivariaten normalverteilten Zufallsgröße mit dem Korrelationskoeffizienten ϱ in Klassen ein und stellen die beobachteten Klassenhäufigkeiten in einer Kontingenztafel dar, so wird mit abnehmender Klassenbreite und wachsender Beobachtungszahl C gegen $|\varrho|$ streben. C liegt immer zwischen Null und Eins, nimmt als maximalen Wert aber nicht Eins an, sondern ist bei einer $r \times s$ -Tafel höchstens gleich

$$\sqrt{\frac{\min(r, s) - 1}{\min(r, s)}}.$$

Man sieht, daß dieser Ausdruck am kleinsten bei einer 2×2 -Tafel ist; und zwar $\sqrt{0,5} = 0,707$, und mit wachsender Tafelgröße, wie zu erwarten, gegen Eins strebt.

B. Der korrigierte Pearsonsche Kontingenzkoeffizient

Deshalb verwendet man gelegentlich den korrigierten Pearsonischen Kontingenzkoeffizienten

$$C_{\text{corr}} = \sqrt{\frac{\min(r, s)}{\min(r, s) - 1}} \cdot \sqrt{\frac{\chi^2}{\chi^2 + n}},$$

der bei jeder Tafelgröße als Maximum den Wert Eins hat. Es ist natürlich immer $C \leq C_{\text{corr}}$.

C. Das Assoziationsmaß von Tschuprow

Die Beziehung $\varrho^2 = \chi^2/n$, die bei der Vierfeldertafel gültig ist, kann zum Anlaß genommen werden, χ^2/n als Assoziationsmaß für beliebige zweidimensionale Kontingenztafeln zu verwenden. Der Maximalwert von χ^2/n ist für eine $r \times s$ -Tafel gleich $\min(r, s) - 1$, wächst also mit der Tafelgröße ins Unendliche. Deshalb nimmt man noch eine Korrektur vor. **Tschuprow** hat vorgeschlagen,

$$T = \sqrt{\frac{\chi^2}{n\sqrt{(r-1)(s-1)}}}$$

als Assoziationsmaß zu verwenden. Dieses Maß ist zwar immer kleiner oder gleich Eins, nimmt aber den Maximalwert Eins nur bei quadratischen Tafeln ($r = s$) an.

D. Das Assoziationsmaß von Cramér

Das von Cramér eingeführte Zusammenhangsmaß

$$V = \sqrt{\frac{\chi^2}{n(\min(r, s) - 1)}}$$

hat dagegen als Maximum stets den Wert Eins. Es gilt $T = V$ bei quadratischen Tafeln und sonst $T < V$. Zum Pearsonschen Kontingenzkoeffizienten C bestehen die Beziehungen

$$C \begin{cases} > \\ = \\ < \end{cases} V, \quad \text{falls} \quad V \begin{cases} < \\ = \\ > \end{cases} \sqrt{1 - \frac{1}{\min(r, s) - 1}}.$$

Tafeln, die nur zwei Zeilen oder zwei Spalten haben, also mit $\min(r, s) = 2$, liefern demnach für C und V stets Werte mit $C < V$. Ist $\min(r, s) = 3$, so gilt $C < V$ nur noch, wenn $V > \sqrt{0,5} = 0,707$ ist; sonst hat man $V \leq C$. Der Bereich, für den $C < V$ gilt, wird mit wachsender Zeilen- und Spaltenzahl immer kleiner und verschwindet bei sehr großen Tafeln fast vollständig.

E. Schätzung der Varianzen der Assoziationsmaße

Um Konfidenzbereiche angeben zu können, benötigen wir noch Schätzungen für die Varianzen der hier vorgestellten Zusammenhangsmaße. Es ist (falls $\chi^2 \neq 0$)

$$n^2 \cdot \sigma_{\chi^2}^2 = 4 \sum_i \sum_j \frac{n_{ij}^3}{n_i^2 n_j^2} - 3 \sum_i \frac{1}{n_i} \left(\sum_j \frac{n_{ij}^2}{n_i n_j} \right)^2 - 3 \sum_j \frac{1}{n_j} \left(\sum_i \frac{n_{ij}^2}{n_i n_j} \right)^2$$

$$+ 2 \sum_i \sum_j \frac{n_{ij}}{n_i n_j} \left(\sum_k \frac{n_{kj}^2}{n_k n_j} \right) \left(\sum_l \frac{n_{il}^2}{n_i n_l} \right),$$

$$\sigma_C^2 = \frac{n^4}{4 \cdot \chi^2 (n + \chi^2)^3} \cdot n^2 \cdot \sigma_{\chi^2}^2,$$

$$\sigma_{C_{corr}}^2 = \frac{n^4 \cdot \min(r, s)}{(\min(r, s) - 1) \cdot 4 \cdot \chi^2 (n + \chi^2)^3} \cdot n^2 \cdot \sigma_{\chi^2}^2,$$

$$\sigma_T^2 = \frac{1}{4(r-1)^2(s-1)^2 \cdot T^2} \cdot n^2 \cdot \sigma_{\chi^2}^2$$

$$\text{und } \sigma_V^2 = \frac{1}{4(\min(r, s) - 1) \cdot V^2} \cdot n^2 \cdot \sigma_{\chi^2}^2.$$

F. Ein Beispiel

In einem Karosseriebetrieb werden zwei verschiedene Verfahren zur Rostschutzbehandlung von Metallteilen angewandt. Es wird eine Untersuchung durchgeführt, mit der man etwas über den Zusammenhang zwischen der gewählten Rostschutzbehandlung und der Rostanfälligkeit des betreffenden Karosserieteiles erfahren will.

Tab. 38: Vergleich zweier Rostschutzverfahren

		Rostanfälligkeit			Σ
		gering	mittel	stark	
Rostschutzverfahren	I	65	103	106	274
	II	74	85	47	206
	Σ	139	188	153	480

Die Ergebnisse sind in einer Kontingenztafel zusammengestellt, die in **Tab. 38** wiedergegeben ist.

Zur Beschreibung des Zusammenhangs zwischen Rostanfälligkeit und verwendetem Rostschutzverfahren wollen wir die oben eingeführten Zusammenhangsmaße heranziehen. Dazu ist es notwendig, zunächst die χ^2 -Teststatistik für den Test auf Unabhängigkeit zu berechnen:

$$\begin{aligned}
 \chi^2 &= \sum_{i} \sum_{j} \frac{\left(n_{ij} - \frac{n_{i \cdot} \cdot n_{\cdot j}}{n} \right)^2}{\frac{n_{i \cdot} \cdot n_{\cdot j}}{n}} \\
 &= \frac{\left(65 - \frac{139 \cdot 274}{480} \right)^2}{\frac{139 \cdot 274}{480}} + \frac{\left(103 - \frac{188 \cdot 274}{480} \right)^2}{\frac{188 \cdot 274}{480}} + \frac{\left(106 - \frac{153 \cdot 274}{480} \right)^2}{\frac{153 \cdot 274}{480}} \\
 &\quad + \frac{\left(74 - \frac{139 \cdot 206}{480} \right)^2}{\frac{139 \cdot 206}{480}} + \frac{\left(85 - \frac{188 \cdot 206}{480} \right)^2}{\frac{188 \cdot 206}{480}} + \frac{\left(47 - \frac{153 \cdot 206}{480} \right)^2}{\frac{153 \cdot 206}{480}} \\
 &= 2,5937 + 0,1736 + 3,98785 + 3,4499 + 0,23095 + 5,3042 \\
 &= 15,74
 \end{aligned}$$

Damit erhält man

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{15,74}{15,74 + 480}} = 0,178,$$

$$C_{corr} = \sqrt{\frac{\min(r, s)}{\min(r, s) - 1}} \cdot \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{2}{2 - 1}} \cdot 0,178 = 0,252,$$

$$T = \sqrt{\frac{\chi^2}{n \sqrt{(r-1)(s-1)}}} = \sqrt{\frac{15,74}{480 \sqrt{1 \cdot 2}}} = 0,152$$

und $V = \sqrt{\frac{\chi^2}{n(\min(r, s) - 1)}} = \sqrt{\frac{15,74}{480(2 - 1)}} = 0,181.$

Nun sollen noch Konfidenzintervalle angegeben werden. Dazu benötigen wir die Varianzen der Assoziationsmaße. Zur Berechnung von $n^2 \cdot \sigma_{\chi^2}^2$ stellen wir zunächst die einzelnen Terme in tabellarischer Form dar, vgl. Tab. 39 bis Tab. 42.

Tab. 39: $\frac{n_{ij}^3}{n_i \cdot n_j}$ mit Werten aus Tab. 38

i \ j	1	2	3	
1	0,0001893	0,0004118	0,0006777	
2	0,0004942	0,0004095	0,0001045	

Tab. 40: $\frac{n_{ij}^2}{n_i \cdot n_j}$ mit Werten aus Tab. 38

i \ j	1	2	3	Σ
1	0,11093	0,20595	0,26802	0,5849
2	0,19124	0,18656	0,07009	0,4479
Σ	0,3022	0,3925	0,3381	

Tab. 41: $\left(\sum_k \frac{n_{kj}^2}{n_k \cdot n_{.j}} \right) \left(\sum_l \frac{n_{il}^2}{n_l \cdot n_{.l}} \right)$ mit Werten aus Tab. 38

i \ j	1	2	3	
1	0,1767	0,2296	0,1977	
2	0,1353	0,1758	0,1514	

Tab. 42: $\frac{n_{ij}}{n_i \cdot n_j}$ mit Werten aus Tab. 38

i \ j	1	2	3	
1	0,0017067	0,0019995	0,0025285	
2	0,0025843	0,0021948	0,0014912	

Damit berechnet sich

$$\begin{aligned} n^2 \cdot \sigma_{\chi^2}^2 &= 4 \cdot 0,002287 - 3 \cdot (0,0012486 + 0,0009738) \\ &\quad - 3 \cdot (0,0006569 + 0,0008195 + 0,000747165) \\ &\quad + 2 \cdot (0,0003016 + 0,0004591 + 0,00049988 \\ &\quad + 0,0003496 + 0,00038585 + 0,00022577) \\ &= 4 \cdot 0,002287 - 3 \cdot 0,0022224 - 3 \cdot 0,00222355 \\ &\quad + 2 \cdot 0,0022218 \\ &= 0,00025, \end{aligned}$$

$$\begin{aligned} \sigma_C^2 &= \frac{n^4}{4 \cdot \chi^2(n + \chi^2)^3} \cdot n^2 \cdot \sigma_{\chi^2}^2 = \frac{480^4}{4 \cdot 15,74 \cdot (480 + 15,74)^3} \cdot 0,00025 \\ &= 0,00173, \end{aligned}$$

$$\sigma_C = 0,0416,$$

$$\sigma_{C_{corr}}^2 = \frac{\min(r, s)}{\min(r, s) - 1} \cdot \sigma_C^2 = 2 \cdot 0,00173 = 0,00346,$$

$$\sigma_{C_{corr}} = 0,0588,$$

$$\begin{aligned} \sigma_T^2 &= \frac{1}{4(r-1)^2(s-1)^2T^2} \cdot n^2 \cdot \sigma_{\chi^2}^2 = \frac{1}{4 \cdot 1 \cdot 4 \cdot 0,152^2} \cdot 0,00025 \\ &= 0,000676, \end{aligned}$$

$$\sigma_T = 0,026,$$

$$\begin{aligned} \sigma_V^2 &= \frac{1}{4(\min(r, s) - 1)V^2} \cdot n^2 \cdot \sigma_{\chi^2}^2 = \frac{1}{4 \cdot (2-1) \cdot 0,181^2} \cdot 0,00025 \\ &= 0,00191 \end{aligned}$$

und $\sigma_V = 0,0437.$

Hiermit erhält man als **95%-Konfidenzintervall** näherungsweise

$$[0,096; 0,26] \text{ für } C,$$

$$[0,14; 0,37] \text{ für } C_{corr},$$

$$[0,10; 0,20] \text{ für } T$$

und $[0,095; 0,27] \text{ für } V.$

3.2.2. Die λ - und die τ -Maße

Alle bis jetzt für eine beliebige zweidimensionale Kontingenztafel besprochenen Assoziationsmaße sind Funktionen von χ^2 . Wir wissen, daß der Wert Null bei Unabhängigkeit und das Maximum bei totaler Abhängigkeit angenommen wird. Es fehlt jedoch an einer wahrscheinlichkeitstheoretischen Interpretation der Zwischenwerte.

Genau dieser Punkt der **Interpretationsmöglichkeit** sämtlicher Werte ist bei den **λ -Maßen**, die zuerst von Guttman (1941) vorgeschlagen wurden, und den **τ -Maßen**, die Goodman/Kruskal (1954) vorgestellt haben, erfüllt; sie sind sogar aus dieser Interpretationsmöglichkeit heraus begründet. Den Ausgangspunkt solcher Überle-

gungen bildet die folgende Frage. Was nützt uns die Kenntnis des einen Merkmals bei der **Vorhersage** des anderen? Bei Unabhängigkeit sicher gar nichts; Abhängigkeiten erhöhen dagegen, wenn wir sie kennen, die Vorhersagegenauigkeit.

Beschäftigen wir uns einmal mit der folgenden Situation. Für die beiden Merkmale A und B sei die Tafel der theoretischen Wahrscheinlichkeiten bekannt, s. **Tab. 43.**

Tab. 43: $r \times s$ -Tafel der theoretischen Wahrscheinlichkeiten p_{ij}

		$b_1 \dots b_s$	Σ
a_1		$p_{11} \dots p_{1s}$	$p_{1.}$
\vdots		\vdots	\vdots
a_r		$p_{r1} \dots p_{rs}$	$p_{r.}$
Σ		$p_{.1} \dots p_{.s}$	1

Diese Information soll nun dazu benutzt werden, B vorherzusagen und zwar einmal mit und einmal ohne die zusätzliche Kenntnis von A. Bei Unkenntnis von A ist, wenn wir für B die Merkmalsausprägung b_j prognostizieren, die Wahrscheinlichkeit einer richtigen Vorhersage $p_{.j}$ und die einer falschen $1 - p_{.j}$. Die Wahrscheinlichkeit einer richtigen Prognose wird demnach am größten, wenn wir diejenige Merkmalsausprägung b_{j_0} vorhersagen, deren Auftreten am wahrscheinlichsten ist. Es ist also j_0 so zu wählen, daß $p_{.j_0} = \max_j p_{.j}$ gilt. Kennt man A schon, hat man etwa a_i beobachtet, so wird das Auftreten des Merkmals B durch die bedingten Wahrscheinlichkeiten $p_{ij}/p_{i.}$ beschrieben. In dieser Situation prognostiziert man deshalb diejenige Merkmalsausprägung $b_{j(i)}$ für die gilt $p_{ij(i)} = \max_j p_{ij}/p_{i.}$.

A. Die λ -Maße λ_A , λ_B und λ

Das Maß λ_B vergleicht nun die **Wahrscheinlichkeit einer falschen Vorhersage** von B bei **Unkenntnis von A**

$$v_{U(A)} = 1 - \max_j p_{.j}$$

mit der **Wahrscheinlichkeit einer falschen Vorhersage** von B bei **Kenntnis von A**

$$v_{K(A)} = 1 - \sum_i p_{i.} \cdot \max_j p_{ij}/p_{i.} = 1 - \sum_i \max_j p_{ij}.$$

$$\begin{aligned} \text{Es ist } \lambda_B &= \frac{(1 - \max_j p_{.j}) - (1 - \sum_i \max_j p_{ij})}{1 - \max_j p_{.j}} = \frac{v_{U(A)} - v_{K(A)}}{v_{U(A)}} \\ &= 1 - \frac{v_{K(A)}}{v_{U(A)}} = \frac{\sum_i \max_j p_{ij} - \max_j p_{.j}}{1 - \max_j p_{.j}} \end{aligned}$$

beziehungsweise

$$\lambda_B = \frac{\sum_{i,j} \max n_{ij} - \max_{j,i} n_{ij}}{n - \max_{j,i} n_{ij}}.$$

Werden die Rollen von A und B vertauscht, wird also A mit und ohne Hilfe von B prognostiziert, so kommt man zu

$$\lambda_A = \frac{\sum_{j,i} \max n_{ij} - \max_{i,j} n_{ij}}{n - \max_{i,j} n_{ij}}.$$

Ein bezüglich der beiden Merkmale symmetrisches Maß, das immer zwischen λ_A und λ_B liegt, wird definiert durch

$$\lambda = \frac{\frac{1}{2} \left(\sum_i \max_{j,i} n_{ij} + \sum_j \max_{i,j} n_{ij} \right) - \frac{1}{2} \left(\max_{j,i} n_{ij} + \max_{i,j} n_{ij} \right)}{n - \frac{1}{2} \left(\max_{j,i} n_{ij} + \max_{i,j} n_{ij} \right)}$$

Sind die λ -Maße weder Null noch Eins, so können die Varianzen dieser asymptotisch normalverteilten Größen wie folgt geschätzt werden. Für λ_B nimmt man

$$\sigma_{\lambda_B}^2 = \frac{\left(n - \sum_i \max_{j,i} n_{ij} \right) \left(\sum_i \max_{j,i} n_{ij} + \max_{j,i} n_{ij} - 2 \cdot \sum_i^* \max_{j,i} n_{ij} \right)}{\left(n - \max_{j,i} n_{ij} \right)^3};$$

\sum_i^* bezeichnet dabei die Summation über solche Zeilen, in denen $\max_{j,i} n_{ij}$ in der gleichen Spalte liegt wie $\max_{j,i} n_{ij}$. Die Varianz von λ_A kann geschätzt werden durch

$$\sigma_{\lambda_A}^2 = \frac{\left(n - \sum_j \max_{i,j} n_{ij} \right) \left(\sum_j \max_{i,j} n_{ij} + \max_{i,j} n_{ij} - 2 \cdot \sum_j^* \max_{i,j} n_{ij} \right)}{\left(n - \max_{i,j} n_{ij} \right)^3};$$

\sum_j^* bezeichnet die Summation über diejenigen Spalten, in denen $\max_{i,j} n_{ij}$ in der gleichen Zeile wie $\max_{i,j} n_{ij}$ liegt. Auf die Angabe einer Schätzung der Varianz von λ wollen wir hier verzichten. Einerseits ist der Ausdruck nämlich recht kompliziert und andererseits wird das Maß λ sowieso selten benutzt.

Für λ_A und λ_B können mit Hilfe der obigen Varianzschätzungen auch Konfidenzintervalle zum Niveau $1 - \alpha$ angegeben werden, nämlich

$$[\lambda_A - u_{1-\alpha/2} \cdot \sigma_{\lambda_A}; \lambda_A + u_{1-\alpha/2} \cdot \sigma_{\lambda_A}]$$

und $[\lambda_B - u_{1-\alpha/2} \cdot \sigma_{\lambda_B}; \lambda_B + u_{1-\alpha/2} \cdot \sigma_{\lambda_B}]$.

Beispiel: Wir wollen uns hier noch einmal mit dem Rostbeispiel beschäftigen, vgl. Tab. 38. Mit den dortigen Daten ist $\max_j n_{1j} = 106$, $\max_j n_{2j} = 85$ und $\max_j n_{ij} = 188$. Damit ergibt sich

$$\lambda_B = \frac{(106 + 85) - 188}{480 - 188} = \frac{3}{292} = 0,0103.$$

Mit $\max_i n_{i1} = 74$, $\max_i n_{i2} = 103$, $\max_i n_{i3} = 106$ und $\max_i n_{ij} = 274$ erhält man

$$\lambda_A = \frac{(74 + 103 + 106) - 274}{480 - 274} = \frac{9}{206} = 0,0437.$$

Für das symmetrische Maß λ ergibt sich

$$\lambda = \frac{0,5(106 + 85 + 74 + 103 + 106) - 0,5(188 + 274)}{480 - 0,5(188 + 274)} = \frac{6}{249} = 0,0241.$$

Für die Angabe von Konfidenzintervallen benötigen wir noch Schätzungen für die Varianzen. Es ist

$$\sigma_B^2 = \frac{(480 - (106 + 85))(106 + 85 + 188 - 2 \cdot 85)}{(480 - 188)^3} = \frac{289 \cdot 209}{292^3}$$

$$= 0,002426, \text{ d.h.}$$

$$\sigma_B = 0,049$$

$$\text{und } \sigma_A^2 = \frac{(480 - (74 + 103 + 106))(74 + 103 + 106 + 274 - 2 \cdot (103 + 106))}{(480 - 274)^3}$$

$$= \frac{197 \cdot 139}{206^3} = 0,0031325, \text{ also}$$

$$\sigma_A = 0,056.$$

Die Grenzen eines 95%-Konfidenzintervalls ergeben sich für λ_B zu

$$0,0103 - 1,96 \cdot 0,049 = -0,086 \text{ und } 0,0103 + 1,96 \cdot 0,049 = 0,106;$$

für λ_A hat man

$$0,0437 - 1,96 \cdot 0,056 = -0,066 \text{ und } 0,0437 + 1,96 \cdot 0,056 = 0,1535.$$

Wir sehen, daß in beiden Fällen die Intervalle größere negative Abschnitte enthalten, obwohl λ_B und λ_A keine negativen Werte annehmen können. Demnach ist wohl noch keine gute Näherung an die Normalverteilung erfolgt. Die hier angegebenen Konfidenzbereiche können also nur zur groben Orientierung dienen. Es ist aber anzunehmen, daß sie etwas zu groß ausgefallen sind.

Bevor wir uns einem weiteren Assoziationsmaß zuwenden, soll noch angemerkt werden, daß bei einer 2×2 -Tafel, deren Randwahrscheinlichkeiten alle gleich $1/2$ sind, die drei Maße λ_A , λ_B und λ identisch sind mit dem Betrag des Yuleschen Assoziationskoeffizienten Y .

B. Die τ -Maße τ_A , τ_B und τ

Zu den λ -Maßen sind wir gekommen, indem wir uns überlegt haben, wie wir **optimale Vorhersagen** treffen können. Dabei wird allerdings die Verteilung der vorhergesagten Klassen im allgemeinen nicht die Verteilung der dann tatsächlich aufgetretenen Klassen widerspiegeln. Um das zu erreichen, müssen wir bei der Vorhersage des einen Merkmals, etwa des Merkmals B, wie folgt verfahren: Ist A unbekannt, so wird die Merkmalsausprägung b_j mit der Wahrscheinlichkeit $p_{.j}$ prognostiziert. Bei bekanntem A, sei etwa a_i aufgetreten, prognostiziere man b_j mit der Wahrscheinlichkeit p_{ij}/p_i . (Praktisch lässt sich das z. B. mit einer Tabelle von Zufallszahlen oder mit Hilfe eines Zufallszahlengenerators durchführen.) Durch diese Vorgehensweise wird bei unbekanntem A die Verteilung von B und bei bekanntem A die gemeinsame Verteilung von A und B rekonstruiert. Die **Wahrscheinlichkeiten für falsche Vorhersagen** sind gegeben durch

$$w_{U(A)} = 1 - \sum_j p_{.j}^2$$

bei Unkenntnis von A und

$$w_{K(A)} = 1 - \sum_i \sum_j p_{ij}^2 / p_i$$

bei Kenntnis von A.

Der Vergleich dieser beiden Wahrscheinlichkeiten führt zu dem Assoziationsmaß

$$\tau_B = \frac{\sum_i \sum_j p_{ij}^2 / p_i - \sum_j p_{.j}^2}{1 - \sum_j p_{.j}^2} = \frac{w_{U(A)} - w_{K(A)}}{w_{U(A)}} = 1 - \frac{w_{K(A)}}{w_{U(A)}}$$

beziehungsweise

$$\tau_B = \frac{\sum_{i=1}^r \frac{1}{n_{i.}} \sum_{j=1}^s n_{ij}^2 - \frac{1}{n} \sum_{j=1}^s n_{.j}^2}{n - \frac{1}{n} \sum_{j=1}^s n_{.j}^2} = \frac{\sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n}\right)^2}{n_{i.}}}{n - \frac{1}{n} \sum_{j=1}^s n_{.j}^2}.$$

Bei **identischen Spaltensummen**, also $n_{.j} = \frac{1}{s} \cdot n$ für alle j, gilt

$$\tau_B = \frac{1}{(s-1)n} \cdot \chi^2.$$

Vertauscht man die Rollen von A und B, so kommt man zu

$$\tau_A = \frac{\sum_{i=1}^r \frac{1}{n_{i.}} \sum_{j=1}^s n_{ij}^2 - \frac{1}{n} \sum_{i=1}^r n_{i.}^2}{n - \frac{1}{n} \sum_i n_{i.}^2} = \frac{\sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n}\right)^2}{n_{.j}}}{n - \frac{1}{n} \sum_i n_{i.}^2}.$$

Hier hat man bei **identischen Zeilensummen** ($n_{i \cdot} = \frac{1}{r} \cdot n$ für alle i) die Beziehung

$$\tau_A = \frac{1}{(r-1)n} \cdot \chi^2.$$

Ein bezüglich A und B **symmetrisches Maß** ist gegeben durch

$$\tau = \frac{\frac{1}{2} \sum_i \sum_j \frac{\left(n_{ij} - \frac{n_{i \cdot} n_{\cdot j}}{n}\right)^2}{\frac{n_{i \cdot} n_{\cdot j}}{n_{i \cdot} + n_{\cdot j}}}}{n - \frac{1}{2n} \sum_i n_{i \cdot}^2 - \frac{1}{2n} \sum_j n_{\cdot j}^2}.$$

Bei beliebigen **2×2 -Tafeln** sind alle τ -Maße **identisch** und es gilt

$$\tau_A = \tau_B = \tau = \frac{1}{n} \cdot \chi^2 = \varrho^2 = \frac{(n_{11} \cdot n_{22} - n_{21} \cdot n_{12})^2}{n_{1 \cdot} \cdot n_{2 \cdot} \cdot n_{\cdot 1} \cdot n_{\cdot 2}}.$$

Wir haben die τ -Maße durch den Vergleich von Wahrscheinlichkeiten richtiger bzw. falscher Vorhersagen eines Merkmals gewonnen. Ein zweiter Zugang zu den τ -Maßen eröffnet sich, wenn wir eine **Quadratsummenzerlegung** ähnlich wie bei der üblichen Varianzanalyse, vgl. Kap. XI, vornehmen. Dazu muß zunächst ein der Summe der Abweichungsquadrate entsprechender Ausdruck für kategoriale Daten gefunden werden. Die übliche Quadratsumme für quantitative Daten läßt sich wie folgt schreiben

$$\sum_{k=1}^n (x_k - \bar{x})^2 = \frac{1}{2n} \sum_k \sum_l (x_k - x_l)^2.$$

Setzen wir nun die kategorialen Daten anstelle von $x_k - x_l$ gleich Null bei Übereinstimmung der beiden Kategorien und Eins bei verschiedenen Kategorien, so erhalten wir

$$\frac{1}{2n} \cdot \sum_{i \neq j} n_i n_j = \frac{n}{2} - \frac{1}{2n} \sum_i n_i^2.$$

Dabei ist n_i die beobachtete Anzahl der Merkmalsausprägungen in der i -ten Kategorie.

Eine **zweidimensionale Kontingenztafel** können wir in Analogie zur **Einfachklassifikation der Varianzanalyse** folgendermaßen beschreiben: Das Merkmal B mit den Merkmalsausprägungen b_1, \dots, b_s wird in r Gruppen beobachtet, die durch das Merkmal A charakterisiert werden. Insgesamt haben wir $n_{\cdot j}$ -mal b_j beobachtet, so daß nach der obigen Formel die **Gesamtvariation** gegeben ist durch

$$\frac{n}{2} - \frac{1}{2n} \sum_j n_{\cdot j}^2.$$

Die Variation innerhalb der i -ten Gruppe ergibt sich zu

$$\frac{n_{i \cdot}}{2} - \frac{1}{2n_{i \cdot}} \sum_j n_{ij}^2.$$

Insgesamt hat man also als **Variation innerhalb der Gruppen**

$$\sum_i \left(\frac{n_{i.}}{2} - \frac{1}{2n} \sum_j n_{ij}^2 \right) = \frac{n}{2} - \frac{1}{2} \sum_i \frac{1}{n_{i.}} \sum_j n_{ij}^2.$$

Durch den Vergleich dieser beiden Quadratsummen kann geprüft werden, ob ein Gruppeneffekt, d. h. ein Einfluß des Merkmals A auf das Merkmal B vorliegt. Als naheliegende Prüfgröße läßt sich

$$\begin{aligned} & \frac{\left(\frac{n}{2} - \frac{1}{2n} \sum_j n_{ij}^2 \right) - \left(\frac{n}{2} - \frac{1}{2} \sum_i \frac{1}{n_{i.}} \sum_j n_{ij}^2 \right)}{\frac{n}{2} - \frac{1}{2n} \sum_j n_{ij}^2} \\ &= \frac{\sum_i \frac{1}{n_{i.}} \sum_j n_{ij}^2 - \frac{1}{n} \sum_j n_{ij}^2}{n - \frac{1}{n} \sum_j n_{ij}^2} = \tau_B \end{aligned}$$

verwenden.

Vertauscht man die Rollen von A und B, so kommt man zu dem Assoziationsmaß τ_A .

Erwähnenswert ist noch, daß unter der Hypothese der **Unabhängigkeit** der beiden Merkmale $(s-1)(n-1)\tau_B$ und $(n-1)(r-1)\tau_A$ approximativ χ^2 -verteilt sind mit $(r-1)(s-1)$ -Freiheitsgraden. Mit dieser Prüfgröße läßt sich also ein **Test auf Unabhängigkeit** durchführen (vgl. Abschnitt 2).

Wir wollen nun noch die Verteilung der τ -Maße bei **Abhängigkeit** der beiden Merkmale angeben. Es handelt sich hierbei um **asymptotisch normalverteilte** Größen, deren Varianzen geschätzt werden können durch

$$\begin{aligned} \sigma_{\tau_B}^2 &= \frac{1}{n^2} \sum_i \sum_j n_{ij} \left(\frac{\psi_{ij}^B - \bar{\psi}^B}{\delta_B^2} \right)^2, \quad \text{wobei} \\ \psi_{ij}^B &= 2(\delta_B - v_B)n_{.j} + 2\delta_B n \frac{n_{ij}}{n_{i.}} - \delta_B n \sum_l \left(\frac{n_{il}}{n_{i.}} \right)^2, \\ \bar{\psi}^B &= 2(\delta_B - v_B) \frac{1}{n} \sum_j n_{.j}^2 + \delta_B \sum_i \frac{1}{n_{i.}} \sum_j n_{ij}^2 \\ &= 2(\delta_B - v_B)(n - \delta_B) + \delta_B(v_B + n - \delta_B), \\ v_B &= \sum_i \frac{1}{n_{i.}} \sum_j n_{ij}^2 - \frac{1}{n} \sum_j n_{.j}^2, \\ \delta_B &= n - \frac{1}{n} \sum_j n_{.j}^2, \end{aligned}$$

$$\text{und } \sigma_{\tau_A}^2 = \frac{1}{n^2} \sum_i \sum_j n_{ij} \left(\frac{\psi_{ij}^A - \bar{\psi}^A}{\delta_A^2} \right)^2, \quad \text{wobei}$$

$$\psi_{ij}^A = 2(\delta_A - v_A)n_{i.} + 2\delta_A n \frac{n_{ij}}{n_{.j}} - \delta_A n \sum_k \left(\frac{n_{kj}}{n_{.j}} \right)^2,$$

$$\begin{aligned}\bar{\psi}^A &= 2(\delta_A - v_A) \frac{1}{n} \sum_i n_{i.}^2 + \delta_A \sum_j \frac{1}{n_{.j}} \sum_i n_{ij}^2, \\ &= 2(\delta_A - v_A)(n - \delta_A) + \delta_A(v_A + n - \delta_A), \\ v_A &= \sum_j \frac{1}{n_{.j}} \sum_i n_{ij}^2 - \frac{1}{n} \sum_i n_{i.}^2, \\ \delta_A &= n - \frac{1}{n} \sum_i n_{i.}^2.\end{aligned}$$

Mit den hier eingeführten Bezeichnungen lassen sich die τ -Maße in der folgenden einfachen Form schreiben:

$$\tau_B = v_B / \delta_B \quad \text{und} \quad \tau_A = v_A / \delta_A.$$

Beispiel: Wir greifen hier noch einmal auf das „Rostschutzbeispiel“ zurück, vgl. Tab. 38. Zur Bestimmung von τ_B berechnen wir mit den dortigen Daten v_B und δ_B . Es ist

$$v_B = 167,52 - 162,65 = 4,87,$$

$$\delta_B = 480 - 162,65 = 317,35,$$

und damit

$$\tau_B = v_B / \delta_B = 0,015.$$

Um ein Konfidenzintervall für τ_B angeben zu können, muß noch $\sigma_{\tau_B}^2$ bestimmt werden. Wir erhalten mit den Daten aus Tab. 38 für $2(\delta_B - v_B)n_{.j}$ die Werte aus

Tab. 44, für $2\delta_B n \frac{n_{ij}}{n_{i.}}$ die Werte aus **Tab. 45**, und es ist

Tab. 44: $2(\delta_B - v_B)n_{.j}$ mit den Daten aus Tab. 38

j	1	2	3
$2(\delta_B - v_B)n_{.j}$	86868	117491	95617

Tab. 45: $2\delta_B n \frac{n_{ij}}{n_{i.}}$ mit den Daten aus Tab. 38

i \ j	1	2	3
1	72272	114523	117858
2	109438	125706	69508

$$\delta_B n \sum_i \left(\frac{n_{1i}}{n_{1.}} \right)^2 = 52895 \quad \text{und} \quad \delta_B n \sum_i \left(\frac{n_{2i}}{n_{2.}} \right)^2 = 53520.$$

Damit ergeben sich für die ψ_{ij}^B die Werte in **Tab. 46** und wir haben

$$\bar{\psi}^B = 154814.$$

Tab. 46: ψ_{ij}^B mit den Daten aus Tab. 38

i	j	1	2	3
1		106245	179118	160580
2		142786	189677	111605

Man erhält schließlich

$$\begin{aligned}\sigma_{\tau_B}^2 &= \frac{1}{480^2} (65 \cdot 0,2326 + 103 \cdot 0,0582 + 106 \cdot 0,0033 \\ &\quad + 74 \cdot 0,0143 + 85 \cdot 0,1198 + 47 \cdot 0,1841) \\ &= 0,000180\end{aligned}$$

also $\sigma_{\tau_B} = 0,0134$.

Die Grenzen eines 95%-Konfidenzintervalls für τ_B ergeben sich damit approximativ zu

$$0,015 - 1,96 \cdot 0,0134 = -0,011 \quad \text{und} \quad 0,015 + 1,96 \cdot 0,0134 = 0,041.$$

Nun zur Berechnung von τ_A . Mit

$$v_A = 252,53 - 244,82 = 7,71$$

und $\delta_A = 480 - 244,82 = 235,18$

erhält man

$$\tau_A = v_A / \delta_A = 0,033.$$

Zur Angabe eines Konfidenzintervalls für τ_A benötigen wir noch $\sigma_{\tau_A}^2$. Es ist $2(\delta_A - v_A)n_1 = 124654$ und $2(\delta_A - v_A)n_2 = 93718$.

Tab. 47: $2\delta_A n \frac{n_{ij}}{n_{.j}}$ mit den Daten aus Tab. 38

i	j	1	2	3
1		105579	123697	156420
2		120198	102080	69356

Tab. 48: $\delta_A n \sum_k \left(\frac{n_{kj}}{n_{.j}} \right)^2$ mit den Daten aus Tab. 38

j	1	2	3
$\delta_A n \sum_k \left(\frac{n_{kj}}{n_{.j}} \right)^2$	56681	56962	64837

Tab. 49: ψ_{ij}^A mit den Daten aus Tab. 38

		j	1	2	3
i					
1		173552	191389	216235	
2		157235	138836	98235	

Für $2\delta_A n \frac{n_{ij}}{n_{i,j}}$ ergeben sich die Werte aus Tab. 47, und für $\delta_A n \sum_k \left(\frac{n_{kj}}{n_{i,j}} \right)^2$ die Werte aus Tab. 48. Daraus berechnet man die ψ_{ij}^A wie in Tab. 49 angegeben; und es ist $\bar{\psi}^A = 170768$.

Schließlich ergibt sich

$$\begin{aligned}\sigma_{\tau_A}^2 &= \frac{1}{480^2} (65 \cdot 0,0025 + 103 \cdot 0,139 + 106 \cdot 0,6757 \\ &\quad + 74 \cdot 0,05986 + 85 \cdot 0,3333 + 47 \cdot 1,72) \\ &= 0,000867 \\ \text{also } \sigma_{\tau_A} &= 0,0294.\end{aligned}$$

Die Grenzen eines 95%-Konfidenzintervalls für τ_A ergeben sich damit approximativ zu $0,033 - 1,96 \cdot 0,0294 = -0,025$ und $0,033 + 1,96 \cdot 0,0294 = 0,091$.

4. Loglineare Modelle und Tests für mehrdimensionale Kontingenztafeln

In diesem Abschnitt sollen die Methoden, die wir in Abschnitt 2 zur Analyse von zweidimensionalen Kontingenztafeln kennengelernt haben, auf mehrdimensionale Tafeln verallgemeinert werden.

Die Darstellungsweise der erwarteten Häufigkeiten in einem **loglinearen Modell** setzt sich dabei in kanonischer Weise fort. Beim Testen von Hypothesen treten jedoch einige Probleme auf. Zum einen erhöht sich mit der Dimension der Kontingenztafel außerordentlich stark die Anzahl der möglichen Hypothesen, und zum anderen sind die Maximum-Likelihood-Schätzer (ML-Schätzer) der unter der Hypothese erwarteten Häufigkeiten nicht immer in geschlossener Form angebar, sondern können zum Teil nur iterativ bestimmt werden.

Wir werden im folgenden zunächst die Parameter des saturierten loglinearen Modells sowie deren Varianzen schätzen. Damit erhält man schon einen ersten Eindruck davon, welche Effekte hauptsächlich wirken. Man bekommt Anhaltspunkte dafür, welches einfachere loglineare Modell zur Beschreibung der vorliegenden Datensituation geeignet sein könnte. Statistisch gesicherte Aussagen darüber lassen sich jedoch nur durch das Testen geeigneter Hypothesen gewinnen. Dazu stellen wir als erstes ein Iterationsverfahren vor, mit dem wir die ML-Schätzungen der unter der jeweiligen Hypothese erwarteten Häufigkeiten bestimmen können. Diese Werte vergleicht man dann mit den beobachteten Häufigkeiten mittels der χ^2 - oder G^2 -Statistik. Den kritischen Wert erhält man als $(1 - \alpha)$ -Quantil der χ^2 -

Verteilung mit den durch die Hypothese bestimmten Freiheitsgraden. Schließlich zeigen wir noch, wieviel die einzelnen im Modell der Hypothese nicht enthaltenen Effekte zum Wert der Teststatistik beitragen.

Wie schon in den vorangegangenen Abschnitten werden wir uns auch hier auf sogenannte **hierarchische Modelle** beschränken. Hierarchisch bedeutet in diesem Zusammenhang, daß das Auftreten von Wechselwirkungstermen höherer Ordnung das Auftreten aller zugehörigen Terme niedrigerer Ordnung impliziert. So werden z. B. bei dem loglinearen Modell für eine 5-dimensionale Kontingenztafel mit der Aufnahme der 4-er Wechselwirkungsterme $u_{1345(iklm)}$ auch alle 3-er und 2-er Wechselwirkungen zwischen den Merkmalen 1, 3, 4 und 5 sowie die zugehörigen einfachen Effekte mit in das Modell aufgenommen. Deshalb reicht es zur Festlegung eines solchen Modells aus, die höchsten Wechselwirkungsterme anzugeben, die noch berücksichtigt werden sollen. So hat man bei einer 4-dimensionalen Kontingenztafel durch die Angabe von u_{123} und u_{14} als höchste Wechselwirkungen das hierarchische, loglineare Modell spezifiziert, in dem für die erwarteten Häufigkeiten m_{ijkl} gilt

$$\ln m_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{12(ij)} + u_{13(ik)} \\ + u_{23(jk)} + u_{14(il)} + u_{123(ijk)},$$

wobei die üblichen Nebenbedingungen für die Parameter erfüllt sein sollen, vgl. Abschnitt 2. Durch die Forderung der Hierarchie müssen mit u_{123} auch u , u_1 , u_2 , u_3 , u_{12} , u_{13} und u_{23} mit in das Modell einbezogen werden. Aus dem gleichen Grund kann die Berücksichtigung der 2-er Wechselwirkung u_{14} nur erfolgen, wenn auch u , u_1 und u_4 im Modell vorhanden sind.

4.1. Die Parameter des satuierten Modells

Wie lassen sich nun die Parameter des loglinearen Modells aus den erwarteten Häufigkeiten berechnen? Wie erhält man z. B. in dem obigen Modell $u_{123(ijk)}$? Dazu mitteln wir zunächst die Logarithmen der erwarteten Häufigkeiten über alle Indizes, von denen der zu berechnende Parameter nicht abhängt. Bei $u_{123(ijk)}$ ist das nur der Index l, so daß man erhält

$$\frac{1}{v} \cdot \sum_{l=1}^v \ln m_{ijkl}.$$

Von diesem Ausdruck müssen nun alle zu u_{123} gehörigen Terme niedrigerer Ordnung subtrahiert werden. So ergibt sich

$$u_{123(ijk)} = \frac{1}{v} \cdot \sum_{l=1}^v \ln m_{ijkl} - u_{12(ij)} - u_{13(ik)} - u_{23(jk)} \\ - u_{1(i)} - u_{2(j)} - u_{3(k)} - u.$$

Entsprechend erhält man z. B. ($i = 1, \dots, r; j = 1, \dots, s; k = 1, \dots, t; l = 1, \dots, v$)

$$u = \frac{1}{r \cdot s \cdot t \cdot v} \cdot \sum_{i,j,k,l} \ln m_{ijkl}, \\ u_{2(j)} = \frac{1}{r \cdot t \cdot v} \cdot \sum_{i,k,l} \ln m_{ijkl} - u \\ = \frac{1}{r \cdot t \cdot v} \cdot \sum_{j,k,l} \ln m_{ijkl} - \frac{1}{r \cdot s \cdot t \cdot v} \cdot \sum_{i,j,k,l} \ln m_{ijkl},$$

$$\begin{aligned}
 u_{12(ij)} &= \frac{1}{t \cdot v} \cdot \sum_{k,l} \ln m_{ijkl} - u_{1(i)} - u_{2(j)} - u \\
 &= \frac{1}{t \cdot v} \cdot \sum_{k,l} \ln m_{ijkl} - \frac{1}{s \cdot t \cdot v} \cdot \sum_{j,k,l} \ln m_{ijkl} \\
 &\quad - \frac{1}{r \cdot t \cdot v} \cdot \sum_{i,k,l} \ln m_{ijkl} + \frac{1}{r \cdot s \cdot t \cdot v} \cdot \sum_{i,j,k,l} \ln m_{ijkl} \text{ „usw.“}
 \end{aligned}$$

A. Schätzen der Parameter des saturierten Modells

Da die in diesen Formeln enthaltenen erwarteten Häufigkeiten m_{ijkl} i. a. unbekannt sind, werden sie durch Schätzwerte ersetzt. Bei einem saturierten Modell, d. h. einem Modell, das alle möglichen Parameter enthält, sind die ML-Schätzungen der erwarteten Häufigkeiten gerade gleich den beobachteten Häufigkeiten n_{ijkl} . Im folgenden Beispiel sollen nun auf diese Weise die einzelnen Parameter des loglinearen Modells einer $r \times s \times t = 3 \times 3 \times 2$ -Tafel geschätzt werden.

Beispiel: Bei einer Untersuchung dreier verschiedener Pkw-Typen von den Herstellern A, B und C wurde das Fahrverhalten mit gutmütig, beherrschbar oder kritisch charakterisiert und der Fahrkomfort mit schlecht oder gut beurteilt. Die Häufigkeiten der einzelnen Beurteilungen lassen sich übersichtlich in einer dreidimensionalen Kontingenztafel darstellen, wie in Tab. 50 geschehen, vgl. auch Abb. 1.

Tab. 50: Fahrverhalten und Komfort von drei Pkw-Typen

		schlechter Komfort ($k = 1$)			
n_{ij1}	Hersteller (j)	A	B	C	$\sum_j n_{ij1}$
Fahr-verhalten (i)					
gutmütig	258	127	376	761	
beherrschbar	207	198	201	606	
kritisch	77	58	26	161	
$\sum_i n_{ij1}$	542	383	603	1528	
		guter Komfort ($k = 2$)			
n_{ij2}	Hersteller (j)	A	B	C	$\sum_j n_{ij2}$
Fahr-verhalten (i)					
gutmütig	507	142	873	1522	
beherrschbar	120	184	172	476	
kritisch	56	35	17	108	
$\sum_i n_{ij2}$	683	361	1062	2106	

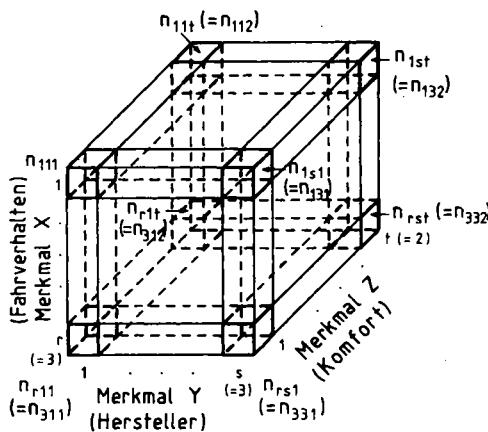


Abb. 1: Veranschaulichung einer dreidimensionalen Kontingenztafel

Die Randhäufigkeiten, die durch Summation über den Index k entstehen, sind in Tab. 51 aufgeführt.

Tab. 51: Randhäufigkeiten zu der Kontingenztafel aus Tab. 50

$n_{ij.}$	Hersteller	A	B	C	$\sum_j n_{ij.}$
Fahrverhalten					
gutmütig		765	269	1249	2283
beherrschbar		327	382	373	1082
kritisch		133	93	43	269
$\sum_i n_{ij.}$		1225	744	1665	3634

Wir wollen nun für das loglineare Modell mit

$$\ln m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)},$$

wobei die üblichen Nebenbedingungen ($\sum u_{1(i)} = 0$, $\sum u_{13(ik)} = 0$ usw.) erfüllt sein sollen, die einzelnen Parameter schätzen. Dazu werden die Logarithmen der beobachteten Häufigkeiten n_{ijk} benötigt, die in Tab. 52 wiedergegeben sind. Tab. 53 enthält noch die Randsummen über den Index k.

Wir erhalten nun

$$u = \frac{1}{r \cdot s \cdot t} \cdot \sum_{i,j,k} \ln n_{ijk} = \frac{1}{3 \cdot 3 \cdot 2} \cdot 87,4335 = 4,8574.$$

Die einfachen Effekte ergeben sich zu

$$u_{1(1)} = \frac{1}{s \cdot t} \cdot \sum_{j,k} \ln n_{1jk} - u = \frac{1}{3 \cdot 2} \cdot 34,283 - 4,8574 = 0,8564,$$

Tab. 52: $\ln n_{ijk}$ mit den Daten aus Tab. 50

$\ln n_{ij1}$	j	1	2	3	$\sum_{j=1}^3 \ln n_{ij1}$
i					
1		5,5530	4,8442	5,9296	16,3268
2		5,3327	5,2883	5,3033	15,9243
3		4,3438	4,0604	3,2581	11,6623
$\sum_{i=1}^3 \ln n_{ij1}$		15,2295	14,1929	14,4910	43,9134

$\ln n_{ij2}$	j	1	2	3	$\sum_{j=1}^3 \ln n_{ij2}$
i					
1		6,2285	4,9558	6,7719	17,9562
2		4,7875	5,2149	5,1475	15,1499
3		4,0254	3,5554	2,8332	10,4140
$\sum_{i=1}^3 \ln n_{ij2}$		15,0414	13,7261	14,7526	43,5201

Tab. 53: Randsummen zu Tab. 52

$\ln n_{ij1} + \ln n_{ij2}$	j	1	2	3	$\sum_{j,k} \ln n_{ijk}$
i					
1		11,7815	9,8000	12,7015	34,2830
2		10,1202	10,5032	10,4508	31,0742
3		8,3692	7,6158	6,0913	22,0763
$\sum_{i,k} \ln n_{ijk}$		30,2709	27,9190	29,2436	87,4335

$$u_{1(2)} = \frac{1}{s \cdot t} \cdot \sum_{j,k} \ln n_{2jk} - u = \frac{1}{3 \cdot 2} \cdot 31,0742 - 4,8574 = 0,3216,$$

$$u_{1(3)} = \frac{1}{s \cdot t} \cdot \sum_{j,k} \ln n_{3jk} - u = \frac{1}{3 \cdot 2} \cdot 22,0763 - 4,8574 = -1,178,$$

$$u_{2(1)} = \frac{1}{r \cdot t} \cdot \sum_{i,k} \ln n_{i1k} - u = \frac{1}{3 \cdot 2} \cdot 30,2709 - 4,8574 = 0,1878,$$

$$u_{2(2)} = \frac{1}{r \cdot t} \cdot \sum_{i,k} \ln n_{i2k} - u = \frac{1}{3 \cdot 2} \cdot 27,919 - 4,8574 = -0,2042,$$

$$u_{2(3)} = \frac{1}{r \cdot t} \cdot \sum_{i,k} \ln n_{i3k} - u = \frac{1}{3 \cdot 2} \cdot 29,2436 - 4,8574 = 0,0165 \quad \text{und}$$

$$u_{3(1)} = \frac{1}{r \cdot s} \cdot \sum_{i,j} \ln n_{ij1} - u = \frac{1}{3 \cdot 3} \cdot 43,9134 - 4,8574 = 0,0219 \\ = -u_{3(2)}$$

Die Wechselwirkungsterme $u_{12(ij)}$ lassen sich mit Hilfe des folgenden Ausdruckes berechnen:

$$u_{12(ij)} = \frac{1}{t} \cdot \sum_k \ln n_{ijk} - u_{1(i)} - u_{2(j)} - u \\ = \frac{1}{t} \cdot \sum_k \ln n_{ijk} - \frac{1}{s \cdot t} \cdot \sum_{j,k} \ln n_{ijk} - \frac{1}{r \cdot t} \cdot \sum_{i,k} \ln n_{ijk} + \frac{1}{r \cdot s \cdot t} \cdot \sum_{i,j,k} \ln n_{ijk}$$

So ergibt sich z.B.

$$u_{12(13)} = \frac{1}{2} \cdot 12,7015 - \frac{1}{3 \cdot 2} \cdot 34,283 - \frac{1}{3 \cdot 2} \cdot 29,2436 + \frac{1}{3 \cdot 3 \cdot 2} \cdot 87,4335 \\ = 0,6204, \\ u_{12(32)} = \frac{1}{2} \cdot 7,6158 - \frac{1}{3 \cdot 2} \cdot 22,0763 - \frac{1}{3 \cdot 2} \cdot 27,919 + \frac{1}{3 \cdot 3 \cdot 2} \cdot 87,4335 \\ = 0,3328;$$

diese und alle übrigen Werte sind in **Tab. 54** zusammengestellt.

Tab. 54: $u_{12(ij)}$ mit den Daten aus Tab. 50

$u_{12(ij)}$	j	1	2	3
i				
1		-0,0108	-0,6096	0,6204
2		-0,3067	0,2768	0,0299
3		0,3175	0,3328	-0,6503

Mit den Formeln

$$u_{13(ik)} = \frac{1}{s} \cdot \sum_j \ln n_{ijk} - u_{1(i)} - u_{3(k)} - u \\ = \frac{1}{s} \cdot \sum_j \ln n_{ijk} - \frac{1}{s \cdot t} \cdot \sum_{j,k} \ln n_{ijk} - \frac{1}{r \cdot s} \cdot \sum_{i,j} \ln n_{ijk} + \frac{1}{r \cdot s \cdot t} \cdot \sum_{i,j,k} \ln n_{ijk}$$

und

$$u_{23(jk)} = \frac{1}{r} \cdot \sum_i \ln n_{ijk} - u_{2(j)} - u_{3(k)} - u \\ = \frac{1}{r} \cdot \sum_i \ln n_{ijk} - \frac{1}{r \cdot t} \cdot \sum_{i,k} \ln n_{ijk} - \frac{1}{r \cdot s} \cdot \sum_{i,j} \ln n_{ijk} + \frac{1}{r \cdot s \cdot t} \cdot \sum_{i,j,k} \ln n_{ijk}$$

ergeben sich die in **Tab. 55** und **Tab. 56** angegebenen Werte für die beiden anderen 2-er Wechselwirkungen.

Tab. 55: $u_{13(ijk)}$ mit den Daten aus Tab. 50

$u_{13(ijk)}$	k	1	2
i			
1		-0,2934	0,2934
2		0,1072	-0,1072
3		0,1862	-0,1862

Tab. 56: $u_{23(jik)}$ mit den Daten aus Tab. 50

$u_{23(jik)}$	k	1	2
j			
1		0,00950	-0,00950
2		0,05595	-0,05595
3		-0,06545	0,06545

Die 3-fachen Wechselwirkungen erhält man wie folgt:

$$\begin{aligned}
 u_{123(ijk)} &= \ln n_{ijk} - u_{12(ij)} - u_{13(ijk)} - u_{23(jik)} - u_{1(i)} - u_{2(j)} - u_{3(k)} - u \\
 &= \ln n_{ijk} - \frac{1}{t} \cdot \sum_k \ln n_{ijk} - \frac{1}{s} \cdot \sum_j \ln n_{ijk} - \frac{1}{r} \cdot \sum_i \ln n_{ijk} \\
 &\quad + \frac{1}{s \cdot t} \cdot \sum_{j,k} \ln n_{ijk} + \frac{1}{r \cdot t} \cdot \sum_{i,k} \ln n_{ijk} + \frac{1}{r \cdot s} \cdot \sum_{i,j} \ln n_{ijk} \\
 &\quad - \frac{1}{r \cdot s \cdot t} \cdot \sum_{i,j,k} \ln n_{ijk}.
 \end{aligned}$$

So ist z. B.

$$\begin{aligned}
 u_{123(121)} &= 4,8442 - \frac{1}{2} \cdot 9,8 - \frac{1}{3} \cdot 16,3268 - \frac{1}{3} \cdot 14,1929 + \frac{1}{3 \cdot 2} \cdot 34,283 \\
 &\quad + \frac{1}{3 \cdot 2} \cdot 27,919 + \frac{1}{3 \cdot 3} \cdot 43,9134 - \frac{1}{3 \cdot 3 \cdot 2} \cdot 87,4335 \\
 &= 0,1598 = -u_{123(122)};
 \end{aligned}$$

alle Ergebnisse sind in **Tab. 57** zusammengestellt.

Tab. 57: $u_{123(ijk)}$ mit Daten aus Tab. 50

$u_{123(ijk)}$ $= -u_{123(ij2)}$	j	1	2	3
i				
1		-0,0758	0,1598	-0,0842
2		0,1340	-0,1483	0,0143
3		-0,0582	-0,0115	0,0699

B. Varianz- und Intervallschätzungen für die Parameter des saturierten Modells

In obigem Beispiel dieses Abschnitts 4.1 haben wir zur Beschreibung der Kontingenztafel ein loglineares Modell zugrundegelegt, in dem alle möglichen Parameter vorhanden sind. Dabei stellt sich sogleich die Frage, ob man nicht mit weniger Parametern auskommen kann. Ist z. B. hier die Berücksichtigung der 3-fachen Wechselwirkung überhaupt notwendig? Anhaltspunkte zur Beantwortung dieser Frage können uns die Schätzungen der Parameter des saturierten Modells geben. Dabei müssen allerdings die Varianzen dieser Schätzwerte ebenfalls bekannt sein bzw. geschätzt werden können. Dann lassen sich nämlich approximative Konfidenzintervalle der asymptotisch normalverteilten Parameterschätzungen angeben. An Hand dieser Konfidenzintervalle ist eine erste Beurteilung des Einflusses der einzelnen Faktoren und Wechselwirkungen mit einer gewissen Grundlage ausgestattet. Wir wollen uns deshalb im folgenden Schätzungen für die Varianzen der Parameter des loglinearen Modells verschaffen.

Eine Schätzung für die Varianz von u ist gegeben durch den Wert des Ausdrückes

$$\left(\frac{\text{reziproke Anzahl der Zellen}}{\text{in der Kontingenztafel}} \right) \cdot \left(\frac{\text{Summe der Kehrwerte der beobachteten Häufigkeiten}}{n} \right) - \frac{1}{n};$$

bei einer 3-dimensionalen Kontingenztafel ist das gerade

$$\text{Var } u = \frac{1}{r \cdot s \cdot t} \cdot \sum_{i,j,k} \frac{1}{n_{ijk}} - \frac{1}{n}.$$

Varianzschätzungen für alle anderen Parameter erhält man durch die folgende Vorgehensweise.

Wir wissen, daß sich die Schätzung jeder dieser Parameter als Linearkombination der Logarithmen der Zellhäufigkeiten darstellen läßt. Dabei sollen nun jeweils alle Summanden, die von der Häufigkeit einer Zelle abhängen, zu einem Term zusammengefaßt sein. Für die 2-fache Wechselwirkung u_{23} der 4-dimensionalen Kontingenztafel sieht das dann z. B. so aus:

$$\begin{aligned} u_{23(31)} &= \frac{1}{r \cdot v} \cdot \sum_{i=1}^r \sum_{l=1}^v \ln n_{i311} - \frac{1}{r \cdot t \cdot v} \cdot \sum_{i=1}^r \sum_{k=1}^t \sum_{l=1}^v \ln n_{i3kl} \\ &\quad - \frac{1}{r \cdot s \cdot v} \cdot \sum_{i=1}^r \sum_{j=1}^s \sum_{l=1}^v \ln n_{ij11} \\ &\quad + \frac{1}{r \cdot s \cdot t \cdot v} \cdot \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t \sum_{l=1}^v \ln n_{ijkl} \\ &= \frac{1}{r \cdot v} \cdot \sum_{i,l} \ln n_{i311} - \frac{1}{r \cdot t \cdot v} \cdot \sum_{i,l} \left(\ln n_{i311} + \sum_{k \neq 1} \ln n_{i3kl} \right) \\ &\quad - \frac{1}{r \cdot s \cdot v} \cdot \sum_{i,l} \left(\ln n_{i311} + \sum_{j \neq 3} \ln n_{ij11} \right) \\ &\quad + \frac{1}{r \cdot s \cdot t \cdot v} \cdot \sum_{i,l} \left(\ln n_{i311} + \sum_{k \neq 1} \ln n_{i3kl} + \sum_{j \neq 3} \ln n_{ij11} + \sum_{j \neq 3} \sum_{k \neq 1} \ln n_{ijkl} \right) \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{1}{r \cdot v} - \frac{1}{r \cdot t \cdot v} - \frac{1}{r \cdot s \cdot v} + \frac{1}{r \cdot s \cdot t \cdot v} \right) \sum_{i,l} \ln n_{i311} \\
&\quad - \left(\frac{1}{r \cdot t \cdot v} - \frac{1}{r \cdot s \cdot t \cdot v} \right) \sum_{i,l} \sum_{k \neq 1} \ln n_{i3kl} \\
&\quad - \left(\frac{1}{r \cdot s \cdot v} - \frac{1}{r \cdot s \cdot t \cdot v} \right) \sum_{i,l} \sum_{j \neq 3} \ln n_{ij11} \\
&\quad + \frac{1}{r \cdot s \cdot t \cdot v} \cdot \sum_{i,l} \sum_{j \neq 3} \sum_{k \neq 1} \ln n_{ijkl} \\
&= \frac{1}{r \cdot v} \left(1 - \frac{1}{s} \right) \left(1 - \frac{1}{t} \right) \sum_{i,l} \ln n_{i311} \\
&\quad - \frac{1}{r \cdot t \cdot v} \left(1 - \frac{1}{s} \right) \sum_{i,l} \sum_{k \neq 1} \ln n_{i3kl} \\
&\quad - \frac{1}{r \cdot s \cdot v} \left(1 - \frac{1}{t} \right) \sum_{i,l} \sum_{j \neq 3} \ln n_{ij11} \\
&\quad + \frac{1}{r \cdot s \cdot t \cdot v} \sum_{i,l} \sum_{j \neq 3} \sum_{k \neq 1} \ln n_{ijkl}.
\end{aligned}$$

Ersetzen wir nun in jedem Term den Logarithmus der Häufigkeiten durch den Kehrwert der Häufigkeiten und quadrieren die Koeffizienten, so ist die Summe all dieser Ausdrücke eine Schätzung der Varianz für den Schätzwert des betreffenden Parameters. Wir erhalten somit für den oben angegebenen Wechselwirkungsterm die Varianzschätzung

$$\begin{aligned}
\text{Var } u_{23(31)} &= \left(\frac{1}{r \cdot v} \right)^2 \left(1 - \frac{1}{s} \right)^2 \left(1 - \frac{1}{t} \right)^2 \sum_{i,l} \frac{1}{n_{i311}} \\
&\quad + \left(\frac{1}{r \cdot t \cdot v} \right)^2 \left(1 - \frac{1}{s} \right)^2 \sum_{i,l} \sum_{k \neq 1} \frac{1}{n_{i3kl}} \\
&\quad + \left(\frac{1}{r \cdot s \cdot v} \right)^2 \left(1 - \frac{1}{t} \right)^2 \sum_{i,l} \sum_{j \neq 3} \frac{1}{n_{ij11}} \\
&\quad + \left(\frac{1}{r \cdot s \cdot t \cdot v} \right)^2 \sum_{i,l} \sum_{j \neq 3} \sum_{k \neq 1} \frac{1}{n_{ijkl}} \\
&= \left(\frac{1}{r \cdot s \cdot t \cdot v} \right)^2 \sum_{i,j,k,l} \frac{1}{n_{ijkl}} \\
&\quad + \left(\left(\frac{1}{r \cdot s \cdot v} \right)^2 \left(1 - \frac{1}{t} \right)^2 - \left(\frac{1}{r \cdot s \cdot t \cdot v} \right)^2 \right) \sum_{i,l} \sum_{j \neq 3} \frac{1}{n_{ij11}}
\end{aligned}$$

$$\begin{aligned}
& + \left(\left(\frac{1}{r \cdot t \cdot v} \right)^2 \left(1 - \frac{1}{s} \right)^2 - \left(\frac{1}{r \cdot s \cdot t \cdot v} \right)^2 \right) \sum_{i,l} \sum_{k \neq 1} \frac{1}{n_{ijk1}} \\
& + \left(\left(\frac{1}{r \cdot v} \right)^2 \left(1 - \frac{1}{s} \right)^2 \left(1 - \frac{1}{t} \right)^2 - \left(\frac{1}{r \cdot s \cdot t \cdot v} \right)^2 \right) \sum_{i,l} \frac{1}{n_{i311}} \\
& = \left(\frac{1}{r \cdot s \cdot t \cdot v} \right)^2 \sum_{i,j,k,l} \frac{1}{n_{ijkl}} \\
& + \left(\frac{1}{r \cdot v} \right)^2 \cdot \frac{1}{s^2} \left(1 - \frac{2}{t} \right) \sum_{i,j,l} \frac{1}{n_{ij1l}} \\
& + \left(\frac{1}{r \cdot v} \right)^2 \cdot \frac{1}{t^2} \left(1 - \frac{2}{s} \right) \sum_{i,k,l} \frac{1}{n_{i3kl}} \\
& + \left(\left(\frac{1}{r \cdot v} \right)^2 \left(1 - \frac{2}{s} + \frac{1}{s^2} \right) \left(1 - \frac{2}{t} + \frac{1}{t^2} \right) - \left(\frac{1}{r \cdot s \cdot t \cdot v} \right)^2 \right. \\
& \quad \left. - \left(\frac{1}{r \cdot v} \right)^2 \cdot \frac{1}{s^2} \left(1 - \frac{2}{t} \right) - \left(\frac{1}{r \cdot v} \right)^2 \cdot \frac{1}{t^2} \left(1 - \frac{2}{s} \right) \right) \sum_{i,l} \frac{1}{n_{i311}} \\
& = \left(\frac{1}{r \cdot v} \right)^2 \left(\frac{1}{s \cdot t} \right)^2 \sum_{i,j,k,l} \frac{1}{n_{ijkl}} \\
& + \left(\frac{1}{r \cdot v} \right)^2 \cdot \frac{1}{s^2} \left(1 - \frac{2}{t} \right) \sum_{i,j,l} \frac{1}{n_{ij1l}} \\
& + \left(\frac{1}{r \cdot v} \right)^2 \cdot \frac{1}{t^2} \left(1 - \frac{2}{s} \right) \sum_{i,k,l} \frac{1}{n_{i3kl}} \\
& + \left(\frac{1}{r \cdot v} \right)^2 \left(1 - \frac{2}{s} \right) \left(1 - \frac{2}{t} \right) \sum_{i,l} \frac{1}{n_{i311}}.
\end{aligned}$$

Dazu nun ein

Beispiel: Wir wollen uns noch einmal mit der Untersuchung über das Fahrverhalten und den Komfort verschiedener Pkw-Typen befassen. Mit den Ergebnissen, die in Tab. 50 wiedergegeben sind, haben wir schon die Parameter des loglinearen Modells geschätzt. Für zwei Wechselwirkungsterme sollen nun auch die Varianzen geschätzt werden. Man kann schreiben

$$\begin{aligned}
u_{12(ij)} &= \frac{1}{t} \left(1 - \frac{1}{r} \right) \left(1 - \frac{1}{s} \right) \sum_{k'} \ln n_{ijk'} - \frac{1}{s \cdot t} \left(1 - \frac{1}{r} \right) \sum_{k'} \sum_{j' \neq j} \ln n_{ij'k'} \\
&\quad - \frac{1}{r \cdot t} \left(1 - \frac{1}{s} \right) \sum_{k'} \sum_{i' \neq i} \ln n_{i'jk'} + \frac{1}{r \cdot s \cdot t} \sum_{k'} \sum_{i' \neq i} \sum_{j' \neq j} \ln n_{i'j'k'}.
\end{aligned}$$

Eine Varianzschätzung für z.B. $u_{12(12)}$ ist gegeben durch

$$\begin{aligned}
\text{Var } u_{12(12)} &= \frac{1}{t^2} \left(1 - \frac{1}{r}\right)^2 \left(1 - \frac{1}{s}\right)^2 \sum_{k'} \frac{1}{n_{12k'}} \\
&\quad + \left(\frac{1}{s \cdot t}\right)^2 \left(1 - \frac{1}{r}\right)^2 \sum_{k'} \sum_{j' \neq 2} \frac{1}{n_{1j'k'}} \\
&\quad + \left(\frac{1}{r \cdot t}\right)^2 \left(1 - \frac{1}{s}\right)^2 \sum_{k'} \sum_{i' \neq 1} \frac{1}{n_{i'2k'}} \\
&\quad + \left(\frac{1}{r \cdot s \cdot t}\right)^2 \sum_{k'} \sum_{i' \neq 1} \sum_{j' \neq 2} \frac{1}{n_{i'j'k'}} \\
&= \frac{1}{2^2} \left(1 - \frac{1}{3}\right)^2 \left(1 - \frac{1}{3}\right)^2 \left(\frac{1}{127} + \frac{1}{142}\right) \\
&\quad + \left(\frac{1}{3 \cdot 2}\right)^2 \left(1 - \frac{1}{3}\right)^2 \left(\frac{1}{258} + \frac{1}{376} + \frac{1}{507} + \frac{1}{873}\right) \\
&\quad + \left(\frac{1}{3 \cdot 2}\right)^2 \left(1 - \frac{1}{3}\right)^2 \left(\frac{1}{198} + \frac{1}{58} + \frac{1}{184} + \frac{1}{35}\right) \\
&\quad + \left(\frac{1}{3 \cdot 3 \cdot 2}\right)^2 \left(\frac{1}{207} + \frac{1}{201} + \frac{1}{77} + \frac{1}{26} + \frac{1}{120} + \frac{1}{172} + \frac{1}{56} + \frac{1}{17}\right) \\
&= \frac{4}{81} \cdot 0,01492 + \frac{1}{81} \cdot 0,00965 + \frac{1}{81} \cdot 0,05630 + \frac{1}{324} \cdot 0,15208 \\
&= 0,00202.
\end{aligned}$$

Damit ergibt sich eine Standardabweichung von 0,0449. Wir erhalten somit als Schranken für ein 95%-Konfidenzintervall von $u_{12(12)}$

$$-0,6096 - 1,96 \cdot 0,0449 = -0,6976 \quad \text{und} \quad -0,6096 + 1,96 \cdot 0,0449 = -0,5216.$$

Es scheint also eine starke Wechselwirkung zwischen den beiden Merkmalen Fahrverhalten und Hersteller vorzuliegen. Zum Vergleich berechnen wir noch

$$\begin{aligned}
\text{Var } u_{23(21)} &= \frac{1}{r^2} \left(1 - \frac{1}{s}\right)^2 \left(1 - \frac{1}{t}\right)^2 \sum_{i'} \frac{1}{n_{i'21}} \\
&\quad + \left(\frac{1}{r \cdot t}\right)^2 \left(1 - \frac{1}{s}\right)^2 \sum_{i'} \sum_{k' \neq 1} \frac{1}{n_{i'2k'}} \\
&= \left(\frac{1}{r \cdot s}\right)^2 \left(1 - \frac{1}{t}\right)^2 \sum_{i'} \sum_{j' \neq 2} \frac{1}{n_{i'j'1}} \\
&\quad + \left(\frac{1}{r \cdot s \cdot t}\right)^2 \sum_{i'} \sum_{j' \neq 2} \sum_{k' \neq 1} \frac{1}{n_{i'j'k'}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{3^2} \left(1 - \frac{1}{3}\right)^2 \left(1 - \frac{1}{2}\right)^2 \left(\frac{1}{127} + \frac{1}{198} + \frac{1}{58}\right) \\
&\quad + \left(\frac{1}{3 \cdot 2}\right)^2 \left(1 - \frac{1}{3}\right)^2 \left(\frac{1}{142} + \frac{1}{184} + \frac{1}{35}\right) \\
&\quad + \left(\frac{1}{3 \cdot 3}\right)^2 \left(1 - \frac{1}{2}\right)^2 \left(\frac{1}{258} + \frac{1}{376} + \frac{1}{207} + \frac{1}{201} + \frac{1}{77} + \frac{1}{26}\right) \\
&\quad + \left(\frac{1}{3 \cdot 3 \cdot 2}\right)^2 \left(\frac{1}{507} + \frac{1}{873} + \frac{1}{120} + \frac{1}{172} + \frac{1}{56} + \frac{1}{17}\right) \\
&= \frac{1}{81} \cdot 0,030166 + \frac{1}{81} \cdot 0,04105 \\
&\quad + \frac{1}{324} \cdot 0,06779 + \frac{1}{324} \cdot 0,093946 \\
&= 0,001378 .
\end{aligned}$$

Wir erhalten hier also die Standardabweichung 0,037, so daß sich für die Schranken eines 95%-Konfidenzintervalls von $u_{23(21)}$

$$0,05595 - 1,96 \cdot 0,037 = -0,0166 \quad \text{und} \quad 0,05595 + 1,96 \cdot 0,037 = 0,1285$$

ergibt. Diese Wechselwirkung scheint demnach nur einen geringen oder sogar keinen Einfluß zu haben. Zum Schluß wollen wir noch die Varianz einer 3-fachen Wechselwirkung berechnen. Es gilt

$$\begin{aligned}
u_{123(ijk)} &= \left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{s}\right) \left(1 - \frac{1}{t}\right) \ln n_{ijk} \\
&\quad - \frac{1}{r} \left(1 - \frac{1}{s}\right) \left(1 - \frac{1}{t}\right) \sum_{i' \neq i} \ln n_{i'jk} \\
&\quad - \frac{1}{s} \left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{t}\right) \sum_{j' \neq j} \ln n_{ij'k} \\
&\quad - \frac{1}{t} \left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{s}\right) \sum_{k' \neq k} \ln n_{ijk'} \\
&\quad + \frac{1}{r \cdot s} \left(1 - \frac{1}{t}\right) \sum_{i' \neq i} \sum_{j' \neq j} \ln n_{i'j'k} \\
&\quad + \frac{1}{r \cdot t} \left(1 - \frac{1}{s}\right) \sum_{i' \neq i} \sum_{k' \neq k} \ln n_{i'jk'} \\
&\quad + \frac{1}{s \cdot t} \left(1 - \frac{1}{r}\right) \sum_{j' \neq j} \sum_{k' \neq k} \ln n_{ij'k'} - \frac{1}{r \cdot s \cdot t} \sum_{i' \neq i} \sum_{j' \neq j} \sum_{k' \neq k} \ln n_{i'j'k'} .
\end{aligned}$$

Wir können nun eine Schätzung der Varianz von z.B. $u_{123(121)}$ angeben durch

$$\begin{aligned}
 \text{Var } u_{123(121)} &= \left(1 - \frac{1}{3}\right)^2 \left(1 - \frac{1}{3}\right)^2 \left(1 - \frac{1}{2}\right)^2 \frac{1}{127} \\
 &\quad + \frac{1}{3^2} \left(1 - \frac{1}{3}\right)^2 \left(1 - \frac{1}{2}\right)^2 \left(\frac{1}{198} + \frac{1}{58}\right) \\
 &\quad + \frac{1}{3^2} \left(1 - \frac{1}{3}\right)^2 \left(1 - \frac{1}{2}\right)^2 \left(\frac{1}{258} + \frac{1}{376}\right) \\
 &\quad + \frac{1}{2^2} \left(1 - \frac{1}{3}\right)^2 \left(1 - \frac{1}{3}\right)^2 \frac{1}{142} \\
 &\quad + \left(\frac{1}{3 \cdot 3}\right)^2 \left(1 - \frac{1}{2}\right)^2 \left(\frac{1}{207} + \frac{1}{201} + \frac{1}{77} + \frac{1}{26}\right) \\
 &\quad + \left(\frac{1}{3 \cdot 2}\right)^2 \left(1 - \frac{1}{3}\right)^2 \left(\frac{1}{184} + \frac{1}{35}\right) \\
 &\quad + \left(\frac{1}{3 \cdot 2}\right)^2 \left(1 - \frac{1}{3}\right)^2 \left(\frac{1}{507} + \frac{1}{873}\right) \\
 &\quad + \left(\frac{1}{3 \cdot 3 \cdot 2}\right)^2 \left(\frac{1}{120} + \frac{1}{172} + \frac{1}{56} + \frac{1}{17}\right) \\
 &= \frac{4}{81} \cdot 0,007874 + \frac{1}{81} \cdot 0,02229 \\
 &\quad + \frac{1}{81} \cdot 0,006536 + \frac{4}{81} \cdot 0,007042 \\
 &\quad + \frac{1}{324} \cdot 0,06125 + \frac{1}{81} \cdot 0,03401 \\
 &\quad + \frac{1}{81} \cdot 0,003118 + \frac{1}{324} \cdot 0,09083 \\
 &= 0,00202.
 \end{aligned}$$

Damit erhalten wir die Standardabweichung 0,045 und somit die folgenden Grenzen für das 95%-Konfidenzintervall von $u_{123(121)}$:

$$0,1598 - 1,96 \cdot 0,045 = 0,0716 \quad \text{und} \quad 0,1598 + 1,96 \cdot 0,045 = 0,248.$$

Danach scheint also eine signifikante 3-fach Wechselwirkung vorzuliegen.

4.2. Testen von Hypothesen über die Parameter des saturierten Modells

Wir können inzwischen sämtliche Parameter des **saturierten loglinearen Modells** sowie deren Varianzen schätzen. Damit haben wir schon einige Anhaltspunkte dafür, welches **loglineare Modell** zur Beschreibung der uns vorliegenden Kontingenztafel geeignet ist. Man erwartet nämlich, daß Parameter, deren Schätzungen im saturierten Modell klein sind im Vergleich zu den zugehörigen Standardabweichungen, im Modell nicht berücksichtigt zu werden brauchen. Solche Hypothesen müssen dann allerdings noch getestet werden. Das wollen wir im folgenden tun. Die allgemeine Vorgehensweise soll dabei an einem Beispiel demonstriert werden. Wir greifen dabei wieder auf die Untersuchung über das Fahrverhalten und den Komfort verschiedener Pkw zurück (vgl. Tab. 50). Es soll hier getestet werden, ob eine 3-fache Wechselwirkung vorhanden ist. Wir haben also als Nullhypothese

$$H_0: u_{123(ijk)} = 0.$$

Das Testverfahren läuft nun folgendermaßen ab. Wir verschaffen uns ML-Schätzungen der unter der Hypothese erwarteten Häufigkeiten m_{ijk} und vergleichen diese mit den beobachteten Häufigkeiten n_{ijk} mittels der Teststatistik

$$\chi^2 = \sum_{i,j,k} \frac{(n_{ijk} - m_{ijk})^2}{m_{ijk}}$$

oder der Statistik

$$G^2 = 2 \cdot \sum_{i,j,k} n_{ijk} \cdot \ln \frac{n_{ijk}}{m_{ijk}}.$$

Die Hypothese H_0 wird zum Niveau α abgelehnt, wenn der Wert der benutzten Prüfgröße das $(1 - \alpha)$ -Quantil einer geeigneten χ^2 -Verteilung überschreitet.

4.2.1. Ein iteratives Verfahren zur Schätzung erwarteter Häufigkeiten unter einer Hypothese

Wir wollen nun zunächst die ML-Schätzungen der unter der Hypothese H_0 erwarteten Häufigkeiten im Beispiel bestimmen. Unter H_0 gilt die Modellannahme

$$\ln m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}.$$

Bei den hier ausschließlich betrachteten hierarchischen Modellen reicht zu deren Festlegung die Angabe der höchsten Wechselwirkungsterme. Das sind in diesem Fall u_{12} , u_{23} und u_{13} . Mit den korrespondierenden Randhäufigkeiten n_{12} , n_{23} und n_{13} , die übereinstimmen mit den entsprechenden Randhäufigkeiten der ML-Schätzungen, lassen sich die gesuchten ML-Schätzer wie folgt iterativ bestimmen.

Ausgangspunkt ist eine Kontingenztafel, in der in jeder Zelle die gleiche Häufigkeit steht; das kann z. B. die Eins sein. Diese Häufigkeiten bezeichnen wir mit $m_{ijk}^{(0)}$. Daraus berechnen wir im **ersten Schritt** eine Tafel mit den Häufigkeiten

$$m_{ijk}^{(1)} = \frac{m_{ijk}^{(0)}}{m_{ij}^{(0)}} \cdot n_{ij}.$$

Im **zweiten Schritt** wird daraus eine Tafel mit den Häufigkeiten

$$m_{ijk}^{(2)} = \frac{m_{ijk}^{(1)}}{m_{jk}^{(1)}} \cdot n_{jk}$$

bestimmt, und im **dritten Schritt** erhalten wir

$$m_{ijk}^{(3)} = \frac{m_{ijk}^{(2)}}{m_{i,k}^{(2)}} \cdot n_{i,k} .$$

Diese drei Schritte werden nun solange wiederholt, bis sich die Ergebnisse stabilisieren. Das sind dann die gesuchten ML-Schätzwerte für die in der Kontingenztafel unter der Nullhypothese erwarteten Häufigkeiten.

Beispiel: Mit den Daten aus Tab. 50 sollen die ML-Schätzungen der m_{ijk} unter H_0 bestimmt werden. Wir setzen

$$m_{ijk}^{(0)} = 1,$$

starten das Iterationsverfahren also mit der Tafel in **Tab. 58**.

Tab. 58: Startwerte für das Iterationsverfahren zur Bestimmung der ML-Schätzwerte der unter H_0 erwarteten Häufigkeiten m_{ijk}

$m_{ij1}^{(0)}$	j	1	2	3	Σ
i					
1		1	1	1	3
2		1	1	1	3
3		1	1	1	3
Σ		3	3	3	9

$m_{ij2}^{(0)}$	j	1	2	3	Σ
i					
1		1	1	1	3
2		1	1	1	3
3		1	1	1	3
Σ		3	3	3	9

$m_{ij.}^{(0)}$	j	1	2	3
i				
1		2	2	2
2		2	2	2
3		2	2	2

Im **ersten Schritt** wird nun mit diesen Werten

$$m_{ijk}^{(1)} = \frac{m_{ijk}^{(0)}}{m_{ij.}^{(0)}} \cdot n_{ij.} = \frac{1}{2} \cdot n_{ij.}$$

berechnet. Die Ergebnisse sind in **Tab. 59** zusammengestellt.

Tab. 59: Erster Iterationsschritt zur Bestimmung von m_{ijk} ($m_{ij1}^{(1)} = m_{ij2}^{(1)} = m_{ijk}^{(1)}$)

$m_{ijk}^{(1)}$	j	1	2	3	Σ
i					
		382,5	134,5	624,5	1141,5
1		163,5	191,0	186,5	541,0
2		66,5	46,5	21,5	134,5
Σ		612,5	372,0	832,5	1817,0

Im zweiten Schritt berechnen wir dann

$$m_{ijk}^{(2)} = \frac{m_{ijk}^{(1)}}{m_{.jk}^{(1)}} \cdot n_{.ik} .$$

Dabei ergibt sich z. B.

$$m_{121}^{(2)} = \frac{134,5}{372,0} \cdot 383 = 138,48 ,$$

$$m_{231}^{(2)} = \frac{186,5}{832,5} \cdot 603 = 135,09 \quad \text{und}$$

$$m_{312}^{(2)} = \frac{66,5}{612,5} \cdot 683 = 74,15 ;$$

sämtliche Ergebnisse sind in **Tab. 60** aufgeführt.

Tab. 60: Zweiter Iterationsschritt zur Bestimmung von m_{ijk}

$m_{ij1}^{(2)}$	j	1	2	3	Σ
i					
		338,47	138,48	452,34	929,29
1		144,68	196,65	135,09	476,42
2		58,85	47,87	15,57	122,29
Σ		542,00	383,00	603,00	1528,00

$m_{ij2}^{(2)}$	j	1	2	3	Σ
i					
		426,53	130,52	796,66	1353,71
1		182,32	185,35	237,91	605,58
2		74,15	45,13	27,43	146,71
Σ		683,00	361,00	1062,00	2106,00

Im **dritten Schritt** ist zu berechnen

$$m_{ijk}^{(3)} = \frac{m_{ijk}^{(2)}}{m_{i,k}^{(2)}} \cdot n_{i,k}.$$

Dabei erhalten wir z.B.

$$m_{211}^{(3)} = \frac{144,68}{476,42} \cdot 606 = 184,03,$$

$$m_{222}^{(3)} = \frac{185,35}{605,58} \cdot 476 = 145,69 \quad \text{und}$$

$$m_{132}^{(3)} = \frac{796,66}{1353,71} \cdot 1522 = 895,70.$$

Alle Ergebnisse dieses dritten Iterationsschrittes sind in **Tab. 61** aufgeführt.

Tab. 61: Dritter Iterationsschritt zur Bestimmung von m_{ijk}

$m_{ij1}^{(3)}$	j	1	2	3	Σ
i					
1		277,18	113,40	370,42	761,00
2		184,03	250,14	171,83	606,00
3		77,48	63,02	20,50	161,00
Σ		538,69	426,56	562,75	1528,00

$m_{ij2}^{(3)}$	j	1	2	3	Σ
i					
1		479,55	146,75	895,70	1522,00
2		143,31	145,69	187,00	476,00
3		54,59	33,22	20,19	108,00
Σ		677,45	325,66	1102,89	2106,00

$m_{ij.}^{(3)}$	j	1	2	3
i				
1		756,73	260,15	1266,12
2		327,34	395,83	358,83
3		132,07	96,24	40,69

Mit diesen Daten führen wir den obigen, aus drei Schritten bestehenden **Iterationszyklus** ein **zweites Mal** durch. Im **ersten Schritt** ist zu berechnen

$$m_{ijk}^{(4)} = \frac{m_{ijk}^{(3)}}{m_{ij.}^{(3)}} \cdot n_{ij.}.$$

So erhält man beispielsweise

$$m_{231}^{(4)} = \frac{171,83}{358,83} \cdot 373 = 178,6,$$

$$m_{112}^{(4)} = \frac{479,55}{756,73} \cdot 765 = 484,8,$$

$$m_{331}^{(4)} = \frac{20,50}{40,69} \cdot 43 = 21,66.$$

Sämtliche Ergebnisse sind in **Tab. 62** aufgeführt.

Tab. 62: Erster Schritt des zweiten Iterationszyklus zur Bestimmung von m_{ijk}

$m_{ij1}^{(4)}$	j	1	2	3	Σ
i					
1		280,20	117,25	365,40	762,85
2		183,90	241,37	178,60	603,87
3		78,03	60,90	21,66	160,59
Σ		542,13	419,52	565,66	1527,31

$m_{ij2}^{(4)}$	j	1	2	3	Σ
i					
1		484,80	151,75	883,60	1520,15
2		143,10	140,63	194,40	478,13
3		54,97	32,10	21,34	108,41
Σ		682,87	324,48	1099,34	2106,69

Im **zweiten Schritt** dieses zweiten Iterationszyklus berechnet man

$$m_{ijk}^{(5)} = \frac{m_{ijk}^{(4)}}{m_{.jk}^{(4)}} \cdot n_{.jk},$$

also z. B.

$$m_{232}^{(5)} = \frac{194,40}{1099,34} \cdot 1062 = 187,8,$$

$$m_{311}^{(5)} = \frac{78,03}{542,13} \cdot 542 = 78 \quad \text{und}$$

$$m_{221}^{(5)} = \frac{241,37}{419,52} \cdot 383 = 220,36.$$

Im **dritten Schritt** ist schließlich zu berechnen

Tab. 63: Dritter Schritt des zweiten Iterationszyklus zur Bestimmung von m_{ijk}

$m_{ij1}^{(6)}$	j	1	2	3	Σ
i					
1		274,50	104,90	381,60	761,00
2		187,30	224,60	194,00	605,90
3		80,20	57,10	23,70	161,00
Σ		542,00	386,60	599,30	1527,90

$m_{ij2}^{(6)}$	j	1	2	3	Σ
i					
1		489,60	170,50	861,90	1522,00
2		139,75	152,80	183,40	475,95
3		53,4	34,64	20,00	108,04
Σ		682,75	357,94	1065,30	2105,99

$m_{ij.}^{(6)}$	j	1	2	3
i				
1		764,10	275,40	1243,50
2		327,05	377,40	377,40
3		133,60	91,74	43,70

$$m_{ijk}^{(6)} = \frac{m_{ijk}^{(5)}}{m_{i..}^{(6)}} n_{i..} .$$

Damit erhält man dann die Ergebnisse aus **Tab. 63**.

Wir führen nun noch einen **dritten Iterationszyklus** durch. Dabei berechnen wir

$$m_{ij.}^{(7)} = \frac{m_{ijk}^{(6)}}{m_{i..}^{(6)}} n_{i..} ,$$

$$m_{.jk}^{(8)} = \frac{m_{ijk}^{(7)}}{m_{.jk}^{(7)}} n_{.jk} \quad \text{und}$$

$$m_{i..}^{(9)} = \frac{m_{ijk}^{(8)}}{m_{i..}^{(8)}} n_{i..} .$$

Damit erhalten wir schließlich mit ausreichender Genauigkeit die gesuchten ML-Schätzwerte für die unter der Nullhypothese erwarteten Häufigkeiten. An den in **Tab. 64** aufgeführten Ergebnissen sieht man auch, daß die Randhäufigkeiten der geschätzten Werte

$$m_{i..}, m_{.jk} \quad \text{und} \quad m_{i..}$$

Tab. 64: ML-Schätzungen der unter der Nullhypothese H_0 erwarteten Häufigkeiten als Ergebnis eines Iterationsverfahrens

m_{ij1}	j	1	2	3	Σ
i					
1		274,4	101,0	385,6	761,0
2		187,5	225,0	193,5	606,0
3		80,1	57,3	23,6	161,0
Σ		542,0	383,3	602,7	1528,0

m_{ij2}	j	1	2	3	Σ
i					
1		490,5	168,8	862,6	1521,9
2		139,4	156,4	180,2	476,0
3		53,0	35,5	19,5	108,0
Σ		682,9	360,7	1062,3	2105,9

$m_{ij.}$	j	1	2	3	Σ
i					
1		764,9	269,8	1248,2	2282,9
2		326,9	381,4	373,7	1082,0
3		133,1	92,8	43,1	269,0
Σ		1224,9	744,0	1665,0	3633,9

praktisch übereinstimmen mit den entsprechenden Randsummen der beobachteten Häufigkeiten

$$n_{ij.}, n_{.jk} \text{ und } n_{i.k}.$$

Wir wollen nun noch an zwei weiteren Beispielen demonstrieren wie man das Iterationsverfahren zur Bestimmung der ML-Schätzwerte erhält. Betrachten wir zunächst für eine vierdimensionale Kontingenztafel die Nullhypothese

$$\begin{aligned} H_0: \ln m_{ijkl} = & u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{12(ij)} \\ & + u_{13(ik)} + u_{23(jk)} + u_{24(jl)} + u_{34(kl)} + u_{123(ijk)}. \end{aligned}$$

Dieses hierarchische loglineare Modell ist schon festgelegt, wenn man weiß, daß u_{24} , u_{34} und u_{123} die höchsten vorkommenden Wechselwirkungen sind. Mit den korrespondierenden Summen der beobachteten Häufigkeiten, $n_{.j.l}$, $n_{..kl}$ und $n_{ijk.}$, werden nun die gesuchten ML-Schätzungen bestimmt. Als Startpunkt für das Iterationsverfahren benutzen wir auch hier eine Tafel mit identischen Zellhäufigkeiten, die wir mit $m_{ijkl}^{(0)}$ bezeichnen. Die Schritte des ersten Iterationszyklus sind gegeben durch

$$\begin{aligned} m_{ijkl}^{(1)} &= \frac{m_{ijkl}^{(0)}}{m_{..j..}^{(0)}} \cdot n_{..j..}, \\ m_{ijkl}^{(2)} &= \frac{m_{ijkl}^{(1)}}{m_{..k..}^{(1)}} \cdot n_{..k..} \quad \text{und} \\ m_{ijkl}^{(3)} &= \frac{m_{ijkl}^{(2)}}{m_{..l..}^{(2)}} \cdot n_{..l..}. \end{aligned}$$

Es werden also nacheinander Kontingenztafeln erzeugt, die die Randsummen $n_{..j..}$, $n_{..k..}$ und $n_{..l..}$ besitzen. Wird dieser aus drei Schritten bestehende Iterationszyklus genügend oft durchgeführt, so erhält man schließlich eine Tafel, die sowohl $n_{..j..}$ als auch $n_{..k..}$ und $n_{..l..}$ als Randsummen besitzt. Haben wir als Hypothese

$$\begin{aligned} H_0: \ln m_{ijkl} &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{12(ij)} \\ &\quad + u_{13(ik)} + u_{23(jk)} + u_{24(jl)} \end{aligned}$$

vorliegen, so hat das Iterationsverfahren zur Bestimmung der ML-Schätzungen der unter dieser Hypothese erwarteten Häufigkeiten einen Zyklus, der vier Schritte umfaßt. In jedem Schritt wird eine Tafel konstruiert, in der einige Randsummen identisch sind mit den entsprechenden Randsummen aus der Tafel der beobachteten Werte. Das sind hier die Randsummen $n_{ij..}$, $n_{i..k..}$, $n_{..jk..}$ und $n_{..j..l..}$. Die dazu korrespondierenden Wechselwirksterme u_{12} , u_{13} , u_{23} und u_{24} sind wie immer diejenigen, die das Modell der Nullhypothese festlegen. Als Startpunkt des Iterationsverfahrens verwenden wir natürlich wieder eine Tafel mit identischen Zellhäufigkeiten, die wir mit $m_{ijkl}^{(0)}$ bezeichnen. Der erste Iterationszyklus umfaßt die folgenden vier Schritte:

$$\begin{aligned} m_{ijkl}^{(1)} &= \frac{m_{ijkl}^{(0)}}{m_{ij..}^{(0)}} \cdot n_{ij..}, \\ m_{ijkl}^{(2)} &= \frac{m_{ijkl}^{(1)}}{m_{i..k..}^{(1)}} \cdot n_{i..k..}, \\ m_{ijkl}^{(3)} &= \frac{m_{ijkl}^{(2)}}{m_{..jk..}^{(2)}} \cdot n_{..jk..} \quad \text{und} \\ m_{ijkl}^{(4)} &= \frac{m_{ijkl}^{(3)}}{m_{..j..l..}^{(3)}} \cdot n_{..j..l..}. \end{aligned}$$

Wird dieser Zyklus genügend oft wiederholt, so erhalten wir schließlich eine Tafel mit den gesuchten ML-Schätzwerten, die die Randsummen $n_{ij..}$, $n_{i..k..}$, $n_{..jk..}$ und $n_{..j..l..}$ besitzt.

Mit dem hier vorgestellten Iterationsverfahren lassen sich für jedes hierarchische loglineare Modell die ML-Schätzungen bestimmen. Bei einem Teil dieser Modelle können außerdem für diese Schätzwerte explizite Formeln angegeben werden. Das Iterationsverfahren liefert in solchen Fällen fast immer schon nach nur einem Zyklus das gewünschte Ergebnis. Explizite Formeln für die ML-Schätzwerte gibt es immer, wenn das betreffende Modell durch weniger als drei Wechselwirkungen festgelegt werden kann. Bei der dreidimensionalen Kontingenztafel gibt es nur ein Modell, für das die ML-Schätzwerte nicht explizit angebbar sind und daher mit

dem Iterationsverfahren bestimmt werden müssen. Es ist dasjenige in dem alle 2-fachen Wechselwirkungen, nicht aber die 3-fache Wechselwirkung vorhanden ist. In Tab. 65 und Tab. 66 sind nun für einige Modelle der 3- und 4-dimensionalen Kontingenztafeln die ML-Schätzwerte der unter diesen Modellen erwarteten Häufigkeiten direkt angegeben.

4.2.2. Die Bestimmung der Freiheitsgrade

Wir sind nun soweit, daß wir die Statistiken χ^2 und G^2 zum Testen der Effekte des loglinearen Modells bestimmen können. Uns fehlen nur noch die Freiheitsgrade. Die erhalten wir folgendermaßen. Wir notieren zunächst sämtliche Terme, die im loglinearen Modell der Nullhypothese nicht auftauchen. Dann berechnen wir für

Tab. 65: ML-Schätzwerte für hierarchische loglineare Modelle 3-dimensionaler Kontingenztafeln

höchste im Modell berücksichtigte Effekte	ML-Schätzung der erwarteten Häufigkeiten
u_{12}, u_{13}	$\frac{n_{ij} \cdot n_{ik}}{n_{..}}$
u_{12}, u_{23}	$\frac{n_{ij} \cdot n_{jk}}{n_{..j}}$
u_{13}, u_{23}	$\frac{n_{ik} \cdot n_{jk}}{n_{..k}}$
u_{12}, u_3	$\frac{n_{ij} \cdot n_{..k}}{n}$
u_{13}, u_2	$\frac{n_{ik} \cdot n_{..j}}{n}$
u_{23}, u_1	$\frac{n_{jk} \cdot n_{..i}}{n}$
u_1, u_2, u_3	$\frac{n_{i..} \cdot n_{..j} \cdot n_{..k}}{n^2}$
u_1, u_2	$\frac{n_{i..} \cdot n_{..j}}{n \cdot t}$
u_1, u_3	$\frac{n_{i..} \cdot n_{..k}}{n \cdot s}$
u_2, u_3	$\frac{n_{..j} \cdot n_{..k}}{n \cdot r}$
u_1	$n_{i..}/(s \cdot t)$
u_2	$n_{..j}/(r \cdot t)$
u_3	$n_{..k}/(r \cdot s)$
u	$n/(r \cdot s \cdot t)$

Tab. 66: ML-Schätzwerte für einige hierarchische loglineare Modelle 4-dimensionaler Kontingenztafeln

höchste im Modell berücksichtigte Effekte	ML-Schätzung der erwarteten Häufigkeiten
u_{123}, u_{124}	$\frac{n_{ijk} \cdot n_{ij..}}{n_{ij..}}$
u_{123}, u_{14}	$\frac{n_{ijk} \cdot n_{i...}}{n_{i...}}$
u_{123}, u_4	$\frac{n_{ijk} \cdot n_{...1}}{n}$
u_{12}, u_{13}, u_{14}	$\frac{n_{ij..} \cdot n_{i..k} \cdot n_{i..l}}{(n_{i...})^2}$
u_{12}, u_{13}, u_{24}	$\frac{n_{ij..} \cdot n_{i..k} \cdot n_{j..l}}{n_{i...} \cdot n_{j..}}$
u_{12}, u_{34}	$\frac{n_{ij..} \cdot n_{..kl}}{n}$
u_{12}, u_{13}, u_4	$\frac{n_{ij..} \cdot n_{i..k} \cdot n_{...l}}{n_{i...} \cdot n}$
u_{12}, u_3, u_4	$\frac{n_{ij..} \cdot n_{..k} \cdot n_{...l}}{n^2}$
u_1, u_2, u_3, u_4	$\frac{n_{i...} \cdot n_{j..} \cdot n_{..k} \cdot n_{...l}}{n^3}$

jeden Term die zugehörigen Freiheitsgrade. Die Summe dieser Werte gibt uns die Freiheitsgrade der Teststatistik χ^2 bzw. G^2 . Die Freiheitsgrade für einen Term ergeben sich als Produkt der um eins verminderten Anzahlen der Merkmalsausprägungen derjenigen Merkmale, deren Zusammenhänge durch eben diesen Term beschrieben werden. So erhalten wir etwa

für u_{134} die Freiheitsgrade $(r - 1)(t - 1)(v - 1)$
 für u_3 $t - 1$ und
 für u_{23} $(s - 1)(t - 1)$.

In dem loglinearen Modell der Nullhypothese

$$H_0: \ln m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$

sind die Terme u_3, u_{13}, u_{23} und u_{123} nicht enthalten. Zu u_3 gehören die Freiheitsgrade $t - 1$, zu u_{13} $(r - 1)(t - 1)$, zu u_{23} $(s - 1)(t - 1)$ und zu u_{123} $(r - 1)(s - 1)(t - 1)$. Die Teststatistiken χ^2 und G^2 zum Testen der obigen Nullhypothese sind somit χ^2 -verteilt mit den Freiheitsgraden $(t - 1) + (r - 1)(t - 1) + (s - 1)(t - 1) + (r - 1)(s - 1)(t - 1) = r \cdot s \cdot (t - 1)$. Die Hypothese H_0 wird also zum Niveau α abgelehnt, wenn der Wert der verwendeten Teststatistik das $(1 - \alpha)$ -Quantil der χ^2 -Verteilung mit $r \cdot s \cdot (t - 1)$ Freiheitsgraden übersteigt.

Beispiel: Bei der Untersuchung über das Fahrverhalten und den Komfort verschiedener Pkws (vgl. Tab. 50) soll getestet werden, ob eine 3-fache Wechselwirkung

vorhanden ist, d.h. wir betrachten die Nullhypothese

$$H_0: u_{123(ijk)} = 0.$$

Die ML-Schätzungen der unter H_0 erwarteten Häufigkeiten haben wir schon früher iterativ berechnet und in Tab. 64 zusammengestellt. Damit erhält man für die χ^2 -Statistik

$$\begin{aligned} \chi^2 &= \sum_{i,j,k} \frac{(n_{ijk} - m_{ijk})^2}{m_{ijk}} \\ &= \frac{(258 - 274,4)^2}{274,4} + \frac{(127 - 101)^2}{101} + \frac{(376 - 385,6)^2}{385,6} + \frac{(207 - 187,5)^2}{187,5} \\ &\quad + \frac{(198 - 225)^2}{225} + \frac{(201 - 193,5)^2}{193,5} + \frac{(77 - 80,1)^2}{80,1} + \frac{(58 - 57,3)^2}{57,3} \\ &\quad + \frac{(26 - 23,6)^2}{23,6} + \frac{(507 - 490,5)^2}{490,5} + \frac{(142 - 168,8)^2}{168,8} + \frac{(873 - 862,6)^2}{862,6} \\ &\quad + \frac{(120 - 139,4)^2}{139,4} + \frac{(184 - 156,4)^2}{156,4} + \frac{(172 - 180,2)^2}{180,2} \\ &\quad + \frac{(56 - 53)^2}{53} + \frac{(35 - 35,5)^2}{35,5} + \frac{(17 - 19,5)^2}{19,5} \\ &= 0,98 + 6,693 + 0,239 + 2,028 + 3,24 + 0,2907 + 0,12 + 0,0086 \\ &\quad + 0,244 + 0,555 + 4,255 + 0,125 + 2,7 + 4,871 + 0,373 + 0,17 \\ &\quad + 0,007 + 0,3205 \\ &= 27,2 \end{aligned}$$

und für die G^2 -Statistik

$$\begin{aligned} G^2 &= 2 \cdot \sum_{i,j,k} n_{ijk} \cdot \ln \frac{n_{ijk}}{m_{ijk}} \\ &= 2 \left(258 \cdot \ln \frac{258}{274,4} + 127 \cdot \ln \frac{127}{101} + 376 \cdot \ln \frac{376}{385,6} + 207 \cdot \ln \frac{207}{187,5} \right. \\ &\quad + 198 \cdot \ln \frac{198}{225} + 201 \cdot \ln \frac{201}{193,5} + 77 \cdot \ln \frac{77}{80,1} + 58 \cdot \ln \frac{58}{57,3} + 26 \cdot \ln \frac{26}{23,6} \\ &\quad + 507 \cdot \ln \frac{507}{490,5} + 142 \cdot \ln \frac{142}{168,8} + 873 \cdot \ln \frac{873}{862,6} + 120 \cdot \ln \frac{120}{139,4} \\ &\quad \left. + 184 \cdot \ln \frac{184}{156,4} + 172 \cdot \ln \frac{172}{180,2} + 56 \cdot \ln \frac{56}{53} + 35 \cdot \ln \frac{35}{35,5} + 17 \cdot \ln \frac{17}{19,5} \right) \\ &= 2(-15,9 + 29,09 - 9,48 + 20,48 - 25,31 + 7,644 - 3,04 + 0,704 \\ &\quad + 2,518 + 16,77 - 24,55 + 10,46 - 17,98 + 29,9 - 8,01 + 3,08 \\ &\quad - 0,496 - 2,33) \\ &= 27,1. \end{aligned}$$

Testen wir zum Niveau $\alpha = 0,05$, so müssen wir diese Werte vergleichen mit dem 0,95-Quantil der χ^2 -Verteilung mit $(r - 1)(s - 1)(t - 1) = (3 - 1)(3 - 1)(2 - 1) = 4$ Freiheitsgraden. Dieses Quantil hat den Wert 9,49. Die Hypothese H_0 wird also abgelehnt.

4.2.3. Die Partitionierung der Teststatistiken

Im obigen Beispiel haben wir nur die 3-fache Wechselwirkung getestet. Der Wert der Teststatistik ist also allein mit diesem Effekt zu erklären. Testen wir dagegen mehrere Effekte, haben wir etwa die Nullhypothese

$$H_0: u_{123(ijk)} = u_{12(ij)} = 0$$

vorliegen, so sind die Anteile beider Effekte in der Teststatistik enthalten. Wir wollen nun den Beitrag der einzelnen Effekte zum Wert der Teststatistik bestimmen. Die Teststatistik für H_0 soll als Summe zweier Ausdrücke geschrieben werden, wobei der Wert des einen nur auf Einflüsse der 2-fachen Wechselwirkung u_{12} und der Wert des anderen nur auf Einflüsse von u_{123} zurückzuführen sein soll. Dazu betrachten wir noch die zwischen H_0 und dem saturierten Modell liegende Hypothese

$$H'_0: u_{123(ijk)} = 0$$

Die Teststatistik G^2 für diese Hypothese, die wir hier mit $G^2(H'_0)$ bezeichnen wollen, ist gegeben durch

$$G^2(H'_0) = 2 \cdot \sum_{i,j,k} n_{ijk} \cdot \ln \frac{n_{ijk}}{m'_{ijk}},$$

wenn m'_{ijk} die unter H'_0 erwarteten Häufigkeiten sind. Vergleichen wir nun die unter den beiden Hypothesen H_0 und H'_0 erwarteten Häufigkeiten m_{ijk} und m'_{ijk} mittels der Statistik

$$G^2(H_0/H'_0) = 2 \cdot \sum_{i,j,k} m'_{ijk} \cdot \ln \frac{m'_{ijk}}{m_{ijk}},$$

so lässt sich die Teststatistik für H_0 ,

$$G^2(H_0) = 2 \cdot \sum_{i,j,k} n_{ijk} \cdot \ln \frac{n_{ijk}}{m_{ijk}},$$

wie folgt schreiben:

$$G^2(H_0) = G^2(H'_0) + G^2(H_0/H'_0).$$

Bei Verwendung der Statistik χ^2 statt G^2 ist diese Beziehung zwar nicht exakt, aber doch meist näherungsweise erfüllt.

Beispiel: Bei der Untersuchung des Fahrverhaltens und des Komforts verschiedener Pkws (vgl. Tab. 50) wollen wir die Nullhypothese

$$H_0: u_{123(ijk)} = u_{23(jk)} = 0$$

testen. Aus Tab. 65 lesen wir für die Schätzungen der unter H_0 erwarteten Häufigkeiten die Formel

$$m_{ijk} = \frac{n_{ij} \cdot n_{i..}}{n_{i..}}$$

ab.

Tab. 67: ML-Schätzungen der erwarteten Häufigkeiten unter der Hypothese, daß $u_{123(ijk)} = u_{23(jk)} = 0$, mit Werten aus Tab. 50.

m_{ij1}	j	1	2	3	Σ
i					
1		255,0	89,7	416,3	761,0
2		183,0	214,0	209,0	606,0
3		79,6	55,7	25,7	161,0
Σ		517,6	359,4	651,0	1528,0

m_{ij2}	j	1	2	3	Σ
i					
1		510,0	179,0	833,0	1522,0
2		143,9	168,0	164,0	475,9
3		53,4	37,3	17,3	108,0
Σ		707,3	384,3	1014,3	2105,9

$m_{ij.}$	j	1	2	3	Σ
i					
1		765,0	268,7	1249,3	2283,0
2		326,9	382,0	373,0	1081,9
3		133,0	93,0	43,0	269,0
Σ		1224,9	743,7	1665,3	3633,9

Die Werte, die man mit den beobachteten Häufigkeiten aus Tab. 50 erhält, sind in **Tab. 67** wiedergegeben.

Nun können die Teststatistiken berechnet werden. Es ist

$$\begin{aligned} \chi^2(H_0) &= \frac{(258 - 255)^2}{255} + \frac{(127 - 89,7)^2}{89,7} + \frac{(376 - 416,3)^2}{416,3} \\ &+ \frac{(207 - 183)^2}{183} + \frac{(198 - 214)^2}{214} + \frac{(201 - 209)^2}{209} \\ &+ \frac{(77 - 79,6)^2}{79,6} + \frac{(58 - 55,7)^2}{55,7} + \frac{(26 - 25,7)^2}{25,7} \\ &+ \frac{(507 - 510)^2}{510} + \frac{(142 - 179)^2}{179} + \frac{(873 - 833)^2}{833} \\ &+ \frac{(120 - 143,9)^2}{143,9} + \frac{(184 - 168)^2}{168} + \frac{(172 - 164)^2}{164} \end{aligned}$$

$$\begin{aligned}
 & + \frac{(56 - 53,4)^2}{53,4} + \frac{(35 - 37,3)^2}{37,3} + \frac{(17 - 17,3)^2}{17,3} \\
 = & 0,0353 + 15,51 + 3,9 + 3,1475 + 1,196 + 0,306 + 0,085 \\
 & + 0,095 + 0,0035 + 0,01765 + 7,648 + 1,921 + 3,97 \\
 & + 1,524 + 0,39 + 0,1266 + 0,142 + 0,005 \\
 = & 40,0
 \end{aligned}$$

und

$$\begin{aligned}
 G^2(H_0) = & 2 \left(258 \ln \frac{258}{255} + 127 \ln \frac{127}{89,7} + 376 \ln \frac{376}{416,3} + 207 \ln \frac{207}{183} \right. \\
 & + 198 \ln \frac{198}{214} + 201 \ln \frac{201}{209} + 77 \ln \frac{77}{79,6} + 58 \ln \frac{58}{55,7} \\
 & + 26 \ln \frac{26}{25,7} + 507 \ln \frac{507}{510} + 142 \ln \frac{142}{179} + 873 \ln \frac{873}{833} \\
 & + 120 \ln \frac{120}{143,9} + 184 \ln \frac{184}{168} + 172 \ln \frac{172}{164} \\
 & \left. + 56 \ln \frac{56}{53,4} + 35 \ln \frac{35}{37,3} + 17 \ln \frac{17}{17,3} \right) \\
 = & 2(3,0176 + 44,16 - 38,28 + 25,509 - 15,386 - 7,845 \\
 & - 2,557 + 2,347 + 0,3017 - 2,99 - 32,88 + 40,945 \\
 & - 21,795 + 16,74 + 8,192 + 2,662 - 2,2275 - 0,2973) \\
 = & 39,2.
 \end{aligned}$$

Der Vergleich der Schätzwerte der unter den Hypothesen

$$H_0: u_{123(ijk)} = u_{23(jk)} = 0 \quad \text{und}$$

$$H'_0: u_{123(ijk)} = 0$$

erwarteten Häufigkeiten m'_{ijk} bzw. m'_{ijk} (vgl. Tab. 67 bzw. Tab. 64), liefert

$$\begin{aligned}
 \chi^2(H_0/H'_0) = & \sum_{i,j,k} \frac{(m'_{ijk} - m_{ijk})^2}{m_{ijk}} \\
 = & \frac{(274,4 - 255)^2}{255} + \frac{(101 - 89,7)^2}{89,7} + \frac{(385,6 - 416,3)^2}{416,3} \\
 & + \frac{(187,5 - 183)^2}{183} + \frac{(225 - 214)^2}{214} + \frac{(193,5 - 209)^2}{209} \\
 & + \frac{(80,1 - 79,6)^2}{79,6} + \frac{(57,3 - 55,7)^2}{55,7} + \frac{(23,6 - 25,7)^2}{25,7} \\
 & + \frac{(490,5 - 510)^2}{510} + \frac{(168,8 - 179)^2}{179} + \frac{(862,6 - 833)^2}{833} \\
 & + \frac{(139,4 - 143,9)^2}{143,9} + \frac{(156,4 - 168)^2}{168} + \frac{(180,2 - 164)^2}{164}
 \end{aligned}$$

$$\begin{aligned}
& + \frac{(53 - 53,4)^2}{53,4} + \frac{(35,5 - 37,3)^2}{37,3} + \frac{(19,5 - 17,3)^2}{17,3} \\
& = 1,48 + 1,42 + 2,26 + 0,11 + 0,565 + 1,15 \\
& + 0,00314 + 0,04596 + 0,1716 + 0,7456 + 0,5812 + 1,052 \\
& + 0,141 + 0,801 + 1,6 + 0,003 + 0,08686 + 0,28 \\
& = 12,5
\end{aligned}$$

und

$$\begin{aligned}
G^2(H_0/H'_0) &= 2 \sum_{i,j,k} m'_{ijk} \cdot \ln \frac{m'_{ijk}}{m_{ijk}} \\
&= 2 \left(274,4 \cdot \ln \frac{274,4}{255} + 101 \cdot \ln \frac{101}{89,7} + 385,6 \cdot \ln \frac{385,6}{416,3} \right. \\
&\quad + 187,5 \cdot \ln \frac{187,5}{183} + 225 \cdot \ln \frac{225}{214} + 193,5 \cdot \ln \frac{193,5}{209} \\
&\quad + 80,1 \cdot \ln \frac{80,1}{79,6} + 57,3 \cdot \ln \frac{57,3}{55,7} + 23,6 \cdot \ln \frac{23,6}{25,7} \\
&\quad + 490,5 \cdot \ln \frac{490,5}{510} + 168,8 \cdot \ln \frac{168,8}{179} + 862,6 \cdot \ln \frac{862,6}{833} \\
&\quad + 139,4 \cdot \ln \frac{139,4}{143,9} + 156,4 \cdot \ln \frac{156,4}{168} + 180,2 \cdot \ln \frac{180,2}{164} \\
&\quad \left. + 53 \cdot \ln \frac{53}{53,4} + 35,5 \cdot \ln \frac{35,5}{37,3} + 19,5 \cdot \ln \frac{19,5}{17,3} \right) \\
&= 2(20,12 + 11,98 - 29,54 + 4,555 + 11,28 - 14,91 \\
&\quad + 0,502 + 1,623 - 2,012 - 19,12 - 9,904 + 30,12 \\
&\quad - 4,43 - 11,19 + 16,975 - 0,3985 - 1,756 + 2,334) \\
&= 12,5.
\end{aligned}$$

Die unter dem saturierten Modell erwarteten Häufigkeiten, die identisch sind mit den beobachteten Häufigkeiten n_{ijk} , haben wir schon früher einmal mit den unter H'_0 erwarteten verglichen, nämlich beim Testen der 3-fachen Wechselwirkung. Es hatte sich dort ergeben, vgl. Abschnitt 4.2.2:

$$\chi^2(H'_0) = \sum_{i,j,k} \frac{(n_{ijk} - m'_{ijk})^2}{m'_{ijk}} = 27,2 \quad \text{und}$$

$$G^2(H'_0) = \sum_{i,j,k} n_{ijk} \cdot \ln \frac{n_{ijk}}{m'_{ijk}} = 27,1.$$

Stellen wir die Ergebnisse nun noch einmal zusammen. Wir haben

$$\begin{aligned}
\chi^2(H_0) &= 40,0, \\
\chi^2(H_0/H'_0) &= 12,5, \\
\chi^2(H'_0) &= 27,2, \\
\chi^2(H_0/H'_0) + \chi^2(H'_0) &= 39,7
\end{aligned}$$

und

$$\begin{aligned} G^2(H_0) &= 39,2, \\ G^2(H_0/H'_0) &= 12,5, \\ G^2(H'_0) &= 27,1, \\ G^2(H_0/H'_0) + G^2(H'_0) &= 39,6. \end{aligned}$$

Man sieht, daß hier sowohl die Beziehung

$$G^2(H_0/H'_0) + G^2(H'_0) = G^2(H_0)$$

als auch die entsprechende Beziehung für χ^2 praktisch erfüllt ist.

Zum Schluß wollen wir noch die Freiheitsgrade angeben. u_{23} schlägt mit $(s-1)(t-1) = (3-1)(2-1) = 2$ und u_{123} mit $(r-1)(s-1)(t-1) = (3-1)(3-1)(2-1) = 4$ Freiheitsgraden zu Buche. Es haben also $G^2(H_0/H'_0)$ und $\chi^2(H_0/H'_0)$ zwei, $G^2(H'_0)$ und $\chi^2(H'_0)$ vier und $G^2(H_0)$ und $\chi^2(H_0)$ sechs Freiheitsgrade.

5. Verteilungsannahmen, Logit-Modell und Adjustieren bei Kontingenztafeln

Wir haben in den bisherigen Abschnitten die Struktur einer Kontingenztafel mittels eines loglinearen Modells beschrieben. Dabei haben wir uns, mit Ausnahme der 2×2 -Tafel, keinerlei Gedanken gemacht über die zugrundeliegenden Verteilungsannahmen. Diese Lücke soll nun geschlossen werden. Wir verschaffen uns damit eine modellhafte Vorstellung vom Erzeugungsmechanismus der in einer Kontingenztafel erfaßten Daten. Das kann bei der Interpretation von Ergebnissen aus dem loglinearen Modell von Nutzen sein.

Neue Interpretationsmöglichkeiten sind auch der Grund dafür, ein loglineares Modell in ein Logit-Modell umzuschreiben bzw. von vornherein ein Logit-Modell anzusetzen. Bei einem solchen Modell bekommt eines der Merkmale mit nur zwei Ausprägungen eine ausgezeichnete Stellung. Es können dann nur noch die Einflüsse der anderen Merkmale auf dieses untersucht werden und nicht mehr wie im allgemeinen loglinearen Modell sämtliche Zusammenhänge zwischen den beobachteten Merkmalen.

Zum Schluß dieses Abschnitts wollen wir noch ein Verfahren vorstellen, mit dem wir eine Kontingenztafel so umformen können, daß die Randsummen vorgegebene Werte annehmen und trotzdem die Wechselwirkungen erhalten bleiben. Damit kann man etwa verschiedene Untersuchungen der gleichen Merkmalszusammenhänge direkt miteinander vergleichen.

5.1. Kontingenztafeln und Verteilungen

5.1.1. Verteilungsannahmen bei Kontingenztafeln

Bei einer Kontingenztafel können einige Häufigkeiten wie z. B. irgendwelche Randsummen oder die gesamte Beobachtungszahl schon im vorhinein festgelegt sein. Aus den verschiedenen Möglichkeiten heraus, so etwas zu tun, resultieren die unterschiedlichen Verteilungsannahmen, die hier vorgestellt werden sollen.

Hat man noch gar nichts festgelegt, also noch nicht einmal die Gesamtzahl der zu

untersuchenden Objekte, so können die in den einzelnen Zellen der Kontingenztafel beobachteten Häufigkeiten als unabhängig und poissonverteilt angesehen werden. Die Parameter dieser **Poissonverteilungen** (vgl. Kap. IV, Abschnitt 3.4) sind gerade die erwarteten Zellhäufigkeiten.

Solch ein Modell eignet sich etwa zur Beschreibung der folgenden Situation. Ein Autohersteller möchte wissen, ob und welche Zusammenhänge es bei der Bestellung der verschiedenen von ihm angebotenen Sonderausstattungen gibt. Oft zusammenbestellte Sonderausstattungen könnten etwa zu Paketen zusammengefaßt werden und so die Produktion vereinfachen. Es werden nun sämtliche in einem bestimmten Zeitraum eingegangenen Bestellungen ausgewertet und die Ergebnisse in einer Kontingenztafel dargestellt. Als Merkmale haben wir dabei die einzelnen Sonderausstattungen wie etwa Radio, getönte Scheiben, Servolenkung, zweiter Außenspiegel usw. Jedes dieser Merkmale hat nur die beiden Ausprägungen „bestellt“ und „nicht bestellt“. Haben wir 3 Sonderausstattungen, so heißt das, daß die Ergebnisse in einer $2 \times 2 \times 2$ Tafel dargestellt werden können. Es sind dann 8 verschiedene Kundentypen vorhanden, von denen im Beobachtungszeitraum der Länge t jeweils n_{ijk} eine Bestellung aufgegeben haben. Die Wahrscheinlichkeit eines bestimmten Bestell-Ergebnisses ist gerade

$$P((n_{111}, \dots, n_{222})) = \prod_{i,j,k} \frac{m_{ijk}^{n_{ijk}} e^{-m_{ijk}}}{n_{ijk}!},$$

wenn die in den einzelnen Zellen beobachteten Häufigkeiten unabhängig voneinander und poissonverteilt sind, d.h.

$$\begin{aligned} P((n_{111}, \dots, n_{222})) &= \prod_{i,j,k} P(n_{ijk}) \quad \text{und} \\ P(n_{ijk}) &= \frac{m_{ijk}^{n_{ijk}} e^{-m_{ijk}}}{n_{ijk}!}. \end{aligned}$$

Die 8 Kundenströme sind dann nämlich durch 8 unabhängige **Poissonprozesse** mit den **Ankunftsralten**

$$\lambda_{ijk} = \frac{m_{ijk}}{t}$$

beschrieben.

Die Gesamtzahl n der Beobachtungen war bei dieser Untersuchung nicht von vornherein festgelegt. Das weiß man erst am Ende des Beobachtungszeitraums. Bei Kenntnis der Gesamtzahl der Kunden wird die Verteilung auf die einzelnen Typen durch eine **Multinomialverteilung** (vgl. Abschnitt 3.3 in Kap. IV) bestimmt,

$$P((n_{111}, \dots, n_{222})) = n! \prod_{i,j,k} \frac{1}{n_{ijk}!} \left(\frac{m_{ijk}}{n} \right)^{n_{ijk}}.$$

Diese Verteilung ist insbesondere auch zur Beschreibung solcher Untersuchungen geeignet, bei denen die Beobachtungszahl von vornherein festliegt. Das ist z.B. bei Umfragen oft der Fall. Wir haben dabei die Vorstellung, daß jede bestimmte Merkmalskombination unabhängig voneinander bei jeder Person mit der gleichen Wahrscheinlichkeit auftaucht. Bei 4 Merkmalen sind diese Wahrscheinlichkeiten gerade gegeben durch

$$P_{ijkl} = \frac{m_{ijkl}}{n}.$$

Haben wir hier nur die Gesamtbeobachtungszahl festgelegt, so ist es an anderen Stellen sinnvoll, zugleich etwa auch festzulegen, wieviele Objekte mit bestimmten Merkmalsausprägungen untersucht werden sollen. Das wollen wir an einem Beispiel näher erklären. Eine Verbraucherorganisation möchte wissen, ob die Qualität von Frischobst davon abhängt, ob man es auf dem Wochenmarkt, beim Lebensmittelhändler oder im Supermarkt einkauft. Bei einer dazu durchgeföhrten Untersuchung wird man sicher vorher bestimmen, wieviele Einkäufe bei jeder Verkaufsform zu tätigen sind. Beurteilen wir die Ware nur mit gut, mittel und schlecht, so können die Ergebnisse in einer 3×3 -Kontingenzttafel dargestellt werden. Wie das prinzipiell aussieht, ist in **Tab. 68** dargestellt. Wie das

Tab. 68: Qualität von Frischobst in Abhängigkeit von der Verkaufsform

Güte der Einkäufe	Verkaufs-form	Wochen-markt	Lebensmittel-händler	Super-markt	Σ
		n_{11}	n_{12}	n_{13}	
gut		n_{21}	n_{22}	n_{23}	n_2
mittel		n_{31}	n_{32}	n_{33}	n_3
Σ		n_1	n_2	n_3	n

Die drei Randsummen n_1 , n_2 und n_3 sind hier schon vorher festgelegt. Wir können annehmen, daß wir drei unabhängige Stichproben mit eben diesen Umfängen vorliegen haben. Bei der Stichprobe vom Wochenmarkt treten die drei Güteklassen mit den Wahrscheinlichkeiten

$$\frac{m_{11}}{n_1}, \frac{m_{21}}{n_1} \text{ und } \frac{m_{31}}{n_1}$$

auf. Die entsprechenden Wahrscheinlichkeiten beim Lebensmittelhändler sind

$$\frac{m_{12}}{n_2}, \frac{m_{22}}{n_2} \text{ und } \frac{m_{32}}{n_2}$$

und beim Supermarkt sind sie

$$\frac{m_{13}}{n_3}, \frac{m_{23}}{n_3} \text{ und } \frac{m_{33}}{n_3}.$$

Wir haben nun

$$P((n_{1j}, n_{2j}, n_{3j})) = n_j! \prod_{i=1}^3 \frac{1}{n_{ij}!} \left(\frac{m_{ij}}{n_j} \right)^{n_{ij}},$$

also für jede Verkaufsform eine Multinomialverteilung über die drei Güteklassen. Hätten wir nur die Gesamtzahl n der Einkäufe festgelegt, so wäre die Beschreibung der Kontingenzttafel durch die Multinomialverteilung erfolgt mit

$$P((n_{11}, \dots, n_{33})) = n! \prod_{i,j} \frac{1}{n_{ij}!} \left(\frac{m_{ij}}{n} \right)^{n_{ij}}.$$

Bedingen wir unter den oben festgelegten drei Randhäufigkeiten, so erhalten wir

$$P((n_{11}, \dots, n_{33})/n_{\cdot 1}, n_{\cdot 2}, n_{\cdot 3}) = \prod_j n_{\cdot j}! \prod_{i=1}^3 \frac{1}{n_{ij}!} \left(\frac{m_{ij}}{n_{\cdot j}} \right)^{n_{ij}}$$

und damit als Randverteilungen die obigen drei Multinomialverteilungen. Unabhängigkeit der beiden Merkmale äußert sich darin, daß diese drei Verteilungen übereinstimmen.

Stellen wir die erwarteten Häufigkeiten m_{ij} in einem loglinearen Modell dar, so müssen wir darauf achten, daß die zu den vorher festgelegten Häufigkeiten korrespondierenden Terme, in diesem Fall $u_{2(1)}$, $u_{2(2)}$ und $u_{2(3)}$, auch im Modell enthalten sind. Haben wir z.B. für eine 3-dimensionale Kontingenztafel ein loglineares Modell mit

$$\ln m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)}$$

aufgestellt, dann können die Randhäufigkeiten $n_{\cdot jk}$ nicht vorher festgelegt werden, weil $u_{23(jk)}$ nicht im Modell enthalten ist. Es können wohl $r \cdot s$ Stichproben mit den Umfängen n_{ij} , oder r Stichproben mit den Umfängen $n_{i\cdot}$ erzeugt werden. Beachtet man diese Regeln, so liefern alle dann noch möglichen Verteilungsannahmen die gleichen ML-Schätzwerte für die erwarteten Häufigkeiten unter der Modellannahme.

5.1.2. Vergleich der Parameter mehrerer diskreter Verteilungen

Tests für Kontingenztafeln lassen sich nun auch auffassen als Tests für die Parameter der zugrundegelegten Verteilungen.

A. Vergleich der Parameter von s Poissonverteilungen

Hat man z.B. eine $1 \times s$ -Tafel, deren Zellhäufigkeiten als unabhängig $Po(\lambda_i)$ -verteilt angesehen werden können, so bedeutet der Test auf Gleichverteilung bei dieser Kontingenztafel das Testen der Hypothese

$$H_0: \lambda_1 = \dots = \lambda_s.$$

Man kann die Gleichheit der Parameter von s unabhängigen Poissonverteilungen also testen mittels der Statistik

$$\chi^2 = \sum_{j=1}^s \frac{(x_j - \bar{x})^2}{\bar{x}}$$

oder der Statistik

$$G^2 = 2 \sum_{j=1}^s x_j \ln \frac{x_j}{\bar{x}}.$$

Dabei sind x_1, \dots, x_s die beobachteten Häufigkeiten.

Die Hypothese H_0 wird zum Niveau α abgelehnt, wenn der Wert der benutzten Teststatistik größer ist als das $1 - \alpha$ Quantil der χ^2 -Verteilung mit $s - 1$ Freiheitsgraden (vgl. Hypothese H_0^4 in Tab. 26; $r = 1$, $m_{1j} \hat{=} \lambda_j$, $n_{1j} \hat{=} x_j$, $\frac{n}{s} \hat{=} \bar{x}$).

Beispiel: Mit den Messungen $x_1 = 37, x_2 = 56, x_3 = 43, x_4 = 65, x_5 = 48, x_6 = 71$ und $x_7 = 62$ soll zum Niveau $\alpha = 0,05$ überprüft werden, ob die Parameter λ_i der zugrundeliegenden Poissonverteilungen identisch sind. Es ergibt sich hier

$$\begin{aligned}\chi^2 &= \frac{(37 - 54,57)^2}{54,57} + \frac{(56 - 54,57)^2}{54,57} + \frac{(43 - 54,57)^2}{54,57} + \frac{(65 - 54,57)^2}{54,57} \\ &\quad + \frac{(48 - 54,57)^2}{54,57} + \frac{(71 - 54,57)^2}{54,57} + \frac{(62 - 54,57)^2}{54,57} \\ &= 5,66 + 0,04 + 2,45 + 1,99 + 0,79 + 4,95 + 1,01 \\ &= 16,89\end{aligned}$$

bzw.

$$\begin{aligned}G^2 &= 2 \left(37 \cdot \ln \frac{37}{54,57} + 56 \cdot \ln \frac{56}{54,57} + 43 \cdot \ln \frac{43}{54,57} + 65 \cdot \ln \frac{65}{54,57} \right. \\ &\quad \left. + 48 \cdot \ln \frac{48}{54,57} + 71 \cdot \ln \frac{71}{54,57} + 62 \cdot \ln \frac{62}{54,57} \right) \\ &= 2(-14,38 + 1,45 - 10,25 + 11,37 - 6,16 + 18,69 + 7,91) \\ &= 2 \cdot 8,63 \\ &= 17,26.\end{aligned}$$

Das 0,95-Quantil der χ^2 -Verteilung mit $s - 1 = 6$ Freiheitsgraden ist $\chi_{0,95}^2 = 12,59$, so daß die Hypothese der Gleichheit der Parameter λ_i abzulehnen ist.

B. Vergleich der Parameter verschiedener Binomialverteilungen

Ähnlich kann auch die Gleichheit der Parameter p_j von s **Binomialverteilungen** $B(n_j, p_j)$ getestet werden. Dazu schreibe man in die erste Zeile einer $2 \times s$ -Tafel jeweils die Anzahl der Erfolge v_j (entspricht m in Abschnitt 3.1 von Kap. IV) und in die zweite Zeile die Anzahl der Mißerfolge $n_j - v_j$. Der Test auf Unabhängigkeit in dieser Kontingenztafel bedeutet dann gerade das Testen der Hypothese

$$H_0: p_1 = \dots = p_s$$

(vgl. Hypothese H_0^1 in Tab. 26; $r = 2$, $m_{1j} \hat{=} p_j \cdot n_j$, $m_{2j} \hat{=} (1 - p_j) n_j$, $n_{1j} \hat{=} v_j$, $n_{2j} \hat{=} n_j - v_j$, $n_{.j} \hat{=} n_j$, $n_{1.} \hat{=} v_{.}$, $n_{2.} \hat{=} n_{.} - v_{.}$, $n \hat{=} n_{.}$).

Die Hypothese kann also getestet werden mit der Statistik

$$\chi^2 = \sum_{j=1}^s \frac{\left(v_j - \frac{v_{.} n_j}{n_{.}}\right)^2}{\frac{v_{.} n_j}{n_{.}}} + \sum_{j=1}^s \frac{\left(n_j - v_j - \frac{(n_{.} - v_{.}) n_j}{n_{.}}\right)^2}{\frac{(n_{.} - v_{.}) n_j}{n_{.}}}$$

oder

$$G^2 = 2 \cdot \sum_{j=1}^s v_j \cdot \ln \left(v_j / \frac{v_{.} n_j}{n_{.}} \right) + 2 \cdot \sum_{j=1}^s (n_j - v_j) \cdot \ln \left((n_j - v_j) / \frac{(n_{.} - v_{.}) n_j}{n_{.}} \right)$$

Man lehnt H_0 zum Niveau α ab, wenn der Wert der benutzten Teststatistik größer ist als das $1 - \alpha$ Quantil der χ^2 -Verteilung mit $s - 1$ Freiheitsgraden.

Beispiel: Es soll zum Niveau $\alpha = 0,01$ überprüft werden, ob die Parameter p_1, \dots, p_8 von Binomialverteilungen $B(n_i, p_i)$ mit $n_1 = 35, n_2 = 41, n_3 = 25, n_4 = 37, n_5 = 39, n_6 = 23, n_7 = 48$ und $n_8 = 56$ identisch sind. Aus den beobachteten Werten $v_1 = 13, v_2 = 16, v_3 = 13, v_4 = 21, v_5 = 27, v_6 = 12, v_7 = 20$ und $v_8 = 22$ erhält man

$$\begin{aligned} \chi^2 &= \frac{\left(13 - \frac{144 \cdot 35}{304}\right)^2}{\frac{144 \cdot 35}{304}} + \frac{\left(16 - \frac{144 \cdot 41}{304}\right)^2}{\frac{144 \cdot 41}{304}} + \frac{\left(13 - \frac{144 \cdot 25}{304}\right)^2}{\frac{144 \cdot 25}{304}} \\ &\quad + \frac{\left(21 - \frac{144 \cdot 37}{304}\right)^2}{\frac{144 \cdot 37}{304}} + \frac{\left(27 - \frac{144 \cdot 39}{304}\right)^2}{\frac{144 \cdot 39}{304}} + \frac{\left(12 - \frac{144 \cdot 23}{304}\right)^2}{\frac{144 \cdot 23}{304}} \\ &\quad + \frac{\left(20 - \frac{144 \cdot 48}{304}\right)^2}{\frac{144 \cdot 48}{304}} + \frac{\left(22 - \frac{144 \cdot 56}{304}\right)^2}{\frac{144 \cdot 56}{304}} \\ &\quad + \frac{\left(35 - 13 - \frac{(304 - 144) \cdot 35}{304}\right)^2}{\frac{(304 - 144) \cdot 35}{304}} + \frac{\left(41 - 16 - \frac{(304 - 144) \cdot 41}{304}\right)^2}{\frac{(304 - 144) \cdot 41}{304}} \\ &\quad + \frac{\left(25 - 13 - \frac{(304 - 144) \cdot 25}{304}\right)^2}{\frac{(304 - 144) \cdot 25}{304}} + \frac{\left(37 - 21 - \frac{(304 - 144) \cdot 37}{304}\right)^2}{\frac{(304 - 144) \cdot 37}{304}} \\ &\quad + \frac{\left(39 - 27 - \frac{(304 - 144) \cdot 39}{304}\right)^2}{\frac{(304 - 144) \cdot 39}{304}} + \frac{\left(23 - 12 - \frac{(304 - 144) \cdot 23}{304}\right)^2}{\frac{(304 - 144) \cdot 23}{304}} \\ &\quad + \frac{\left(48 - 20 - \frac{(304 - 144) \cdot 48}{304}\right)^2}{\frac{(304 - 144) \cdot 48}{304}} + \frac{\left(56 - 22 - \frac{(304 - 144) \cdot 56}{304}\right)^2}{\frac{(304 - 144) \cdot 56}{304}} \\ &= 0,77 + 0,6 + 0,11 + 0,69 + 3,94 + 0,11 + 0,33 + 0,77 \\ &\quad + 0,7 + 0,54 + 0,10 + 0,62 + 3,54 + 0,10 + 0,30 + 0,70 \\ &= 13,92 \end{aligned}$$

bzw.

$$\begin{aligned} G^2 &= 2(-3,16 - 3,10 + 1,21 + 3,80 + 10,25 + 1,16 - 2,57 - 4,12 \\ &\quad + 3,91 + 3,68 - 1,11 - 3,14 - 6,44 - 1,05 + 2,88 + 4,86) \\ &= 2 \cdot 7,06 \\ &= 14,12. \end{aligned}$$

Das $1 - \alpha = 0,99$ Quantil der χ^2 -Verteilung mit $s - 1 = 7$ Freiheitsgraden hat den Wert $\chi^2_{7, 0,99} = 18,48$, so daß die Hypothese der Gleichheit der p_j nicht abgelehnt werden kann.

C. Vergleich der Parameter mehrerer Multinomialverteilungen

Wir wollen nun noch angeben, wie man bei s **Multinomialverteilungen** $M(n_1, p_{11}, \dots, p_{r1}), \dots, M(n_s, p_{1s}, \dots, p_{rs})$ aufgrund der Beobachtungsreihen x_{1j}, \dots, x_{sj} für $j = 1, \dots, s$ die Hypothese

$$H_0: p_{11} = \dots = p_{1s}, p_{21} = \dots = p_{2s}, \dots, p_{r1} = \dots = p_{rs}$$

testet (vgl. Hypothese H_0^1 in Tab. 26, $m_{ij} \hat{=} p_{ij} n_j$, $n_{ij} \hat{=} x_{ij}$, $n_i \hat{=} x_i$, $n_j \hat{=} n_j = x_{.j}$, $n \hat{=} n$). Man berechnet zunächst die Teststatistik

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{\left(\frac{x_{ij}}{n_j} - \frac{x_i \cdot n_j}{n} \right)^2}{\frac{x_i \cdot n_j}{n}} \quad \text{oder} \quad G^2 = 2 \cdot \sum_{i=1}^r \sum_{j=1}^s x_{ij} \cdot \ln \left(\frac{x_{ij}}{\frac{x_i \cdot n_j}{n}} \right).$$

Ist der dabei gewonnene Wert größer als das $1 - \alpha$ Quantil der χ^2 -Verteilung mit $(r - 1)(s - 1)$ Freiheitsgraden, so muß H_0 zum Niveau α verworfen werden.

Wir wollen an dieser Stelle noch einmal darauf hinweisen, daß die Teststatistiken nur approximativ χ^2 -verteilt sind und deshalb diese Tests nur bei genügend umfangreichem Beobachtungsmaterial angewandt werden sollten.

5.2. Das Logit-Modell bei Kontingenztafeln

Beschreiben wir eine Kontingenztafel durch ein loglineares Modell, so können wir sämtliche Wechselwirkungen zwischen den vorkommenden Merkmalen erfassen. Oft interessieren aber gar nicht alle diese Zusammenhänge, sondern nur, wie ein ganz bestimmtes Merkmal von den anderen beeinflußt wird. Sehen wir uns dazu einmal als Beispiel die Ergebnisse einer Untersuchung an, bei der der Erfolg zweier Operationsmethoden in 3 verschiedenen Kliniken untersucht wurde. Die erhobenen Daten sind in Tab. 69 wiedergegeben.

Wir haben es hier mit einer 3-dimensionalen Tafel zu tun, für die das saturierte loglineare Modell gegeben ist durch

$$\ln m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}.$$

Wie man sieht, werden hier sämtliche Wechselwirkungen berücksichtigt. Bei einer gezielten Fragestellung, etwa wie die Erfolgsaussichten der Operation von der behandelnden Klinik und der Operationstechnik abhängen, interessieren jedoch nicht alle diese Wechselwirkungen; so z. B. nicht diejenige zwischen Klinik und Operationstechnik, die durch $u_{12(ij)}$ beschrieben wird. Für solch eine spezielle Situ-

Tab. 69: Vergleich zweier Operationstechniken

		Erfolg			
		1	2	3	Σ
Operations-technik	Klinik				
	1	37	128	97	262
	2	21	82	157	260
Σ		58	210	254	522
Mißerfolg					
Operations-technik	Klinik				
	1	12	15	8	35
	2	3	9	17	29
Σ		15	24	25	64

ation wollen wir ein sogenanntes **Logit-Modell** aufstellen, das sich aus dem allgemeinen loglinearen Modell wie folgt gewinnen läßt: Im saturierten loglinearen Modell gilt für die erwartete Anzahl erfolgreicher Operationen

$$\ln m_{ij1} = u + u_{1(i)} + u_{2(j)} + u_{3(1)} + u_{12(ij)} + u_{13(i1)} + u_{23(j1)} + u_{123(ij1)}$$

und für die erwartete Anzahl von Mißerfolgen

$$\ln m_{ij2} = u + u_{1(i)} + u_{2(j)} + u_{3(2)} + u_{12(ij)} + u_{13(i2)} + u_{23(j2)} + u_{123(ij2)}.$$

Den Quotienten dieser beiden erwarteten Häufigkeiten können wir als Maß für die Erfolgsaussichten betrachten. Dafür gilt dann

$$\begin{aligned} \ln \frac{m_{ij1}}{m_{ij2}} &= u_{3(1)} - u_{3(2)} + u_{13(i1)} - u_{13(i2)} + u_{23(j1)} - u_{23(j2)} \\ &\quad + u_{123(ij1)} - u_{123(ij2)}. \end{aligned}$$

Man sieht, daß hier nur noch solche Terme auftauchen, die mit dem Merkmal Erfolgsaussicht in Verbindung stehen. Wir setzen

$$\begin{aligned} w &= u_{3(1)} - u_{3(2)}, \quad w_{1(i)} = u_{13(i1)} - u_{13(i2)}, \\ w_{2(j)} &= u_{23(j1)} - u_{23(j2)} \quad \text{und} \quad w_{12(ij)} = u_{123(ij1)} - u_{123(ij2)}; \end{aligned}$$

damit erhalten wir das **Logit-Modell**

$$\ln \frac{m_{ij1}}{m_{ij2}} = w + w_{1(i)} + w_{2(j)} + w_{12(ij)}.$$

Bei der Behandlung quantitativer Daten findet man ein Analogon im Modell der zweifachen Varianzanalyse, vgl. Kap. XI. Dieses Logit-Modell für eine 3-dimensio-

nale Kontingenztafel hat die gleiche Form wie ein loglineares Modell für eine 2-dimensionale Tafel und auch die Parameter werden entsprechend berechnet.

Beispiel: Mit den Daten aus Tab. 69 berechnen wir zunächst die in Tab. 70 aufgeführten Quotienten sowie deren Logarithmen, die in Tab. 71 zu finden sind.

Tab. 70: m_{ij1}/m_{ij2} mit den Daten aus Tab. 69

$\frac{m_{ij1}}{m_{ij2}}$	j	1	2	3
i				
1		3,083	8,533	12,125
2		7,000	9,111	9,235

Tab. 71: $\ln m_{ij1}/m_{ij2}$ mit den Daten aus Tab. 69

$\ln \frac{m_{ij1}}{m_{ij2}}$	j	1	2	3	Σ
i					
1		1,126	2,144	2,495	5,765
2		1,946	2,209	2,223	6,378
Σ		3,072	4,353	4,718	12,143

Es ist nun

$$w = \frac{1}{r \cdot s} \sum_{i,j} \ln \frac{m_{ij1}}{m_{ij2}} = \frac{1}{6} \cdot 12,143 = 2,024,$$

$$w_{1(1)} = \frac{1}{s} \sum_j \ln \frac{m_{1j1}}{m_{1j2}} - w = \frac{1}{3} \cdot 5,765 - \frac{1}{6} \cdot 12,143 = -0,102 = -w_{1(2)},$$

$$w_{2(1)} = \frac{1}{r} \sum_i \ln \frac{m_{i11}}{m_{i12}} - w = \frac{1}{2} \cdot 3,072 - \frac{1}{6} \cdot 12,143 = -0,488,$$

$$w_{2(2)} = \frac{1}{r} \sum_i \ln \frac{m_{i21}}{m_{i22}} - w = \frac{1}{2} \cdot 4,353 - \frac{1}{6} \cdot 12,143 = 0,153,$$

$$w_{2(3)} = \frac{1}{r} \sum_i \ln \frac{m_{i31}}{m_{i32}} - w = \frac{1}{2} \cdot 4,718 - \frac{1}{6} \cdot 12,143 = 0,335,$$

$$\begin{aligned} w_{12(11)} &= \ln \frac{m_{111}}{m_{112}} - w_{1(1)} - w_{2(1)} - w \\ &= 1,126 - (-0,102) - (-0,488) - 2,024 = -0,308 = -w_{12(21)}, \end{aligned}$$

$$\begin{aligned} w_{12(12)} &= \ln \frac{m_{121}}{m_{122}} - w_{1(1)} - w_{2(2)} - w \\ &= 2,144 - (-0,102) - 0,153 - 2,024 = 0,07 = -w_{12(22)}, \end{aligned}$$

$$\begin{aligned} w_{12(1,3)} &= \ln \frac{m_{131}}{m_{132}} - w_{1(1)} - w_{2(3)} - w \\ &= 2,495 - (-0,102) - 0,335 - 2,024 = 0,238 = -w_{12(23)}. \end{aligned}$$

Diese Ergebnisse sind noch einmal übersichtlich in **Tab. 72** zusammengefaßt.

Tab. 72: Parameter des Logit-Modells mit den Daten aus Tab. 69

$w_{12(ij)}$	j	1	2	3	$w_{1(i)}$
i					
1		-0,308	0,07	0,238	-0,102
2		0,308	-0,07	-0,238	0,102
$w_{2(j)}$		-0,488	0,153	0,335	$w = 2,024$

5.3. Adjustieren von Kontingenztafeln

Werden Untersuchungen über Merkmalszusammenhänge mehrfach durchgeführt, etwa in verschiedenen Jahren, an verschiedenen Orten usw., so möchte man die in Kontingenztafeln dargestellten Ergebnisse im Hinblick auf Zusammenhänge der beobachteten Merkmale direkt miteinander vergleichen können. Das wäre leicht möglich, wenn alle Tafeln die gleichen Randsummen hätten. Die dann vorhandenen Unterschiede könnten nämlich nur noch von den Wechselwirkungen herrühren, wenn man von zufälligen Effekten einmal absieht. Wir werden nun ein Verfahren vorstellen, mit dem man, ohne die Wechselwirkungen zu verändern, die Randsummen einer Kontingenztafel an vorgegebene Werte anpassen kann. Es handelt sich dabei um das Iterationsverfahren, das wir in Abschnitt 4.2.1 zur Bestimmung von ML-Schätzwerten kennengelernt haben. Hier wird jedoch als Startwert nicht eine Tafel mit identischen Häufigkeiten benutzt, sondern die Tafel der beobachteten Häufigkeiten. Die Vorgehensweise soll nun an einem Beispiel demonstriert werden.

Beispiel: Zur Untersuchung des Zusammenhangs zweier Merkmale liegen Daten aus zwei verschiedenen Jahren vor, die in **Tab. 73** und **Tab. 74** aufgeführt sind.

Tab. 73: Erhebung aus dem Jahr 1976

Merkmal 1		1	2	3	Σ
Merkmal 2					
1		178	229	68	475
2		127	362	192	681
Σ		305	591	260	1156

Die Kontingenztafel aus **Tab. 73** soll nun unter Beibehaltung der Wechselwirkungen in eine Tafel mit den Randsummen aus **Tab. 74** überführt werden. Als Startwerte für das Iterationsverfahren werden die Häufigkeiten $n_{ij}^{(0)}$ aus **Tab. 73**

Tab. 74: Erhebung aus dem Jahr 1978

		Merkmal 1	1	2	3	Σ
		Merkmal 2	1	2	3	Σ
Merkmal 2	1	362	421	212	995	
	2	256	497	381	1134	
Σ	618	918	593	2129		

genommen. Wir haben zwei Randsummen anzupassen, so daß jeder Iterationszyklus aus zwei Schritten besteht. Sind n_{ij} die Häufigkeiten aus Tab. 74, dann berechnet man im ersten Zyklus

$$n_{ij}^{(1)} = \frac{n_{ij}^{(0)}}{n_{.j}^{(0)}} \cdot n_{.j} \quad \text{und} \quad n_{ij}^{(2)} = \frac{n_{ij}^{(1)}}{n_{i.}^{(1)}} \cdot n_{i.} .$$

Dabei ergibt sich z. B. für

$$n_{12}^{(1)} = \frac{n_{12}^{(0)}}{n_{.2}^{(0)}} \cdot n_{.2} = \frac{229}{591} \cdot 918 = 355,7 \quad \text{und}$$

$$n_{23}^{(1)} = \frac{n_{23}^{(0)}}{n_{.3}^{(0)}} \cdot n_{.3} = \frac{192}{260} \cdot 593 = 437,91 .$$

Sämtliche Ergebnisse dieses 1. Schrittes vom 1. Zyklus sind in Tab. 75 aufgeführt.

Tab. 75: 1. Schritt im 1. Iterationszyklus zur Anpassung der Daten aus Tab. 73 an die Randsummen aus Tab. 74

$n_{ij}^{(1)}$	j	1	2	3	Σ
i					
1	j	360,67	355,70	155,09	871,46
	i	257,33	562,29	437,91	1257,53
Σ		618,0	917,99	593,0	2128,99

Mit diesen Werten kann nun der 2. Schritt des 1. Zyklus durchgeführt werden. Es ist z. B.

$$n_{11}^{(2)} = \frac{n_{11}^{(1)}}{n_{1.}^{(1)}} \cdot n_{1.} = \frac{360,67}{871,46} \cdot 995 = 411,8 \quad \text{und}$$

$$n_{21}^{(2)} = \frac{n_{21}^{(1)}}{n_{2.}^{(1)}} \cdot n_{2.} = \frac{257,33}{1257,53} \cdot 1134 = 232,05 .$$

Sämtliche Werte sind in Tab. 76 zu finden.

Tab. 76: 2. Schritt im 1. Iterationszyklus zur Anpassung der Daten aus Tab. 73 an die Randsummen aus Tab. 74

$n_{ij}^{(2)}$	j	1	2	3	Σ
i					
1		411,8	406,12	177,08	995
2		232,05	507,06	394,9	1134,01
Σ		643,85	913,18	571,98	2129,01

Im 2. Zyklus wird nun berechnet

$$n_{ij}^{(3)} = \frac{n_{ij}^{(2)}}{n_{\cdot j}^{(2)}} n_{\cdot j} \quad \text{und} \quad n_{ij}^{(4)} = \frac{n_{ij}^{(3)}}{n_{i \cdot}^{(3)}} n_{i \cdot} .$$

Das wiederholt man solange bis man eine Tafel mit den geforderten Randsummen erhält. In diesem Beispiel ist das praktisch schon nach dem 2. Iterationszyklus der Fall. Die Ergebnisse findet man in **Tab. 77**.

Tab. 77: Anpassung der Daten aus Tab. 73 an die Randsummen aus Tab. 74

$n_{ij}^{(4)}$	j	1	2	3	Σ
i					
1		398,4	411,5	185,1	995
2		221,2	506,2	406,6	1134
Σ		619,6	917,7	591,7	2129

Jetzt sind die Kontingenztafeln aus Tab. 74 und 77 gut zu vergleichen: Unterschiedliche Zellhäufigkeiten kann man auf veränderte Wechselwirkungen (vom Jahr 1976 auf das Jahr 1978) zurückführen.

Kapitel VIII: Vergleich zweier Meßreihen (Stichproben)

In Kap. IV haben wir Tests, Punkt- und Intervallschätzungen bei Vorliegen einer einzelnen Meßreihe (Stichprobe) kennengelernt. In diesem Kapitel wollen wir uns nun mit dem Vergleich zweier Meßreihen beschäftigen; bzgl. des Vergleichs mehrerer Meßreihen sei auf Abschnitt 1 in Kapitel XI verwiesen.

In Abschnitt 1 betrachten wir zunächst Konfidenzintervalle und Tests zum Vergleich zweier unabhängiger Meßreihen mit n_1 bzw. n_2 unabhängigen Beobachtungen bei normalverteilten Grundgesamtheiten und dann noch einige nichtparametrische Lokations- und Dispersionsvergleiche. Weiterhin wird ein Test vorgestellt, der prüft, ob zwei unabhängige Meßreihen einem gleichartigen Trend unterliegen.

In Abschnitt 2 werden wir schließlich zwei Meßreihen betrachten, die voneinander abhängig sind, und zwar in der Weise, daß jeweils eine Beobachtung der einen in direkter Beziehung zu einer Beobachtung der zweiten Meßreihe steht (etwa Sehstärke linkes Auge – rechtes Auge).

Qualitative und quantitative Analysen der Beziehungen zwischen zwei den Meßreihen zugrundeliegenden Merkmalen erfolgen mit Methoden der Korrelations- und Regressionsrechnung, vgl. Kap. IX und Kap. X. Für speziell auf diskrete Merkmale zugeschnittene Verfahren sei auf Kapitel VII hingewiesen. Insbesondere umfaßt der dortige Abschnitt 5.2 den Vergleich der Parameter mehrerer Poisson-, Binomial- und Multinomialverteilungen.

Aus Notationsgründen werden wir auch hier in den Bezeichnungen nicht immer streng zwischen Zufallsvariablen und ihren Realisationen unterscheiden.

1. Vergleich zweier unabhängiger Meßreihen

In diesem Abschnitt wollen wir der Einfachheit halber und zur Abkürzung des Abschnitts fast immer von folgendem Beispiel ausgehen:

Beispiel: Bei gleichen Umgebungsbedingungen wird an zwei verschiedenen Ampermetern die Stromstärke abgelesen. Am 1. Ampermeter wird der Versuch 10 mal und am zweiten 15 mal wiederholt. Die Messungen ergaben [in Ampère] die in Tab. 1 angegebenen Werte.

1.1. Lokationsvergleiche bei normalverteilter Grundgesamtheit

Betrachtet man die Ergebnisse x_{11}, \dots, x_{1n_1} der 1. Meßreihe als Beobachtungen aus einer $N(\mu_1, \sigma_1^2)$ -verteilten Grundgesamtheit und die Beobachtungen x_{21}, \dots, x_{2n_2} der 2. Meßreihe als aus einer $N(\mu_2, \sigma_2^2)$ -verteilten Grundgesamtheit stammende Größen, so kann man Konfidenzintervalle und Tests für die Differenz $\mu_1 - \mu_2$ der Mittelwerte der beiden Grundgesamtheiten bestimmen. Dabei müssen mehrere Fälle unterschieden werden.

1.1.1. Tests und Konfidenzintervalle bei bekannten Varianzen σ_1^2 und σ_2^2 der Grundgesamtheiten

Da in diesem Fall beide Meßreihen normalverteilten Grundgesamtheiten entstammen, wissen wir aus Abschnitt 7.2 in Kap. II, daß die Größe mit der Realisation

Tab. 1: Meßergebnisse des Stromstärkenversuchs

Nr. der Messung j	1. Ampèremeter x_{1j}	2. Ampèremeter x_{2j}
1	1,26	1,20
2	1,23	1,25
3	1,28	1,24
4	1,27	1,29
5	1,22	1,26
6	1,26	1,21
7	1,24	1,24
8	1,27	1,28
9	1,21	1,29
$n_1 = 10$	1,25	1,25
11	—	1,27
12	—	1,23
13	—	1,24
14	—	1,27
$n_2 = 15$	—	1,29

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

standardnormalverteilt ist, wobei \bar{x}_1 und \bar{x}_2 die Mittelwerte der ersten bzw. zweiten Meßreihe sind. Ein $1 - \alpha$ Konfidenzintervall für $\mu_1 - \mu_2$ bestimmt sich somit (vgl. Abschnitt 1.3.2 in Kap. IV) aus der Gleichung

$$P(u_{\alpha/2} \leq Z \leq u_{1-\alpha/2}) = 1 - \alpha$$

zu

$$\left[\bar{x}_1 - \bar{x}_2 - u_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} ; \quad \bar{x}_1 - \bar{x}_2 + u_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right].$$

Wissen wir, daß gilt

$$\sigma_1^2 = \sigma_2^2,$$

so vereinfacht sich die Berechnung dieses Konfidenzintervalls:

$$\left[\bar{x}_1 - \bar{x}_2 - u_{1-\alpha/2} \cdot \sigma_1 \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} ; \quad \bar{x}_1 - \bar{x}_2 + u_{1-\alpha/2} \cdot \sigma_1 \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right].$$

Wie stets bezeichnet u_γ hier das γ -Quantil der Standardnormalverteilung, das für verschiedene γ im Anhang vertafelt ist.

Beispiel: In unserem Beispiel ist $n_1 = 10$ und $n_2 = 15$. Die Standardabweichungen der Ampèremeter sind vom Hersteller mit $\sigma_1 = 0,021$ bzw. $\sigma_2 = 0,028$ angegeben. Wegen

$$\bar{x}_1 = \frac{1}{10} \sum_{j=1}^{10} x_{1j} = 1,249 \quad \text{und}$$

$$\bar{x}_2 = \frac{1}{15} \sum_{j=1}^{15} x_{2j} = 1,254$$

ergibt sich ein Konfidenzintervall für $\mu_1 - \mu_2$ zum Niveau $1 - \alpha = 0,95$ zu:

$$\begin{aligned} & \left[\bar{x}_1 - \bar{x}_2 - u_{0,975} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} ; \quad \bar{x}_1 - \bar{x}_2 + u_{0,975} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] \\ & = [-0,005 - 1,96 \cdot 0,0098 ; -0,005 + 1,96 \cdot 0,0098] \\ & = [-0,024 ; 0,014]. \end{aligned}$$

Wir kommen nun zum Testen von Hypothesen über die Differenz $\mu_1 - \mu_2$ bei bekannten Varianzen σ_1^2 und σ_2^2 .

Die möglichen Hypothesen der Niveau- α -Tests über die Differenz $\mu_1 - \mu_2$ bei vorgegebenem (bekanntem) Wert d sind

$$\begin{array}{lll} H_{01}: \mu_1 - \mu_2 \leq d & \text{gegen} & H_{11}: \mu_1 - \mu_2 > d, \\ H_{02}: \mu_1 - \mu_2 \geq d & \text{gegen} & H_{12}: \mu_1 - \mu_2 < d, \\ H_{03}: \mu_1 - \mu_2 = d & \text{gegen} & H_{13}: \mu_1 - \mu_2 \neq d. \end{array}$$

Ein häufiger Spezialfall ist $d = 0$, was z. B. im zweiseitigen Testproblem gerade einem Test auf Gleichheit der Mittelwerte μ_1 und μ_2 entspricht.

Wie wir aus Abschnitt 1.1 in Kap. IV wissen, ist die zufällige Größe, deren Realisation

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \left(= \frac{\bar{x}_1 - \bar{x}_2 - d}{\sigma_1 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \quad \text{falls } \sigma_1^2 = \sigma_2^2 \right)$$

ist, im Fall $\mu_1 - \mu_2 = d$ gerade $N(0,1)$ -verteilt, so daß wir hier den Einstichproben-Gaußtest auf $(\mu_1 - \mu_2)$ anwenden können. Damit ergeben sich die nachfolgenden Entscheidungsregeln:

Die Hypothese H_{01} wird verworfen, falls gilt:

$$Z > u_{1-\alpha}.$$

Die Nullhypothese H_{02} wird dann verworfen, wenn gilt:

$$Z < u_\alpha.$$

Beim zweiseitigen Testproblem H_{03} gegen H_{13} wird H_{03} verworfen, falls gilt:

$$|Z| > u_{1-\alpha/2}.$$

(Das gilt wegen der Symmetrie der Dichte der Standardnormalverteilung.)

Beispiel: Nehmen wir wie schon bei dem Beispiel zu den Konfidenzintervallen an, daß die Standardabweichungen der beiden Ampèremeter vom Hersteller mit $\sigma_1 = 0,021$ und $\sigma_2 = 0,028$ angegeben sind.

Wir wollen zum Niveau $\alpha = 0,05$ die Hypothese

$$H_{02}: \mu_1 - \mu_2 \geq 0 \quad \text{gegen} \quad H_{12}: \mu_1 - \mu_2 < 0$$

testen. Der Wert der Prüfgröße ist dann

$$Z = -0,51$$

und das 0,05-Quantil der $N(0,1)$ -Verteilung ist

$$u_{0,05} = -1,645.$$

Wegen $-0,51 > -1,645$ kann die Hypothese $H_{02}: \mu_1 - \mu_2 \geq 0$ nicht verworfen werden, d.h. μ_2 ist zum 5%-Niveau nicht signifikant größer als μ_1 .

1.1.2. Tests und Konfidenzintervalle bei unbekannten aber gleichen Varianzen σ_1^2 und σ_2^2 der beiden Grundgesamtheiten

Wissen wir lediglich, daß die Varianzen σ_1^2 und σ_2^2 gleich sind, so ist, vgl. Abschnitt 1.2.2 in Kap. IV, die Größe, deren Realisation

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ist, $t_{n_1+n_2-2}$ -verteilt. Dabei bezeichnet s_p die Stichprobenstandardabweichung:

$$s_p^2 = \frac{\sum_{i=1}^{n_1} x_{1i}^2 - \frac{1}{n_1} \left(\sum_{i=1}^{n_1} x_{1i} \right)^2 + \sum_{j=1}^{n_2} x_{2j}^2 - \frac{1}{n_2} \left(\sum_{j=1}^{n_2} x_{2j} \right)^2}{n_1 + n_2 - 2}.$$

Aus Abschnitt 1.3.2 in Kap. IV wissen wir dann, daß sich ein $1 - \alpha$ **Konfidenzintervall für die Differenz** $\mu_1 - \mu_2$ der Mittelwerte der Grundgesamtheiten aus der Gleichung

$$P(t_{n_1+n_2-2; \alpha/2} \leq t \leq t_{n_1+n_2-2; 1-\alpha/2}) = 1 - \alpha$$

bestimmen läßt. Es ergibt sich gerade

$$\left[\bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2; 1-\alpha/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} ; \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2; 1-\alpha/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

als das gewünschte Konfidenzintervall.

Beispiel: Nehmen wir an, daß in unserem Beispiel lediglich bekannt ist, daß die Varianzen beider Meßgeräte gleich sind, so ergibt sich wegen

$$\begin{aligned} s_p^2 &= \frac{\sum_{i=1}^{10} x_{1i}^2 - \frac{1}{10} \left(\sum_{i=1}^{10} x_{1i} \right)^2 + \sum_{j=1}^{15} x_{2j}^2 - \frac{1}{15} \left(\sum_{j=1}^{15} x_{2j} \right)^2}{10 + 15 - 2} \\ &= \frac{0,00489 + 0,01116}{23} = 0,0007 \quad \text{und} \end{aligned}$$

$$t_{n_1+n_2-2; 0,975} = t_{23; 0,975} = 2,069$$

ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ für die Differenz $\mu_1 - \mu_2$ zu:

$$\left[\bar{x}_1 - \bar{x}_2 - t_{23; 0,975} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} ; \quad \bar{x}_1 - \bar{x}_2 + t_{23; 0,975} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

$$= [-0,005 - 2,069 \cdot 0,011; -0,005 + 2,069 \cdot 0,011]$$

$$= [-0,027; 0,017].$$

Wie schon in Abschnitt 1.1.1 wollen wir nun die einseitigen Hypothesen

$$H_{01}: \mu_1 - \mu_2 \leq d \quad \text{gegen} \quad H_{11}: \mu_1 - \mu_2 > d,$$

$$H_{02}: \mu_1 - \mu_2 \geq d \quad \text{gegen} \quad H_{12}: \mu_1 - \mu_2 < d$$

sowie die zweiseitige Hypothese

$$H_{03}: \mu_1 - \mu_2 = d \quad \text{gegen} \quad H_{13}: \mu_1 - \mu_2 \neq d$$

zum Niveau α testen.

In Analogie zu den Aussagen über Konfidenzintervalle ist für $\mu_1 - \mu_2 = d$ die Zufallsvariable, deren Realisation

$$t = \frac{\bar{x}_1 - \bar{x}_2 - d}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

ist, $t_{n_1+n_2-2}$ -verteilt. Diese Größe t wählen wir als Prüfgröße der α -Tests über die Differenz $\mu_1 - \mu_2$ der Mittelwerte der beiden Grundgesamtheiten und kommen somit zu folgenden Entscheidungsregeln:

Die Nullhypothese H_{01} wird genau dann verworfen, wenn

$$t > t_{n_1+n_2-2; 1-\alpha} \quad \text{ist.}$$

Die Hypothese H_{02} wird verworfen, falls gilt

$$t < t_{n_1+n_2-2; \alpha},$$

und die Hypothese im zweiseitigen Testproblem H_{03} gegen H_{13} muß verworfen werden, wenn

$$|t| > t_{n_1+n_2-2; 1-\alpha/2}.$$

Beispiel: Kommen wir noch einmal zum Beispiel für das Konfidenzintervall aus diesem Abschnitt zurück. Wir wollen jetzt die Hypothese

$$H_{02}: \mu_1 - \mu_2 \geq 1$$

gegen die Alternative

$$H_{12}: \mu_1 - \mu_2 < 1$$

zum Niveau $\alpha = 0,05$ testen.

Wie wir schon berechnet haben, ist

$$s_p^2 = 0,0007,$$

so daß der Wert unserer Prüfgröße t gerade

$$t = -93,045 \quad \text{ist.}$$

Das zu bestimmende Quantil der t -Verteilung ist nun gerade

$$t_{23; 0,05} = -1,714,$$

so daß wir die Nullhypothese verwerfen müssen: Die Differenz $\mu_1 - \mu_2$ ist also zum Niveau 0,05 signifikant kleiner als 1.

1.1.3. Tests und Konfidenzintervalle bei unbekannten und ungleichen Varianzen σ_1^2 und σ_2^2 der beiden Grundgesamtheiten

In diesem Fall ist die standardisierte Zufallsvariable mit Realisation

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

approximativ t-verteilt mit (Lösungsansatz zum Behrens-Fisher-Problem)

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\left(\frac{s_1^2}{n_1} \right)^2 / (n_1 - 1) + \left(\frac{s_2^2}{n_2} \right)^2 / (n_2 - 1)}$$

Freiheitsgraden. Die Stichprobenvarianzen s_1^2 und s_2^2 der beiden Meßreihen sind dabei

$$\begin{aligned} s_1^2 &= \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 \quad \text{und} \\ s_2^2 &= \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2. \end{aligned}$$

Das (approximative) $1 - \alpha$ Konfidenzintervall für die Differenz $\mu_1 - \mu_2$ der Mittelwerte der beiden betrachteten Grundgesamtheiten läßt sich dann aus der Gleichung

$$P(t_{v, \alpha/2} \leq t \leq t_{v, 1-\alpha/2}) = 1 - \alpha$$

bestimmen, und zwar ergibt es sich zu

$$\left[\bar{x}_1 - \bar{x}_2 - t_{v, 1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad \bar{x}_1 - \bar{x}_2 + t_{v, 1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right].$$

Beispiel: In unserem Beispiel mit den Ampèremetern ergibt sich:

$$\begin{aligned} s_1^2 &= \frac{1}{9} \sum_{i=1}^{10} (x_{1i} - \bar{x}_1)^2 = 0,00054 \quad \text{und} \\ s_2^2 &= \frac{1}{14} \sum_{i=1}^{15} (x_{2i} - \bar{x}_2)^2 = 0,00080. \end{aligned}$$

Somit bestimmt sich der approximative Freiheitsgrad wegen

$$v = 21,85$$

zu 21 (der Wert von v wird stets auf die nächstkleinere ganze Zahl abgerundet).

Ein Konfidenzintervall zum Niveau $1 - \alpha = 0,9$ ergibt sich somit wegen

$$t_{v, 1-\alpha/2} = t_{21, 0,95} = 1,72$$

zu

$$\left[\bar{x}_1 - \bar{x}_2 - t_{v; 1-\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} ; \quad \bar{x}_1 - \bar{x}_2 + t_{v; 1-\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right] \\ = [-0,005 - 1,72 \cdot 0,0104; -0,005 + 1,72 \cdot 0,0104] \\ = [-0,023; 0,013].$$

Wollen wir nun α -Tests über die Differenz $\mu_1 - \mu_2$ durchführen so sind wieder folgende Testprobleme gegeben:

$$H_{01}: \mu_1 - \mu_2 \leq d \quad \text{gegen} \quad H_{11}: \mu_1 - \mu_2 > d,$$

$$H_{02}: \mu_1 - \mu_2 \geq d \quad \text{gegen} \quad H_{12}: \mu_1 - \mu_2 < d$$

und

$$H_{03}: \mu_1 - \mu_2 = d \quad \text{gegen} \quad H_{13}: \mu_1 - \mu_2 \neq d.$$

Unter $\mu_1 - \mu_2 = d$ ist natürlich die zu

$$t = \frac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

gehörige Zufallsvariable (vgl. Konfidenzintervall) approximativ t-verteilt mit v Freiheitsgraden. Wir werden also die Nullhypothese H_{01} zum Niveau α verwerfen, falls gilt:

$$t > t_{v; 1-\alpha},$$

und die Hypothese H_{02} verwerfen, falls

$$t < t_{v; \alpha}$$

ist. Im zweiseitigen Testproblem H_{03} gegen H_{13} müssen wir die Nullhypothese H_{03} verwerfen, wenn gilt:

$$|t| > t_{v; 1-\alpha/2}.$$

Beispiel: Gehen wir davon aus, daß wir nichts über die Varianzen unserer Ampèremeter wissen; wir wollen dann die Hypothese

$$H_{03}: \mu_1 - \mu_2 = -0,5$$

gegen die Alternativhypothese

$$H_{13}: \mu_1 - \mu_2 \neq -0,5$$

zum Niveau $\alpha = 0,05$ testen. Es ergibt sich wegen

$$s_1^2 = 0,00054, \quad s_2^2 = 0,00080 \quad \text{und} \quad v = 21,85 \quad (\text{vgl. Konfidenzintervall})$$

$$|t| = 47,60 \quad \text{und} \quad t_{21; 0,975} = 2,08,$$

so daß unsere Nullhypothese über die Differenz $\mu_1 - \mu_2$ verworfen werden muß.

1.1.4. Bestimmung von Stichprobenumfängen bei Tests und Konfidenzintervallen

Sind die Varianzen σ_1^2 und σ_2^2 der beiden Grundgesamtheiten bekannt, so kann man absichern, daß das Konfidenzintervall zum Niveau $1 - \alpha$ eine vorgegebene Breite d nicht überschreitet, indem man zuvor die Anzahl der Beobachtungen in beiden

Grundgesamtheiten groß genug wählt. Dabei verlangt man der Einfachheit halber, daß die Stichprobenumfänge aus beiden Grundgesamtheiten gleich groß sein sollen, also $n_1 = n_2 = n$. Da die Breite eines $1 - \alpha$ Konfidenzintervalls für $\mu_1 - \mu_2$ gerade

$$d = 2 u_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}}$$

ist, muß man

$$n \geq \left(\frac{2 u_{1-\alpha/2}}{d} \right)^2 \cdot (\sigma_1^2 + \sigma_2^2)$$

wählen. Sind σ_1^2 und σ_2^2 unbekannt, so ist diese Formel zur Bestimmung des Stichprobenumfangs n noch approximativ richtig, wenn aus Vorstichproben gute Abschätzungen für die unbekannten Varianzen vorhanden sind.

Beispiel: Das Konfidenzintervall für die Differenz $\mu_1 - \mu_2$ der Erwartungswerte zweier Grundgesamtheiten zum Niveau $1 - \alpha = 0,95$ soll höchstens die Breite $d = 2,4$ haben. Die Varianzen der Grundgesamtheiten sind, wie man weiß, gerade

$$\sigma_1^2 = 1,6 \quad \text{und} \quad \sigma_2^2 = 1,9.$$

Wieviele Beobachtungen muß man machen, damit die Breite $d = 2,4$ nicht überschritten wird?

Wir müssen in jeder Grundgesamtheit

$$n \geq \left(\frac{2 u_{0,975}}{2,4} \right)^2 \cdot (1,6 + 1,9) = \left(\frac{2 \cdot 1,96}{2,4} \right)^2 \cdot 3,5 = 9,34,$$

also mindestens 10 Beobachtungen machen.

Sind die Varianzen σ_1^2 und σ_2^2 zweier normalverteilter Grundgesamtheiten bekannt und gleich ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) oder ist aus Vorerhebungen eine genügend gute Abschätzung der gemeinsamen Varianz σ^2 (dann allerdings nur noch approximativ) vorhanden, so kann man den **nötigen Stichprobenumfang** n (je n aus beiden Grundgesamtheiten) bestimmen, um **beim Testen der Differenz** $\mu_1 - \mu_2$ der Mittelwerte der Grundgesamtheiten einen Fehler 1. Art α und einen Fehler 2. Art β nicht zu überschreiten. Will man bei den einseitigen Tests für

$$\begin{aligned} H_{01}: \mu_1 - \mu_2 &\leq d \quad \text{gegen} \quad H_{11}: \mu_1 - \mu_2 > d \\ (\text{bzw. } H_{02}: \mu_1 - \mu_2 &\geq d \quad \text{gegen} \quad H_{12}: \mu_1 - \mu_2 < d) \end{aligned}$$

einen Fehler 1. Art α und an einer festgewählten Stelle

$$d_1 > d \quad (\text{bzw. } d_1 < d)$$

einen Fehler 2. Art β absichern, so muß man

$$n \geq 2 \left(\frac{(u_{1-\alpha} + u_{1-\beta})\sigma}{d_1 - d} \right)^2$$

wählen. Im zweiseitigen Testproblem

$$H_{03}: \mu_1 - \mu_2 = d \quad \text{gegen} \quad H_{13}: \mu_1 - \mu_2 \neq d$$

kann man dagegen

$$n \geq 2 \left(\frac{(u_{1-\alpha/2} + u_{1-\beta})\sigma}{d_1 - d} \right)^2$$

wählen, um das Niveau α und den Fehler 2. Art β – an der Stelle $d_1 \neq d$ – abzusichern.

Beispiel: Wir wollen zum Niveau $\alpha = 0,05$

$$H_{02}: \mu_1 - \mu_2 \geq 0 \quad \text{gegen} \quad H_{12}: \mu_1 - \mu_2 < 0$$

testen und dabei an der Stelle $d_1 = -1$ einen Fehler 2. Art $\beta = 0,1$ absichern. Die Varianz der beiden Grundgesamtheiten ist als $\sigma = 0,5$ bekannt. Wie viele Beobachtungen aus jeder Grundgesamtheit müssen wir machen? Es ergibt sich wegen

$$n \geq \left(\frac{2(u_{0,95} + u_{0,9}) \cdot \sqrt{0,5}}{-1 - 0} \right)^2 = \left(\frac{2(1,645 + 1,282)}{-1} \right)^2 \cdot 0,5 = 17,13,$$

daß mindestens 18 Beobachtungen aus jeder Grundgesamtheit nötig sind, um die vorgegebenen Fehlerwahrscheinlichkeiten abzusichern.

1.2. Verteilungsfreie Lokationsvergleiche

Ausgehend von zwei unabhängigen Meßreihen x_{11}, \dots, x_{1n_1} und x_{21}, \dots, x_{2n_2} werden wir den Wilcoxon-Rangsummentest und den Kolmogoroff-Smirnov-Test zum Vergleich der zugrundeliegenden Verteilungen kennenlernen. Dabei werden die zugehörigen (unbekannten) Verteilungsfunktionen als stetig angenommen.

1.2.1. Der Wilcoxon-Rangsummentest, der U-Test von Mann-Whitney

Beispiel: Um die Wirkung von Vitamin B₁ auf die Wachstumsförderung von Pilzen festzustellen, wurden von 24 Pilzen 13 zufällig ausgewählte Pilze mit Vitamin B₁ behandelt, während die restlichen 11 ohne Behandlung blieben (Kontrolle). Es ergaben sich die Daten der Tab. 2.

Tab. 2: Gewicht der Pilze (vgl. Schopfer/Blumer (1943), „Zur Wirkstoffphysiologie von Trichophyton album Sab.“, Ber. Schweiz. Botan. Ges. 53, S. 409–456)

Behandlung	27	34	20,5	29,5	20	28	20	26,5	22	24,5	34	35,5	19
Kontrolle	18	14,5	13,5	12,5	23	24	21	17	18,5	9,5	14	–	–

Die Frage, die sich hier aufdrängt, lautet: Führt die Behandlung mit Vitamin B₁ zu (signifikant) höheren Gewichten?

Bezeichnen wir mit X_B die Zufallsvariable, die das Gewicht eines behandelten Pilzes angibt, und mit X_K die eines nichtbehandelten, so interessiert uns, ob X_B „im Durchschnitt größere Werte“ annimmt als X_K . Ist $F_1(x) = P(X_B \leq x)$, $F_2(x) = P(X_K \leq x)$, so stehen zur Beantwortung unserer Frage eine Stichprobe vom Umfang $n_1 = 13$, x_{11}, \dots, x_{1n_1} , die gemäß F_1 verteilt ist, und eine vom Umfang $n_2 = 11$, x_{21}, \dots, x_{2n_2} , die gemäß F_2 verteilt ist, zur Verfügung.

Nimmt X_B „im Durchschnitt größere Werte an“ als X_K , so läßt sich das so interpretieren, daß für alle x

$$P(X_B > x) \geq P(X_K > x)$$

gelten muß. Aus dieser Ungleichung folgt sofort

$$F_1(x) = P(X_B \leq x) \leq P(X_K \leq x) = F_2(x).$$

Man sagt dann auch: Die Zufallsvariable X_B ist „stochastisch größer“ als X_K . Unsere statistischen Hypothesen lauten demnach

$$H_{01}: X_B \text{ ist stochastisch kleiner als } X_K \text{ gegen}$$

$$H_{11}: X_B \text{ ist stochastisch größer als } X_K$$

oder nach den obigen Überlegungen:

$$H_{01}: F_1(x) \geq F_2(x) \text{ für alle } x \text{ gegen}$$

$$H_{11}: F_1(x) \leq F_2(x) \text{ für alle } x \text{ und für mindestens ein } x \text{ gilt}$$

$$F_1(x) < F_2(x)$$

Man setzt dabei stets voraus, daß die Verteilungen der beiden Grundgesamtheiten überhaupt vergleichbar sind, d. h. daß sich die zugehörigen Verteilungsfunktionen F_1 und F_2 nicht schneiden.

Können wir annehmen, daß der Effekt der Behandlung nur in einer Verschiebung besteht, d.h.

$$F_1(x) = F_2(x - \Delta),$$

so spricht man von einem **additiven Lokationsmodell** und Δ wird **Lokationsparameter** genannt. Die obige Identität bedeutet nämlich wegen

$$P(X_B \leq x) = F_1(x) = F_2(x - \Delta) = P(X_K \leq x - \Delta) = P(X_K + \Delta \leq x),$$

daß X_B dieselbe Verteilung wie $X_K + \Delta$ hat. Die Konstante Δ läßt sich als additiver Behandlungseffekt interpretieren und die korrespondierenden Hypothesen lauten gerade

$$H_{01}: \Delta \leq 0 \text{ gegen } H_{11}: \Delta > 0.$$

Das Modell eines additiven Behandlungseffekts – das in vielen Anwendungen zu restriktiv sein dürfte – ordnet sich dem der stochastischen Ordnung also unter.

Interessiert man sich dafür, ob überhaupt ein Behandlungseffekt besteht, so wird man

$$H_{02}: F_1 = F_2 \text{ gegen } H_{12}: F_1 \neq F_2$$

testen. Unterstellt man ein additives Modell, so erhält man analog

$$H_{02}: \Delta = 0 \text{ gegen } H_{12}: \Delta \neq 0.$$

Wie üblich bezeichnet man (H_{01}, H_{11}) als einseitige, (H_{02}, H_{12}) als zweiseitige Hypothesen. Die zu (H_{01}, H_{11}) dualen Hypothesen lauten

$$H_{03}: F_1(x) \leq F_2(x) \text{ gegen}$$

$$H_{13}: F_1(x) \geq F_2(x) \text{ und für mindestens ein } x \text{ gilt } F_1(x) > F_2(x).$$

Da wir in obigem Beispiel nichts über die konkreten Verteilungen der Grundgesamtheiten wissen, wollen wir Hypothesen über Unterschiede zwischen den Grundgesamtheiten mit Hilfe eines nichtparametrischen Tests, nämlich des **Wilcoxon-Rangsummentests** (**Wilcoxon Rank Sum Test**) überprüfen. Dieser Test benutzt nicht die Beobachtungswerte $x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}$ selbst, sondern nur die den beiden Meßreihen gemeinsam zugeordneten Rangzahlen zwischen 1 und $n_1 + n_2$:

$$r_1, r_2, \dots, r_{n_1}, r_{n_1+1}, \dots, r_{n_1+n_2}.$$

Dabei bezeichnet r_i für $i = 1, \dots, n_1$ die der Beobachtung x_{1i} (in der zusammengefaßten Stichprobe) zugeordnete Rangzahl, und für $k = 1, \dots, n_2$ ist r_{n_1+k} die der Beobachtung x_{2k} entsprechend zugeordnete Rangzahl; hier nehmen wir zunächst einmal an, daß keine Bindungen (gleiche Beobachtungswerte) vorliegen, vgl. Abschnitt 7 in Kap. III. Wie die Beobachtungen x_{ji} Realisationen von zugrundeliegenden Zufallsvariablen X_{ji} sind ($j = 1, 2; i = 1, \dots, n_j$), so sind die Rangzahlen r_i bzw. r_{n_1+k} Realisationen der zugehörigen, auf die Gesamtstichprobe bezogenen Rangstatistiken R_i bzw. R_{n_1+k} , vgl. auch Kap. III. Jede Folge von Realisationen der Rangstatistiken

$$R_1, R_2, \dots, R_{n_1}, R_{n_1+1}, \dots, R_{n_1+n_2}$$

ist nun eine Permutation der Zahlen $1, 2, \dots, n_1 + n_2$.

Ist $F_1 = F_2$, entstammen die Beobachtungen also derselben stetigen Verteilung, so wird jede solche Rangpermutation mit der gleichen Wahrscheinlichkeit auftreten:

$$P(R_1 = r_1, \dots, R_{n_1+n_2} = r_{n_1+n_2}) = \frac{1}{(n_1 + n_2)!}.$$

Ist dagegen die Verteilungsfunktion der ersten Stichprobe kleiner als die der zweiten, so wird man erwarten, daß bei der ersten Stichprobe die größeren Ränge auftreten. Es liegt also nahe, die Prüfgröße

$$\sum_{i=1}^{n_1} R_i$$

zum Testen der Hypothese H_{01} gegen H_{11} zu verwenden und dabei (zum Niveau α) H_{01} zu verwerfen, wenn die Summe der Ränge der ersten Meßreihe

$$W_{n_1, n_2} = \sum_{i=1}^{n_1} r_i > c_{1-\alpha}$$

ist. Der kritische Wert $c_{1-\alpha}$ wird hierbei festgelegt als kleinster Wert c , für den unter Annahme der Gültigkeit von $F_1 = F_2$ gilt:

$$P\left(\sum_{i=1}^{n_1} R_i > c\right) \leq \alpha,$$

d.h. der Anteil der Rangpermutationen mit

$$W_{n_1, n_2} > c_{1-\alpha}$$

soll höchstens α betragen, wenn $F_1 = F_2$ gilt. Wegen

$$\sum_{i=1}^{n_1} R_i = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - \sum_{i=n_1+1}^{n_1+n_2} R_i$$

können wir auch

$$\sum_{i=n_1+1}^{n_1+n_2} R_i$$

zum Testen verwenden; in diesem Falle sind kleine Werte signifikant.

Treten bei der Durchführung des Wilcoxon-Tests **Bindungen (ties)** auf, d.h. sind mehrere Meßwerte gleich, so wird diesen Werten das arithmetische Mittel der in

Frage kommenden Ränge zugewiesen (Midranks), vgl. auch Abschnitt 7 von Kap. III. Liegen die Bindungen alle in derselben Gruppe, wie im obigen Beispiel der Fall, so ist die Vergabe von Midranks überflüssig: es ist dann völlig gleichgültig, wie man den Bindungen die Ränge zuordnet.

Wir kommen zurück zum

Beispiel: Es ergibt sich für die Rangzahlen der Gewichte der Pilze die Tab. 3.

Tab. 3: Gewichte der Pilze (Rangzahlen in Klammern)

Behandlung	Kontrolle	Behandlung	Kontrolle
27 (19)	18 (7)	26,5 (18)	17 (6)
34 (22,5)	14,5 (5)	22 (14)	18,5 (8)
20,5 (12)	13,5 (3)	24,5 (17)	9,5 (1)
29,5 (21)	12,5 (2)	34 (22,5)	14 (4)
20 (10,5)	23 (15)	35,5 (24)	—
28 (20)	24 (16)	19 (9)	—
20 (10,5)	21 (13)		

Die kritischen Werte c_{γ} werden hier mit $W_{n_1, n_2; \gamma}$ bezeichnet, und aus der noch folgenden Tab. 4 ergibt sich, daß

$$\begin{aligned} c_{0,95} &= W_{n_1, n_2; 0,95} = n_1(n_1 + n_2 + 1) - W_{n_1, n_2; 0,05} \\ &= 13 \cdot (13 + 11 + 1) - W_{13, 11; 0,05} \\ &= 325 - 134 = 191 \end{aligned}$$

ist. Da

$$W_{n_1, n_2} = 220 > 191 = W_{n_1, n_2; 0,95}$$

gilt, ist also die Nullhypothese H_0 zum 5%-Niveau zu verwerfen, d.h. die Behandlung führt hier zu signifikanten Erfolgen.

Für größere Stichprobenumfänge bedient man sich auch der folgenden Approximation durch die Standardnormalverteilung. Unter der Annahme der Gültigkeit von $F_1 = F_2$ ist nämlich

$$E\left(\sum_{i=1}^{n_1} R_i\right) = \frac{1}{2} n_1 \cdot (n_1 + n_2 + 1),$$

$$\text{Var}\left(\sum_{i=1}^{n_1} R_i\right) = \frac{1}{12} n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)$$

und somit

$$\left(\sum_{i=1}^{n_1} R_i - \frac{1}{2} n_1 \cdot (n_1 + n_2 + 1) \right) / \sqrt{\frac{1}{12} n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}$$

approximativ standardnormalverteilt.

Auch unter Verwendung dieser Approximation muß in unserem Beispiel die Nullhypothese H_0 zum 5%-Niveau verworfen werden, denn

$$\left(220 - \frac{1}{2} \cdot 13 \cdot (13 + 11 + 1) \right) / \sqrt{\frac{1}{12} \cdot 13 \cdot 11 \cdot (13 + 11 + 1)} \\ = (220 - 162,5) / \sqrt{297,917} = 3,331 > 1,645 = u_{0,95} = u_{1-\alpha};$$

die Quantile u_γ der $N(0,1)$ -Verteilung sind im Anhang vertafelt.

Wir wollen die Vorgehensweise bei der Durchführung des Wilcoxon-Rangsummentests noch einmal zusammenfassend darstellen.

Sind x_1, \dots, x_{n_1} und x_{21}, \dots, x_{n_2} zwei unabhängige Meßreihen bzw. Stichproben, deren zugehörige Grundgesamtheiten die Verteilungsfunktionen $F_1(x)$ und $F_2(x)$ besitzen, so soll getestet werden:

$H_{01}: F_1(x) \geq F_2(x)$ gegen $H_{11}: F_1(x) \leq F_2(x)$, „ $<$ “ für mindestens ein x

$H_{02}: F_1 = F_2$ gegen $H_{12}: F_1 \neq F_2$

$H_{03}: F_1(x) \leq F_2(x)$ gegen $H_{13}: F_1(x) \geq F_2(x)$, „ $>$ “ für mindestens ein x .

Den Stichprobenwerten insgesamt werden die Ränge r_i zugeschrieben, und die Summe der Ränge in der ersten Stichprobe berechnet:

$$W_{n_1, n_2} = \sum_{i=1}^{n_1} r_i.$$

Die Entscheidungsregeln für den Wilcoxon-Rangsummentest zum Niveau α sind dann:

(a) H_{01} wird verworfen, falls gilt:

$$W_{n_1, n_2} > W_{n_1, n_2; 1-\alpha},$$

(b) H_{02} wird verworfen, falls

$$W_{n_1, n_2} < W_{n_1, n_2; \alpha/2} \quad \text{oder}$$

$$W_{n_1, n_2} > W_{n_1, n_2; 1-\alpha/2} \quad \text{gilt, und}$$

(c) H_{03} wird verworfen, falls gilt:

$$W_{n_1, n_2} < W_{n_1, n_2; \alpha}.$$

Kritische Werte $W_{n_1, n_2; \alpha}$ sind für $\alpha = 0,05$ und $\alpha = 0,10$ bis $n_1, n_2 = 20$ in der Tabelle 4 vertafelt. Die oberen kritischen Werte erhält man durch

$$W_{n_1, n_2; 1-\alpha} = n_1(n_1 + n_2 + 1) - W_{n_1, n_2; \alpha}.$$

Hat man keine kritischen Werte für W_{n_1, n_2} , so verwende man das folgende approximative Verfahren:

Mittels der standardisierten Größe

$$T = \frac{W_{n_1, n_2} - \frac{1}{2}n_1(n_1 + n_2 + 1)}{\sqrt{\frac{1}{12}n_1 n_2(n_1 + n_2 + 1)}}$$

gehe man nach folgenden Entscheidungsregeln für einen Test zum Niveau α vor, wobei u_γ das γ -Quantil der Standardnormalverteilung bezeichnet,

(a) H_{01} wird verworfen, falls $T > u_{1-\alpha}$,

(b) H_{02} wird verworfen, falls $|T| > u_{1-\alpha/2}$,

(c) H_{03} wird verworfen, falls $T < u_\alpha$.

Tab. 4: Kritische Werte $W_{n_1, n_2; \alpha}$ des Wilcoxon-Rangsummen-Tests für $\alpha = 0,05$ (jeweils 1. Zeile) und für $\alpha = 0,10$ (2. Zeile). Es ist $W_{n_1, n_2; 1-\alpha} = n_1(n_1 + n_2 + 1) - W_{n_1, n_2; \alpha}$.

$n_1 \backslash n_2$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	3	3	3	4	4	4	5	5	5	5	6	6	7	7	7	7	8	8	
	3	4	4	5	5	6	6	7	7	8	8	8	9	9	10	10	11	11	
3	6	7	7	8	9	9	10	11	11	12	12	13	14	14	15	16	16	17	18
	7	8	8	9	10	11	12	12	13	14	15	16	17	17	18	19	20	21	22
4	10	11	12	13	14	15	16	17	18	19	20	21	22	23	25	26	27	28	29
	11	12	14	15	16	17	18	20	21	22	23	24	26	27	28	29	31	32	33
5	16	17	18	20	21	22	24	25	27	28	29	31	32	34	35	36	38	39	41
	17	18	20	21	23	24	26	28	29	31	33	34	36	38	39	41	43	44	46
6	22	24	25	27	29	30	32	34	36	38	39	41	43	45	47	48	50	52	54
	23	25	27	29	31	33	35	37	39	41	43	45	47	49	51	53	56	58	60
7	29	31	33	35	37	40	42	44	46	48	50	53	55	57	59	62	64	66	68
	30	33	35	37	40	42	45	47	50	52	55	57	60	62	65	67	70	72	75
8	38	40	42	45	47	50	52	55	57	60	63	65	68	70	73	76	78	81	84
	39	42	44	47	50	53	56	59	61	64	67	70	73	76	79	82	85	88	91
9	47	50	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97	100
	48	51	55	58	61	64	68	71	74	77	81	84	87	91	94	98	101	104	108
10	57	60	63	67	70	73	76	80	83	87	90	93	97	100	104	107	111	114	118
	59	62	66	69	73	77	80	84	88	92	95	99	103	107	110	114	118	122	126
11	68	72	75	79	83	86	90	94	98	101	105	109	113	117	121	124	128	132	136
	70	74	78	82	86	90	94	98	103	107	111	115	119	124	128	132	136	140	145
12	81	84	88	92	96	100	105	109	113	117	121	126	130	134	139	143	147	151	156
	83	87	91	96	100	105	109	114	118	123	128	132	137	142	146	151	156	160	165
13	94	98	102	107	111	116	120	125	129	134	139	143	148	153	157	162	167	172	176
	96	101	105	110	115	120	125	130	135	140	145	150	155	160	166	171	176	181	186
14	109	113	117	122	127	132	137	142	147	152	157	162	167	172	177	183	188	193	198
	110	116	121	126	131	137	142	147	153	158	164	169	175	180	186	191	197	203	208
15	124	128	133	139	144	149	154	160	165	171	176	182	187	193	198	204	209	215	221
	126	131	137	143	148	154	160	166	172	178	184	189	195	201	207	213	219	225	231
16	140	145	151	156	162	167	173	179	185	191	198	204	211	217	223	230	236	243	244
	142	148	154	160	166	173	179	185	191	198	204	211	217	223	230	236	243	249	256
17	157	163	169	174	180	187	193	199	205	211	218	224	231	237	243	250	256	263	269
	160	166	172	179	185	192	199	206	212	219	226	233	239	246	253	260	267	274	281
18	176	181	188	194	200	207	213	220	227	233	240	247	254	260	267	274	281	288	295
	178	185	192	199	206	213	220	227	234	241	249	256	263	270	278	285	292	300	307
19	195	201	208	214	221	228	235	242	249	256	263	271	278	285	292	300	307	314	321
	198	205	212	219	227	234	242	249	257	264	272	280	288	295	303	311	319	326	334
20	215	222	229	236	243	250	258	265	273	280	288	295	303	311	318	326	334	341	349
	218	226	233	241	249	257	265	273	281	289	297	305	313	321	330	338	346	354	362

Beispiel: Wir wollen auch hier die an unseren Ampèremetern vorgenommenen Messungen betrachten (In 1.1 hatten wir hier eine Normalverteilung unterstellt, der Kürze halber wollen wir jedoch dieses Beispiel auch hier verwenden.), vgl. Tab. 1.

Geprüft werden soll, ob signifikante Unterschiede zwischen den Ampèremetern bestehen. Das Testproblem, das wir zum Niveau $\alpha = 0,10$ betrachten wollen, ist also

$$H_{02}: F_1 = F_2 \quad \text{gegen} \quad H_{12}: F_1 \neq F_2.$$

Zunächst ordnen wir die 25 Meßwerte der Größe nach und verteilen die Ränge, vgl. Tab. 5.

Tab. 5: Arbeitstabelle zum Wilcoxon-Rangsummentest im Beispiel

Meßreihe	Meßwert	Rang
2	1,20	1
2	1,21	2,5
1	1,21	2,5
1	1,22	4
1	1,23	5,5
2	1,23	5,5
1	1,24	8,5
2	1,24	8,5
2	1,24	8,5
2	1,24	8,5
1	1,25	12
2	1,25	12
2	1,25	12
1	1,26	15
1	1,26	15
2	1,26	15
1	1,27	18,5
1	1,27	18,5
2	1,27	18,5
2	1,27	18,5
1	1,28	21,5
2	1,28	21,5
2	1,29	24
2	1,29	24
2	1,29	24

Es ergibt sich hier

$$W_{n_1, n_2} = W_{10, 15} = \sum (\text{Ränge in der ersten Meßreihe}) = 121.$$

Wir können also wegen

$$W_{n_1, n_2; \alpha/2} = W_{10, 15; 0,05} = 100,$$

$$W_{n_1, n_2; 1 - \alpha/2} = n_1(n_1 + n_2 + 1) - W_{n_1, n_2; \alpha/2} = 10(10 + 15 + 1) - 100 \\ = 160$$

zum 10%-Niveau keine signifikanten Unterschiede zwischen den Ampèremetern feststellen.

Eine andere plausible Möglichkeit für die Behandlung dieses Testproblems be-

stunde darin, die Prüfgröße

$$U = \text{Anzahl aller Paare } (x_{1i}, x_{2j}) \text{ mit } x_{1i} < x_{2j}$$

zu benutzen. Wegen

$$W_{n_1, n_2} = n_1 \cdot n_2 + \frac{1}{2} n_1 (n_1 + 1) - U$$

sind die beiden Teststatistiken äquivalent. Die Größe U nennt man die **Mann-Whitney-Statistik** und daher wird der hier betrachtete Test auch als **U-Test von Mann-Whitney** bezeichnet.

1.2.2. Der Kolmogoroff-Smirnov-Test

Für die Testprobleme des Abschnitts 1.2.1 liegt es nahe, die empirischen Verteilungsfunktionen der beiden Stichproben für die Ermittlung einer Prüfgröße heranzuziehen. Intuitiv einleuchtend als Maß für die Übereinstimmung ist der maximale Abstand der empirischen Verteilungsfunktionen. Es bezeichne $S_{n_1, 1}$ die empirische Verteilungsfunktion von x_{11}, \dots, x_{1n_1} , $S_{n_2, 2}$ die von x_{21}, \dots, x_{2n_2} .

Dabei ist für $i = 1, 2$

$$S_{n_i, i}(x) = \begin{cases} 0, & x < x_{ij} \text{ für alle } j = 1, \dots, n_i \\ \frac{k}{n_i}, & \text{falls } x \geq x_{ij} \text{ für genau } k \text{ der } j \text{ aus } \{1, \dots, n_i\} \\ 1, & x \geq x_{ij} \text{ für alle } j = 1, \dots, n_i. \end{cases}$$

Der größte vertikale Abstand von $S_{n_1, 1}$ und $S_{n_2, 2}$, wenn erstere Funktion oberhalb der zweiten liegt, sei bezeichnet mit

$$D_{n_1, n_2}^+(1, 2) = \max_x (S_{n_1, 1}(x) - S_{n_2, 2}(x));$$

entsprechend sei

$$D_{n_1, n_2}^-(1, 2) = D_{n_2, n_1}^+(2, 1) = \max_x (S_{n_2, 2}(x) - S_{n_1, 1}(x)), \text{ und}$$

$$D_{n_1, n_2}(1, 2) = \max_x |S_{n_1, 1}(x) - S_{n_2, 2}(x)| = \max(D_{n_1, n_2}^+(1, 2), D_{n_1, n_2}^-(1, 2)).$$

Sind $F_1(x)$ und $F_2(x)$ wieder die unbekannten Verteilungsfunktionen der Grundgesamtheiten aus denen die erste und zweite Stichprobe stammen, so werden die ein- und zweiseitigen Hypothesen

$$H_{01}: F_1(x) \leqq F_2(x) \text{ gegen } H_{11}: F_1(x) \geqq F_2(x) \text{ und } „>“ \text{ für mindestens ein } x,$$

$$H_{02}: F_1(x) = F_2(x) \text{ gegen } H_{12}: F_1(x) \neq F_2(x) \text{ für mindestens ein } x$$

und

$$H_{03}: F_1(x) \geqq F_2(x) \text{ gegen } H_{13}: F_1(x) \leqq F_2(x) \text{ und } „<“ \text{ für mindestens ein } x$$

zum Niveau α mittels der folgenden Entscheidungsregeln getestet.

Ist $n_1 = n_2 = n$, so wird

H_{01} verworfen, falls gilt

$$Z_{2,n}^+ > z_{2,n; 1-2\alpha}, \quad \text{mit } Z_{2,n}^+ = n D_{n,n}^+(1,2);$$

H_{02} verworfen, falls gilt

$$Z_{2,n} > z_{2,n; 1-\alpha}, \quad \text{mit } Z_{2,n} = n D_{n,n}(1,2);$$

H_{03} verworfen, falls gilt

$$Z_{2,n}^- > z_{2,n; 1-2\alpha}, \quad \text{mit } Z_{2,n}^- = n D_{n,n}^-(1,2).$$

Dabei sind die kritischen Werte $z_{2,n; 1-\alpha}$ für einige α in Tabelle 6 so vertafelt, daß in der entsprechenden α -Spalte die Sprungstellen n zu dem in der ersten Spalte zu findenden $z_{2,n; 1-\alpha}$ angegeben sind. Für $n > 40$ sind in der letzten Zeile derselben α -Spalte die kritischen Werte $z_{2,n; 1-\alpha}$ angegeben. Für nicht angegebene $n \leq 40$ gilt der kritische Wert vom nächstkleineren aufgeführten n .

Tab. 6: Kritische Werte $z_{2,n; 1-\alpha}$ des Kolmogoroff-Smirnov-Tests

$z_{2,n; 1-\alpha}$	$n \leq 40$				
	$\alpha = 0,20$	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,02$	$\alpha = 0,01$
2	3	3			
3	4	4	4		
4	7	6	5	5	5
5	11	9	7	6	6
6	16	13	10	9	8
7	22	17	14	11	10
8	28	22	18	15	13
9	36	28	23	18	16
10		34	28	22	20
11			33	27	24
12			40	32	28
13				37	33
14					38

$n > 40$	$\sqrt{n} \cdot 1,527$	$\sqrt{n} \cdot 1,739$	$\sqrt{n} \cdot 1,923$	$\sqrt{n} \cdot 2,150$	$\sqrt{n} \cdot 2,305$
$z_{2,n; 1-\alpha}$					

Beispiel:

$$n = 11, \quad \alpha = 0,02, \quad \text{so gilt } z_{2,n; 1-\alpha} = z_{2,11; 0,98} = 7;$$

$$n = 25, \quad \alpha = 0,10, \quad \text{so gilt } z_{2,n; 1-\alpha} = z_{2,25; 0,90} = 8 = z_{2,22; 0,90};$$

$$n = 70, \quad \alpha = 0,05, \quad \text{so gilt } z_{2,n; 1-\alpha} = z_{2,70; 0,95} = \sqrt{70} \cdot 1,923 = 16,1.$$

Ist $n_1 \neq n_2$, so wird die Hypothese

H_{01} verworfen, falls gilt

$$K_{n_1, n_2}^+ > k_{n_1, n_2; 1-2\alpha}, \quad \text{mit } K_{n_1, n_2}^+ = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \cdot D_{n_1, n_2}^+(1,2);$$

H_{02} verworfen, falls gilt

$$K_{n_1, n_2} > k_{n_1, n_2; 1-\alpha}, \quad \text{mit} \quad K_{n_1, n_2} = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \cdot D_{n_1, n_2}(1, 2);$$

H_{03} verworfen, falls gilt

$$K_{n_1, n_2}^- > k_{n_1, n_2; 1-2\alpha}, \quad \text{mit} \quad K_{n_1, n_2}^- = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \cdot D_{n_1, n_2}^-(1, 2).$$

Einige approximative (konservative) kritische Werte $k_{n_1, n_2; 1-\alpha}$ sind für einige α in Tab. 7 angegeben; dabei bezeichnet $n_{(1)}$ die kürzere und $n_{(2)}$ die längere Meßreihe: Zum Beispiel für:

Tab. 7: Approximative kritische Werte $k_{n_1, n_2; 1-\alpha}$ des Kolmogoroff-Smirnov-Tests bei ungleichen Reihenlängen

$\alpha = 0,20$		
$n_{(1)}$	$n_{(2)}$	$k_{n_{(1)}, n_{(2)}; 1-\alpha}$
von 2 bis 4	von 5 bis 40	1,02
von 5 bis 15	von 5 bis 40	1,03
	sonst	1,08

$\alpha = 0,10$		
$n_{(1)}$	$n_{(2)}$	$k_{n_{(1)}, n_{(2)}; 1-\alpha}$
von 2 bis 3	von 3 bis 12	1,10
von 4 bis 8	von 5 bis 9	1,12
von 4 bis 16	von 10 bis 20	1,16
	sonst	1,23

$\alpha = 0,05$		
$n_{(1)}$	$n_{(2)}$	$k_{n_{(1)}, n_{(2)}; 1-\alpha}$
von 2 bis 4	von 3 bis 15	1,22
von 5 bis 16	von 6 bis 20	1,30
	sonst	1,36

$$n_1 = 5, \quad n_2 = 10, \quad \alpha = 0,05 \quad \text{erhält man} \quad k_{n_1, n_2; 1-\alpha} = 1,30;$$

$$n_1 = 7, \quad n_2 = 4, \quad \alpha = 0,10 \quad \text{erhält man} \quad k_{n_1, n_2; 1-2\alpha} = 1,02;$$

$$n_1 = 20, \quad n_2 = 18, \quad \alpha = 0,05 \quad \text{erhält man} \quad k_{n_1, n_2; 1-\alpha} = 1,36.$$

Der Kolmogoroff-Smirnov-Test ist exakt, wenn die zugrundeliegenden Verteilungsfunktionen $F_1(x)$ und $F_2(x)$ stetig sind; andernfalls bleibt er zwar gültig, ist jedoch konservativ, d.h. er bevorzugt eine Entscheidung zugunsten der jeweiligen Nullhypothese.

Beispiel: Für unsere Meßreihen an den Ampèremetern erstellen wir zunächst eine Tabelle der empirischen Verteilungsfunktionen sowie der Werte $S_{10,1}(x) - S_{15,2}(x)$, $S_{15,2}(x) - S_{10,1}(x)$ und $|S_{10,1}(x) - S_{15,2}(x)|$, vgl. Tab. 8 und Abb. 1.

Tab. 8: Arbeitstabelle zum Kolmogoroff-Smirnov-Test

x_{1j}	x_{2j}	$S_{10,1}(x)$	$S_{15,2}(x)$	$S_{10,1}(x) - S_{15,2}(x)$	$S_{15,2}(x) - S_{10,1}(x)$	$ S_{10,1}(x) - S_{15,2}(x) $
	1,20	0	1/15	-1/15	1/15	1/15
1,21	1,21	1/10	2/15	-1/30	1/30	1/30
1,22		2/10	2/15	2/30	-2/30	2/30
1,23	1,23	3/10	3/15	3/30	-3/30	3/30
1,24	1,24	4/10	6/15	0	0	0
1,25	1,25	5/10	8/15	-1/30	1/30	1/30
1,26	1,26	7/10	9/15	3/30	-3/30	3/30
1,27	1,27	9/10	11/15	5/30	-5/30	5/30
1,28	1,28	1	12/15	6/30	-6/30	6/30
1,29		1	1	0	0	0

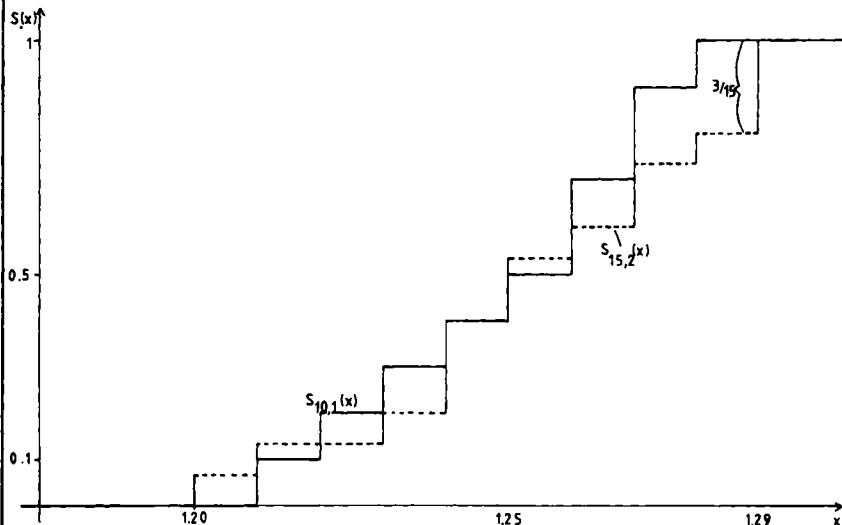


Abb. 1: Empirische Verteilungsfunktionen der an zwei Ampèremetern gemessenen Stromstärken und Prüfgröße des Kolmogoroff-Smirnov-Tests

Wir wollen nun zum Niveau $\alpha = 0,05$ die Hypothese, daß keine Unterschiede zwischen den beiden Ampèremetern bestehen, also

$$H_{02}: F_1 = F_2,$$

gegen die Alternative

$$H_{12}: F_1 \neq F_2$$

testen. Der Wert der Teststatistik ergibt sich zu

$$\begin{aligned} K_{n_1, n_2} = K_{10, 15} &= \sqrt{\frac{10 \cdot 15}{10 + 15}} \cdot \max_x |S_{10,1}(x) - S_{15,2}(x)| = \sqrt{\frac{150}{25}} \cdot \frac{3}{15} \\ &= 0,49 < 1,30 = k_{10, 15; 0,95}, \end{aligned}$$

so daß wir die Hypothese nicht verwerfen können.

Wenn nur Ranginformationen vorliegen, kann man mit dem Kolmogoroff-Smirnov-Test überprüfen, ob die Verteilungen der Ränge gleich sind.

Die Verwendung des Kolmogoroff-Smirnov-Tests ist insbesondere dann angezeigt, wenn man über die Art der Alternative (z. B. Lagealternative) überhaupt keine Vorstellungen hat. Er wird allgemein auch als **Omnibus-Test** bezeichnet.

1.3. Dispersionvergleiche bei normalverteilten Grundgesamtheiten- Tests und Konfidenzintervalle

In diesem Abschnitt wollen wir ein Konfidenzintervall und einen Test für den Quotienten σ_1^2/σ_2^2 der Varianzen zweier unabhängiger Meßreihen aus normalverteilten Grundgesamtheiten angeben. Stammen die Beobachtungen der ersten Meßreihe x_{11}, \dots, x_{1n_1} aus einer $N(\mu_1, \sigma_1^2)$ -Verteilung und die der zweiten Meßreihe x_{21}, \dots, x_{2n_2} aus einer $N(\mu_2, \sigma_2^2)$ -Verteilung, so wissen wir aus Abschnitt 1.3 in Kap. IV, daß der Quotient

$$F = \frac{s_1^2 \sigma_2^2}{s_2^2 \sigma_1^2}$$

gerade die Realisation einer F_{n_1-1, n_2-1} -verteilten Zufallsvariablen ist. Daher können wir aus der Gleichung

$$P(F_{n_1-1, n_2-1; \alpha/2} \leq F \leq F_{n_1-1, n_2-1; 1-\alpha/2}) = 1 - \alpha$$

ein $1 - \alpha$ Konfidenzintervall für den Quotienten $\frac{\sigma_1^2}{\sigma_2^2}$ der Varianzen zweier Meßreihen bestimmen. Es ergibt sich:

$$\left[\frac{s_1^2}{s_2^2} F_{n_2-1, n_1-1; \alpha/2}; \frac{s_1^2}{s_2^2} F_{n_2-1, n_1-1; 1-\alpha/2} \right],$$

wobei sich die Stichprobenvarianzen s_1^2 und s_2^2 in üblicher Weise zu

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2,$$

$$s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2$$

bestimmen lassen.

Beispiel: Wir bestimmen für den Quotienten σ_1^2/σ_2^2 der Varianzen zweier Ampèremeter, vgl. Tab. 1, ein Konfidenzintervall zum Niveau 0,95. Es ist

$$s_1^2 = 0,00054 \quad \text{und} \quad s_2^2 = 0,00080,$$

$$F_{n_2-1, n_1-1; \alpha/2} = \frac{1}{F_{9, 14; 0, 975}} = \frac{1}{3,21} = 0,3115$$

und

$$F_{n_2-1, n_1-1; 1-\alpha/2} = F_{14, 9; 0, 975} = 3,79.$$

Damit ergibt sich das gesuchte Konfidenzintervall zu

$$\left[\frac{s_1^2}{s_2^2} \cdot F_{n_2-1, n_1-1; \alpha/2}; \frac{s_1^2}{s_2^2} \cdot F_{n_2-1, n_1-1; 1-\alpha/2} \right] = [0,210; 2,558].$$

Wir wollen nun zum Testen von Hypothesen über den Quotienten der Varianzen zweier unabhängiger Meßreihen kommen. Von Interesse sind bei diesen α -Tests über σ_1^2/σ_2^2 die Hypothesen

$$H_{01}: \frac{\sigma_1^2}{\sigma_2^2} \leq 1 \quad (\text{d.h. } \sigma_1^2 \leq \sigma_2^2) \quad \text{gegen} \quad H_{11}: \frac{\sigma_1^2}{\sigma_2^2} > 1,$$

$$H_{02}: \frac{\sigma_1^2}{\sigma_2^2} \geq 1 \quad (\text{d.h. } \sigma_1^2 \geq \sigma_2^2) \quad \text{gegen} \quad H_{12}: \frac{\sigma_1^2}{\sigma_2^2} < 1,$$

sowie

$$H_{03}: \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad (\text{d.h. } \sigma_1^2 = \sigma_2^2) \quad \text{gegen} \quad H_{13}: \frac{\sigma_1^2}{\sigma_2^2} \neq 1.$$

Wie wir schon gesehen haben ist die Zufallsvariable mit Realisation

$$F = \frac{s_1^2 \sigma_2^2}{s_2^2 \sigma_1^2}$$

F_{n_1-1, n_2-1} -verteilt, so daß im Falle der Gültigkeit von $\sigma_1^2 = \sigma_2^2$ der Quotient der Stichprobenvarianzen $\frac{s_1^2}{s_2^2}$ Realisation einer F_{n_1-1, n_2-1} - verteilten Zufallsvariablen ist.

Testen wir also zum Niveau α die Nullhypothese H_{01} , so verwerfen wir diese, falls gilt:

$$\frac{s_1^2}{s_2^2} > F_{n_1-1, n_2-1; 1-\alpha}.$$

Wir müssen H_{02} zum Niveau α verwerfen, wenn gilt

$$\frac{s_1^2}{s_2^2} < F_{n_1-1, n_2-1; \alpha},$$

und wir müssen die Hypothese H_{03} verwerfen, falls gilt

$$\frac{s_1^2}{s_2^2} < F_{n_1-1, n_2-1; \alpha/2} \quad \text{oder} \quad \frac{s_1^2}{s_2^2} > F_{n_1-1, n_2-1; 1-\alpha/2}.$$

Beispiel: Wir wollen testen, ob die Varianz des zweiten Ampèremeters, vgl. auch Tab. 1, zum Niveau $\alpha = 0,01$ signifikant größer ist als die des ersten Ampèremeters, d.h. wir testen

$$H_{02}: \frac{\sigma_1^2}{\sigma_2^2} \geq 1 \quad \text{gegen} \quad H_{12}: \frac{\sigma_1^2}{\sigma_2^2} < 1.$$

Wegen

$$\frac{s_1^2}{s_2^2} = \frac{0,00054}{0,00080} = 0,675 \quad \text{und}$$

$$F_{n_1-1, n_2-1; \alpha} = F_{9, 14; 0,01} = \frac{1}{F_{14, 9; 0,99}} = \frac{1}{5,00} = 0,2$$

können wir die Nullhypothese, daß das zweite Ampèremeter eine kleinere Varianz hat als das erste, nicht verwerfen.

1.4 Verteilungsfreie Dispersionsvergleiche

In Abschnitt 1.2 hatten wir als Verteilungsfreie Tests zum Vergleich zweier Behandlungen den Wilcoxon-Rangsummentest und den Kolmogoroff-Smirnov-Zweistichprobentest kennengelernt. Der Kolmogoroff-Smirnov-Test wurde dort als Omnibus-Test bezeichnet, der darauf angelegt ist, jede Abweichung zwischen den Verteilungen der beiden Stichproben zu entdecken. Der Wilcoxon-Test eignet sich besonders gut für die Entdeckung eines (im wesentlichen) additiven Behandlungseffekts. Das Ergebnis einer Behandlung kann sich aber auch in der Veränderung der Variabilität der Stichprobe ausdrücken, ohne daß eine Lokationsverschiebung stattfindet. Tests, die darauf ausgerichtet sind, solche Unterschiede zu entdecken, nennt man **Dispersionstests**. Ihre Verwendung ist angezeigt, wenn eine Behandlung zu einer Veränderung der Dispersion führt oder man sich für den Vergleich der Variabilität zweier unabhängiger Stichproben interessiert.

Wir werden für den Fall (bis auf eine additive Konstante) bekannter bzw. gleicher Lokationsparameter in beiden Stichproben den **Ansari-Bradley-Freund-Test** und für den Fall unbekannter eventuell verschiedener Lokationsparameter den **Test von Moses** betrachten.

1.4.1. Der Test von Ansari-Bradley-Freund, der Siegel-Tukey-Test

Wir gehen aus von zwei Stichproben $x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}$, die gemäß Verteilungsfunktionen $F_1(x)$ bzw. $F_2(x)$ verteilt sind. Weiterhin unterstellen wir, daß eine Verteilungsfunktion $F(x)$ existiert mit

$$F_1(x) = F\left(\frac{x - \mu_1}{\sigma_1}\right), \quad F_2(x) = F\left(\frac{x - \mu_2}{\sigma_2}\right).$$

Es wird also vorausgesetzt, daß F_1 und F_2 den gleichen „shape“ haben.

Ist $\mu_1 - \mu_2 = \theta$ bekannt, so bilden wir

$$x'_{1i} = x_{1i} - \theta \quad x'_{2i} = x_{2i}$$

und erhalten die den x'_{1i} bzw. x'_{2i} zugrundeliegende Verteilungsfunktion

$$F\left(\frac{x - \mu_2}{\sigma_1}\right) \quad \text{bzw.} \quad F\left(\frac{x - \mu_2}{\sigma_2}\right).$$

Ist $\mu_1 = \mu_2 = \mu$, so ist $F_1(x) = F\left(\frac{x - \mu}{\sigma_1}\right)$ und $F_2(x) = F\left(\frac{x - \mu}{\sigma_2}\right)$.

Die interessierenden Nullhypotesen und Alternativen sind in beiden Fällen

$$H_{01}: \sigma_1 \geq \sigma_2 \quad \text{gegen} \quad H_{11}: \sigma_1 < \sigma_2$$

$$H_{02}: \sigma_1 \leq \sigma_2 \quad \text{gegen} \quad H_{12}: \sigma_1 > \sigma_2$$

$$H_{03}: \sigma_1 = \sigma_2 \quad \text{gegen} \quad H_{13}: \sigma_1 \neq \sigma_2$$

Wir werden die Testprozedur hier nur für den Fall gleicher, aber unbekannter μ_1 und μ_2 vorführen. Im anderen Fall wird dieselbe Prozedur mit den transformierten Daten x'_{1i} und x'_{2i} durchgeführt. Zunächst ordnen wir die $n_1 + n_2$ Beobachtungen

Tab. 9: Vergabe der Ränge zum Ansari-Bradley-Freund-Test

Rangzahlen R_i für	geordnete Beobachtungen									
	$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n_1+n_2-1)} \leq x_{(n_1+n_2)}$									
$n_1 + n_2$ gerade	1	2	$\frac{n_1 + n_2}{2}$	2	1			
$n_1 + n_2$ ungerade	1	2	$\frac{n_1 + n_2 - 1}{2}$	$\frac{n_1 + n_2 + 1}{2}$	$\frac{n_1 + n_2 - 1}{2}$	2	1		

unserer beiden Meßreihen der Größe nach und verteilen Ränge (treten Bindungen auf, so müssen die zugehörigen Ränge ermittelt werden), wie in **Tab. 9** angezeigt.

Nun berechnen wir die Größe

$$A = \sum_{i=1}^{n_1+n_2} c_i R_i \quad \text{mit} \quad c_i = \begin{cases} 0, & \text{falls } x_{(i)} \text{ zur 2. Meßreihe gehört} \\ 1, & \text{falls } x_{(i)} \text{ zur 1. Meßreihe gehört} \end{cases}$$

Wir verwerfen dann die Nullhypothese H_{01} , falls gilt

$$A \geq a_2(\alpha, n_2, n_1)$$

und die Nullhypothese H_{02} , falls gilt

$$A \leq a_1(\alpha, n_2, n_1).$$

Die Hypothese H_{03} verwerfen wir schließlich, falls

$$A \leq a_1(\alpha/2, n_2, n_1) \quad \text{oder} \quad A \geq a_2(\alpha/2, n_2, n_1).$$

Die kritischen Werte $a_1(\alpha, n_2, n_1)$ und $a_2(\alpha, n_2, n_1)$ des Ansari-Bradley-Freund-Tests sind z. B. bei Hollander/Wolfe (1973) vertafelt. Wir wollen hier lediglich eine Approximation angeben: analoger Vergleich mit den Quantilen der Standardnormalverteilung ($a_2(\alpha, n_2, n_1) \approx u_{1-\alpha}$, $a_1(\alpha, n_2, n_1) \approx u_\alpha$). Als Teststatistik müssen wir dabei folgende Größen verwenden:

(a) falls keine Bindungen auftreten

$$A^* = \begin{cases} \frac{A - n_1(n_1 + n_2 + 2)/4}{(n_2 n_1 (n_1 + n_2 + 2) (n_1 + n_2 - 2)/(48(n_1 + n_2 + 1)))^{1/2}}, & \text{für } n_1 + n_2 \text{ gerade} \\ \frac{A - n_1(n_1 + n_2 + 1)^2/(4(n_1 + n_2))}{(n_2 n_1 (n_1 + n_2 + 1) (3 + (n_1 + n_2)^2)/(48(n_1 + n_2)^2))^{1/2}}, & \text{für } n_1 + n_2 \text{ ungerade} \end{cases}$$

(b) falls Bindungen auftreten mit

g = Anzahl der gebundenen Gruppen; vgl. Kap. III und Kap. IV, Abschnitt 5.4

t_j = Anzahl der Elemente der j -ten gebundenen Gruppe

r_j = mittlerer Rang der j -ten gebundenen Gruppe

$$A^{**} = \begin{cases} \frac{A - n_1(n_1 + n_2 + 2)/4}{[n_2 n_1 (16 \sum_{j=1}^g t_j r_j^2 - (n_1 + n_2)(n_1 + n_2 + 2)^2)/(16(n_1 + n_2)(n_1 + n_2 - 1))]^{1/2}}, & \text{falls } n_1 + n_2 \text{ gerade} \\ \frac{A - n_1(n_1 + n_2 + 1)^2/(4(n_1 + n_2))}{[n_2 n_1 (16(n_1 + n_2) \sum_{j=1}^g t_j r_j^2 - (n_1 + n_2 + 1)^4)/(16(n_1 + n_2)^2(n_1 + n_2 - 1))]^{1/2}}, & \text{falls } n_1 + n_2 \text{ ungerade.} \end{cases}$$

Beispiel: Wir können davon ausgehen, daß die Mediane der Meßwerte an zwei Ampèremetern, vgl. Tab. 1, gleich sind. Daher ist der Ansari-Bradley-Freund-Test anwendbar. Zunächst verteilen wir die Ränge, vgl. Tab. 10.

Tab. 10: Arbeitstabelle zum Ansari-Bradley-Freund-Test

Gruppe	Meßwert	Rang
2	1,20	1
2	1,21	2,5
1	1,21	2,5
1	1,22	4
1	1,23	5,5
2	1,23	5,5
1	1,24	8,5
2	1,24	8,5
2	1,24	8,5
2	1,24	8,5
1	1,25	12
2	1,25	12
2	1,25	12
1	1,26	11
1	1,26	11
2	1,26	11
1	1,27	7,5
1	1,27	7,5
2	1,27	7,5
2	1,27	7,5
1	1,28	4,5
2	1,28	4,5
2	1,29	2
2	1,29	2
2	1,29	2

Wir wollen hier die Hypothese, daß die Dispersionen der beiden Ampèremeter gleich sind, zum Niveau $\alpha = 0,05$ testen, d.h. wir testen

$$H_{03}: \frac{\sigma_2}{\sigma_1} = 1 \quad \text{gegen} \quad H_{13}: \frac{\sigma_2}{\sigma_1} \neq 1.$$

Dabei wollen wir einmal die Approximation für große Stichproben verwenden (diese wird wegen $n_1 = 10, n_2 = 15$ natürlich noch nicht besonders gut sein).

Da Bindungen auftreten, müssen wir zunächst g , t_j und r_j bestimmen:

g = „Anzahl verschiedener Meßwerte“ = 10 und

$$t_1 = 1, t_2 = 2, t_3 = 1, t_4 = 2, t_5 = 4, t_6 = 3, t_7 = 3, t_8 = 4, t_9 = 2, t_{10} = 3,$$

$$r_1 = 1, r_2 = 2,5, r_3 = 4, r_4 = 5,5, r_5 = 8,5, r_6 = 12, r_7 = 11, r_8 = 7,5,$$

$$r_9 = 4,5, r_{10} = 2.$$

Damit ist

$$\begin{aligned} \sum_{j=1}^{10} t_j r_j^2 &= 1 + 12,5 + 16 + 60,5 + 289 + 432 + 363 + 225 + 40,5 + 12 \\ &= 1451,5, \end{aligned}$$

und es ist

$$\begin{aligned} A &= \sum_{i=1}^{25} c_i R_i = 2,5 + 4 + 5,5 + 8,5 + 12 + 11 + 11 + 7,5 + 7,5 + 4,5 \\ &= 74. \end{aligned}$$

Da $n_1 + n_2 = 25$ ungerade ist, bestimmt sich die Prüfgröße des Tests zu

$$\begin{aligned} A^{**} &= \frac{74 - 10 \cdot 26^2 / (4 \cdot 25)}{[15 \cdot 10 [16 \cdot 25 \cdot 1451,5 - 26^4] / (16 \cdot 25^2 \cdot 24)]^{1/2}} \\ &= \frac{74 - 67,6}{[150 \cdot 123624 / 240000]^{1/2}} = \frac{6,4}{\sqrt{77,265}} = 0,728. \end{aligned}$$

Damit ist

$$|A^{**}| = 0,728 < u_{1-\alpha/2} = u_{0,975} = 1,96,$$

so daß wir die Hypothese „beide Ampèremeter haben gleiche Dispersion“ nicht verwerfen können.

Vergibt man an den kleinsten Wert der vereinigten Stichprobe den Rang 1, an den größten den Rang 2, an den zweitgrößten den Rang 3, an den zweitkleinsten den Rang 4, usw., also

$$1 \ 4 \ 5 \ 8 \ 9 \ 12 \dots 11 \ 10 \ 7 \ 6 \ 3 \ 2,$$

und bildet die Statistik

$$W_{n_1, n_2} = \text{Summe der Ränge in der ersten Meßreihe},$$

so erhält man den **Siegel-Tukey-Test** zur Prüfung obiger Hypothesen; dessen Prüfverteilung ist gerade die des Wilcoxon-Rangsummentests, vgl. Abschnitt 1.2.

1.4.2. Der Test von Moses

Weiß man nichts über die Lokationsparameter μ_1 und μ_2 der beiden Grundgesamtheiten und möchte Aussagen über den Quotienten σ_1/σ_2 machen, so bietet sich der Quasi-Rangtest von Moses an.

Die unabhängigen Beobachtungen zweier unabhängiger Meßreihen x_{11}, \dots, x_{1n_1} und x_{21}, \dots, x_{2n_2} müssen dabei aus Grundgesamtheiten mit stetigen Verteilungen stammen. Bei der Prüfung der Hypothesen

$$H_{01}: \frac{\sigma_1^2}{\sigma_2^2} \leq 1 \quad \text{gegen} \quad H_{11}: \frac{\sigma_1^2}{\sigma_2^2} > 1,$$

$$H_{02}: \frac{\sigma_1^2}{\sigma_2^2} \geq 1 \quad \text{gegen} \quad H_{12}: \frac{\sigma_1^2}{\sigma_2^2} < 1$$

und

$$H_{03}: \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad \text{gegen} \quad H_{13}: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

gehen wir wie folgt vor:

Wir wählen uns eine beliebige Zahl $k \geq 2$, so daß n_1 und n_2 ganzzahlige Vielfache von k sind, und unterteilen dann die n_1 bzw. n_2 Beobachtungen rein zufällig in n'_1 bzw. n'_2 Untergruppen à k Beobachtungen. In diesen $n'_1 + n'_2$ Untergruppen bestimmen wir dann das $k - 1$ -fache der empirischen Varianzen:

$$s_{11} = \sum_{i=1}^k (x_{1i} - \bar{x}_{1(1)})^2, \dots, s_{1n'_1} = \sum_{i=n_1-k+1}^{n_1} (x_{1i} - \bar{x}_{1(n'_1)})^2,$$

$$s_{21} = \sum_{i=1}^k (x_{2i} - \bar{x}_{2(1)})^2, \dots, s_{2n'_2} = \sum_{i=n_2-k+1}^{n_2} (x_{2i} - \bar{x}_{2(n'_2)})^2,$$

wobei die Mittelwerte $\bar{x}_{1(j)}$ bzw. $\bar{x}_{2(j)}$ gerade über die Beobachtungswerte in den jeweiligen Gruppen gebildet werden.

Auf diese $n'_1 + n'_2$ Werte $s_{11}, \dots, s_{1n'_1}, s_{21}, \dots, s_{2n'_2}$ wenden wir dann den in Abschnitt 1.2.1 beschriebenen Rangsummentest von Wilcoxon an (mit $x_{1i} \leftrightarrow s_{1i}$, $x_{2i} \leftrightarrow s_{2i}$).

Beispiel: (vgl. Bughi et al. (1969), „A measurement of sodium and potassium in erythrocytes and whole blood“, Clin. Chem. 15, S. 712–719).

In leichter Abwandlung einer in oben erwähnter Arbeit beschriebenen Methode wollen wir Aussagen über den Quotienten σ_1^2/σ_2^2 der Varianzen von Sodium-Ionen-Gehalt in Blutplasma und roten Blutkörperchen machen. Die Meßergebnisse aus Tab. 11 liegen vor. Diese jeweils 30 Beobachtungen unterteilen wir in $n'_1 = n'_2 = 6$ Untergruppen à 5 Beobachtungen. Die Zuteilung ist ebenfalls Tab. 11 zu entnehmen.

Wir wollen nun zum Niveau $\alpha = 0,10$ die Hypothese

$$H_{02}: \frac{\sigma_1^2}{\sigma_2^2} \geq 1 \quad \text{gegen} \quad H_{12}: \frac{\sigma_1^2}{\sigma_2^2} < 1$$

testen. Wir berechnen zunächst die Werte s_{ij} :

$$\begin{array}{ll} s_{11} = 7,2 & s_{21} = 176,06 \\ s_{12} = 175,3 & s_{22} = 119,248 \\ s_{13} = 25,2 & s_{23} = 278,348 \\ s_{14} = 234,5 & s_{24} = 299,768 \\ s_{15} = 19,2 & s_{25} = 139,888 \\ s_{16} = 290,0 & s_{26} = 301,768 \end{array}$$

Tab. 11: Sodium-Ionen-Gehalt in Blutplasma und roten Blutkörperchen

i	x_{1i} (Plasma)	Gruppe	x_{2i} (rote Blutkörperchen)	Gruppe
1	147,0	1	10,3	3
2	147,0	5	11,0	5
3	144,0	3	10,7	6
4	147,0	4	12,2	2
5	142,0	2	16,0	3
6	145,0	1	19,4	1
7	146,0	6	16,5	2
8	144,0	1	19,2	5
9	144,0	3	16,5	1
10	145,0	5	19,3	4
11	146,0	4	9,7	3
12	162,5	4	12,6	6
13	146,5	2	8,3	2
14	163,0	6	9,7	2
15	147,0	1	8,5	4
16	161,0	4	15,2	5
17	145,0	5	31,4	4
18	145,5	3	26,8	3
19	141,0	5	27,0	5
20	144,0	6	31,4	6
21	142,0	3	32,2	1
22	146,5	1	26,3	3
23	145,0	6	12,7	4
24	157,0	6	17,2	5
25	145,0	5	17,5	4
26	154,5	2	16,9	1
27	148,5	2	19,0	6
28	153,5	4	21,7	2
29	139,0	3	11,4	6
30	137,0	2	17,5	1

und wenden darauf den Rangsummentest von Wilcoxon, vgl. Abschnitt 1.2.1, an. Die Summe der Rangzahlen in der ersten Gruppe und somit der Wert der Teststatistik des Rangsummentests von Wilcoxon ist

$$W_{n'_1, n'_2} = 1 + 2 + 3 + 6 + 8 + 10 = 30.$$

Da

$$W_{n'_1, n'_2; \alpha} = W_{6, 6; 0, 10} = 31$$

ist, vgl. Tab. 4, muß die Hypothese H_{02} verworfen werden, d.h. die Varianz der ersten Meßreihe ist zum 10%-Niveau signifikant kleiner als die der zweiten.

Weitere Verfahren für den Vergleich von Dispersionen von p Meßreihen sind in Kap. XI zu finden. Dort findet man z.B. den Levene-Test, für p = 2 auch Pfanzagl-Test genannt.

1.5 Test auf Trend

Liegen zwei unabhängige Meßreihen mit zeitlich geordneten Beobachtungen x_{11}, \dots, x_{1n_1} und x_{21}, \dots, x_{2n_2} vor, so interessiert man sich häufig dafür, ob in den Meßreihen ein Aufwärtstrend vorhanden ist. Man möchte also die Nullhypothese

H_0 : beide Reihen haben keinen Aufwärtstrend
zum Niveau α testen.

Ein natürliches Vorgehen besteht dann darin, die Beobachtungen in jeder der beiden Reihen (der Größe nach $x_{i(1)} \leq \dots \leq x_{i(n_i)}$ für $i = 1, 2$) zu ordnen und ihnen die Rangzahlen $1, \dots, n_i$ zuzuweisen. Für jede Reihe berechnet man dann die Teststatistik:

$$D_i = \frac{1}{3} n_i (n_i + 1) (2n_i + 1) - 2 \sum_{j=1}^{n_i} j R_{ij} \quad \text{für } i = 1, 2$$

(R_{ij} bezeichnet hier die Rangzahl der j -ten Beobachtung in der i -ten Meßreihe.) sowie deren Erwartungswert unter der Nullhypothese

$$E(D_i) = \frac{n_i^3 - n_i}{6}$$

und deren Varianz unter H_0

$$\text{Var}(D_i) = \frac{n_i^2 (n_i + 1)^2 (n_i - 1)}{36}.$$

Unter der Annahme, daß die Reihen trendlos sind, gilt approximativ

$$\begin{aligned} & D_1 + D_2 - \frac{n_1^3 - n_1}{6} - \frac{n_2^3 - n_2}{6} \\ & \sqrt{\frac{n_1^2 (n_1 + 1)^2 (n_1 - 1)}{36} + \frac{n_2^2 (n_2 + 1)^2 (n_2 - 1)}{36}} \\ & = \frac{D_1 + D_2 - \frac{1}{6} (n_1^3 - n_1 + n_2^3 - n_2)}{\frac{1}{6} \sqrt{n_1^2 (n_1 + 1)^2 (n_1 - 1) + n_2^2 (n_2 + 1)^2 (n_2 - 1)}} = Z \end{aligned}$$

ist standardnormalverteilt, so daß wir die Nullhypothese verwerfen müssen, falls gilt

$$Z < u_\alpha.$$

Ein Test auf Abwärtstrend ergibt sich gerade, wenn man die kleinste Rangzahl dem größten Beobachtungswert, ..., die größte Rangzahl dem kleinsten Beobachtungswert zuordnet.

Beispiel: (vgl. Vital Statistics-Special Reports 47(8) (1960): „Illegitimate birth in the United States 1938–1957“.)

Wir wollen überprüfen, ob die Zahl der unehelichen Geburten in den Staaten Alabama und Louisiana in den Jahren 1938 bis 1957 signifikant angestiegen ist. Die Zahlen der unehelichen Geburten und die zugehörigen Rangzahlen sind in Tab. 12 zusammengestellt. Der Test soll für $\alpha = 0,10$ mittels Normalapproximation durchgeführt werden, obwohl relativ wenige Beobachtungen vorliegen. Wegen

$$D_1 = 182 - 150 = 32 \text{ und } D_2 = 182 - 178 = 4$$

ergibt sich

Tab. 12: Uneheliche Geburten (und Rangzahlen) in Alabama und Louisiana

i	1	2	3	4	5	6
Jahr	1940	1945	1947	1950	1955	1957
x_{1j} (Alabama)	660	745	724	642	661	740
R_{1j}	2	6	4	1	3	5
x_{2j} (Louisiana)	520	517	598	560	691	726
R_{2j}	2	1	4	3	5	6

$$Z = \frac{32 + 4 - 35 - 35}{\frac{1}{6}\sqrt{8820 + 8820}} = \frac{-34}{22,136} = -1,54.$$

Weiter ist

$$u_\alpha = u_{0,10} = -1,282 > Z,$$

und wir müssen die Nullhypothese, daß die Anzahl der unehelichen Geburten in beiden Staaten keinen Aufwärtstrend aufweist, verwerfen.

Analog läßt sich dieser Test auch auf $p > 2$ Meßreihen erweitern.

2. Vergleich zweier abhängiger Meßreihen

Wir kommen nun zu solchen Situationen, wo zwar die Beobachtungen innerhalb jeder Meßreihe voneinander unabhängig sind, jedoch nicht die beiden Meßreihen. Vielmehr sind hier die Meßreihen so untereinander gekoppelt, daß gerade jeder Beobachtung der einen Meßreihe genau eine Beobachtung der anderen Meßreihe zugeordnet werden kann.

Beispiel: Die Sehstärke von n Personen wird am linken und rechten Auge festgestellt. Die erste Meßreihe entspricht den Beobachtungen am linken, die zweite Meßreihe den Beobachtungen am rechten Auge. Die Sehstärken linkes-rechtes Auge werden natürlich bei einer Person nicht unabhängig voneinander sein.

Hat man solche Meßreihen vorliegen, so wird man die natürlichen Beobachtungspaare betrachten. Für zwei auf die beschriebene Art und Weise voneinander abhängigen Meßreihen mit je n Beobachtungen wollen wir im folgenden Lokations- und Dispersionsvergleiche durchführen. Um den Abschnitt nicht zu lang werden zu lassen, werden wir, obwohl das in der Praxis nicht zu empfehlen ist, alle Verfahren an folgendem Beispiel erläutern.

Beispiel: Zwei verschiedene Schlafmittel werden an $n = 15$ Patienten eines Krankenhauses auf ihre Wirksamkeit untersucht.

Am ersten Tag erhalten die Patienten kein Medikament, am zweiten Tag das erste Schlafmittel und am dritten Tag das zweite Schlafmittel. Beobachtet werden für beide Schlafmittel die Abweichungen von der Schlafdauer am 1. Tag. Die gemessenen Abweichungen in Stunden sind in **Tab. 13** zusammengestellt.

Die natürlichen Beobachtungspaare sind hier natürlich die Messungen an der i -ten Person bei Schlafmittel 1 und 2.

Tab. 13: Wirksamkeitsprüfung von 2 Schlafmitteln an 15 Personen

Patient (i)	Medikament 1 (x_{1i})	Medikament 2 (x_{2i})	Diff. ($d_i = x_{1i} - x_{2i}$)
1	+3,6	-0,4	+4,0
2	+1,2	+0,6	+0,6
3	+0,3	+4,2	-3,9
4	-0,8	+1,7	-2,5
5	+4,2	-3,5	+7,7
6	-1,3	-2,1	+0,8
7	-2,6	+2,4	-5,0
8	+3,2	+0,9	+2,3
9	+0,9	-0,9	+1,8
10	-0,3	+0,1	-0,4
11	-0,1	+2,7	-2,8
12	0,0	-0,7	+0,7
13	+2,5	+3,4	-0,9
14	-1,1	-1,8	+0,7
15	-0,7	+1,3	-2,0

2.1. Lokationsvergleiche bei normalverteilten Grundgesamtheiten – Tests, Konfidenzintervalle und Bestimmung der Stichprobenumfänge

Wir wollen hier Tests und Konfidenzintervalle für die Differenz $\mu_1 - \mu_2$ der Mittelwerte zweier Grundgesamtheiten angeben. Die Beobachtungen mögen dabei in oben beschriebener Weise voneinander abhängig sein. Dabei unterstellen wir, daß die Beobachtungen der ersten Meßreihe einer $N(\mu_1, \sigma_1^2)$ -Verteilung, die der zweiten Meßreihe einer $N(\mu_2, \sigma_2^2)$ -Verteilung entstammen. Die n Beobachtungspaare sollen alle die gleiche Kovarianz σ_{12} haben. Wir betrachten die sogenannten **gepaarten Differenzen**

$$d_i = x_{1i} - x_{2i} \quad \text{für } i = 1, \dots, n,$$

die dann natürlich aus einer $N(\mu_1 - \mu_2, \sigma_d^2)$ -Verteilung stammen. Dabei ist die Varianz σ_d^2 der gepaarten Differenzen gerade

$$\sigma_d^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$$

(vgl. Kap. II, Abschnitt 8.2.A und 8.4).

Wollen wir nun Tests und Konfidenzintervalle für die Differenzen $\mu_1 - \mu_2$ angeben, so müssen wir zwei Fälle unterscheiden:

- A. Die Varianz σ_d^2 ist bekannt.
- B. Die Varianz σ_d^2 ist unbekannt.

A. Die Varianz σ_d^2 ist bekannt

Kennen wir die Varianz σ_d^2 , d.h. kennen wir die Varianz beider Grundgesamtheiten sowie die Kovarianz der natürlichen Beobachtungspaare, so wissen wir aus Abschnitt 1 in Kap. IV, daß die Größe mit Realisation

$$Z = \frac{\bar{d} - (\mu_1 - \mu_2)}{\sigma_d} \sqrt{n}, \quad \text{wobei } \bar{d} = \frac{1}{n} \sum_{i=1}^n d_i,$$

standardnormalverteilt ist. Wir können also bei der Bestimmung eines $1 - \alpha$ Konfidenzintervalls für die Differenz $\mu_1 - \mu_2$ der Mittelwerte unserer Grundgesamtheiten genauso vorgehen, wie im Einstichprobenfall:

Aus der Gleichung

$$P(u_{\alpha/2} \leq Z \leq u_{1-\alpha/2}) = 1 - \alpha$$

können wir das gesuchte $1 - \alpha$ Konfidenzintervall zu

$$\left[\bar{d} - \frac{\sigma_d}{\sqrt{n}} u_{1-\alpha/2}, \bar{d} + \frac{\sigma_d}{\sqrt{n}} u_{1-\alpha/2} \right]$$

bestimmen.

Beispiel: Ist uns bekannt, daß die gepaarten Differenzen d_1, \dots, d_{15} der Schlafabweichungen unserer 15 Patienten, vgl. Tab. 13, die Varianz $\sigma_d^2 = 9$ haben, so ist wegen

$$\bar{d} = \frac{1}{15} \sum_{i=1}^n d_i = 0,073$$

das Intervall

$$\begin{aligned} \left[\bar{d} - \frac{\sigma_d}{\sqrt{n}} u_{0,975}, \bar{d} + \frac{\sigma_d}{\sqrt{n}} u_{0,975} \right] &= [0,073 - 1,518; 0,073 + 1,518] \\ &= [-1,445; 1,591] \end{aligned}$$

ein Konfidenzintervall zum Niveau 0,95 für die Differenz $\mu_1 - \mu_2$.

Interessieren wir uns nun dafür, ob die Differenz $\mu_1 - \mu_2$ in einer gewissen Beziehung zu einem festen Wert μ_0 steht, d.h. wollen wir die Hypothesen

$$\begin{array}{ll} H_{01}: \mu_1 - \mu_2 \leq \mu_0 & \text{gegen } H_{11}: \mu_1 - \mu_2 > \mu_0, \\ H_{02}: \mu_1 - \mu_2 \geq \mu_0 & \text{gegen } H_{12}: \mu_1 - \mu_2 < \mu_0, \\ H_{03}: \mu_1 - \mu_2 = \mu_0 & \text{gegen } H_{13}: \mu_1 - \mu_2 \neq \mu_0 \end{array}$$

überprüfen, so müssen wir einen Test zum Niveau α für die Differenz der Mittelwerte $\mu_1 - \mu_2$ durchführen. Ein häufiger Spezialfall ist in der Praxis

$$\mu_0 = 0,$$

also zum Beispiel im zweiseitigen Testproblem H_{03} gegen H_{13} die Prüfung der Hypothese, ob die Mittelwerte beider Grundgesamtheiten gleich sind.

Wie wir aus Abschnitt 1.1 in Kap. IV wissen, ist

$$Z = \frac{\bar{d} - \mu_0}{\sigma_d} \sqrt{n}$$

bei Gültigkeit von $\mu_1 - \mu_2 = \mu_0$ Realisation einer standardnormalverteilten Zufallsvariablen, so daß wir den Einstichproben-Gaußtest hier auf die gepaarten Differenzen anwenden können.

Wir müssen demnach die Hypothese H_{01} verwerfen, wenn

$$Z > u_{1-\alpha},$$

und die Hypothese H_{02} verwerfen, falls

$$Z < u_\alpha$$

ist. Im zweiseitigen Problem verwerfen wir die Hypothese H_{03} im Falle

$$|Z| > u_{1-\alpha/2}.$$

Beispiel: Wir wollen in unserem Beispiel, wo wir wiederum $\sigma_d^2 = 9$ als bekannt annehmen wollen, die Hypothese

$$H_{01}: \mu_1 - \mu_2 \leq -2$$

gegen die Alternative

$$H_{11}: \mu_1 - \mu_2 > -2$$

zum Niveau $\alpha = 0,01$ testen. Es ist wegen $\bar{d} = 0,073$

$$Z = \frac{0,073 + 2}{\sqrt{15}} = 2,676.$$

Da

$$u_{1-\alpha} = u_{0,99} = 2,33$$

ist, muß die Hypothese verworfen werden. Zum Niveau 0,01 liefert der Test also, daß die Schlafverlängerung bei Mittel 2 signifikant weniger als zwei Stunden gegenüber Mittel 1 beträgt.

B. Die Varianz σ_d^2 ist unbekannt

Ist σ_d^2 unbekannt, so müssen wir σ_d^2 aus unseren Beobachtungen schätzen:

$$s_d^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2.$$

Wie wir in Abschnitt 1 von Kap.IV schon gesehen haben, ist dann

$$t = \frac{\bar{d} - (\mu_1 - \mu_2)}{s_d} \sqrt{n}$$

die Realisation einer t-verteilten Zufallsvariablen mit $n-1$ Freiheitsgraden. Hiervon ausgehend, können wir nun ein $(1-\alpha)$ -Konfidenzintervall für die Differenz $\mu_1 - \mu_2$ der Mittelwerte unserer Grundgesamtheiten – d.h. für den Mittelwert der gepaarten Differenzen – aus der Gleichung

$$P(t_{n-1; \alpha/2} \leq t \leq t_{n-1; 1-\alpha/2}) = 1 - \alpha$$

bestimmen. Es ergibt sich gerade das Konfidenzintervall

$$\left[\bar{d} - \frac{s_d}{\sqrt{n}} t_{n-1; 1-\alpha/2}, \bar{d} + \frac{s_d}{\sqrt{n}} t_{n-1; 1-\alpha/2} \right].$$

Beispiel: Wir wollen ein Konfidenzintervall zum Niveau $1-\alpha = 0,99$ für die Differenz $\mu_1 - \mu_2$, also für die Unterschiede unserer beiden Schläfmittel bestimmen, vgl. Tab.13. Wie wir aus Abschnitt A bereits wissen, ist

$$\bar{d} = 0,073.$$

Hiermit läßt sich die Varianz σ_d^2 durch

$$s_d^2 = \frac{1}{14} \sum_{i=1}^{15} (d_i - 0,073)^2 = 10,356$$

schätzen. Das gesuchte Konfidenzintervall ergibt sich nun wegen

$$t_{n-1; 1-\alpha/2} = t_{14; 0,995} = 2,977$$

zu

$$\left[0,073 - \sqrt{\frac{10,356}{15}} 2,977; 0,073 + \sqrt{\frac{10,356}{15}} 2,977 \right] = [-2,401; 2,547].$$

Nun wollen wir unter der Annahme, daß σ_d^2 unbekannt ist, Tests zum Niveau α über die Differenz $\mu_1 - \mu_2$ der Mittelwerte zweier abhängiger Populationen herleiten. Unsere Hypothesen haben dabei die gleiche Gestalt wie in Abschnitt A. Da die Teststatistik mit Realisation

$$t = \frac{\bar{d} - \mu_0}{s_d} \sqrt{n}$$

bei Gültigkeit von $\mu_1 - \mu_2 = \mu_0$ gerade t_{n-1} -verteilt ist, können wir hier den schon aus Abschnitt 1 in Kap. IV bekannten Einstichproben-t-Test auf die gepaarten Differenzen d_1, \dots, d_n anwenden, d.h. wir müssen die Hypothese $H_{01}: \mu_1 - \mu_2 \leq \mu_0$ zum Niveau α verwerfen, falls gilt:

$$t > t_{n-1; 1-\alpha}.$$

Wir müssen weiter $H_{02}: \mu_1 - \mu_2 \geq \mu_0$ verwerfen, falls

$$t < t_{n-1; \alpha}$$

ist, und wir müssen die Nullhypothese $H_{03}: \mu_1 - \mu_2 = \mu_0$ verwerfen, wenn

$$|t| > t_{n-1; 1-\alpha/2}$$

ist.

Beispiel: Kommen wir wieder zu den beiden Schlafmitteln, vgl. Tab. 13, zurück. Wir testen zum Niveau $\alpha = 0,05$ die Hypothese

$$H_{02}: \mu_1 - \mu_2 \geq 0,5 \quad \text{gegen} \quad H_{12}: \mu_1 - \mu_2 < 0,5.$$

Mit

$$\bar{d} = 0,073 \quad \text{und}$$

$$s_d^2 = 10,356,$$

ergibt sich

$$t = \frac{0,073 - 0,5}{\sqrt{10,356}} \cdot \sqrt{15} = -0,514.$$

Wegen

$$t_{n-1; \alpha} = t_{14; 0,05} = -1,761$$

können wir unsere Hypothese zu diesem Niveau nicht verwerfen.

Natürlich läßt sich auch bei abhängigen Meßreihen der Stichprobenumfang n bestimmen, um etwa eine vorgegebene Breite b von Konfidenzintervallen für die Differenz $\mu_1 - \mu_2$ der Erwartungswerte einhalten zu können oder beim Testen der Differenz $\mu_1 - \mu_2$ vorgegebene α - und β -Fehler abzusichern. Wir können hier alle

in Kapitel IV, Abschnitt 1.4, beschriebenen Verfahren anwenden, wenn wir nicht mehr die Beobachtungen selbst, sondern die Differenzen

$$x_{1i} - x_{2i} = d_i$$

betrachten.

Beispiel: Zur Einhaltung der Breite $b = 1,2$ des Konfidenzintervalls zum Niveau $1 - \alpha = 0,95$ für die Differenz $\mu_1 - \mu_2$ der Mittelwerte der beiden normalverteilten Grundgesamtheiten müssen wir – bei bekannter Varianz $\sigma_d^2 = 0,7$ der gepaarten Differenzen –

$$n \geq \left(\frac{2u_{1-\alpha/2} \cdot \sigma_d}{b} \right)^2 = \left(\frac{2 \cdot u_{0,975}}{1,2} \right)^2 \cdot 0,7 = \left(\frac{2 \cdot 1,96}{1,2} \right)^2 \cdot 0,7 = 7,47$$

Beobachtungspaare – also mindestens $n = 8$ Beobachtungspaare – aus den Grundgesamtheiten betrachten.

2.2. Dispersionsvergleiche bei normalverteilten Grundgesamtheiten

In diesem Abschnitt wollen wir Konfidenzintervalle und Tests für den unbekannten Streuungsparameter σ_d^2 der Verteilung $N(\mu_1 - \mu_2, \sigma_d^2)$ der gepaarten Differenzen zweier abhängiger Meßreihen angeben. Dabei können wir natürlich wieder auf die in Kapitel IV, Abschnitt 1 vorgestellten Verfahren zurückgreifen. Wie wir wissen ist der zufällige Quotient mit der Realisation

$$\chi^2 = \frac{(n-1)s_d^2}{\sigma_d^2} \quad \left(\text{bzw. } \chi^2 = \frac{\sum_{i=1}^n (d_i - \mu_1 + \mu_2)^2}{\sigma_d^2} \right)$$

χ^2 -verteilt mit $n-1$ (bzw. n) Freiheitsgraden. Daher können wir aus der Gleichung

$$P(\chi_{n-1; \alpha/2}^2 \leq \chi^2 \leq \chi_{n-1; 1-\alpha/2}^2) = 1 - \alpha$$

$$(\text{bzw. } P(\chi_{n; \alpha/2}^2 \leq \chi^2 \leq \chi_{n; 1-\alpha/2}^2) = 1 - \alpha)$$

ein $1 - \alpha$ Konfidenzintervall für die unbekannte Varianz σ_d^2 bestimmen. Es ergibt sich gerade zu

$$\left[\frac{(n-1)s_d^2}{\chi_{n-1; 1-\alpha/2}^2}; \frac{(n-1)s_d^2}{\chi_{n-1; \alpha/2}^2} \right] \quad \left(\text{bzw. } \left[\frac{\sum_{i=1}^n (d_i - \mu_1 + \mu_2)^2}{\chi_{n-1; 1-\alpha/2}^2}; \frac{\sum_{i=1}^n (d_i - \mu_1 + \mu_2)^2}{\chi_{n; \alpha/2}^2} \right] \right)$$

Beispiel: Wir wollen wissen, in welchem Bereich die Varianz σ_d^2 der gepaarten Differenzen bei unserem Versuch mit den beiden Schlafmitteln mit 98%-tiger Sicherheit liegt, d.h. wir suchen ein 0,98 Konfidenzintervall für σ_d^2 . Dabei kennen wir in diesem Fall den Wert $\mu_1 - \mu_2$ nicht. Wegen $s_d^2 = 10,356$,

$$\chi_{n-1; \alpha/2}^2 = \chi_{14; 0,01}^2 = 4,660 \quad \text{und} \quad \chi_{n-1; 1-\alpha/2}^2 = \chi_{14; 0,99}^2 = 29,141$$

ergibt sich

$$\left[\frac{14 \cdot 10,356}{29,141}; \frac{14 \cdot 10,356}{4,660} \right] = [4,975; 31,112]$$

gerade als das gesuchte Intervall.

Wollen wir nun **Tests zum Niveau α über die Varianz σ_d^2** für die Testprobleme

$$\begin{aligned} H_{01}: \sigma_d^2 &\leq \sigma_0^2 & \text{gegen } H_{11}: \sigma_d^2 > \sigma_0^2, \\ H_{02}: \sigma_d^2 &\geq \sigma_0^2 & \text{gegen } H_{12}: \sigma_d^2 < \sigma_0^2 \end{aligned}$$

und

$$H_{03}: \sigma_d^2 = \sigma_0^2 \quad \text{gegen} \quad H_{13}: \sigma_d^2 \neq \sigma_0^2$$

herleiten (σ_0^2 ist eine zu wählende Zahl), so kommt uns folgende Tatsache zugute:
Falls $\sigma_d^2 = \sigma_0^2$ gültig ist, so ist

$$\chi^2 = \frac{(n - 1)s_d^2}{\sigma_0^2}$$

die Realisation einer χ_{n-1}^2 -verteilten Zufallsgröße.
(Ist $\mu_1 - \mu_2$ bekannt, so verwendet man hier

$$\chi^2 = \frac{\sum_{i=1}^n (d_i - \mu_1 + \mu_2)^2}{\sigma_0^2} ;$$

die zugrundeliegende Zufallsvariable ist χ^2 -verteilt mit n Freiheitsgraden.)

Wir wissen somit, daß wir die Hypothese H_{01} verwerfen, falls

$$\chi^2 > \chi_{n-1; 1-\alpha}^2 \quad (\text{bzw. } \chi^2 > \chi_{n; 1-\alpha}^2)$$

ist, und daß wir die Hypothese H_{02} verwerfen müssen, wenn gilt

$$\chi^2 < \chi_{n-1; \alpha}^2 \quad (\text{bzw. } \chi^2 < \chi_{n; \alpha}^2).$$

Im zweiseitigen Testproblem müssen wir schließlich die Nullhypothese H_{03} verwerfen, falls gilt:

$$\begin{aligned} \chi^2 &> \chi_{n-1; 1-\alpha/2}^2 \quad \text{oder} \quad \chi^2 < \chi_{n-1; \alpha/2}^2 \\ (\text{bzw. } \chi^2 &> \chi_{n; 1-\alpha/2}^2 \quad \text{oder} \quad \chi^2 < \chi_{n; \alpha/2}^2). \end{aligned}$$

Beispiel: Wir wollen testen, ob die Varianz in der Gesamtheit der gepaarten Differenzen d_i bei den beiden Schlafmitteln signifikant größer als 6 ist. Hier kennen wir die Größe $\mu_1 - \mu_2$ nicht. Als Niveau dieses Tests von

$$H_{01}: \sigma_d^2 \leq 6 \quad \text{gegen} \quad H_{11}: \sigma_d^2 > 6$$

wählen wir hier einmal $\alpha = 0,05$, so daß sich wegen

$$\chi_{n-1; 1-\alpha}^2 = \chi_{14; 0,95}^2 = 23,685 < 24,164 = \frac{14 \cdot 10,356}{6} = \frac{(n-1) \cdot s_d^2}{\sigma_0^2} = \chi^2$$

ein Verwerfen der Nullhypothese rechtfertigen läßt. Somit ist die Varianz der gepaarten Differenzen zum 5%-Niveau signifikant größer als 6.

2.3. Verteilungsfreie Lokationsvergleiche

Wir wollen hier noch den Vergleich zweier abhängiger Meßreihen anhand des **Vorzeichenrangtests von Wilcoxon** und anhand des **Zeichentests** vorstellen. Diese Tests,

die wir hier natürlich auf die gepaarten Differenzen der beiden vorliegenden Meßreihen

$$d_i = x_{1i} - x_{2i} \quad \text{für } i = 1, \dots, n$$

anwenden werden, sind uns schon als Einstichprobentests aus Kap. IV, Abschnitt 5.4, bekannt. Daher werden wir beide Tests nur noch anhand unseres Beispiels mit den beiden Schlafmitteln vorführen.

Beim **Zeichentest** hatten wir schon von vornherein nur eine Approximation der Verteilung der Teststatistik unter der Nullhypothese durch die Normalverteilung kennengelernt. Dort in Abschnitt 5.4.1 wurde angegeben, daß diese Approximation für $n > 20$ hinreichend genau ist. Trotz dieser Einschränkung wollen wir hier auch für $n = 15$ Patienten die Approximation verwenden.

Beispiel: Wir wollen mit Hilfe des Zeichentests zum Niveau $\alpha = 0,05$ überprüfen, ob die Abweichungen von der normalen Schlafdauer bei beiden Schlafmitteln den gleichen Median haben. Das heißt, wir testen die Hypothese

$$H_{03}: \xi_{0,5} = 0$$

(der Median der gepaarten Differenzen ist 0) gegen die Alternative

$$H_{13}: \xi_{0,5} \neq 0.$$

Die 15 gepaarten Differenzen sowie die zugehörigen y_i -Werte

$$y_i = \begin{cases} 1, & \text{falls } d'_i > 0 \\ \frac{1}{2}, & \text{falls } d'_i = 0 \\ 0, & \text{falls } d'_i < 0 \end{cases}$$

sind in **Tab. 14** zusammengestellt. Eine wirkliche Transformation $d'_i = d_i - \xi_0$ entfällt hier, da ja gerade $\xi_0 = 0$ gesetzt wurde.

Tab. 14: Gepaarte Differenzen des Schlafmittelversuchs (Zeichentest)

i	$d'_i = d_i = x_{1i} - x_{2i}$	y_i
1	4,0	1
2	0,6	1
3	-3,9	0
4	-2,5	0
5	7,7	1
6	0,8	1
7	-5,0	0
8	2,3	1
9	1,8	1
10	-0,4	0
11	-2,8	0
12	0,7	1
13	-0,9	0
14	0,7	1
15	-2,0	0

Der Wert der Teststatistik ergibt sich damit zu

$$Z = \frac{\sum_{i=1}^{15} y_i - \frac{1}{2} \cdot 15}{\frac{1}{2} \cdot \sqrt{15}} = \frac{8 - 7,5}{0,5 \cdot \sqrt{15}} = 0,26.$$

Da

$$u_{\alpha/2} = u_{0,025} = -1,96 \leq 0,26 \leq 1,96 = u_{0,975} = u_{1-\alpha/2}$$

ist, kann die Hypothese nicht verworfen werden.

Kommen wir nun zum **Vorzeichenrangtest von Wilcoxon**. Hier hatten wir in Kap. IV gesehen, daß die Wilcoxon-Signed-Rank-Statistik T^+ eine eigene Tafel für die Quantile, vgl. Kap. IV, Abschnitt 5.4.2, Tab. 40, benötigt. In diesem Abschnitt nun wollen wir die in Kap. IV angegebene Approximation im Beispiel verwenden. Bezeichnen, wie dort näher erläutert, ζ das Symmetriezentrum der den gepaarten Differenzen $d_i = x_{1i} - x_{2i}$ zugrundeliegenden Verteilung und \tilde{R}_i die signierten Rangstatistiken zu den $d'_i = d_i - \zeta_0$, so haben wir bei Gültigkeit von $\zeta = \zeta_0$ für

$$T^+ = \sum_{i=1}^n c_i \tilde{R}_i \quad \text{mit} \quad c_i = \begin{cases} 1, & \text{falls } \tilde{R}_i > 0 \\ 0, & \text{falls } \tilde{R}_i \leq 0 \end{cases}$$

gerade

$$E T^+ = \frac{n(n+1)}{4} \quad \text{und} \quad \text{Var } T^+ = \frac{n(n+1)(2n+1)}{24},$$

so daß die Statistik

$$T^* = \frac{T^+ - E T^+}{\sqrt{\text{Var } T^+}} = \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

approximativ standardnormalverteilt ist, falls in den

$$d'_i = d_i - \zeta_0$$

keine Bindungen auftreten. Treten Bindungen auf, so ist die Approximation an die $N(0,1)$ -Verteilung besser mit

$$T^{**} = \frac{T^+ - \frac{n(n+1)}{4}}{\left(\frac{1}{24} \left[n(n+1)(2n+1) - \frac{1}{2} \sum_{j=1}^g t_j(t_j-1)(t_j+1) \right] \right)^{1/2}},$$

wobei g die Anzahl verschiedener Werte d'_1, \dots, d'_n und t_j für $j = 1, \dots, g$ die Anzahl der d'_i in der j -ten Gruppe ist.

Beispiel: Wir müssen hier voraussetzen, daß die gepaarten Differenzen bei unserem Schlafmittelversuch symmetrisch um einen Punkt verteilt sind und wollen die Hypothese testen, daß dieses Symmetriezentrum gerade der Punkt $\zeta_0 = 0$ ist, d.h. wir wollen überprüfen, ob die beiden Schlafmittel im Mittel die gleiche Wirkung erzielen.

Diesen Test für

$$H_{03}: \zeta = 0 \quad \text{gegen} \quad H_{13}: \zeta \neq 0$$

wollen wir zum Niveau $\alpha = 0,05$ durchführen. Die in Abschnitt 5.4 von Kap. IV angegebene Transformation der Beobachtungswerte entfällt hier natürlich wegen

$$d'_i = d_i - \zeta_0 = d_i,$$

so daß wir in Tab. 15 nur die gepaarten Differenzen sowie die Rangzahlen angeben. Den Test wollen wir dann, obwohl n natürlich relativ klein ist, anhand der Approximation durchführen.

Tab. 15: Schlafmittelversuch; gepaarte Differenzen und Rangzahlen zum Wilcoxon-Rangsummentest

i	$d'_i = d_i$	$R_i = \tilde{R}_i $	\tilde{R}_i
1	4,0	13	13
2	0,6	2	2
3	-3,9	12	-12
4	-2,5	10	-10
5	7,7	15	15
6	0,8	5	5
7	-5,0	14	-14
8	2,3	9	9
9	1,8	7	7
10	-0,4	1	-1
11	-2,8	11	-11
12	0,7	3,5	3,5
13	-0,9	6	-6
14	0,7	3,5	3,5
15	-2,0	8	-8

Aus Tabelle 15 entnehmen wir

$$T^+ = \sum_{i=1}^{15} c_i \tilde{R}_i = 58, \quad g = 14 \quad \text{und}$$

$$t_1 = t_2 = t_4 = t_5 = t_6 = t_7 = t_8 = t_9 = t_{10} = t_{11} = t_{12} = t_{13} = t_{14} = 1 \\ \text{sowie } t_3 = 2.$$

Da Bindungen vorliegen, müssen wir nun T^{**} berechnen.

Es ergibt sich gerade

$$T^{**} = \frac{58 - \frac{15 \cdot 16}{4}}{\left(\frac{1}{24} \left[15 \cdot 16 \cdot 31 - \frac{1}{2} (13 \cdot 1 \cdot 0 \cdot 2 + 2 \cdot 1 \cdot 3) \right] \right)^{1/2}}$$

$$= \frac{58 - 60}{\sqrt{(7440 - 3)/24}} = - \frac{2}{\sqrt{309,875}} = - 0,114.$$

Nun wenden wir den zweiseitigen Einstichproben-Gaußtest an: Es ist

$u_{1-\alpha/2} = u_{0,975} = 1,96 > |T^{**}| = 0,114$,
so daß wir die Nullhypothese nicht verwerfen können.

Natürlich können auch alle anderen in Kapitel IV, Abschnitt 5, vorgestellten Verfahren angewandt werden, indem man nicht die Beobachtungen x_{11}, \dots, x_{1n} , x_{21}, \dots, x_{2n} der beiden abhängigen Meßreihen, sondern die gepaarten Differenzen

$d_i = x_{i1} - x_{i2}$ für $i = 1, \dots, n$
betrachtet.

Kapitel IX: Die Korrelation von Merkmalen

Eines der Hauptanliegen bei der Analyse statistischen Datenmaterials ist das Erkennen von Abhängigkeiten und Zusammenhängen von verschiedenen Merkmalen. Zum einen ist man natürlich daran interessiert, qualitative Aussagen zu machen, zum anderen möchte man die Zusammenhänge auch quantifizieren.

Die Spezifizierung der Zusammenhänge erfolgt mittels Methoden der Regressionsrechnung, vgl. Kap. X; in diesem Kapitel wollen wir uns mit der **Korrelation von Merkmalen** beschäftigen, die den Grad der Abhängigkeit verschiedener Merkmale mißt.

Die **Korrelation zwischen zwei Merkmalen** gibt vornehmlich den Grad des linearen Zusammenhangs zwischen den beiden Merkmalen wieder. Während in Kapitel I, Abschnitt 9 bereits ausführlich auf die empirische Bedeutung der Korrelation eingegangen worden ist, werden nun auch **Tests auf Signifikanz eines Zusammenhangs** („Unabhängigkeitstests“) sowie **Vergleiche zwischen den Korrelationen verschiedener Merkmalskombinationen** durchgeführt. Dabei werden neben dem für metrisch skalierte (normalverteilte) Merkmale geeigneten **Pearson'schen Korrelationskoeffizienten**, vgl. Abschnitt 1, auch die **Rangkorrelationskoeffizienten von Spearman und Kendall**, vgl. Abschnitt 2, verwandt, die sich für zumindest ordinal skalierte Merkmale eignen.

Für qualitative, insbesondere nominalskaulierte Merkmale sei auf die **Zusammenhangsanalysen und -maße (Assoziationsmaße)** in Kap. VII hingewiesen.

Bei der Berechnung von Korrelationen sollte man stets darauf achten, daß die beiden Merkmale, für die man aufgrund einer Beobachtungsreihe vom Umfang n eine Korrelation schätzen möchte, auch in einem sachlogischen Zusammenhang stehen, da sonst inhaltliche „**Nonsense-Korrelationen**“ berechnet werden, wie z. B. die Korrelation zwischen der Anzahl der Störche und der Anzahl der Kindergeburten. Auch sollte man sich vor **Scheinkorrelationen** hüten, die dadurch entstehen, daß zwei an und für sich nicht korrelierte Merkmale mit einem dritten Merkmal korreliert sind: z. B. eine scheinbar recht hohe Korrelation zwischen Hämoglobin gehalt und mittlerer Oberfläche der Erythrozyten im menschlichen Blut, die lediglich aufgrund einer starken Abhängigkeit vom Geschlecht zustandekommt.

In solchen Fällen eignen sich **partielle Korrelationskoeffizienten**, vgl. Abschnitt 3, zur Abhängigkeitsanalyse, da sie den Einfluß des dritten Merkmals konstant halten. In Abschnitt 4 wird noch die **bipartiale Korrelation** zwischen zwei Merkmalen vorgestellt, die sich ebenfalls mit Abhängigkeitsanalysen unter Ausschaltung anderer Einflüsse befaßt. Mit der Abhängigkeitsanalyse eines Merkmals von verschiedenen anderen Merkmalen, der **multiplen Korrelationsanalyse**, werden wir uns in Abschnitt 5 befassen. Die multiple Korrelation ist ein Maß dafür, wie gut sich ein Merkmal Y durch p andere Merkmale X_1, \dots, X_p erklären läßt.

Schließlich wird in Abschnitt 6 noch ein **Test auf paarweise Unabhängigkeit von p Meßreihen**, die auch die Ausprägungen von p verschiedenen Merkmalen darstellen können, vorgestellt.

Auf die **multivariate Korrelationsanalyse (kanonische Korrelation)**, die sich mit Abhängigkeiten zwischen 2 Gruppen Y_1, \dots, Y_q und X_1, \dots, X_p von Merkmalen befaßt, kann hier nicht eingegangen werden, vgl. aber Hartung/Elpelt (1984, Kap. III).

1. Die Korrelation zweier normalverteilter Merkmale

In Kapitel II, Abschnitt 8.4 haben wir als Kenngröße von Zufallsvariablen die Kovarianz $\text{Cov}(X, Y)$ zwischen den Zufallsvariablen X und Y kennengelernt. Diese Größe haben wir, um sie besser interpretieren zu können, so normiert, daß der Wert der normierten Größe, die **Korrelation der Zufallsvariablen X und Y**

$$\text{Corr}(X, Y) = \rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}X \cdot \text{Var}Y}},$$

stets zwischen -1 und $+1$ liegt. Wir haben dort gesehen, daß die Korrelation von zwei unabhängigen Zufallsvariablen X und Y stets Null ist, d. h. zwei unabhängige Zufallsvariable sind stets **unkorreliert**. Umgekehrt können wir aber i. a. nicht aus der Unkorreliertheit von X und Y deren Unabhängigkeit folgern. Wir haben aber gesehen, daß speziell zwei unkorrelierte normalverteilte Zufallsvariablen X und Y auch unabhängig sind. Daher ist im Normalverteilungsfall die Korrelation ρ ein Maß für die Stärke der Abhängigkeit zweier Zufallsvariablen.

Wir wollen nun die Korrelation zwischen zwei in einer Grundgesamtheit normalverteilten Merkmalen anhand einer Stichprobe (bzw. Beobachtungsreihe) vom Umfang n aus dieser Grundgesamtheit schätzen, um so ein Maß für die Abhängigkeit zwischen den Merkmalen zu erhalten. Als **Schätzer für die Korrelation ρ zweier normalverteilter Zufallsvariablen X und Y** können wir natürlich die **Stichprobenkorrelation**

$$\begin{aligned} r_{XY} &= \frac{s_{XY}}{s_X \cdot s_Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \\ &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2\right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2\right)}} \end{aligned}$$

verwenden, die auch als **Pearson'scher Korrelationskoeffizient** bezeichnet wird, wobei die x_1, \dots, x_n Realisationen von X und die y_1, \dots, y_n Realisationen von Y sind und paarweise in der Form $(x_1, y_1), \dots, (x_n, y_n)$ erhoben werden.

Beispiel: Aufgrund einer Stichprobe vom Umfang $n = 15$ wollen wir die Korrelation – den Zusammenhang – zwischen Alter (X) und Blutdruck (Y) bei Frauen schätzen. Wir können davon ausgehen, daß beide Merkmale in der Gesamtheit der Frauen normalverteilt sind. Die Daten sind in **Tab. 1** angegeben.

Da das mittlere Alter bzw. der mittlere Blutdruck in unserer Beobachtungsreihe vom Umfang $n = 15$

$$\bar{x} = \frac{1}{15} \sum_{i=1}^{15} x_i = \frac{705}{15} \quad \text{bzw.} \quad \bar{y} = \frac{1}{15} \sum_{i=1}^{15} y_i = \frac{2011}{15}$$

ist, ergibt sich als Schätzer für die Korrelation ρ von Alter und Blutdruck bei Frauen

Tab. 1: Alter und Blutdruck von 15 zufällig ausgewählten Frauen

i	Alter (x_i)	Blutdruck (y_i)
1	47	129
2	52	139
3	30	112
4	35	119
5	59	145
6	44	133
7	63	152
8	38	117
9	49	145
10	41	136
11	32	115
12	55	137
13	46	134
14	51	141
15	63	157

$$\begin{aligned}
 r_{XY} &= \frac{\sum_{i=1}^n x_i y_i - 15 \cdot \frac{705}{15} \cdot \frac{2011}{15}}{\sqrt{\left(\sum_{i=1}^{15} x_i^2 - 15 \cdot \left(\frac{705}{15}\right)^2\right) \left(\sum_{i=1}^{15} y_i^2 - 15 \cdot \left(\frac{2011}{15}\right)^2\right)}} \\
 &= \frac{96387 - 94517}{\sqrt{(34685 - 33135)(272175 - 269608,07)}} = \frac{1870}{\sqrt{3978741,5}} \\
 &= 0,9375.
 \end{aligned}$$

Sind die Zufallsvariablen X und Y normalverteilt, so können wir aufgrund von n Realisationen x_1, \dots, x_n von X und y_1, \dots, y_n von Y mit Hilfe der Korrelation **testen**, ob X und Y unabhängig sind: Hat der unbekannte Korrelationskoeffizient ρ den Wert Null, so ist

$$t = \frac{r_{XY} \cdot \sqrt{n-2}}{\sqrt{1-r_{XY}^2}}$$

die Realisation einer t-verteilten Zufallsvariablen mit $n-2$ Freiheitsgraden. Dabei bezeichnet r_{XY} die Pearsonsche Schätzung für ρ . Wollen wir also die Hypothese

$$H_0: \rho = 0$$

gegen die Alternative

$$H_1: \rho \neq 0$$

zum Niveau α testen, so müssen wir die Hypothese H_0 verwerfen, falls gilt

$$|t| > t_{n-2; 1-\alpha/2}.$$

In den einseitigen Testproblemen

$$H'_0: \rho \geq 0 \quad \text{gegen} \quad H'_1: \rho < 0$$

$$\text{bzw.} \quad H''_0: \rho \leq 0 \quad \text{gegen} \quad H''_1: \rho > 0$$

verwerfen wir die Nullhypothese zum Niveau α , falls gilt

$$\begin{aligned} t &< t_{n-2; \alpha} \\ \text{bzw. } t &> t_{n-2; 1-\alpha} \end{aligned}$$

Wollen wir aber für festes $\rho_0 \neq 0$ die Hypothese

$$H_0: \rho = \rho_0$$

gegen die Alternative

$$H_1: \rho \neq \rho_0$$

zum Niveau α exakt testen, so müssen wir die Prüfgröße r_{XY} verwenden. Die Verteilung von r_{XY} für verschiedene Werte ρ_0 ist z.B. bei David (1954) vertafelt.

Einen approximativen Test erhält man, wenn man zunächst die **Fishersche Z-Transformation**

$$z = \operatorname{arctanh} r_{XY} = \frac{1}{2} \ln \frac{1 + r_{XY}}{1 - r_{XY}}, \text{ d.h. } r_{XY} = \frac{e^{2z} - 1}{e^{2z} + 1},$$

durchführt. Zur Umrechnung von r_{XY} in z und umgekehrt verwende man **Tab. 2**.

Tab. 2: Die Fishersche Z-Transformation [$-r_{XY} \rightarrow -z$, $-z \rightarrow -r_{XY}$]

r_{XY}	$r_{XY} \rightarrow z$									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,310	0,321	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,497	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,662	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,775	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,950	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647

z	$z \rightarrow r_{XY}$									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,119	0,129	0,139	0,149	0,159	0,168	0,178	0,188
0,2	0,197	0,207	0,217	0,226	0,235	0,245	0,254	0,264	0,273	0,282
0,3	0,291	0,300	0,310	0,319	0,327	0,336	0,345	0,354	0,363	0,371
0,4	0,380	0,388	0,397	0,405	0,414	0,422	0,430	0,438	0,446	0,454
0,5	0,462	0,470	0,478	0,485	0,493	0,501	0,508	0,515	0,523	0,530
0,6	0,537	0,544	0,551	0,558	0,565	0,572	0,578	0,585	0,592	0,598
0,7	0,604	0,611	0,617	0,623	0,629	0,635	0,641	0,647	0,653	0,658
0,8	0,664	0,670	0,675	0,680	0,686	0,691	0,696	0,701	0,706	0,711
0,9	0,716	0,721	0,726	0,731	0,735	0,740	0,744	0,749	0,753	0,757
1,	0,762	0,800	0,834	0,862	0,885	0,905	0,922	0,935	0,947	0,956
2,	0,964	0,970	0,976	0,980	0,984	0,987	0,989	0,991	0,993	0,994

Ist etwa $r_{XY} = 0,43$, so gilt $z = 0,46$; ist $r_{XY} = -0,61$, so gilt $z = -0,709$. Ist $z = 0,25$, so ist $r_{XY} = 0,245$; ist $z = -1,3$, so ist $r_{XY} = -0,862$.

Die Größe z (eigentlich die z zugrundeliegende Zufallsvariable) ist auch für kleine n recht gut approximativ normalverteilt mit

$$Ez = \frac{1}{2} \ln \frac{1 + \varrho}{1 - \varrho} + \frac{\varrho}{2(n - 1)} = \zeta \quad \text{und} \quad \text{Var}z = \frac{1}{n - 3}.$$

Daher ist

$$(z - \zeta) \cdot \sqrt{n - 3}$$

standardnormalverteilt (approximativ), so daß wir den gewöhnlichen Einstichproben-Gaußtest, vgl. Kap. III, verwenden können. Wir müssen also z. B. die Hypothese

$$H_0: \varrho = \varrho_0$$

im Test zum Niveau α gegen die Alternative

$$H_1: \varrho \neq \varrho_0$$

verwerfen, falls mit dem $(1 - \alpha/2)$ -Quantil $u_{1-\alpha/2}$ der Standardnormalverteilung gilt:

$$|(z - \zeta_0) \cdot \sqrt{n - 3}| > u_{1-\alpha/2}.$$

Dabei ist

$$\zeta_0 = \frac{1}{2} \ln \frac{1 + \varrho_0}{1 - \varrho_0} + \frac{\varrho_0}{2(n - 1)}.$$

Analog verfährt man bei den einseitigen Hypothesen H'_0 und H''_0 für $\varrho_0 \neq 0$; vgl. nachstehendes Beispiel.

Die Z-Transformation kann auch zur Bestimmung eines $1 - \alpha$ -Konfidenzintervalls für ϱ benutzt werden. Für $\zeta - \varrho/(2(n - 1))$ ist approximativ

$$\left[\frac{1}{2} \ln \frac{1 + r_{XY}}{1 - r_{XY}} - \frac{r_{XY}}{2(n - 1)} - \frac{u_{1-\alpha/2}}{\sqrt{n - 3}}, \frac{1}{2} \ln \frac{1 + r_{XY}}{1 - r_{XY}} - \frac{r_{XY}}{2(n - 1)} + \frac{u_{1-\alpha/2}}{\sqrt{n - 3}} \right] = [\zeta_1; \zeta_2]$$

ein $1 - \alpha$ Konfidenzintervall. Rechnen wir nun mit Hilfe der Tabelle 2 die Intervallgrenzen ζ_1 und ζ_2 in Intervallgrenzen r_1 und r_2 um, so erhalten wir das gesuchte Konfidenzintervall für ϱ , vgl. auch Kap. III in Hartung/Elpelt (1984).

Beispiel: Wir wollen bei Vorliegen der Beobachtungen aus dem vorherigen Beispiel die Hypothese, daß Blutdruck und Alter bei Frauen unkorreliert sind, zum Niveau $\alpha = 0,1$ testen. Als Schätzer für die Korrelation hatten wir

$$r_{XY} = 0,9375$$

berechnet. Testen wir nun die Hypothese

$$H_0: \varrho = 0 \quad \text{gegen} \quad H_1: \varrho \neq 0,$$

so müssen wir wegen

$$t = \frac{0,9375 \cdot \sqrt{15 - 2}}{\sqrt{1 - 0,9375^2}} = 9,714 > 1,771 = t_{15-2; 0,95} = t_{13; 0,95}$$

die Hypothese verwerfen.

Nun wollen wir noch mit Hilfe des approximativen Tests die Hypothese

$$H_0': \rho \leq 0,90 \quad \text{gegen} \quad H_1': \rho > 0,90$$

zum Niveau $\alpha = 0,05$ testen. Für

$$r_{XY} = 0,9375$$

ergibt sich

$$z = 1,717 \quad \text{und} \quad \zeta_0 = \frac{1}{2} \ln \frac{1,90}{0,10} + \frac{0,90}{2 \cdot 14} = 1,47 + 0,32 = 1,79.$$

Da

$$(z - \zeta_0) \cdot \sqrt{n - 3} = (1,717 - 1,79) \cdot \sqrt{12} = -0,21 < 1,645 = u_{0,95}$$

ist, können wir diese Hypothese nicht verwerfen.

Zum Niveau 0,05 ist also nicht signifikant gesichert, daß die Korrelation zwischen Alter und Blutdruck bei Frauen größer als 0,90 ist.

Wir wollen nun ein Konfidenzintervall zum Niveau $1 - \alpha = 0,95$ für diese Korrelation ρ bestimmen. Mit

$$\begin{aligned} & \left[\frac{1}{2} \ln \frac{1,9375}{0,0625} - \frac{0,9375}{28} - \frac{u_{0,975}}{\sqrt{12}} ; \frac{1}{2} \ln \frac{1,9375}{0,0625} - \frac{0,9375}{28} + \frac{u_{0,975}}{\sqrt{12}} \right] \\ &= \left[1,717 - 0,033 - \frac{1,96}{\sqrt{12}} ; 1,717 - 0,033 + \frac{1,96}{\sqrt{12}} \right] \\ &= [1,118; 2,250] \end{aligned}$$

ergeben sich die Größen

$$\zeta_1 = 1,118 \quad \text{und} \quad \zeta_2 = 2,250,$$

die wir nun zurücktransformieren. Es ergibt sich

$$r_1 = 0,807 \quad \text{und} \quad r_2 = 0,978,$$

so daß

$$[0,807; 0,978]$$

das gesuchte Konfidenzintervall für die Korrelation ρ zwischen Alter und Blutdruck bei Frauen zum Niveau 0,95 ist.

Sind X_1, \dots, X_n normalverteilte Zufallsvariable, so faßt man die geschätzten Korrelationen zwischen je zwei Merkmalen in der **Korrelationsmatrix R** zusammen:

$$R = \begin{bmatrix} 1 & r_{X_1 X_2} & \dots & r_{X_1 X_n} \\ r_{X_1 X_2} & 1 & & \vdots \\ \vdots & & & r_{X_{n-1} X_n} \\ r_{X_1 X_n} & \dots & r_{X_{n-1} X_n} & 1 \end{bmatrix}.$$

Beispielsweise ist die Matrix der geschätzten Korrelationen von 3 Zufallsvariablen X, Y und Z:

$$R = \begin{bmatrix} 1 & r_{XY} & r_{XZ} \\ r_{XY} & 1 & r_{YZ} \\ r_{XZ} & r_{YZ} & 1 \end{bmatrix}.$$

Möchte man überprüfen, ob die Korrelationen ρ_1 und ρ_2 von je zwei Merkmalen zum Signifikanzniveau α identisch sind, so bestimmt man zunächst die transformierten Werte

$$z_1 = \operatorname{arctanh} r_1 \quad \text{und} \quad z_2 = \operatorname{arctanh} r_2,$$

wobei r_1 die aufgrund einer Stichprobe vom Umfang n_1 geschätzte Korrelation für ρ_1 und r_2 die aufgrund einer Stichprobe vom Umfang n_2 geschätzte Korrelation für ρ_2 ist.

Es ist dann

$$T = \frac{z_1 - z_2}{\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}}$$

im Falle der Gültigkeit von $\rho_1 = \rho_2$ die Realisation einer approximativ standard-normalverteilten Zufallsvariablen, d. h. im Testproblem

$$H_0: \rho_1 = \rho_2 \quad \text{gegen} \quad H_1: \rho_1 \neq \rho_2$$

muß man die Nullhypothese zum Niveau α verwerfen, falls gilt

$$|T| > u_{1-\alpha/2}.$$

Natürlich kann man auch die einseitigen Testprobleme zum Niveau α

$$H'_0: \rho_1 \geq \rho_2 \quad \text{gegen} \quad H'_1: \rho_1 < \rho_2$$

$$\text{und} \quad H''_0: \rho_1 \leq \rho_2 \quad \text{gegen} \quad H''_1: \rho_1 > \rho_2$$

betrachten. Man muß H'_0 verwerfen, wenn gilt

$$T < u_\alpha,$$

bzw. H''_0 verwerfen, falls gilt

$$T > u_{1-\alpha}.$$

Wird die Hypothese H_0 der Gleichheit der Korrelationen ρ_1 und ρ_2 zum Niveau α nicht verworfen, so kann man einen neuen gemeinsamen Schätzer für ρ_1 und ρ_2 bestimmen. Der gemeinsame Schätzer für ζ ist

$$\tilde{z} = \frac{(n_1 - 3)z_1 + (n_2 - 3)z_2}{n_1 + n_2}.$$

In Tab. 2 sucht man dann den zu \tilde{z} gehörigen Wert \tilde{r} und hat damit einen gemeinsamen Schätzer für ρ_1 und ρ_2 gefunden.

Beispiel: Aufgrund einer Stichprobe vom Umfang $n_1 = 17$ hat sich als Schätzer der Korrelation ρ_1

$$r_1 = 0,72$$

und aufgrund einer Stichprobe vom Umfang $n_2 = 26$ hat sich als Schätzer für ρ_2

$$r_2 = 0,58$$

ergeben. Wir wollen nun zum Niveau $\alpha = 0,10$ die Nullhypothese

$$H_0: \rho_1 = \rho_2 \text{ gegen } H_1: \rho_1 \neq \rho_2$$

testen. Aus Tab. 2 ergibt sich

$$\text{für } r_1 = 0,72 : z_1 = 0,908$$

$$\text{und für } r_2 = 0,58 : z_2 = 0,662.$$

Somit ist

$$|T| = \left| \frac{0,908 - 0,662}{\sqrt{1/(17-3) + 1/(26-3)}} \right| = \frac{0,246}{0,339} = 0,726 < 1,645 = u_{0,95},$$

so daß wir die Hypothese der Gleichheit von ρ_1 und ρ_2 zum Niveau 0,10 nicht verwerfen können. Nun wollen wir noch den gemeinsamen Schätzer \tilde{r} für ρ_1 und ρ_2 bestimmen.

Es ist

$$\tilde{z} = \frac{(17-3) \cdot 0,908 + (26-3) \cdot 0,662}{17+26} = \frac{27,938}{43} = 0,650$$

und somit nach Tab. 2

$$\tilde{r} = 0,572.$$

Will man nicht nur die Gleichheit von zwei Korrelationen ρ_1, ρ_2 überprüfen, sondern vielmehr die Homogenität, d. h. die Gleichheit, von k Korrelationen ρ_1, \dots, ρ_k , so bestimmt man zunächst aufgrund von Stichproben der Umfänge n_1, \dots, n_k Schätzer r_1, \dots, r_k für diese unbekannten Korrelationen, die man dann mit Hilfe von Tab. 2 in Größen z_1, \dots, z_k transformiert. Im Test zum Niveau α von

$H_0: \rho_1 = \rho_2 = \dots = \rho_k$ gegen $H_1: \text{es gibt ein Paar } (i, j) \text{ mit } \rho_i \neq \rho_j$ verwirft man die Nullhypothese H_0 , falls

$$T = \sum_{i=1}^k z_i^2 (n_i - 3) - \frac{\left(\sum_{i=1}^k z_i (n_i - 3) \right)^2}{\sum_{i=1}^k n_i} > \chi_{k-1; 1-\alpha}^2$$

gilt, mit dem $(1-\alpha)$ -Quantil $\chi_{k-1; 1-\alpha}^2$ der χ^2 -Verteilung mit $k-1$ Freiheitsgraden. Wird H_0 zum Niveau α nicht verworfen, so kann man, wie schon beim Test auf Gleichheit von zwei Korrelationen, einen gemeinsamen Schätzer \tilde{r} für ρ_1, \dots, ρ_k bestimmen, indem man mit Hilfe der Tab. 2 den Wert

$$\tilde{z} = \frac{\sum_{i=1}^k z_i (n_i - 3)}{\sum_{i=1}^k n_i}$$

in den Wert \tilde{r} transformiert.

Beispiel: Bei Stichprobenumfängen $n_1 = 11, n_2 = 15, n_3 = 7$ und $n_4 = 22$ ergaben sich als Schätzer für ρ_1, ρ_2, ρ_3 und ρ_4 die Werte $r_1 = 0,37, r_2 = 0,26, r_3 = 0,19$ und $r_4 = 0,32$. Wir wollen nun zum Niveau $\alpha = 0,10$ die Hypothese

$$H_0: \rho_1 = \rho_2 = \rho_3 = \rho_4$$

testen. Es ergibt sich aus Tab. 2 für

$$r_1 = 0,37 : z_1 = 0,388,$$

$$r_2 = 0,26 : z_2 = 0,266,$$

$$r_3 = 0,19 : z_3 = 0,192,$$

$$r_4 = 0,32 : z_4 = 0,332.$$

Es ist

$$\sum_{i=1}^4 n_i = 11 + 15 + 7 + 22 = 55,$$

$$\sum_{i=1}^4 z_i(n_i - 3) = 0,388 \cdot 8 + 0,266 \cdot 12 + 0,192 \cdot 4 + 0,332 \cdot 19 = 13,372$$

$$\text{und } \sum_{i=1}^4 z_i^2(n_i - 3) = 0,388^2 \cdot 8 + 0,266^2 \cdot 12 + 0,192^2 \cdot 4 + 0,332^2 \cdot 19 \\ = 4,295.$$

Damit können wir wegen

$$T = 4,295 - \frac{13,372^2}{55} = 1,0439 < 6,251 = \chi^2_{3;0,90} = \chi^2_{k-1;1-\alpha}$$

die Hypothese der Homogenität der 4 Korrelationen ρ_1, ρ_2, ρ_3 und ρ_4 nicht verworfen. Der gemeinsame Schätzer \tilde{r} ergibt sich nun wegen

$$\tilde{z} = \frac{13,372}{55} = 0,24 \quad \text{zu} \quad \tilde{r} = 0,235.$$

2. Die Rangkorrelation zweier Merkmale

Möchte man die Korrelation von Zufallsvariablen X und Y, die nicht normalverteilt sind, schätzen, so kann dies mit Hilfe der **Rangkorrelationskoeffizienten von Spearman** (r_s) und **Kendall** (τ) geschehen. Wie der Name Rangkorrelationskoeffizient schon sagt, schätzt man die Korrelation nur aufgrund von Ranginformationen. Man muß dabei lediglich wissen, ob z. B. eine Realisation x_i größer oder kleiner oder gleich einer anderen Realisation x_j der Zufallsvariablen X ist.

2.1. Der Spearmansche Rangkorrelationskoeffizient

Vergeben wir die Rangzahlen $R(x_i)$ bzw. $R(y_i)$ für die je n Realisationen der Zufallsvariablen X und Y so, daß in jeder der beiden Reihen die kleinste („schlechteste“) Realisation den Rang 1, . . . , die größte („beste“) Realisation den Rang n erhält, oder in beiden Reihen genau umgekehrt, so ergibt sich der Spearmansche Korrelationskoeffizient r_s durch Einsetzen der Rangzahlen – anstelle der Realisationen – in die Formel zur Berechnung von r_{XY} :

$$\begin{aligned}
 r_s &= \frac{\sum_{i=1}^n (R(x_i) - \bar{R}(x))(R(y_i) - \bar{R}(y))}{\sqrt{\sum_{i=1}^n (R(x_i) - \bar{R}(x))^2 \cdot \sum_{i=1}^n (R(y_i) - \bar{R}(y))^2}} \\
 &\left(= 1 - \frac{6 \sum_{i=1}^n (R(x_i) - R(y_i))^2}{n(n^2 - 1)} \right. \\
 &\quad \left. = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \text{ mit } d_i = R(x_i) - R(y_i), \text{ falls keine Bindungen auftreten.} \right)
 \end{aligned}$$

Treten Bindungen auf, d. h. sind mehrere Realisationen einer Zufallsvariablen genau gleich, so nehmen wir jeweils als Rang das arithmetische Mittel der in Frage stehenden Ränge („midranks“), vgl. Abschnitt 7 in Kap. III.

Beispiel: Zwei Personen stellen eine Bewertungstabelle für 8 zufällig ausgewählte Gemälde auf. Sie dürfen 1 bis 8 Punkte und auch, falls ihnen mehrere Gemälde gleich gut gefallen, mittlere Punkte vergeben. Die Punkteskala gibt natürlich nur eine Rangfolge an, so daß wir zur Schätzung der Korrelation zwischen den „Geschmäckern“ der Personen A und B nur einen Rangkorrelationskoeffizienten berechnen können. Die Punktzahlen sind in Tab. 3 angegeben.

Tab. 3: Bewertung von 8 Gemälden und Arbeitstabelle zur Bestimmung des Spearmanschen Rangkorrelationskoeffizienten

Gemälde i	Punkte von A = $R(x_i)$	Punkte von B = $R(y_i)$	$R(x_i) - \bar{R}(x)$	$R(y_i) - \bar{R}(y)$
1	8	6	3,5	1,5
2	6	7,5	1,5	3,0
3	5	4	0,5	-0,5
4	3,5	1	-1,0	-3,5
5	1	2	-3,5	-2,5
6	2	3	-2,5	-1,5
7	3,5	5	-1,0	0,5
8	7	7,5	2,5	3,0

Mit diesen Werten ergibt sich als Schätzung für die Korrelation ρ

$$r_s = \frac{32,5}{\sqrt{41,5 \cdot 41,5}} = \frac{32,5}{41,5} = 0,783 .$$

Häufig findet der Spearmansche Rangkorrelationskoeffizient in der Qualitätskontrolle und auch in der Psychologie Verwendung. Im Rahmen der Qualitätskontrolle ist man oft darauf angewiesen, Produkte zu beurteilen bzw. deren Lebensdauer zu bestimmen, die sehr haltbar sind. Man kann dann nicht abwarten, bis ein Produkt defekt ist, um seine Lebensdauer zu schätzen, sondern muß einen kurzzeitig durchführbaren Versuch konstruieren, der eine gute Aussage über die Lebensdauer macht, vgl. Abschnitt 4 in Kap. XIII und nachstehendes Beispiel.

In der Psychologie tritt beispielsweise bei der Konstruktion von Eignungstests das Problem auf, den Eignungstest so festzulegen, daß er die wirkliche Eignung (die sich ja erst viel später herausstellt) einer Person etwa für eine bestimmte Arbeit möglichst gut widerspiegelt.

Beispiel: (vgl. R. C. Stoll: "An improved multipurpose abrasion tester and its application for the evalution of the wear resistance of textiles", Text. Res. Journal, 19, 1949, S. 394–419) Um einen technologischen Kurzzeitversuch zur Prüfung der Haltbarkeit von Textilien so zu entwickeln, daß dieser die wirkliche Haltbarkeit der Textilien möglichst gut widerspiegelt, wurden an 7 verschiedenen Textilien Trageversuche und verschiedene Kurzzeitversuche durchgeführt. Die Kurzzeitversuche waren hier Scheuerversuche wie z. B. Biegescheuerung in Kett- und Schußrichtung. Bei den Scheuerversuchen wurde die Anzahl der Scheuertouren bis zum Bruch der Gewebe gezählt. Bei den Trageversuchen konnte man lediglich eine Rangfolge in der Haltbarkeit der Textilien festlegen. Um die Korrelation zwischen Trage- und Scheuerversuchen schätzen zu können, wurden auch die Ergebnisse der Scheuerversuche in Rangzahlen umgewandelt, vgl. **Tab. 4**.

Tab. 4: Ergebnisse der Trage- und Scheuerversuche bei 7 Textilien

Ge- webe- art	Rangzahl R_i im Trage- versuch	Scheuer- touren bis zum Gewebe- bruch in Kettrichtung	Rangzahl R'_i im Scheuer- versuch/ Kettrichtung	Scheuer- touren bis zum Gewebe- bruch in Schußrichtung	Rangzahl R''_i im Scheuer- versuch/ Schußrichtung
1	6	72	7	115	2
2	5	85	4	78	6
3	7	78	6	91	5
4	4	82	5	102	4
5	3	91	3	67	7
6	2	138	2	111	3
7	1	154	1	136	1

Wir wollen nun die Spearmanschen Rangkorrelationen $r_s^{(1)}$ zwischen Trageversuch und Scheuerversuch in Kettrichtung, $r_s^{(2)}$ zwischen Trageversuch und Scheuerversuch in Schußrichtung sowie $r_s^{(3)}$ zwischen Scheuerversuch in Kettrichtung und Scheuerversuch in Schußrichtung berechnen.

Es ist

$$r_s^{(1)} = 1 - \frac{6 \sum_{i=1}^7 (R_i - R'_i)^2}{7(7^2 - 1)} = 1 - \frac{6(1 + 1 + 1 + 1 + 0 + 0 + 0)}{7 \cdot 48}$$

$$= 1 - \frac{24}{336} = 0,9286,$$

$$r_s^{(2)} = 1 - \frac{6 \sum_{i=1}^7 (R_i - R''_i)^2}{7(7^2 - 1)} = 1 - \frac{6(16 + 1 + 4 + 0 + 16 + 1 + 0)}{336}$$

$$= 1 - \frac{228}{336} = 0,3214,$$

$$\text{und } r_s^{(3)} = 1 - \frac{6 \sum_{i=1}^7 (R'_i - R''_i)^2}{7(7^2 - 1)} = 1 - \frac{6(25 + 4 + 1 + 1 + 16 + 1 + 0)}{336} \\ = 1 - \frac{288}{336} = 0,1429.$$

Die Rangkorrelation $r_s^{(3)}$ zwischen den beiden Arten von Scheuerversuchen ist sehr gering, und die Korrelationen $r_s^{(1)}$ und $r_s^{(2)}$ sind recht unterschiedlich. Wir sehen, daß der Scheuerversuch in Kettrichtung die Haltbarkeit im Trageversuch recht gut widerspiegelt, so daß man diesen wohl als Kurzzeitversuch wählen kann.

Basierend auf dem Spearmanschen Rangkorrelationskoeffizienten r_s kann auch die Hypothese der Unabhängigkeit zweier Meßreihen $x_1, \dots, x_n, y_1, \dots, y_n$ zum Niveau α getestet werden.

Man verwendet anstelle von

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

die sogenannte **Hotelling-Pabst-Statistik**

$$D = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (R(x_i) - R(y_i))^2$$

als Prüfgröße. Die Hypothese

H_0 : die Meßreihen sind unabhängig

muß im Test gegen die Alternative

H_1 : die Meßreihen sind nicht unabhängig

zum Niveau α verworfen werden, falls

$$D < h_{n; \alpha/2} \quad \text{oder} \quad D > h_{n; 1-\alpha/2}$$

ist. Für $n \leq 30$ sind kritische Werte $h_{n; \alpha}$ in **Tab. 5** angegeben. Es ist weiter

$$h_{n; 1-\alpha} = \frac{1}{3} n(n^2 - 1) - h_{n; \alpha}.$$

Stehen keine kritischen Werte $h_{n; \alpha}$ zur Verfügung, so macht man sich die Tatsache zunutze, daß unter H_0 gilt

$$E(D) = \frac{1}{6}(n^3 - n) - \frac{1}{12} \sum_{j=1}^p (d_{1j}^3 - d_{1j}) - \frac{1}{12} \sum_{k=1}^q (d_{2k}^3 - d_{2k}),$$

$$\text{Var}(D) = \frac{(n-1)(n+1)^2 n^2}{36} \cdot \left(1 - \frac{\sum_{j=1}^p (d_{1j}^3 - d_{1j})}{n^3 - n} \right)$$

$$\cdot \left(1 - \frac{\sum_{k=1}^q (d_{2k}^3 - d_{2k})}{n^3 - n} \right).$$

Tab. 5: Kritische Werte $h_{n,\alpha}$ der Hotelling-Pabst-Statistik

n	$\alpha = 0,01$	$\alpha = 0,025$	$\alpha = 0,05$	$\alpha = 0,10$
1	—	—	—	—
2	—	—	—	—
3	—	—	—	—
4	—	—	2	2
5	2	2	4	6
6	4	6	8	14
7	8	14	18	26
8	16	24	32	42
9	28	38	50	64
10	44	60	74	92
11	66	86	104	128
12	94	120	144	172
13	130	162	190	226
14	172	212	246	290
15	224	270	312	364
16	284	340	390	450
17	356	420	480	550
18	438	512	582	664
19	532	618	696	790
20	638	738	826	934
21	758	870	972	1092
22	892	1020	1134	1270
23	1042	1184	1312	1464
24	1208	1366	1510	1678
25	1390	1566	1726	1912
26	1590	1786	1960	2168
27	1808	2024	2216	2444
28	2046	2284	2494	2744
29	2306	2564	2796	3068
30	2584	2868	3120	3416

Dabei bezeichnet p bzw. q die Zahl der unterschiedlichen Werte x_i bzw. y_j in der ersten bzw. zweiten Meßreihe, d_{1j} bzw. d_{2j} die Anzahl der Beobachtungen, die den j-ten dieser unterschiedlichen Werte in der ersten bzw. zweiten Meßreihe haben. Man weiß dann, daß unter H_0

$$T = \frac{D - E(D)}{\sqrt{\text{Var}(D)}}$$

approximativ standardnormalverteilt ist, d. h. man muß die Hypothese H_0 verwerfen, falls gilt:

$$|T| > u_{1-\alpha/2}.$$

Kann man die Hypothese H_0 der Unabhängigkeit nicht verwerfen, so sollte man sich davor hüten zu folgern, daß die Meßreihen wirklich unabhängig sind, denn der Fehler 2. Art (Nichtverwerfen von H_0 , obwohl H_1 richtig ist) kann sehr hoch sein.

Beispiel: (aus: Karelitz, Fisichelli, Costa, Rosenfeld: "Relation of crying activity in early infancy to speech and intellectual development at age three years", Child Development, 35 (1964), S. 769–777).

Um zu untersuchen, ob ein Zusammenhang zwischen Schreiintensität und spätem Intelligenzquotienten besteht, wurde die Anzahl der Schreie von 22 fünf Tage alten Säuglingen mit ihrem IQ (Stanford-Binet-IQ) im Alter von 3 Jahren verglichen. In Tab. 6 sind die Versuchsergebnisse und die zugehörigen Rangzahlen zusammengestellt.

Tab. 6: Anzahl von Schreien und Intelligenzquotient von 22 Kindern

i	Schreie x_{1i}	IQ x_{2i}	R_{1i}	R_{2i}	$(R_{1i} - R_{2i})^2$
1	19	103	14	4,5	90,25
2	12	119	1	15	196,00
3	18	124	11,5	17	30,25
4	16	133	6,5	19	156,25
5	26	155	21	21	0,00
6	15	112	4	10,5	42,25
7	27	108	22	7	225,00
8	23	103	19,5	4,5	225,00
9	20	90	16	1	225,00
10	21	114	17,5	13	20,25
11	19	120	14	16	4,00
12	15	100	4	3	1,00
13	17	109	9	8,5	0,25
14	15	112	4	10,5	42,25
15	21	157	17,5	22	20,25
16	16	118	6,5	14	56,25
17	23	113	19,5	12	56,25
18	17	94	9	2	49,00
19	14	106	2	6	16,00
20	18	109	11,5	8,5	9,00
21	17	141	9	20	121,00
22	19	132	14	18	16,00

Aus dieser Tabelle ergibt sich dann

$$D = \sum_{i=1}^n (R_{1i} - R_{2i})^2 = 1601,50.$$

Wegen

$$h_{n; \alpha/2} = h_{22; 0,025} = 1020 < 1601,5 < 2522 = \frac{1}{3}n(n^2 - 1) - h_{n; \alpha/2} = h_{n; 1 - \alpha/2}$$

können wir die Hypothese der Unabhängigkeit von Schreiintensität und Intelligenzquotient zum Niveau 5% nicht verwerfen. Wir wollen hier auch noch die Durchführung des approximativen Tests vorführen.

Da in der ersten Meßreihe x_1, \dots, x_n (Anzahl der Schreie) 12 verschiedene Werte auftreten, ist $p = 12$, und da in der zweiten Meßreihe y_1, \dots, y_n (IQ) 19 verschiedene Werte auftreten, ist $q = 19$. Die Werte $d_{1,1}, \dots, d_{1,12}$ und $d_{2,1}, \dots, d_{2,19}$ sind in Tab. 7 zusammengestellt, so daß wir nun $E(D)$ und $\text{Var}(D)$ berechnen können. Es ist

$$\begin{aligned} \sum_{j=1}^{12} (d_{1,j}^3 - d_{1,j}) &= 3 \cdot (27 - 3) + 4 \cdot (8 - 2) = 3 \cdot 24 + 4 \cdot 6 \\ &= 72 + 24 = 96, \end{aligned}$$

Tab. 7: Anzahl der Bindungen bei verschiedenen Werten der Meßreihen aus Tab. 6

Anzahl der Schreie	12	14	15	16	17	18	19	20	21	23	26	27
d_{1j}	1	1	3	2	3	2	3	1	2	2	1	1
IQ	90	94	100	103	106	108	109	112	113	114	118	119
d_{2k}	1	1	1	2	1	1	2	2	1	1	1	1
IQ	132	133	141	155	157							
d_{2k}	1	1	1	1	1							

$$\sum_{k=1}^{19} (d_{2k}^3 - d_{2k}) = (8 - 2) + (8 - 2) + (8 - 2) = 3 \cdot 6 = 18$$

und somit

$$E(D) = \frac{1}{6}(22^3 - 22) - \frac{1}{12} \cdot 96 - \frac{1}{12} \cdot 18 = 1771 - 8 - 1,5 = 1761,5,$$

$$\begin{aligned} \text{Var}(D) &= \frac{(22-1)(22+1)^2 22^2}{36} \left(1 - \frac{96}{22^3 - 22}\right) \left(1 - \frac{18}{22^3 - 22}\right) \\ &= 147754,28. \end{aligned}$$

Daraus ergibt sich

$$|T| = \left| \frac{D - E(D)}{\sqrt{\text{Var}(D)}} \right| = \left| \frac{1601,50 - 1761,50}{\sqrt{147754,28}} \right| = \left| \frac{-160}{384,388} \right| = 0,416.$$

Da $|T| = 0,416 < 1,96 = u_{0,975} = u_{1-\alpha/2}$

ist, können wir auch bei approximativem Vorgehen die Hypothese H_0 : „Schreintensität und IQ sind unabhängig“ zum 5%-Niveau nicht verwerfen.

2.2. Der Kendallsche Rangkorrelationskoeffizient

Der Kendallsche Rangkorrelationskoeffizient τ wird basierend auf Rangzahlen, die genau wie bei Spearmans r_s vergeben werden, berechnet. Die n Realisationen der Zufallsvariablen X und Y bilden, wie wir schon gesehen haben, n natürliche Beobachtungspaares $(x_1, y_1), \dots, (x_n, y_n)$. Haben wir nun die Rangzahlen $R(x_i)$ und $R(y_i)$ vergeben, so ordnen wir zur Berechnung von Kendalls τ die Beobachtungspaares so an, daß im 1. Paar der Rang der Realisation von X gerade 1 ist, im 2. Paar der Rang der Realisation von X gerade 2 ist, ..., im n -ten Paar der Rang der Realisation von X gerade n ist. Dadurch haben wir, wie wir im Beispiel noch sehen werden, auch eine Reihenfolge der Rangzahlen der Realisationen y_1, \dots, y_n der Zufallsvariablen Y festgelegt. In dieser Reihenfolge bestimmen wir nun für jede Rangzahl $R(y_i)$ die Anzahl q_i der Rangzahlen $R(y_j)$, die kleiner oder gleich $R(y_i)$ sind und in der Reihenfolge hinter $R(y_i)$ stehen. Das Kendallsche τ ergibt sich dann zu

$$\tau = 1 - \frac{4 \sum_{i=1}^n q_i}{n(n-1)}.$$

Beispiel: Kommen wir noch einmal auf das Beispiel der Gemäldebewertung, vgl. Tab. 3, zurück. Wir wollen nun die Korrelation τ der „Geschmäcke“ der Personen A und B nach Kendall bestimmen. Dazu stellen wir zunächst ausgehend von Tab. 3 die **Tab. 8** auf. (Die Anordnung der Gemälde 4 und 7 erfolgte hier zufällig.)

Tab. 8: Arbeitstabelle zur Berechnung von Kendalls τ bei der Gemäldebewertung (Tab. 3)

i	R(x _i)	R(y _i)	q _i
5	1	2	1
6	2	3	1
7	3,5	5	2
4	3,5	1	0
3	5	4	0
2	6	7,5	2
8	7	7,5	1
1	8	6	0

Nun können wir τ bestimmen:

$$\tau = 1 - \frac{4 \sum_{i=1}^8 q_i}{n(n-1)} = 1 - \frac{4(1+1+2+2+1)}{8 \cdot 7} = 1 - \frac{4 \cdot 7}{8 \cdot 7} = 1 - \frac{1}{2} = 0,5.$$

Ein Test auf Unabhängigkeit zweier Meßreihen $x_1, \dots, x_n, y_1, \dots, y_n$ lässt sich auch basierend auf Kendalls τ durchführen, denn

$$K = \frac{1}{2} n(n-1) - 2 \sum q_i = \frac{n(n-1)}{2} \tau$$

ist gerade die Kendallsche K-Statistik, deren Verteilung wir schon in Kap. IV, Abschnitt 5.4 kennengelernt haben. Die kritischen Werte $K_{n; 1-\alpha}$ sind in der dortigen Tab. 44, S. 249 vertafelt.

Wir verwerfen die Hypothese

$$H_0: \text{die Meßreihen sind unabhängig}$$

im Test zum Niveau α gegen die Alternative

$$H_1: \text{die Meßreihen sind nicht unabhängig},$$

falls gilt:

$$|K| > K_{n; 1-\alpha/2}.$$

Stehen keine kritischen Werte $K_{n; 1-\alpha}$ zur Verfügung, so verwendet man die unter H_0 approximativ standardnormalverteilte Prüfgröße

$$K^* = \frac{K}{\sqrt{n(n-1)(2n+5)/18}}$$

und verwirft die Nullhypothese H_0 , falls gilt

$$|K^*| > u_{1-\alpha/2}.$$

Beispiel: Im Beispiel der Gemälde Sammlung wollen wir zum Niveau $\alpha = 0,05$ die Hypothese

H_0 : der Geschmack der Person A ist unabhängig vom Geschmack der Person B

gegen die Alternative

H_1 : der Geschmack der Person A ist nicht unabhängig vom Geschmack der Person B

testen. Die Anzahl der Beobachtungen war $n = 8$, und

$$\sum_{i=1}^8 q_i = 7$$

hatten wir schon berechnet, vgl. Tab. 8. Da

$$|K| = |\frac{1}{2} \cdot 8 \cdot 7 - 2 \cdot 7| = 14 < 16 = K_{8;0,975} = K_{n;1-\alpha/2}$$

gilt, können wir die Nullhypothese H_0 nicht verwerfen.

Bei Verwendung des approximativen Tests ergibt sich

$$|K^*| = \frac{14}{\sqrt{8 \cdot 7 \cdot 21/18}} = \frac{14}{8,083} = 1,732 < 1,96 = u_{0,975} = u_{1-\alpha/2},$$

und wir kommen hier zum gleichen Ergebnis.

3. Die partielle Korrelation

Oft läßt sich eine Korrelation zwischen zwei Merkmalen X und Y nur deshalb finden, weil beide Merkmale mit einem dritten Merkmal U korreliert sind. Die Korrelation von X und Y ist dann eine reine **Scheinkorrelation**; vgl. auch Kap. I, Abschnitt 9.

Daher ist es vielmals von Interesse, eine **Korrelation zwischen X und Y unter Partialisierung eines Merkmals U**, d. h. die Korrelation von X und Y, die ohne den Einfluß von U vorhanden ist, zu bestimmen. Eine solche Korrelation $\rho_{(X,Y)/U}$ heißt auch kurz **partielle Korrelation von X und Y unter (Konstanthaltung von) U**. Sie ist gegeben als

$$\rho_{(X,Y)/U} = \frac{\rho_{XY} - \rho_{XU}\rho_{YU}}{\sqrt{(1 - \rho_{XU}^2)(1 - \rho_{YU}^2)}}.$$

3.1. Die partielle Korrelation zwischen normalverteilten Merkmalen

Sind die Merkmale X, Y und U in der interessierenden Grundgesamtheit (gemeinsam) normalverteilt, so kann man $\rho_{(X,Y)/U}$ aufgrund von je n Realisationen $x_1, \dots, x_n, y_1, \dots, y_n$ und u_1, \dots, u_n schätzen, indem man die einfachen Korrelationen ρ_{XY} , ρ_{XU} und ρ_{YU} mittels der Pearsonschen Korrelationskoeffizienten schätzt. Der Schätzer für die Korrelation zwischen den Merkmalen X und Y bei Partialisierung des Merkmals U ist dann

$$r_{(X,Y)/U} = \frac{r_{XY} - r_{XU}r_{YU}}{\sqrt{(1 - r_{XU}^2)(1 - r_{YU}^2)}}.$$

Ausgehend von diesem Schätzwert für die partielle Korrelation kann auch ein Test zum Niveau α auf partielle Unkorreliertheit bzw. Unabhängigkeit von X und Y unter U durchgeführt werden. Die Hypothese

$$H_0: \rho_{(X,Y)/U} = 0$$

wird im Test gegen die Alternative

$$H_1: \rho_{(X,Y)/U} \neq 0$$

zum Niveau α verworfen, falls

$$\left| \frac{r_{(X,Y)/U} \cdot \sqrt{n-3}}{\sqrt{1 - r_{(X,Y)/U}^2}} \right| > t_{n-3; 1-\alpha/2}$$

gilt. Die Quantile $t_{v,\gamma}$ der t-Verteilung mit v Freiheitsgraden sind im Anhang verfasst.

Beispiel: Bei der Untersuchung von $n = 142$ Frauen wurden drei als i. w. gemeinsam normalverteilt angenommene Merkmale beobachtet: Blutdruck (X), Cholesterin-Konzentration im Blut (Y) und Alter (U). Der Versuch, entnommen aus Swanson et al. (1955): "Blood values of women: Cholesterol" (Journal Gerontology, 10, S. 41), ergab für die einzelnen Korrelationen folgende Schätzwerte:

$$r_{XY} = 0,2495, \quad r_{XU} = 0,3332 \quad \text{und} \quad r_{YU} = 0,5029.$$

Wie man sieht, ist sowohl der Blutdruck als auch die Cholesterin-Konzentration mit dem Alter der Versuchspersonen korreliert. Daher muß man, um die eigentliche Korrelation zwischen Blutdruck und Cholesterin-Konzentration, die hier hauptsächlich von Interesse ist, zu schätzen, das Merkmal Alter partialisieren:

$$\begin{aligned} r_{(X,Y)/U} &= \frac{r_{XY} - r_{XU}r_{YU}}{\sqrt{(1 - r_{XU}^2)(1 - r_{YU}^2)}} = \frac{0,2495 - 0,3332 \cdot 0,5029}{\sqrt{(1 - 0,3332^2)(1 - 0,5029^2)}} \\ &= \frac{0,0819}{0,8150} = 0,1005. \end{aligned}$$

Die wirkliche Korrelation zwischen Blutdruck und Cholesterin-Konzentration wird nach „Eliminierung“ bzw. „Konstanthaltung“ des Einflusses des Merkmals Alter also viel geringer geschätzt als vorher. Zum Niveau $\alpha = 0,05$ soll nun die Hypothese

$$H_0: \rho_{(X,Y)/U} = 0 \quad \text{gegen} \quad H_1: \rho_{(X,Y)/U} \neq 0$$

getestet werden. Es ist

$$\left| \frac{0,1005 \cdot \sqrt{139}}{\sqrt{1 - 0,1005^2}} \right| = \left| \frac{1,1849}{0,9949} \right| = 1,191 < 1,97 = t_{139; 0,975},$$

so daß die Hypothese, daß Blutdruck und Cholesterin-Konzentration unter Partialisierung des Alters unabhängig sind, nicht verworfen werden kann.

Dagegen ist die einfache Korrelation zwischen Blutdruck und Cholesterin-Konzentration signifikant zum 5%-Niveau, denn, vgl. Abschnitt 1,

$$\left| \frac{r_{XY} \cdot \sqrt{n-2}}{\sqrt{1 - r_{XY}^2}} \right| = \left| \frac{0,2495 \cdot \sqrt{140}}{\sqrt{1 - 0,2495^2}} \right| = 3,048 > 1,97 = t_{140, 0,975}.$$

3.2. Der partielle Rangkorrelationskoeffizient nach Kendall

Sind die Merkmale X und Y und das zu partialisierende Merkmal U in der interessierenden Grundgesamtheit zumindest ordinal skaliert, so läßt sich die partielle Korrelation $\rho_{(X,Y)/U}$ mittels des partiellen Kendallschen Rangkorrelationskoeffizienten $\tau_{(X,Y)/U}$ schätzen. Den je n Beobachtungen $x_1, \dots, x_n, y_1, \dots, y_n$ und u_1, \dots, u_n werden die Rangzahlen zugeordnet: der jeweils kleinste (schlechteste) Wert erhält den Rang 1, ..., der jeweils größte (beste) Beobachtungswert erhält den Rang n. Man berechnet dann, vgl. Abschnitt 2.2, für die drei Merkmalspaare (X, Y) , (X, U) und (Y, U) Kendalls τ . Diese Schätzwerte seien hier mit $\tau_{X,Y}$, $\tau_{X,U}$ und $\tau_{Y,U}$ bezeichnet. Es ist dann

$$\tau_{(X,Y)/U} = \frac{\tau_{X,Y} - \tau_{X,U} \cdot \tau_{Y,U}}{\sqrt{(1 - \tau_{X,U}^2)(1 - \tau_{Y,U}^2)}}$$

der gesuchte Schätzer für die partielle Korrelation zwischen den Merkmalen X und Y unter U.

4. Die bi-partielle Korrelation

Ist ein Merkmal X mit einem Merkmal U, ein Merkmal Y mit einem Merkmal V korreliert und sind ferner die Merkmale U und V korreliert, so wird die Korrelation zwischen den Merkmalen X und Y von U und V stark beeinflußt.

Eine Untersuchung ergab zum Beispiel, daß in einer Gruppe von Schülern die Korrelation zwischen räumlichem Denkvermögen X und sprachlich-analytischer Begabung Y auf eine Korrelation der Leistungen in Mathematik U und Latein V zurückzuführen war. Die Korrelation zwischen X und V bzw. Y und U hingegen war viel geringer.

Will man dann aufgrund einer Stichprobe von n Objekten die eigentliche Korrelation zwischen X und Y schätzen, so empfiehlt es sich, das Merkmal X von U und das Merkmal Y von V zu partialisieren. Diese Korrelation

$$\rho_{X/U, Y/V} = \frac{\rho_{XY} - \rho_{XU}\rho_{YU} - \rho_{XV}\rho_{YV} + \rho_{XU}\rho_{UV}\rho_{YV}}{\sqrt{(1 - \rho_{XU}^2)(1 - \rho_{YV}^2)}}$$

heißt **bi-partielle Korrelation von X unter Partialisierung von U und Y unter Partialisierung von V**.

Sind die Merkmale X, Y, U und V in der interessierenden Grundgesamtheit (gemeinsam) normalverteilt, so kann die bi-partielle Korrelation $\rho_{X/U, Y/V}$ aufgrund von je n Realisationen $x_1, \dots, x_n, y_1, \dots, y_n, u_1, \dots, u_n$ und v_1, \dots, v_n mittels einfacher Pearson'scher Korrelationskoeffizienten geschätzt werden:

$$r_{X/U, Y/V} = \frac{r_{XY} - r_{XU}r_{YU} - r_{XV}r_{YV} + r_{XU}r_{UV}r_{YV}}{\sqrt{(1 - r_{XU}^2)(1 - r_{YV}^2)}}.$$

Zum Niveau α läßt sich dann auch die Hypothese

$$H_0: \rho_{X/U, Y/V} = 0,$$

daß X und Y bi-partiell unabhängig sind, gegen die Alternative

$$H_1: \rho_{X/U, Y/V} \neq 0$$

testen. Bei diesem nur approximativen Test wird die Nullhypothese H_0 zum Niveau α verworfen, falls gilt:

$$\left| \frac{r_{X/U, Y/V} \sqrt{n-3}}{\sqrt{1 - r_{X/U, Y/V}^2}} \right| > t_{n-3; 1-\alpha/2}.$$

Beispiel: An $n = 60$ Objekten wurden die Merkmale X, Y, U und V beobachtet; die einfachen Korrelationen zwischen je zweien dieser Merkmale ergaben sich zu

$$r_{XY} = -0,73, \quad r_{XU} = 0,89, \quad r_{YU} = -0,17, \quad r_{XV} = -0,31, \\ r_{YV} = 0,92 \quad \text{und} \quad r_{UV} = 0,42.$$

Berechnet man die bi-partielle Korrelation zwischen den Merkmalen X und Y unter Konstanthaltung der Einflüsse von U bzw. V

$$\begin{aligned} r_{X/U, Y/V} &= \frac{r_{XY} - r_{XU} r_{YU} - r_{XV} r_{YV} + r_{XU} r_{UV} r_{YV}}{\sqrt{(1 - r_{XU}^2)(1 - r_{YV}^2)}} \\ &= \frac{-0,73 - 0,89 \cdot (-0,17) - (-0,31) \cdot 0,92 + 0,89 \cdot 0,42 \cdot 0,92}{\sqrt{(1 - 0,89^2)(1 - 0,92^2)}} \\ &= \frac{0,0504}{\sqrt{0,0319}} = 0,2822, \end{aligned}$$

so zeigt sich, daß die scheinbar hohe negative Korrelation $r_{XY} = -0,73$ stark auf die Einflüsse von U und V zurückzuführen ist: Schaltet man diese Einflüsse aus, so sind die Merkmale X und Y schwach positiv korreliert. Ein Test von

$$H_0: \rho_{X/U, Y/V} = 0$$

zum 1% Niveau ergibt, daß diese bi-partielle Korrelation dann nicht signifikant verschieden von Null ist, denn es ist

$$\left| \frac{r_{X/U, Y/V} \sqrt{n-3}}{\sqrt{1 - r_{X/U, Y/V}^2}} \right| = \frac{0,2822 \sqrt{60-3}}{\sqrt{1 - 0,2822^2}} = 2,221 < 2,665 = t_{57; 0,995}.$$

5. Die multiple Korrelation

Die multiple Korrelation ist ein Maß für die Abhängigkeit eines Merkmals Y von p anderen Merkmalen X_1, \dots, X_p ; dabei wird vorausgesetzt, daß alle interessierenden Merkmale in der Grundgesamtheit (gemeinsam) normalverteilt sind. Sie ist definiert als die betragsmäßig größte (einfache) Korrelation unter den Korrelationen zwischen dem Merkmal Y und allen möglichen Linearkombinationen

$$a_1 X_1 + a_2 X_2 + \dots + a_p X_p$$

der Merkmale X_1, \dots, X_p mit beliebigen Gewichten a_1, \dots, a_p .

Beispiel: In der Fertigung eines Betriebes interessiert man sich für die Abhängigkeit der Lieferbereitschaft Y eines bestimmten Artikels von den Auftragseingängen X_1, \dots, X_p einer Sparte verwandter Artikel in der vorausgehenden Planungsperiode; dies führt natürlich auf eine multiple Korrelation.

Will man die multiple Korrelation $\rho_{Y,(X_1, \dots, X_p)}$ zwischen Y und X_1, \dots, X_p aufgrund einer Stichprobe vom Umfang n aus der interessierenden Grundgesamtheit schätzen, so mißt man die Ausprägungen von Y, X_1, \dots, X_p an jedem der n Objekte aus der Stichprobe und schätzt dann zunächst alle möglichen einfachen Korrelationen ρ_{YX_i} und $\rho_{X_i X_j}$ mittels der Pearsonschen Korrelationskoeffizienten.

Speziell für $p = 1$ wird die multiple Korrelation zwischen Y und X_1 durch deren betragliche einfache Korrelation

$$r_{Y,(X_1)} = |r_{YX_1}|,$$

für $p = 2$ durch

$$r_{Y,(X_1, X_2)} = \sqrt{\frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{YX_1}r_{YX_2}r_{X_1 X_2}}{1 - r_{X_1 X_2}^2}}$$

und für $p = 3$ durch

$$r_{Y,(X_1, X_2, X_3)} = \sqrt{\frac{a \cdot (b + c + d)}{1 + 2r_{X_1 X_2} \cdot r_{X_1 X_3} \cdot r_{X_2 X_3} - r_{X_1 X_2}^2 - r_{X_1 X_3}^2 - r_{X_2 X_3}^2}}$$

$$\text{mit } a = r_{YX_1}^2 \cdot (1 - r_{X_2 X_3}^2) + r_{YX_2}^2 \cdot (1 - r_{X_1 X_3}^2) + r_{YX_3}^2 \cdot (1 - r_{X_1 X_2}^2),$$

$$b = 2r_{YX_1} \cdot r_{YX_2} \cdot (r_{X_1 X_2} - r_{X_1 X_3} \cdot r_{X_2 X_3}),$$

$$c = 2r_{YX_1} \cdot r_{YX_3} \cdot (r_{X_1 X_3} - r_{X_1 X_2} \cdot r_{X_2 X_3}),$$

$$d = 2r_{YX_2} \cdot r_{YX_3} \cdot (r_{X_2 X_3} - r_{X_1 X_2} \cdot r_{X_1 X_3})$$

geschätzt. Allgemein ergibt sich als Schätzwert für die multiple Korrelation zwischen Y und X_1, \dots, X_p gerade

$$r_{Y,(X_1, \dots, X_p)} = \sqrt{r_{YX}^T R_{XX}^{-1} r_{YX}}$$

$$= \left[\begin{matrix} [r_{YX_1}, \dots, r_{YX_p}] & \begin{bmatrix} 1 & r_{X_1 X_2} & \dots & r_{X_1 X_p} \\ r_{X_1 X_2} & 1 & \dots & r_{X_2 X_p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{X_1 X_p} & r_{X_2 X_p} & \dots & 1 \end{bmatrix}^{-1} & \begin{bmatrix} r_{YX_1} \\ r_{YX_2} \\ \vdots \\ r_{YX_p} \end{bmatrix} \end{matrix} \right]^{1/2}.$$

Die Größe

$$B = B_{Y,(X_1, \dots, X_p)} = r_{Y,(X_1, \dots, X_p)}^2$$

nennt man auch das **Bestimmtheitsmaß** der multiplen Regression von Y auf X_1, \dots, X_p . Es ist ein Maß dafür, wie gut das Merkmal Y durch die Merkmale X_1, \dots, X_p erklärt wird, vgl. auch Kap. X, in dem auch eine andere Berechnungsmöglichkeit für das Bestimmtheitsmaß mittels Regressionskoeffizienten angegeben wird. B gibt den Anteil an der Varianz von Y wieder, der durch die Merkmale X_1, \dots, X_p erklärt werden kann.

Die multiple Korrelation zwischen einem Merkmal Y und den p Merkmalen X_1, \dots, X_p ist gerade dann Null, wenn alle einfachen Korrelationen $\rho_{YX_1}, \dots, \rho_{YX_p}$ gleich Null sind. Will man die Hypothese

$$H_0: \rho_{Y,(X_1, \dots, X_p)} (= \rho_{YX_1} = \dots = \rho_{YX_p}) = 0$$

zum Niveau α gegen die Alternative

$$H_1: \text{es gibt ein } \rho_{YX_i} \neq 0$$

testen, so verwendet man als Prüfgröße

$$F = \frac{r_{Y,(X_1, \dots, X_p)}^2(n - 1 - p)}{p(1 - r_{Y,(X_1, \dots, X_p)}^2)},$$

wobei n die Anzahl der Objekte in der Stichprobe bezeichnet.

Die Hypothese H_0 wird zum Niveau α verworfen, falls

$$F > F_{p, n-1-p, 1-\alpha}$$

gilt, wobei $F_{p, v, \gamma}$ das γ -Quantil der $F_{p, v}$ -Verteilung bezeichnet. Diese Quantile sind im Anhang vertafelt.

Beispiel: (vgl. Haiger (1978): „Biometrische Methoden der Tierproduktion“. Österreichischer Agrarverlag, Wien) Ausgehend von einer Stichprobe 48 kastrierter männlicher Schweine werden Korrelationen zwischen den Merkmalen Y Fleisch-Fettverhältnis im Schlachtkörper, X_1 Rückenspeckdicke und X_2 Karreefläche geschätzt. Es ergab sich

$$r_{YX_1} = -0,52 \quad r_{YX_2} = 0,47 \quad \text{und} \quad r_{X_1 X_2} = -0,07.$$

Hieraus ersieht man, daß Schweine mit gutem Fleisch-Fettverhältnis im Durchschnitt eine geringere Rückenspeckdicke und eine große Karreefläche haben. Rückenspeckdicke und Karreefläche sind dagegen praktisch unkorreliert. Die multiple Korrelation zwischen Fleisch-Fettverhältnis im Schlachtkörper und den Merkmalen Rückenspeckdichte und Karreefläche ergibt sich nun zu

$$\begin{aligned} r_{Y,(X_1, X_2)} &= \sqrt{\frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{YX_1}r_{X_1 X_2}r_{YX_2}}{1 - r_{X_1 X_2}^2}} \\ &= \sqrt{\frac{(-0,52)^2 + 0,47^2 - 2(-0,52)(-0,07) \cdot 0,47}{1 - (-0,07)^2}} \\ &= \sqrt{\frac{0,4571}{0,9951}} = 0,68. \end{aligned}$$

Wegen

$$F = \frac{0,68^2 \cdot (48 - 1 - 2)}{2(1 - 0,68^2)} = 19,35 > 3,21 = F_{2, 45, 0,95}$$

erweist sich die multiple Korrelation zwischen Fleisch-Fettverhältnis und (Rückenspeckdicke, Karreefläche) als zum 5% Niveau signifikant verschieden von Null.

Berechnet man nun die Bestimmtheitsmaße

$$B_{Y, X_1} = r_{YX_1}^2 = (-0,52)^2 = 0,27, \quad B_{Y, X_2} = r_{YX_2}^2 = 0,47^2 = 0,22$$

und das multiple Bestimmtheitsmaß

$$B_{Y,(x_1 x_2)} = r_{Y,(x_1 x_2)}^2 = 0,68^2 = 0,46,$$

so läßt sich sagen, daß 27% der Schwankungen im Fleisch-Fettverhältnis durch Unterschiede in der Rückenspeckdicke und 22% der Schwankungen durch Unterschiede in der Karreefläche erkläbar sind. Insgesamt werden durch Rückenspeckdicke und Karreefläche gemeinsam sogar 46% der Varianz des Fleisch-Fettverhältnisses erklärt.

Will man testen, ob $\rho_{Y,(x_1 \dots x_p)}$ gleich einem Wert $\rho_0 \neq 0$ ist, so muß man eine Tafel der Quantile der nichtzentralen F-Verteilung oder eine spezielle Tafel der Verteilung des multiplen Korrelationskoeffizienten verwenden. Eine solche findet man etwa bei Graybill (1976).

6. Ein Test auf Unabhängigkeit von p Meßreihen

In diesem Abschnitt wird, nachdem im Abschnitt 1 schon Tests auf Unabhängigkeit von 2 Meßreihen behandelt wurden, ein Unabhängigkeitstest für p Meßreihen der Länge n vorgestellt.

Die Beobachtungen x_{i1}, \dots, x_{in} der i-ten Meßreihe mögen einer $N(\mu_i, \sigma_i^2)$ -Verteilung entstammen (für $i = 1, \dots, p$) und die Meßreihen i und k für $i \neq k$ die Korrelation ρ_{ik} besitzen, dann können mittels **multipler Vergleiche nach dem Prinzip von Holm** zum Niveau α die Nullhypotesen

$$H_{0i}: \rho_{ik} = 0 \quad \text{gegen} \quad H_{1i}: \rho_{ik} \neq 0$$

für $1 \leq i < k \leq p$ getestet werden. Es handelt sich hier somit um einen simultanen Test auf paarweise Unabhängigkeit der p Meßreihen. Für $1 \leq i < k \leq p$ berechnet man den üblichen Schätzer (Pearson'scher Korrelationskoeffizient)

$$r_{ik} = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)}{\sqrt{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \cdot \sum_{j=1}^n (x_{kj} - \bar{x}_k)^2}}$$

und ordnet dann die $p(p-1)/2$ Prüfgrößen

$$R_{ik} = |r_{ik}| \sqrt{\frac{n-2}{1-r_{ik}^2}}$$

nach absteigender Größe

$$R_{i_1 k_1} \geq R_{i_2 k_2} \geq \dots \geq R_{i_{p(p-1)/2} k_{p(p-1)/2}}.$$

Dann überprüft man sukzessive für $m = 1, \dots, p(p-1)/2$, ob

$$R_{i_m k_m} < t_{n-2; 1-\alpha/(p(p-1)+2-2m)}$$

gilt. Ist dies für ein m das erste Mal der Fall, so sind die Korrelationen $\rho_{i_1 k_1}, \dots, \rho_{i_{m-1} k_{m-1}}$ als zum Niveau α signifikant verschieden von Null zu betrachten. Im Gegensatz dazu sind alle verbleibenden Korrelationen nicht signifikant verschieden von Null.

Die Paare $(i_1, k_1), \dots, (i_{m-1}, k_{m-1})$ von je zwei Meßreihen sind demnach (mit einer Gesamt-Irrtumswahrscheinlichkeit α) nicht unabhängig.

Läßt sich ein solches m jedoch nicht finden, so sind alle Meßreihen als paarweise korreliert zu betrachten.

Beispiel: Zum Schleifen optischer Linsen stehen $p = 5$ Maschinen zur Auswahl. An jeder der Maschinen werden $n = 10$ Linsen geschliffen und jeweils die Abweichung vom Sollwert gemessen. Die Meßergebnisse x_{ij} sind in Tab. 9 angegeben. Alle Aussagen sollen zum 5% Niveau gesichert werden.

Tab. 9: Versuchsergebnisse für das Schleifen optischer Linsen

Linse j	Maschine i				
	1	2	3	4	5
1	5	2	7	3	-2
2	4	0	5	4	1
3	7	-1	-9	-2	3
4	-2	1	-3	5	-3
5	0	0	0	-4	2
6	3	1	-6	-1	-5
7	6	-3	5	3	0
8	1	4	3	7	4
9	1	2	-1	-4	-2
10	-5	0	0	-3	1

In Tab. 10 sind alle Korrelationen r_{ik} sowie die Prüfgrößen R_{ik} für die multiplen Vergleiche nach Holm angegeben.

Tab. 10: Geschätzte Korrelationen r_{ik} und Prüfgrößen R_{ik} für multiple Vergleiche nach Holm zu den Daten aus Tab. 9

i	k	r_{ik}	R_{ik}
1	2	-0,3130	0,9321
1	3	0,0581	0,1646
1	4	0,1760	0,5057
1	5	0,0522	0,1478
2	3	0,0963	0,2736
2	4	0,2632	0,7716
2	5	-0,0494	0,1399
3	4	0,4790	1,5434
3	5	0,1307	0,3729
4	5	0,0752	0,2133

Bei den multiplen Vergleichen nach Holm ist für $m = 1$

$$R_{i_1 k_1} = R_{34} = 1,5434 < 3,8325 = t_{8,1-0,05/20} = t_{y,0,9975};$$

somit kann keine der Korrelationen r_{ik} signifikant verschieden von Null sein, d. h. nach diesem simultanen Verfahren sind zum Gesamt-Niveau von 5% in keinem Meßreihenpaar die Reihen signifikant abhängig.

Kapitel X: Regressionsanalyse

In der **Korrelationsanalyse**, vgl. Kap. IX, erhält man ein **quantitatives Maß für einen Zusammenhang**, während in der **Regressionsanalyse** ein **funktionaler Zusammenhang spezifiziert** wird. Ziele sind dabei unter anderem

- Nachweis einer bekannten Beziehung,
- Schätzen der Parameter einer bekannten funktionalen Beziehung,
- Erkennen eines funktionalen Zusammenhangs,
- Empirische Repräsentation großer Datenmengen,
- Interpolation fehlender bzw. Prognose zukünftiger Werte.

Interessiert man sich nur für den Zusammenhang zweier Merkmale X und Y mit den **Ausprägungsvariablen** x und y, **Regressor** und **Regressand** genannt, so wird eine funktionale Beziehung $y = f(x)$ spezifiziert. Man spricht dann auch von einer **Regression von Y auf X**. Dazu zieht man aus der interessierenden Grundgesamtheit, welche die Merkmale X und Y trägt, eine Stichprobe vom Umfang n, die allerdings auch die gesamte Grundgesamtheit darstellen kann – wie etwa bei kontrollierten Versuchen oder Totalerhebungen. Die Elemente dieser Stichprobe, unsere Untersuchungsobjekte, tragen dann die Ausprägungskombinationen $(x_1, y_1), \dots, (x_n, y_n)$. Die Ausprägungen x_1, \dots, x_n in dieser Stichprobe können vorgegeben werden und somit die Auswahl bedingt haben (geschichtete Stichprobe) oder miterhoben werden (einfache Zufallsauswahl). Mittels dieser Daten wird dann obiger funktionaler Zusammenhang $y = f(x)$ geschätzt.

Der Name **Regression** wurde von Galton (1889) in seinen Studien zur Vererbung geprägt. Als „*Gesetz der universalen Regression*“ formulierte er:

„Each peculiarity in a man is shared by his kinsman *but on the average* in a less degree“,

d. h. jede vom „Normalen“ abweichende Eigenschaft eines Menschen wird in der nachfolgenden Generation zwar übernommen, aber im Durchschnitt in einem geringeren Maße. Es tritt also bzgl. dieser Eigenschaft ein Rückschritt (Regression) ein.

Galtions Freund Karl Pearson hat in 1078 Familien die Größe von Vater und Sohn untersucht, vgl. Pearson/Lee (1903). In Abb. 1 ist die Regression der Größe des Sohnes auf die des Vaters dargestellt. Obwohl große Väter dazu neigen, große Söhne zu haben, ist jedoch die durchschnittliche Größe von Söhnen zu großer Vätern geringer als die ihrer Väter. Ebenso wird die Besonderheit der Kleinheit von Vätern nicht in vollem Maße an die Söhne vererbt, denn diese sind durchschnittlich größer als ihre Väter. Es ist also eine Regression oder ein Rückschritt in Bezug auf die Besonderheiten in der Eigenschaft Größe bei den Söhnen sichtbar. Dies wird deutlich am Regressionskoeffizienten 0,516, der wesentlich kleiner ist als 1, was besagt: Ist ein Vater um 1 cm größer als ein anderer, so wird im Durchschnitt sein Sohn nur um 0,516 cm größer sein als der des anderen Vaters, bzw. umgekehrt ist der Vater um 1 cm kleiner (in Bezug auf dessen Vater), so wird sein Sohn im Durchschnitt nur 0,516 cm kleiner werden, als wenn sein Vater nicht um 1 cm kleiner wäre, also ein Rückschritt in der Verkleinerung.

Einige weitere typische Beispiele seien hier genannt.

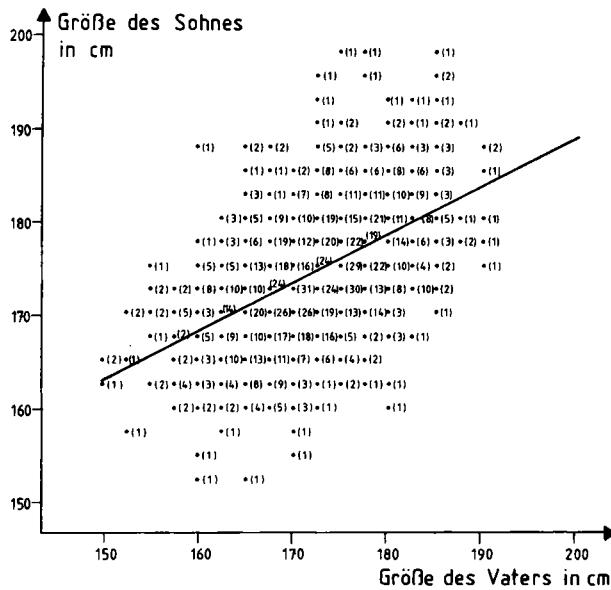


Abb. 1: Galton-Pearson-Regression: Zusammenhang zwischen der Größe von Vätern (x) und ihren Söhnen (y), $y = 85,6742 + 0,561x$

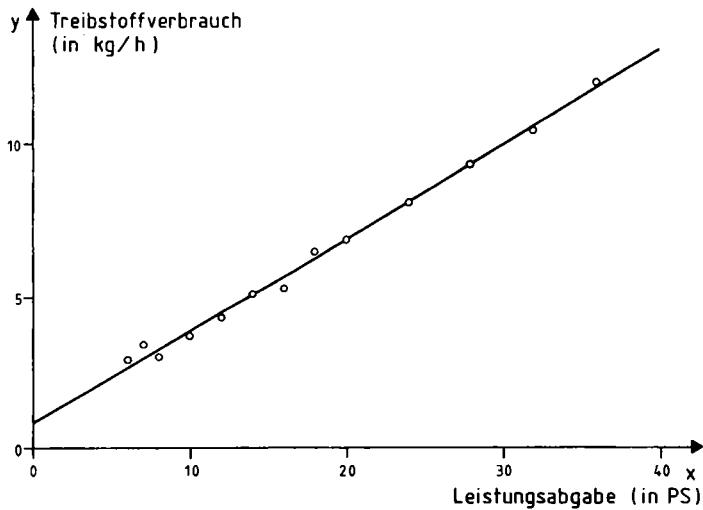


Abb. 2: Zusammenhang zwischen Treibstoffverbrauch (in kg pro Std) und Leistungsabgabe (in PS) eines Motors bei konstanter Drehzahl: $y = 0,854 + 0,302x$

In Abb. 2 ist die Regression von Treibstoffverbrauch (in kg/h) auf die Leistungsabgabe (in PS) eines Motors bei konstanter Drehzahl dargestellt. Der Regressionskoeffizient 0,302 zeigt, daß der Treibstoffverbrauch nur um 0,302 kg/h ansteigt, wenn die Leistungsabgabe um ein PS steigt. Bei großer Leistung wird also relativ wenig Treibstoff verbraucht.

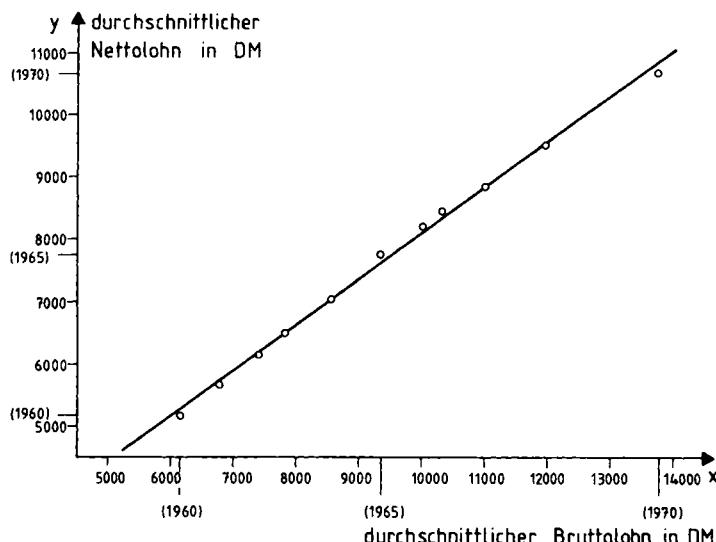


Abb. 3: Beziehung zwischen durchschnittlichem Bruttolohn (-gehalt) x und Nettolohn (-gehalt) y bei Beschäftigten für die Jahre 1960 bis 1970 in der Bundesrepublik (Daten: Statistisches Bundesamt); Regressionsgerade $y = 806,985 + 0,727x$

Auf der Datenbasis der Jahre 1960 bis 1970 zeigt **Abb. 3** die Regression vom durchschnittlichen Nettolohn(-gehalt) auf den durchschnittlichen Bruttolohn (-gehalt) bei Beschäftigten in der Bundesrepublik Deutschland. Der Regressionskoeffizient 0,727 macht deutlich, daß der Nettolohn in relativ geringerem Maße ansteigt als der Bruttolohn: Steigt der Bruttolohn um 1000 DM, so steigt der Nettolohn nur um 727 DM; man vgl. auch Abschnitt 1.2.3 in Kap. XII.

Die gesamtwirtschaftliche **Konsumfunktion** für die Bundesrepublik Deutschland (Daten: Statistisches Bundesamt) für die Jahre 1960 bis 1970, d.h. die Regression von privatem Konsum (y) auf das verfügbare private Einkommen (x), liefert die Regressionsgerade (in Millionen DM)

$$y = 18784,341 + 0,8227039x$$

Der Regressionskoeffizient 0,8227039 besagt, daß der private Konsum in erheblich geringerem Maße ansteigt als das verfügbare private Einkommen: Steigt das verfügbare private Einkommen um 1 Million DM, so steigt der private Konsum lediglich um 822703,9 DM.

Als weitere ökonomische Regressionsbeispiele seien hier **Kostenfunktionen** für eine Strumpffabrik und für Stahlerzeugung vorgestellt. Aufgrund von monatlichen Daten einer Strumpffabrik der Jahre 1935 bis 1939 stellte Dean (Dean (1941): „Statistical cost functions of a hosiery mill“, Studies in Business Administration, School of Business, University of Chicago, Vol. 11, 4, S. 19ff) die Kostenfunktion

$$y = 2935,59 + 1,998x$$

auf. Es handelt sich hier um eine Regression von den Gesamtkosten y (in Dollar) auf die erzeugte Menge x (in Dutzend Strümpfen): Wird ein Dutzend Strümpfe zusätzlich hergestellt, so steigen die Gesamtkosten um 1,998 Dollar.

Basierend auf jährlichen Daten schätzte Yntema (Yntema (1940): „Steel prices, volume and cost“, United States Steel Corporation, T.N.E.C. Papers, Band 1, S. 231ff) eine Kostenfunktion

$$y = 182,1 + 55,734x$$

für die Stahlerzeugung bei der US Steel Corporation. Dabei bezeichnet y die Gesamtkosten der Stahlerzeugung (in Mill. Dollar) und x die Menge des verschifften Stahls (in Mill. Tonnen): Wird eine Million Tonnen zusätzlich erzeugt, so steigen die Gesamtkosten um 55,734 Millionen Dollar.

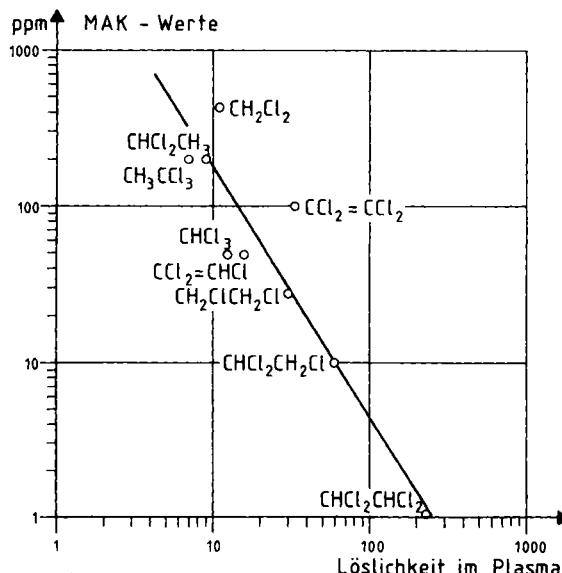


Abb. 4: Abhängigkeit der chronischen Toxizität (MAK-Werte) chlorierter aliphatischer Kohlenwasserstoffe von der Löslichkeit in Plasma (nach: Forth/Henschler/Rumme (1977), Pharmakologie und Toxikologie, BI Wissenschaftsverlag, Mannheim)

In Abb. 4 ist eine Regressionsgerade eingezeichnet, die, anders als bei den vorherigen Beispielen, lediglich zur Überprüfung dient und die Abhängigkeit der chronischen Toxizität (MAK-Werte) chlorierter aliphatischer Kohlenwasserstoffe von der Löslichkeit in Plasma veranschaulichen soll.

Ausführlich wird in Abschnitt 1 auf die Regression von Gewicht auf Körpergröße beim Menschen eingegangen.

Bis jetzt haben wir nur Beispiele zur **linearen Regression**, die in Abschnitt 1 behandelt wird, kennengelernt. In Abschnitt 2 wird die **Residualanalyse** als Instrument zur Überprüfung eines Regressionsansatzes vorgestellt.

Auch sei schon an dieser Stelle erwähnt, daß sich viele nichtlineare Funktionen $y = f(x)$ linearisieren lassen, vgl. Abschnitt 3. Manchmal kann es nützlich sein, Daten vor der Regression zu transformieren. Solche Datentransformationen sind in Kapitel V, Abschnitt 4.5 beschrieben.

Mit Hilfe der Regressionsrechnung können auch **nichtlineare Zusammenhänge**

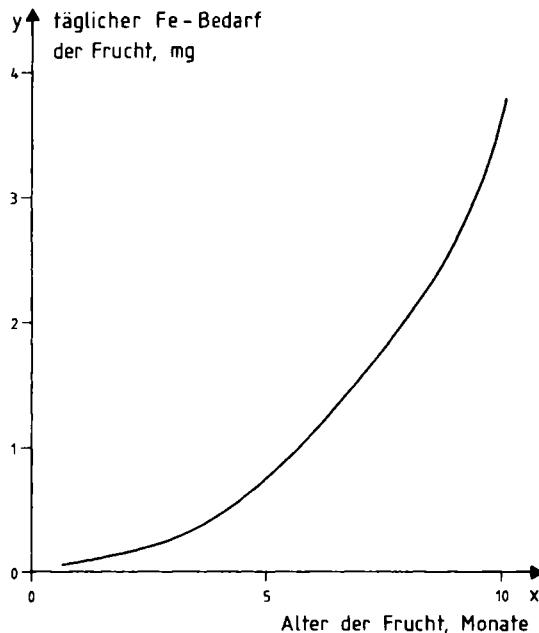


Abb. 5: Beziehung zwischen Eisenbedarf der Frucht und Alter der Frucht (nach: Pribilla et al. (1958) in "Iron in chemical medicine", Hrsg.: Wallenstein / Mettier, Univ. Calif. Press, Los Angeles)

zwischen zwei Merkmalen beschrieben werden. Als Beispiel zeigt **Abb. 5** die Regression vom täglichen Eisenbedarf y der Frucht in mg auf das Alter x der Frucht. Vor allem Ertrags- und Leistungsfunktionen zeigen häufig einen Verlauf wie in **Abb. 6**, wo die Regression vom Körnerertrag y auf die Dosierung x einer Düngung dargestellt ist: Bis zu einer bestimmten Düngungsgrenze steigt der Körnerertrag an, um dann wieder abzufallen (Daten aus Laur (1930): „Wirtschaftslehre des Landbaus“, 2. Auflage, Berlin, S. 143).

Solche Regressionsprobleme kann man mit Hilfe der **polynomialen Regression**, vgl. Abschnitt 4, wo **nichtlineare Regressionsprobleme** behandelt werden, lösen. In Abschnitt 5 schließlich wird noch auf die **multiple Regression** zur Analyse des funktionalen Zusammenhangs zwischen $k + 1$ Merkmalen Y, X_1, \dots, X_k eingegangen; man vgl. aber auch Abschnitt 1.2, 3.5 und 5 in Kap. XII.

Die in diesem Kapitel auftauchenden Quantile $t_{v,y}$ der t_v -Verteilung, $\chi^2_{v,y}$ der χ^2_v -Verteilung und $F_{v_1,v_2,y}$ der F_{v_1,v_2} -Verteilung sind im Anhang vertafelt.

1. Lineare Regression

Vermutet man einen zumindest näherungsweise linearen Zusammenhang zwischen zwei Merkmalen in einer Grundgesamtheit, so kann mittels **linearer Regressionsrechnung** dieser Zusammenhang näher spezifiziert und untersucht werden. Dazu werden in einem Experiment die beiden interessierenden Merkmale X und Y an n Objekten aus der Grundgesamtheit beobachtet. Die $2n$ Realisationen x_1, \dots, x_n

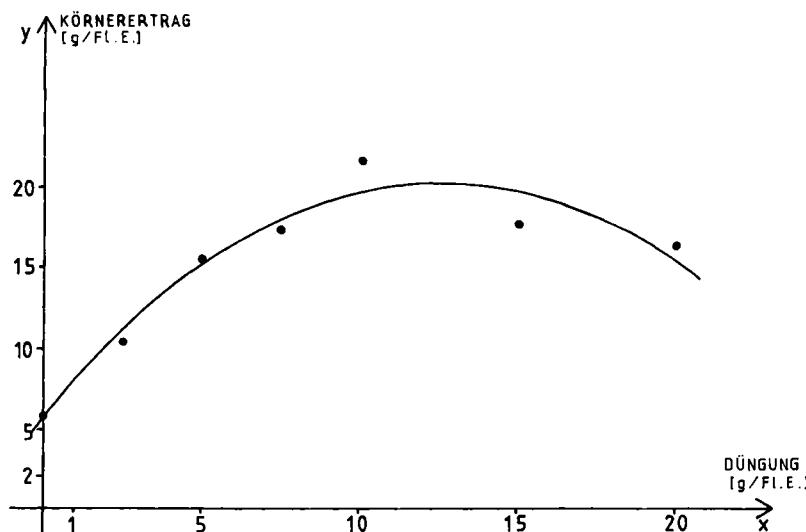


Abb. 6: Quadratische Regressionsfunktion für den Zusammenhang zwischen Körnerertrag y und Düngung x: $y = 5,778 + 2,299x - 0,091x^2$

und y_1, \dots, y_n werden dann zur Untersuchung der linearen Beziehung verwandt: Man geht davon aus, daß gilt

$$y_i = \alpha + \beta x_i + e_i \quad \text{für } i = 1, \dots, n,$$

wobei α das **Absolutglied** und β den **Steigungsparameter** der linearen Beziehung darstellen und e_1, \dots, e_n **zufällige Fehler** sind. Im folgenden werden u.a. Punkt-, Intervallschätzungen und Tests für die Parameter α und β des Regressionsproblems angegeben, und die zufälligen Fehler e_i werden näher analysiert.

1.1. Die Methode der kleinsten Quadrate

Die Punktschätzungen a und b für die Parameter α und β eines linearen Regressionsproblems wollen wir so bestimmen, daß durch die **Regressionsgerade**

$$\hat{y} = a + bx$$

eine „möglichst gute Schätzung“ \hat{y} für die Ausprägung y des Merkmals Y eines Objektes, das die Ausprägung x des Merkmals X trägt, bestimmt ist. Als Kriterium für die Güte dieser Schätzung ist die **Summe der Abweichungsquadrat (Residual-Quadratsumme)**

$$S^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

ein geeignetes Maß: Die Schätzer a und b für α und β sollen so bestimmt werden, daß die Summe S^2 der vertikalen quadratischen Abweichungen der gemessenen Werte y_i von den durch die Regressionsgerade an den Stellen x_i gelieferten Werten $\hat{y}_i = a + bx_i$ minimal wird (**Methode der kleinsten Quadrate**), vgl. auch Abb. 7.

Somit ergeben sich a und b als Lösungen des Normalengleichungssystems

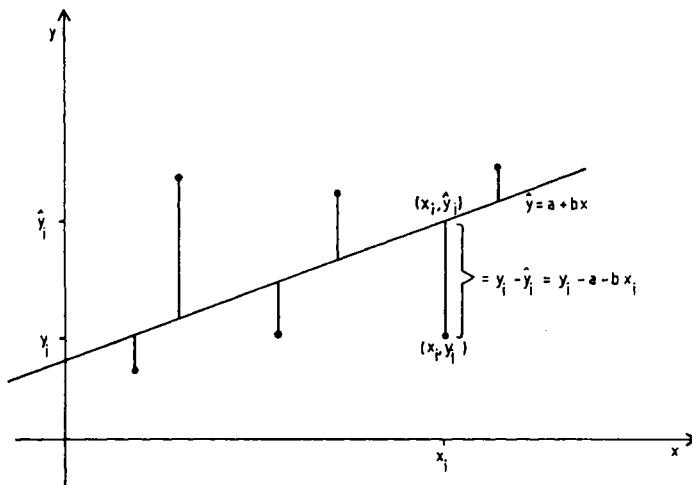


Abb. 7: Graphische Veranschaulichung der Methode der kleinsten Quadrate

$$\frac{\partial S^2}{\partial a} = -2 \sum_{i=1}^n (y_i - a - bx_i) \stackrel{!}{=} 0,$$

$$\frac{\partial S^2}{\partial b} = -2 \sum_{i=1}^n x_i(y_i - a - bx_i) \stackrel{!}{=} 0;$$

d.h. der Schätzer für \$\beta\$ ist

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

und der Schätzwert für \$\alpha\$ ist

$$a = \bar{y} - b\bar{x};$$

a und b heißen auch **Kleinste-Quadrate-Schätzer (least squares estimators)** für \$\alpha\$ bzw. \$\beta\$; natürlich bezeichnen hier \$\bar{x}\$ und \$\bar{y}\$ die Mittelwerte der \$x_i\$ bzw. der \$y_i\$.

Beispiel: Bei einer Stichprobe von \$n = 40\$ Personen wurden Körpergröße X und Gewicht Y gemessen. Da man eine lineare Beziehung zwischen Gewicht und Körpergröße vermutet, möchte man die Parameter \$\alpha\$ und \$\beta\$ dieser Beziehung schätzen. Die beobachteten Merkmalsausprägungen sind in Tab. 1 angegeben und in Abb. 8 graphisch dargestellt; aus ihnen ergeben sich die Schätzwerte für a und b nach der Methode der kleinsten Quadrate zu

$$b = \frac{\sum_{i=1}^{40} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{40} (x_i - \bar{x})^2} = \frac{2906,30}{3015,1} = 0,9639 \quad \text{und}$$

$$a = \bar{y} - b\bar{x} = 68,05 - 0,9639 \cdot 170,35 = -96,1504.$$

Wir können also das Gewicht einer Person mit Körpergröße x schätzen durch

$$\hat{y} = a + bx = -96,1504 + 0,9639x.$$

Diese Regressionsgerade ist in Abb. 9 eingetragen; sie unterstützt die Faustregel: Gewicht = Größe - 100.

Tab. 1: Körpergröße x_i , Gewicht y_i , geschätztes Gewicht \hat{y}_i , Residuen \hat{e}_i und normierte Residuen d_i von 40 Personen
(Die Größen d_i werden erst in Abschnitt 2 benötigt)

i	x_i	y_i	\hat{y}_i	$\hat{e}_i = y_i - \hat{y}_i$	$d_i = \hat{e}_i/s$
1	175	75	72,532	2,468	0,311
2	175	73	72,532	0,468	0,059
3	184	74	81,207	-7,207	-0,909
4	180	82	77,352	4,648	0,586
5	173	77	70,604	6,396	0,807
6	173	70	70,604	-0,604	-0,076
7	184	88	81,207	6,793	0,857
8	179	68	76,388	-8,388	-1,058
9	168	60	65,785	-5,785	-0,730
10	183	82	80,243	1,757	0,222
11	178	79	75,424	3,576	0,451
12	185	86	82,171	3,829	0,483
13	180	64	77,352	-13,352	-1,684
14	175	70	72,532	-2,532	-0,319
15	173	64	70,604	-6,604	-0,833
16	180	91	77,352	13,648	1,721
17	185	84	82,171	1,829	0,231
18	170	81	67,712	13,288	1,676
19	178	72	75,424	-3,424	-0,432
20	166	67	63,857	3,143	0,396
21	168	66	65,785	0,215	0,027
22	160	45	58,074	-13,074	-1,649
23	157	47	55,182	-8,182	-1,032
24	158	61	56,146	4,854	0,612
25	163	57	60,965	-3,965	-0,500
26	168	57	65,785	-8,785	-1,108
27	170	69	67,712	1,288	0,162
28	163	76	60,965	15,035	1,896
29	158	54	56,146	-2,146	-0,271
30	166	61	63,857	-2,857	-0,360
31	167	62	64,821	-2,821	-0,356
32	161	51	59,038	-8,038	-1,014
33	172	62	69,640	-7,640	-0,964
34	156	50	54,218	-4,218	-0,532
35	162	63	60,001	2,999	0,378
36	168	56	65,785	-9,785	-1,234
37	160	72	58,074	13,926	1,757
38	156	60	54,218	5,782	0,729
39	164	81	61,929	19,071	2,405
40	173	65	70,604	-5,604	-0,707

Eine nach der Methode der kleinsten Quadrate geschätzte Regressionsgerade schneidet stets den Punkt (\bar{x}, \bar{y}) . Hat man, wie von jetzt an unterstellt, daß die Beobachtungswerte x_1, \dots, x_n fehlerfrei gemessen sind und die Fehlerterme e_1, \dots, e_n unabhängig sind mit Erwartungswert 0 und Varianz σ^2 , so sind die Kleinst-Quadrat-Schätzer a und b erwartungstreu mit den Varianzen

$$\sigma_a^2 = \left(\frac{1}{n} + \bar{x}^2 \left| \sum_{i=1}^n (x_i - \bar{x})^2 \right. \right) \sigma^2 \quad \text{bzw.} \quad \sigma_b^2 = \sigma^2 \left| \sum_{i=1}^n (x_i - \bar{x})^2 \right.$$

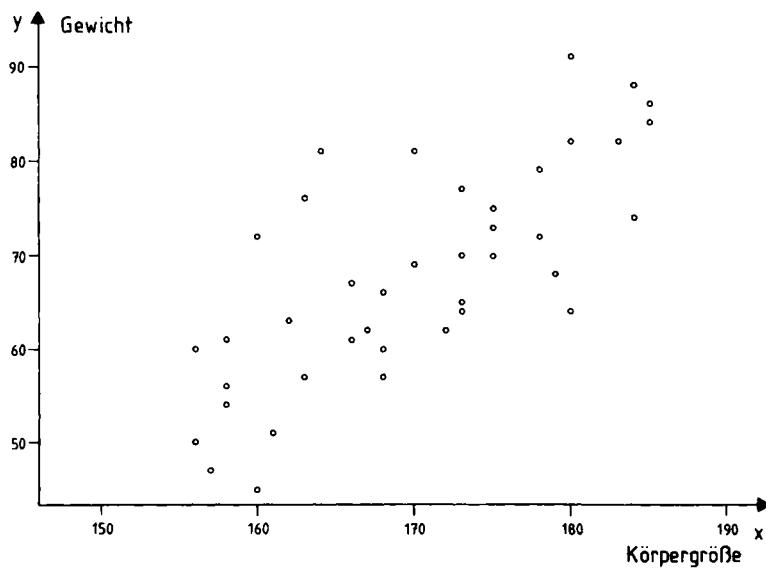


Abb. 8: Größe x und Gewicht y von 40 Personen

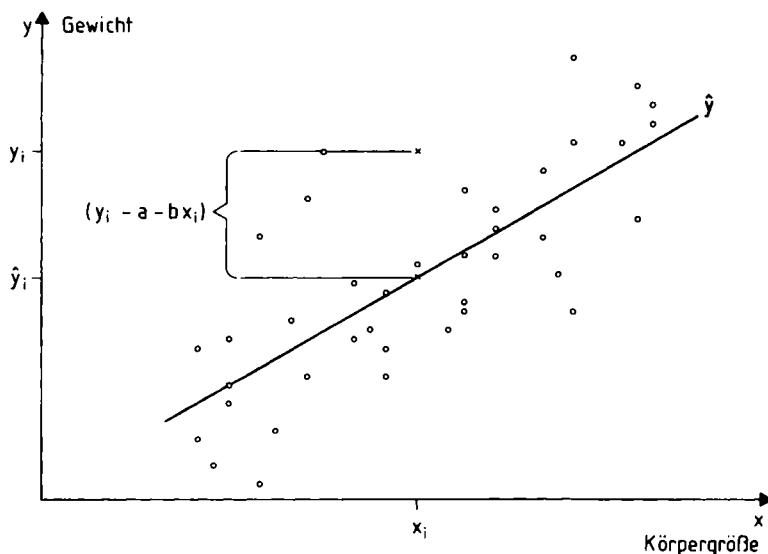


Abb. 9: Gewicht und Körpergröße von 40 Personen und die daraus geschätzte Regressionsgerade

Gilt zusätzlich – was wir auch im Folgenden annehmen – daß die Fehlerterme normalverteilt sind, so sind auch die Schätzer a und b normalverteilt:

$$a \sim N(\alpha, \sigma_a^2) \quad \text{und} \quad b \sim N(\beta, \sigma_b^2).$$

Ihre Kovarianz ist

$$\text{Cov}(a, b) = -\bar{x}\sigma^2 \left/ \sum_{i=1}^n (x_i - \bar{x})^2 \right.$$

und die Korrelation zwischen a und b , vgl. Abschnitt 1 in Kap.IX, ist

$$\rho_{a,b} = -n\bar{x}/\sqrt{n \sum x_i^2},$$

Beispiel: In unserem Beispiel, vgl. Tab. 1, ergibt sich unter der Annahme

$$e_i \sim N(0, \sigma^2) \quad \text{für } i = 1, \dots, 40$$

gerade

$$b \sim N(\beta, \sigma^2/3015,1), \quad a \sim N(\alpha, 9,65\sigma^2),$$

$$\text{Cov}(a, b) = -\frac{170,35}{3015,1} \sigma^2 = -0,0565\sigma^2 \quad \text{und}$$

$$\rho_{a,b} = -6814/6822,84 = -0,9987.$$

1.2. Schätzen der Fehlervarianz σ^2

Da die Varianzen σ_a^2 und σ_b^2 der Kleinsten-Quadrate-Schätzer von der Varianz σ^2 der zufälligen Fehler e_1, \dots, e_n abhängen, möchte man σ^2 i. a. erwartungstreu schätzen. Man schätzt dazu für $i = 1, \dots, n$ die **theoretischen Residuen**

$$e_i = y_i - \alpha - \beta x_i \quad \text{durch} \quad \hat{e}_i = y_i - a - bx_i$$

und nimmt als Schätzung für σ^2 die Größe

$$\begin{aligned} s^2 &= \sum_{i=1}^n (y_i - a - bx_i)^2 / (n - 2) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - 2) \\ &= \sum_{i=1}^n \hat{e}_i^2 / (n - 2) = S^2 / (n - 2), \end{aligned}$$

die unter allen erwartungstreuen Schätzern für σ^2 die kleinste Varianz hat.

Beispiel: In unserem Beispiel sind die Größen

$$\hat{e}_i = y_i - \hat{y}_i = y_i - a - bx_i = y_i + 96,1504 - 0,9639x_i$$

bereits in Tab. 1 mit angegeben; die \hat{y}_i sind natürlich die Schätzwerte der Gewichte der 40 Personen aus der Regressionsgeraden.

Hieraus ergibt sich als Schätzer für σ^2 gerade

$$s^2 = S^2 / (n - 2) = 2388,519 / (40 - 2) = 62,856,$$

so daß

$$s_a^2 = 9,65 s^2 = 606,560, \quad s_b^2 = s^2 / 3015,1 = 0,021 \quad \text{bzw.}$$

$$s_{a,b} = -0,0565 s^2 = -3,551$$

Schätzungen für die Varianzen σ_a^2 und σ_b^2 bzw. die Kovarianz $\text{Cov}(a, b)$ sind.

1.3. Zusammenhang zwischen Regressions- und Korrelationsrechnung; das Bestimmtheitsmaß

Der Pearson'sche Korrelationskoeffizient r_{XY} zweier Merkmale X und Y , den wir in Kap.IX kennengelernt haben, steht in engem Zusammenhang zum Schätzer b des Steigungsparameters β : Es ist nämlich

$$\begin{aligned}
 b &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= \left(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right) \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \\
 &\quad \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= r_{XY} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \\
 &= r_{XY} s_Y / s_X.
 \end{aligned}$$

Aus der Korrelation r_{XY} kann also der Steigungsparameter β der linearen Regression von Y auf X geschätzt werden und umgekehrt.

Als Maß für die Güte der Anpassung, die eine Regression erzielt, dient das **Bestimmtheitsmaß der Regression**:

$$\begin{aligned}
 B_{Y,X} &= 1 - \sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum_{i=1}^n (y_i - \bar{y})^2 \\
 &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / \sum_{i=1}^n (y_i - \bar{y})^2 \\
 &= s_Y^2 / s_X^2,
 \end{aligned}$$

also das Verhältnis der Varianz der geschätzten Werte \hat{y}_i zur Varianz der beobachteten Werte y_i , mithin der Anteil an der Varianz von Y, der durch das Merkmal X (bzw. die Regression) erklärt werden kann.

Es gilt stets

$$0 \leq B_{Y,X} \leq 1;$$

eine optimale Anpassung ist natürlich erreicht, wenn gilt $B_{Y,X} = 1$, denn dann wird Y durch die Regression vollständig erklärt.

Die Größe

$$U_{Y,X} = 1 - B_{Y,X}$$

nennt man auch das **Unbestimmtheitsmaß der Regression**.

Speziell bei der linearen Regression ist das Bestimmtheitsmaß identisch mit dem Quadrat der Korrelation der Merkmale X und Y:

$$B_{Y,X} = r_{XY}^2.$$

Beispiel: In Abschnitt 1.1 haben wir als Schätzer b für den Steigungsparameter der linearen Regression von Gewicht auf Körpergröße, vgl. auch Tab. 1,

$$b = 0,9639$$

berechnet; d.h. die Korrelation zwischen Gewicht und Körpergröße wird durch

$$r_{XY} = 0,9639 \cdot \sqrt{3015,1} / \sqrt{5189,9} = 0,9639 \cdot 0,7622 = 0,735$$

geschätzt. Das Bestimmtheitsmaß der Regression von Gewicht auf Körpergröße ergibt sich gerade zu

$$B_{Y,X} = r_{XY}^2 = 0,735^2 = 0,54 \quad \text{und} \quad U_{Y,X} = 1 - 0,54 = 0,46.$$

1.4. Konfidenzintervalle und Testen von Hypothesen über die unbekannten Parameter α , β und σ^2

Wir wollen uns nun zunächst mit dem Testen von Hypothesen zum Niveau γ über die unbekannten Parameter α , β und σ^2 der linearen Regression beschäftigen. Die Entscheidungsregeln zum Testen von Hypothesen H_0 gegen Alternativen H_1 sind in **Tab. 2** zusammengestellt. Dort bezeichnen α_0 , β_0 und σ_0^2 durch die Nullhypothese vorgegebene (bekannte) Zahlen und $t_{v,\gamma}$ Quantile der t_v -Verteilung, $\chi_{v,\gamma}^2$ Quantile der χ_v^2 -Verteilung, die im Anhang zu finden sind.

Tab. 2: Entscheidungsregeln für Tests zum Niveau γ über die unbekannten Parameter α , β und σ^2

Hypothese H_0	Alternative H_1	H_0 wird verworfen, falls
$\alpha = \alpha_0$	$\alpha \neq \alpha_0$	$ (\alpha - \alpha_0)/s_a > t_{n-2, 1-\gamma/2}$
$\alpha \leq \alpha_0$	$\alpha > \alpha_0$	$(\alpha - \alpha_0)/s_a > t_{n-2, 1-\gamma}$
$\alpha \geq \alpha_0$	$\alpha < \alpha_0$	$(\alpha - \alpha_0)/s_a < t_{n-2, \gamma}$
$\beta = \beta_0$	$\beta \neq \beta_0$	$ (\beta - \beta_0)/s_b > t_{n-2, 1-\gamma/2}$
$\beta \leq \beta_0$	$\beta > \beta_0$	$(\beta - \beta_0)/s_b > t_{n-2, 1-\gamma}$
$\beta \geq \beta_0$	$\beta < \beta_0$	$(\beta - \beta_0)/s_b < t_{n-2, \gamma}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$(n-2)s^2/\sigma_0^2 < \chi_{n-2, \gamma/2}^2$ oder $(n-2)s^2/\sigma_0^2 > \chi_{n-2, 1-\gamma/2}^2$
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$(n-2)s^2/\sigma_0^2 > \chi_{n-2, 1-\gamma}^2$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$(n-2)s^2/\sigma_0^2 < \chi_{n-2, \gamma}^2$

Beispiel: Für die Daten aus Tab. 1 wollen wir zum 5% Niveau

$$H_0: \alpha = -100$$

testen. Da gilt

$$|(\alpha - \alpha_0)/s_a| = |(-96,1504 + 100)/\sqrt{606,560}| = 0,156 < 2,02 = t_{38, 0,975}$$

kann diese Hypothese nicht verworfen werden, d.h. es besteht keine signifikante Abweichung des Absolutgliedes α von $\alpha_0 = -100$.

Dagegen muß zum 10% Niveau wegen

$$(n-2)s^2/40 = 38 \cdot 62,856/40 = 59,71 > 53,38 = \chi_{38, 0,95}^2$$

die Hypothese

$$H_0: \sigma^2 \leq 40$$

über die unbekannte Fehlervarianz σ^2 verworfen werden.

Für den Regressionskoeffizienten β kann man ein $1 - \gamma$ Konfidenzintervall berechnen als

$$[b - s_b t_{n-2; 1-\gamma/2}; \quad b + s_b t_{n-2; 1-\gamma/2}]$$

und für das Absolutglied α ist

$$[a - s_a t_{n-2; 1-\gamma/2}; \quad a + s_a t_{n-2; 1-\gamma/2}]$$

ein Konfidenzintervall zum Niveau $1 - \gamma$; diese Intervalle überdecken den unbekannten Parameter α bzw. β eines linearen Regressionsproblems mit Wahrscheinlichkeit $1 - \gamma$. Für die Fehlervarianz σ^2 ist schließlich

$$[(n-2)s^2/\chi^2_{n-2; 1-\gamma/2}; \quad (n-2)s^2/\chi^2_{n-2; \gamma/2}]$$

ein Konfidenzintervall zum Niveau $1 - \gamma$.

Beispiel: Wir wollen Konfidenzintervalle zum 90% Niveau für die Parameter α , β und σ^2 der Regression von Gewicht auf Körpergröße, vgl. Tab. 1, bestimmen. Als Konfidenzintervall für α ergibt sich

$$\begin{aligned} & [-96,1504 - \sqrt{606,560} \cdot 1,68; -96,1504 + \sqrt{606,560} \cdot 1,68] \\ & = [-137,526; -54,775]; \end{aligned}$$

das gesuchte Konfidenzintervall für β ist

$$\begin{aligned} & [0,9639 - \sqrt{0,021} \cdot 1,68; \quad 0,9639 + \sqrt{0,021} \cdot 1,68] \\ & = [0,720; 1,207] \end{aligned}$$

und schließlich ist

$$\begin{aligned} & [38 \cdot 62,856/53,38; 38 \cdot 62,856/24,90] \\ & = [44,746; 95,925] \end{aligned}$$

ein 90% Konfidenzintervall für die Varianz σ^2 der Fehlerterme.

Simultane Konfidenzintervalle zum Niveau $1 - \gamma$ für die Parameter α und β sind nun

$$\begin{aligned} & [a - \sqrt{2s_a^2 F_{2,n-2; 1-\gamma}}; \quad a + \sqrt{2s_a^2 F_{2,n-2; 1-\gamma}}] \quad \text{und} \\ & [b - \sqrt{2s_b^2 F_{2,n-2; 1-\gamma}}; \quad b + \sqrt{2s_b^2 F_{2,n-2; 1-\gamma}}]. \end{aligned}$$

Das erste Intervall überdeckt den Parameter α , das zweite „zugleich“ den Parameter β , wobei insgesamt das Niveau $1 - \gamma$ eingehalten wird.

Beispiel: Mit Wahrscheinlichkeit $1 - \gamma = 0,90$ wird bei der Regression von Gewicht auf Körpergröße, vgl. Tab. 1, der Parameter α vom Intervall

$$\begin{aligned} & [-96,1504 - \sqrt{2 \cdot 606,560 \cdot 2,45}; \quad -96,1504 + \sqrt{2 \cdot 606,560 \cdot 2,45}] \\ & = [-150,668; -41,633] \end{aligned}$$

und gleichzeitig der Parameter β vom Intervall

$$\begin{aligned} & [0,9639 - \sqrt{2 \cdot 0,021 \cdot 2,45}; \quad 0,9639 + \sqrt{2 \cdot 0,021 \cdot 2,45}] \\ & = [0,643; 1,285] \end{aligned}$$

überdeckt. Natürlich sind die simultanen Konfidenzintervalle breiter als die entsprechenden einzelnen Konfidenzintervalle.

1.5. Konfidenz- und Prognosestreifen

Bisher haben wir nur Konfidenzintervalle für die Parameter eines linearen Regressionsproblems betrachtet. Man ist jedoch oftmals daran interessiert, **Konfidenzintervalle für den erwarteten Wert $E(y|x_0)$ von y an einer Stelle x_0** zu bestimmen. Mit

$$C = s \cdot t_{n-2; 1-\gamma/2} \cdot \sqrt{\frac{1}{n} + (\bar{x} - x_0)^2 / \sum_{i=1}^n (x_i - \bar{x})^2}$$

ist ein solches $1 - \gamma$ Konfidenzintervall durch

$$[a + bx_0 - C; a + bx_0 + C]$$

gegeben. Berechnet man dieses Konfidenzintervall für y an jeder Stelle x_0 , so ergibt sich ein Konfidenzstreifen für die gesamte Regressionsgerade, der überall verschiedene Breite hat und an der Stelle $x_0 = \bar{x}$ am schmalsten ist, vgl. Abb. 10. Dieses Konfidenzintervall überdeckt lediglich den Erwartungswert von y an der Stelle x_0 mit Wahrscheinlichkeit $1 - \gamma$. Ist man hingegen daran interessiert, ein Intervall zu bestimmen, daß eine Realisation y_0 von y an der Stelle x_0 mit Wahrscheinlichkeit $1 - \gamma$ überdeckt, so muß man ein **Prognoseintervall für y_0 an der Stelle x_0 mit Trefferwahrscheinlichkeit $1 - \gamma$** bestimmen, das stets breiter ist als das entsprechende Konfidenzintervall; ein solches Prognoseintervall ist

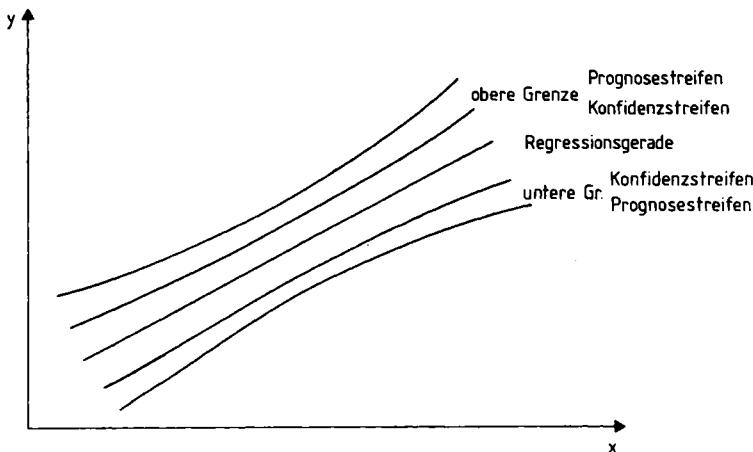


Abb. 10: Konfidenz- und Prognosestreifen für eine Regressionsgerade

$$[a + bx_0 - D; a + bx_0 + D]$$

mit

$$D = s \cdot t_{n-2; 1-\gamma/2} \cdot \sqrt{1 + \frac{1}{n} + (x_0 - \bar{x})^2 / \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Ein Prognosestreifen, vgl. Abb. 10, ergibt sich, wenn man an jeder Stelle x_0 ein Prognoseintervall berechnet.

Will man einen **simultanen Konfidenzstreifen für die Regressionsgerade**, d.h. einen Streifen, der zugleich an allen Stellen x_0 den Erwartungswert $E(y|x_0)$ von y mit Wahrscheinlichkeit $1 - \gamma$ überdeckt, bestimmen, so berechnet man an jeder Stelle

x_0 das Intervall

$$\left[a + b x_0 - \sqrt{2 s^2 F_{2,n-2;1-\gamma} \cdot \left(\frac{1}{n} + (\bar{x} - x_0)^2 / \sum_{i=1}^n (x_i - \bar{x})^2 \right)}; \right.$$

$$\left. a + b x_0 + \sqrt{2 s^2 F_{2,n-2;1-\gamma} \cdot \left(\frac{1}{n} + (\bar{x} - x_0)^2 / \sum_{i=1}^n (x_i - \bar{x})^2 \right)} \right],$$

vgl. auch Abb. 11.

Beispiel: Aus den Daten der Tabelle 1 wollen wir nun Prognose- und Konfidenzstreifen für das Körpergewicht bestimmen. Als Niveau wählen wir $1 - \gamma = 0,95$. In Tab. 3 findet man die Grenzen für Prognose-, Konfidenz- und simultane Konfidenzintervalle an einigen Stellen x_0 . In Abb. 11 sind die drei Streifen dargestellt.

Tab. 3: Wertetabelle zum Konfidenz-, Prognose- und simultanen Konfidenzstreifen zu den Daten aus Tab. 1, vgl. auch Abb. 11

x_0	$E(y x_0)$	y_0	$E(y x_0)$ simultan
150	41,982 bis 54,887	31,169 bis 65,701	40,303 bis 56,566
155	48,111 bis 58,398	36,433 bis 70,075	46,772 bis 59,736
160	54,134 bis 62,014	41,581 bis 74,566	53,108 bis 63,039
165	59,919 bis 65,867	46,604 bis 79,182	59,145 bis 66,641
170	65,178 bis 70,247	51,498 bis 83,927	64,519 bis 70,906
175	69,660 bis 75,405	56,262 bis 88,803	68,912 bis 76,152
180	73,566 bis 81,138	60,895 bis 93,808	72,581 bis 82,123
185	77,204 bis 87,138	65,404 bis 98,939	75,912 bis 88,430
190	80,725 bis 93,256	69,794 bis 104,188	79,095 bis 94,886
195	84,188 bis 99,432	74,074 bis 109,546	82,205 bis 101,416

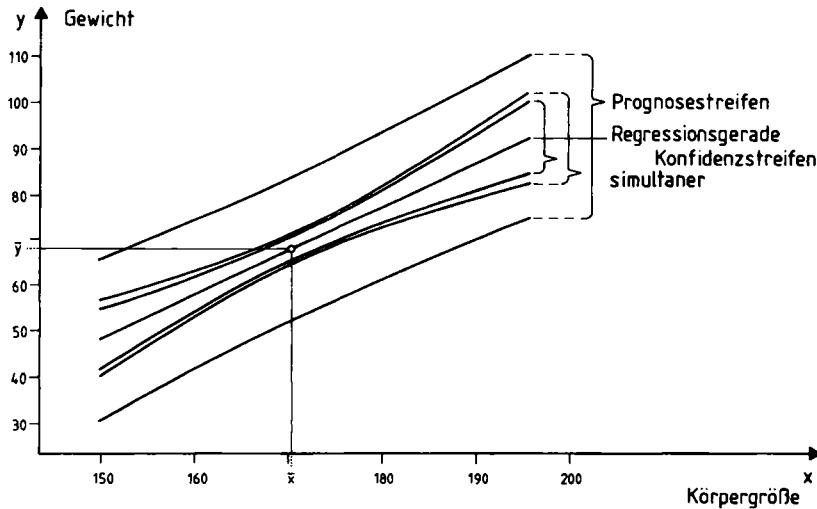


Abb. 11: Konfidenz- und Prognosestreifen zum Niveau $1 - \gamma = 0,95$ im Beispiel, vgl. Tab. 1 und Tab. 3

Wir sehen aus der Abb. 11 sehr deutlich, daß der Prognosestreifen wesentlich breiter ist als der Konfidenzstreifen zum selben Niveau. Der simultane Konfidenzstreifen ist nur unwesentlich breiter als der Konfidenzstreifen.

1.6. Regression durch einen vorgegebenen Punkt, Regression ohne Absolutglied (eigentlich-lineare Regression)

Es gibt Problemstellungen, bei denen es von vornherein sinnvoll erscheint zu verlangen, daß die Regressionsgerade durch einen bestimmten Punkt (x_0, y_0) geht. Will man z. B. das Gewicht von Küken in den ersten Lebenstagen durch eine Regressionsgerade schätzen, so empfiehlt es sich, von allen Beobachtungswerten zunächst den Mittelwert zu subtrahieren, denn dadurch werden Unterschiede in den Schlupfgewichten ausgeglichen, die eine unterschiedliche Gewichtszunahme in den ersten Lebenstagen zur Folge haben. Bei einer Regressionsgeraden, die keinen spezifizierten Wert schneiden soll, wissen wir, daß diese stets durch den Punkt (\bar{x}, \bar{y}) geht. Verlangen wir nun im vorhinein, daß eine Regressionsgerade durch den Punkt (x_0, y_0) geht, so können wir die Parameter α und β dieser Regressionsgeraden

$$y = \alpha + \beta x$$

schätzen, indem wir in den gewöhnlichen Schätzformeln (aus Abschnitt 1.1) \bar{x} durch x_0 und \bar{y} durch y_0 ersetzen (was gerade bedeutet, daß wir alle Beobachtungswerte so transformieren, daß der Punkt (x_0, y_0) dem Ursprung $(0,0)$ des Koordinatensystems entspricht), β schätzen und dann durch die umgekehrte Verschiebung α schätzen:

$$b = \frac{\sum_{i=1}^n (y_i - y_0)(x_i - x_0)}{\sum_{i=1}^n (x_i - x_0)^2} \quad \text{und} \quad a = y_0 - bx_0.$$

Verlangt man nun speziell, daß die Regressionsgerade durch den Ursprung $(0,0)$ des Koordinatensystems (**Regression ohne Absolutglied**) geht, so entfallen natürlich beide Verschiebungen und es ist

$$b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \quad \text{und} \quad a = 0.$$

Natürlich ist auch hier

$$\hat{y}_i = a + bx_i,$$

so daß die Varianz σ^2 des Fehlers durch

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

geschätzt wird.

Zu beachten ist bei der Regression durch einen vorgegebenen Punkt (x_0, y_0) , daß die Varianz der Fehler e_i der Beobachtungen nur noch in der Nähe von (x_0, y_0) als konstant angesehen werden kann, denn mit wachsender Entfernung von (x_0, y_0) wird in praxi die Varianz größer, so daß man mitunter Methoden der Regression bei

ungleicher Varianz der Beobachtungen verwenden muß, vgl. auch Schach/Schäfer (1978).

Beispiel: siehe Abschnitt 3.

2. Residualanalyse

Im Beispiel aus Abschnitt 1.1 haben wir neben den Beobachtungswerten y_1, \dots, y_n auch die geschätzten Werte $\hat{y}_1, \dots, \hat{y}_n$ sowie die Differenzen $y_i - \hat{y}_i = \hat{e}_i$, die man **Residuen** nennt, angegeben. Normiert man die n Residuen mit dem Schätzer s für die Standardabweichung σ der Beobachtungen, also

$$d_i = \frac{y_i - \hat{y}_i}{s} \quad \text{für } i = 1, \dots, n,$$

so sind die **normierten Residuen** d_i , falls der Modellansatz mit normalverteilten Fehlern e_i richtig ist, approximativ unabhängig identisch $N(0,1)$ -verteilt. Stellt man die d_i nun graphisch in Abhängigkeit von \hat{y}_i dar, so kann man aus dieser Darstellung Schlüsse über die Güte des Modellansatzes der Regression ziehen. In Abb. 12 sind einige idealisierte solche Verläufe dargestellt. Abb. 12 (a) zeigt den idealen Verlauf, d.h. den Verlauf, bei dem keine Abweichungen von den Modellvoraussetzungen vorhanden sind. Ein Verlauf wie in Abb. 12 (b) deutet auf einen Rechenfehler hin, jedoch ist möglicherweise auch ein inadäquater Ansatz gewählt worden. Bei Verläufen wie in Abb. 12 (c) müssen entweder ungleiche Varianzen der Beobachtun-

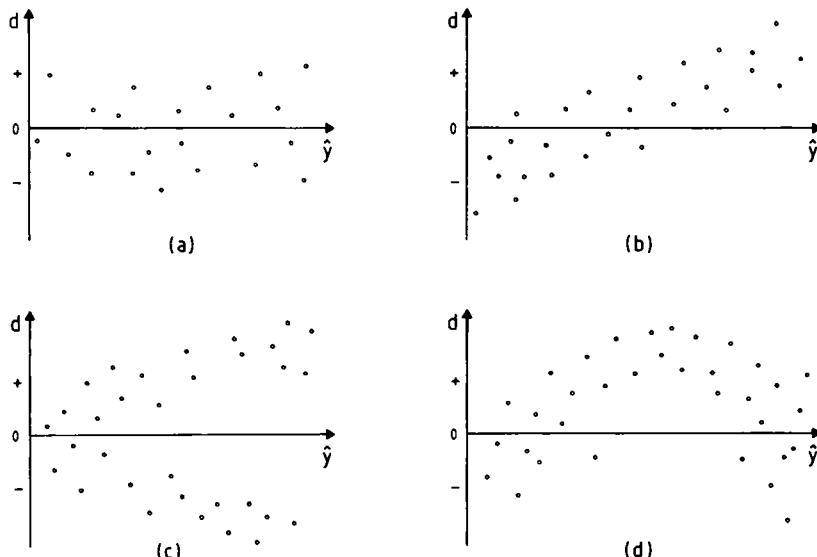


Abb. 12: Graphische Darstellung der normierten Residuen in Abhängigkeit von den geschätzten Werten \hat{y} des Regressanden: (a) idealer Verlauf; (b) linearer Trend, auf einen Rechenfehler hindeutend; (c) ansteigende Varianzen (mit \hat{y}), Wechsel zu einem Modell für ungleiche Varianz der Beobachtungen eventuell angebracht; (d) nichtlinearer Verlauf der Residuen, inadäquates Modell, d.h. Transformation der Daten oder Änderung der Regressionsfunktion angezeigt

gen in den Modellansatz aufgenommen werden, vgl. Kap. XII, 5, oder es muß eine Varianzstabilisierende Transformation, vgl. Abschnitt 4.5 in Kapitel V, der Daten vorgenommen werden, da hier die Varianz der Beobachtungen mit wachsendem \hat{y} wächst. Zeigen die normierten Residuen einen Verlauf wie in Abb. 12 (d), so ist angezeigt, daß ein quadratischer Regressionsatz, vgl. Abschnitt 4, zu wählen ist.

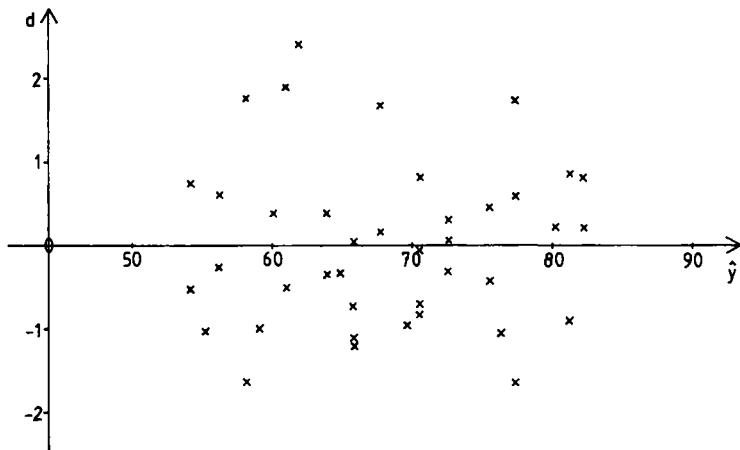


Abb. 13: Normierte Residuen d_i für die Regression von Gewicht auf Körpergröße, vgl. auch Tab. 1

Beispiel: In Abb. 13 sind die normierten Residuen für das Beispiel der linearen Regression von Gewicht auf Körpergröße aus Abschnitt 1.1, Tab. 1, eingetragen. Es zeigt sich, daß die Residuen recht gleichmäßig um die Nullachse verteilt sind, d.h. der lineare Ansatz ist hier – beurteilt in den Residuen – sehr gut.

Man kann normierte Residuen auch in Abhängigkeit von der zeitlichen Reihenfolge der Beobachtungen graphisch darstellen. Einige idealisierte Verläufe bei dieser Darstellung sind in Abb. 14 dargestellt. Natürlich kann man die d_i auch in Abhängigkeit der x_i , was häufig identisch ist mit der Darstellung in Abhängigkeit von der Zeit, darstellen. Die Interpretation der Verläufe ist in diesem Fall identisch mit der Darstellung in Abhängigkeit von den geschätzten Regressanden. Weiterhin besteht die Möglichkeit, die normierten Residuen d_i in Abhängigkeit von miterhobenen **Dritt faktoren** darzustellen, um zu untersuchen, ob diese Faktoren einen Einfluß auf die Werte des Regressanden y haben.

Mit Hilfe der graphischen Darstellung der d_i ist es auch möglich, sogenannte **Ausreißer** in den Beobachtungen zu lokalisieren und zu eliminieren. Eine grobe Regel, wann eine Beobachtung y_i als Ausreißer betrachtet werden kann, ist über die absolute Größe von $|d_i|$ gegeben:

Ist $|d_i| > 3$, so kann man y_i als Ausreißer eliminieren.

Ein approximativer Test auf Ausreißer wird beschrieben bei Stefansky (1971, 1972), Goldsmith/Boddy (1973) und Williams (1973), vgl. auch Abschnitt 4.4 in Kap. V.

Eine sehr einfache Methode zur Überprüfung der Normalität der d_i bieten die

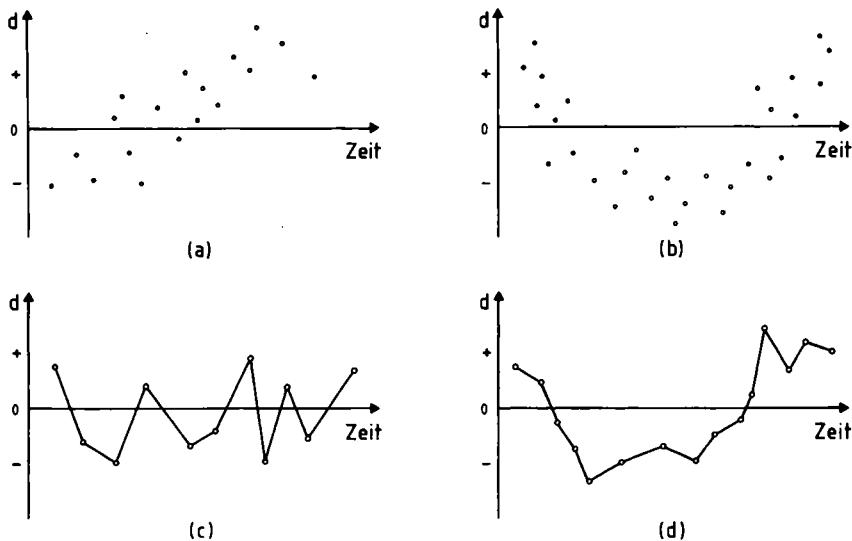


Abb. 14: Graphische Darstellung der normierten Residuen in Abhängigkeit von einem Dritt faktor, hier die Zeit: (a) linearer Zeittrend; (b) nichtlinearer Zeittrend, z. B. quadratischer Trend; (c) negative Korrelation der zeitlich aufeinanderfolgenden Residuen; (d) positive Korrelation der zeitlich aufeinanderfolgenden Residuen

Verwendung von Normalverteilungspapier, vgl. Kapitel IV, Abschnitt 1.7, oder der Q-Q-Plot, vgl. Kapitel XIV, Abschnitt 1.9. Diese können dann zur Überprüfung der Modellannahmen herangezogen werden. Ist die Grundgesamtheit, aus der die Daten y_i stammen, nicht normalverteilt, so können zunächst Datentransformationen vorgenommen werden, wie in Abschnitt 4.5 von Kapitel V beschrieben wurde.

3. Transformationen auf Linearität

Viele – bzgl. der unbekannten Parameter – nichtlineare Funktionen, von denen einige in Tab. 4 zusammengestellt sind, lassen sich in lineare Funktionen transformieren. In solchen Fällen reduziert sich bei passender Fehlerannahme das Problem der Parameterschätzung der nichtlinearen Funktionen auf die Schätzung der Parameter einer linearen Funktion. Sofern dies möglich ist, sollte man zur Parameterschätzung jedoch die ursprüngliche Funktion heranziehen. Für einige spezielle Funktionen wird die Vorgehensweise im Abschnitt 1.2 des Kap. XII demonstriert. Bei Anwendung solcher Transformationen ändern sich die Parameterschätzungen durch die Methode der kleinsten Quadrate in der gleichen Weise, wie sich die Parameter selbst ändern.

Beispiel: Eine chemische Reaktion erster Ordnung wird durch die Gleichung

$$\ln \frac{c}{a} = -kt,$$

wobei t die Zeit, a die Konzentration zur Zeit 0, c die Konzentration zur Zeit t und k die Reaktionskonstante ist, beschrieben.

Tab. 4: Nichtlineare, in lineare transformierbare Funktionen

Funktion	transformierte Funktion
$y = \alpha x^\beta$	$\ln y = \ln \alpha + \beta \ln x$
$y = \alpha e^{\beta x}$	$\ln y = \ln \alpha + \beta x$
$y = (\alpha + \beta x)^{-1}$	$1/y = \alpha + \beta x$
$y = \frac{x}{\alpha x + \beta}$	$1/y = \alpha + \beta \frac{1}{x}$
$y = \alpha e^{\beta/x}$	$\ln y = \ln \alpha + \beta \frac{1}{x}$
$y = \frac{1}{\alpha + \beta e^{-x}}$	$1/y = \alpha + \beta e^{-x}$
$y = \alpha x^\beta e^{\gamma x}$	$\ln y = \ln \alpha + \beta \ln x + \gamma x$

Die Dekomposition von Nitrogenpentoxid ist eine Reaktion 1. Ordnung:



(vgl. Daniels/Johnston (1921): „The thermal decomposition of gaseous nitrogen pentoxide: A monomolecular reaction“, Journal Amer. Chem. Society, 43, S. 53–71). Die Größe c/a kann hier beschrieben werden durch:

$$1 - \frac{p}{p_\infty},$$

wobei p den Druck des sich bildenden Oxygens O₂ zur Zeit t und p_∞ den Druck nach Beendigung der Reaktion angibt. Setzt man nun

$$\ln \left(1 - \frac{p}{p_\infty} \right) = y \quad \text{und} \quad -kt = \beta x,$$

so kann man, wie Daniels/Johnston (s.o.) es getan haben, aufgrund eines Experiments die Reaktionskonstante k als negative Steigung der Regressionsgeraden ohne Absolutglied

$$y = \beta x$$

schätzen. In Tab. 5 sind Messungen des Drucks p_i zu verschiedenen Zeiten t = x_i bei der Reaktion angegeben. Zusätzlich wurde dort zu jeder Zeit

$$y_i = \ln \left(1 - \frac{p_i}{p_\infty} \right)$$

berechnet. Im Experiment ergab sich für

$$p_\infty = 331,3.$$

Wir wollen nun den Parameter β, also die negative Reaktionskonstante, mit Hilfe dieser Daten schätzen:

$$b = \frac{\sum_{i=1}^{16} x_i y_i}{\sum_{i=1}^{16} x_i^2} = \frac{-429,7062}{4804} = -0,08945.$$

Tab. 5: Druck des Oxygens bei der Reaktion von Nitrogenpentoxid

i	x_i	p_i	$y_i = \ln\left(1 - \frac{p_i}{331,3}\right)$	\hat{p}_i	$p_i - \hat{p}_i$
1	3	72,9	-0,2485155	77,975	-5,075
2	4	95,0	-0,3379221	99,651	-4,651
3	5	115,9	-0,4305275	119,472	-3,572
4	6	135,2	-0,5243996	137,597	-2,397
5	7	151,9	-0,6134064	154,172	-2,272
6	8	168,0	-0,7074353	169,328	-1,328
7	9	182,8	-0,8024394	183,187	-0,387
8	10	195,7	-0,8933149	195,861	-0,161
9	12	217,8	-1,0712215	218,047	-0,247
10	14	236,4	-1,2502006	236,599	-0,199
11	16	252,2	-1,4323114	252,112	0,088
12	18	265,9	-1,6225021	265,084	0,816
13	22	285,7	-1,9831166	285,001	0,699
14	26	299,3	-2,3372884	298,927	0,373
15	30	309,0	-2,6984376	308,664	0,336
16	38	320,0	-3,3782216	320,233	-0,233

Die Reaktionskonstante k kann also durch 0,08945 geschätzt werden.

Um nun zu einer vorgegebenen Zeit $t = x$ den Druck p des Oxygens schätzen zu können, müssen wir die Gleichung

$$\hat{y} = \ln\left(1 - \frac{\hat{p}}{331,3}\right) = -0,08945x$$

nach \hat{p} auflösen. Es ergibt sich

$$\hat{p} = p_\infty - p_\infty e^{-kx} = 331,3 - 331,3 \cdot e^{-0,08945x}.$$

In Tabelle 5 sind zu den Zeiten x_i auch noch diese geschätzten Druckwerte angegeben.

Auch kann man in Situationen, wo eine einfache lineare Regression nicht angebracht ist, Methoden der multiplen Regression, vgl. Abschnitt 5, anwenden. In Abschnitt 4 dieses Kapitels wird zudem noch die Regression bei quadratischen Funktionen explizit vorgestellt.

4. Nichtlineare Regression und Schätzen des Maximums (Minimums) einer quadratischen Regressionsfunktion

Bisher haben wir nur lineare Regressionsprobleme betrachtet, d.h. wir haben uns mit der Schätzung und Untersuchung von Regressionsgeraden beschäftigt. Oft weiß man jedoch bzw. stellt anhand von Beobachtungen fest, daß es angebracht ist, eine **nichtlineare Regressionsfunktion**, z.B. eine **Regressionsparabel**, zu bestimmen. Wir wollen uns hier speziell mit **quadratischen Regressionsproblemen** beschäftigen.

Kann man aufgrund von n Beobachtungen annehmen, daß zwischen zwei Merkmalen X und Y ein quadratischer Zusammenhang besteht, so wird man aus dem Ansatz

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + e_i, \quad \text{für } i = 1, \dots, n,$$

die Regressionsfunktion

$$\hat{y} = a + b_1 x + b_2 x^2$$

schätzen. Um die Bestimmung von Konfidenzintervallen und das Testen von Hypothesen zu ermöglichen, unterstellen wir, daß die Fehlerterme e_1, \dots, e_n unabhängig identisch $N(0, \sigma^2)$ -verteilt sind und bestimmen zunächst nach der Methode der kleinsten Quadrate durch Minimierung der Summe der quadratischen Residuen

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

die Schätzer a , b_1 und b_2 für die unbekannten Parameter α , β_1 und β_2 des Regressionsproblems. Dazu berechnen wir die Hilfsgröße

$$K = n(\sum x_i^2)(\sum x_i^4) + 2(\sum x_i)(\sum x_i^2)(\sum x_i^3) - (\sum x_i^2)^3 - n(\sum x_i^3)^2 \\ - (\sum x_i)^2(\sum x_i^4)$$

und damit die Größen

$$A = [\sum x_i^2](\sum x_i^4) - (\sum x_i^3)^2]/K, \quad B = [n \sum x_i^4 - (\sum x_i^2)^2]/K, \\ C = [n \sum x_i^2 - (\sum x_i)^2]/K, \quad D = [\sum x_i^3](\sum x_i^2) - (\sum x_i)(\sum x_i^4)]/K, \\ E = [\sum x_i](\sum x_i^3) - (\sum x_i^2)^2]/K, \quad G = [\sum x_i](\sum x_i^2) - n \sum x_i^3]/K,$$

und erhalten die Kleinst-Quadrat-Schätzer

$$a = A \sum y_i + D \sum x_i y_i + E \sum x_i^2 y_i, \\ b_1 = D \sum y_i + B \sum x_i y_i + G \sum x_i^2 y_i, \\ b_2 = E \sum y_i + G \sum x_i y_i + C \sum x_i^2 y_i.$$

Damit haben wir die Regressionsfunktion \hat{y} bestimmt und können die Varianz σ^2 der Fehler e_1, \dots, e_n schätzen durch

$$s^2 = \frac{1}{n-3} \sum (y_i - a - b_1 x_i - b_2 x_i^2)^2 = \frac{1}{n-3} \sum (y_i - \hat{y}_i)^2.$$

Die empirischen Varianzen und die Korrelationen der Schätzer a , b_1 und b_2 sind dann

$$s_a^2 = s^2 A, \quad s_{b_1}^2 = s^2 B, \quad s_{b_2}^2 = s^2 C, \\ \rho_{ab_1} = D/\sqrt{AB}, \quad \rho_{ab_2} = E/\sqrt{AC}, \quad \rho_{b_1 b_2} = G/\sqrt{BC}.$$

Nun können wir auch Konfidenzintervalle zum Niveau $1 - \gamma$ für den Erwartungswert von y an der Stelle x_0 bestimmen, die dann einen **Konfidenzbereich für die Regressionskurve** liefern. Mit

$$g(x_0) = A + 2Dx_0 + (B + 2E)x_0^2 + 2Gx_0^3 + Cx_0^4$$

ist

$$[a + b_1 x_0 + b_2 x_0^2 - s \sqrt{g(x_0)} \cdot t_{n-3; 1-\gamma/2};$$

$$a + b_1 x_0 + b_2 x_0^2 + s \sqrt{g(x_0)} \cdot t_{n-3; 1-\gamma/2}]$$

ein $1 - \gamma$ Konfidenzintervall für den Erwartungswert $E(y|x_0)$ von y an der Stelle x_0 .

Sind α_0 , β_{10} und β_{20} vorgegebene Werte, so lassen sich auch **Hypothesen über die unbekannten Parameter α , β_1 und β_2** zum Niveau γ testen; die Entscheidungsregeln sind in **Tab. 6** zusammengestellt.

Tab. 6: Entscheidungsregeln für Niveau- γ -Tests im quadratischen Regressionsproblem

Hypothese H_0	Alternative H_1	H_0 wird verworfen, falls
$\alpha = \alpha_0$	$\alpha \neq \alpha_0$	$ (\alpha - \alpha_0)/s_\alpha > t_{n-3, 1-\gamma/2}$
$\beta_1 = \beta_{10}$	$\beta_1 \neq \beta_{10}$	$ (\beta_1 - \beta_{10})/s_{\beta_1} > t_{n-3, 1-\gamma/2}$
$\beta_2 = \beta_{20}$	$\beta_2 \neq \beta_{20}$	$ (\beta_2 - \beta_{20})/s_{\beta_2} > t_{n-3, 1-\gamma/2}$

Von besonderem Interesse ist ein Test über die Hypothese

$$H_0: \beta_2 = 0,$$

denn so kann überprüft werden, ob der quadratische Ansatz zur Erklärung der Beziehung zwischen X und Y notwendig ist: Wird H_0 zum Niveau γ verworfen, so ist er signifikant notwendig, wird H_0 nicht verworfen, so wird man, da man die Beziehung zwischen den Merkmalen X und Y ja möglichst einfach darstellen will, zum linearen Ansatz aus Abschnitt 1 übergehen.

Ist der quadratische Ansatz signifikant, so ist man vielmals auch an der **Schätzung des Maximums (bzw. Minimums) x_M** der Regressionsfunktion interessiert. Ein Punktschätzer für x_M ist

$$\hat{x}_M = -b_1/(2b_2)$$

und ein Konfidenzintervall für x_M zum Niveau $1 - \gamma$ ist

$$[(-K_2 - \sqrt{K_2^2 - K_1 K_3})/K_1; \quad (-K_2 + \sqrt{K_2^2 - K_1 K_3})/K_1],$$

mit den Hilfsgrößen

$$K_1 = 4b_2^2 - 4Cs^2 F_{1,n-3; 1-\gamma}, \quad K_2 = 2b_1 b_2 - 2Gs^2 F_{1,n-3; 1-\gamma} \quad \text{und}$$

$$K_3 = b_1^2 - Bs^2 F_{1,n-3; 1-\gamma}.$$

Allerdings existiert dieses Konfidenzintervall nur, wenn $K_1 > 0$ gilt, d.h. wenn $H_0: \beta_2 = 0$ zum Niveau γ verworfen wird.

Beispiel: In **Tab. 7** sind die Leistungswerte y_i (für $i = 1, \dots, 16$) von Studenten bei unterschiedlicher Dosierung x_i eines Stimulans angegeben. Stellt man diese Leistungswerte graphisch dar, vgl. **Abb. 15**, so zeigt sich, daß ihre Verbindungslien in etwa einen quadratischen Verlauf angibt. Daher wollen wir ausgehend vom Regressionsansatz

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + e_i, \quad \text{für } i = 1, \dots, 16,$$

die quadratische Regressionsfunktion schätzen. Mit

$$K = 16 \cdot 49,6 \cdot 285,2992 + 2 \cdot 24 \cdot 49,6 \cdot 115,2 - (49,6)^3 - 16 \cdot (115,2)^2 \\ - (24)^2 \cdot 285,2992 = 1988,6899,$$

$$A = (49,6 \cdot 285,2992 - (115,2)^2)/1988,6899 = 0,44240,$$

$$B = (16 \cdot 285,2992 - (49,6)^2)/1988,6899 = 1,05830,$$

Tab. 7: Beobachtete und geschätzte Leistungswerte, vgl. Abb. 15 und 16

x_i	y_i	\hat{y}_i	$y_i - \hat{y}_i$
0,0	3,2	3,380	-0,180
0,2	5,0	4,447	0,553
0,4	4,7	5,334	-0,634
0,6	5,8	6,041	-0,241
0,8	7,1	6,568	0,532
1,0	7,3	6,914	0,386
1,2	6,5	7,080	-0,580
1,4	7,5	7,066	0,434
1,6	6,9	6,872	0,028
1,8	6,4	6,497	-0,097
2,0	5,3	5,942	-0,642
2,2	5,8	5,207	0,593
2,4	3,9	4,292	-0,392
2,6	3,5	3,196	0,304
2,8	1,7	1,920	-0,220
3,0	0,6	0,464	0,136

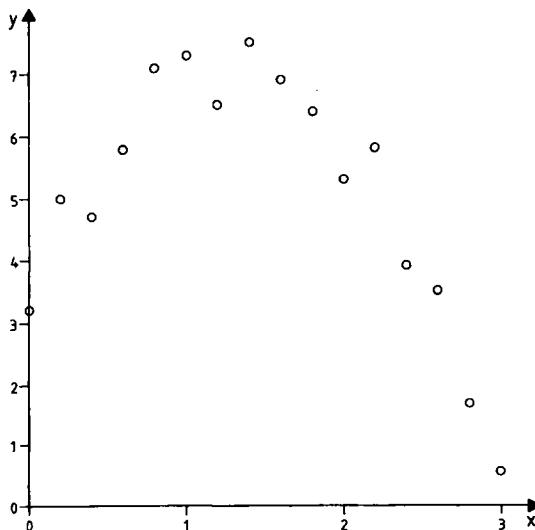


Abb. 15: Wirkung einer Stimulanz auf das Leistungsvermögen

$$C = (16 \cdot 49,6 - (24)^2)/1988,6899 = 0,10942,$$

$$D = (115,2 \cdot 49,6 - 24 \cdot 285,2992)/1988,6899 = -0,56985,$$

$$E = (24 \cdot 115,2 - (49,6)^2)/1988,6899 = 0,15319 \text{ und}$$

$$G = (24 \cdot 49,6 - 16 \cdot 115,2)/1988,6899 = -0,32826$$

ergeben sich die Kleinstes-Quadrat-Schätzer zu

$$a = 0,44240 \sum y_i - 0,56985 \sum x_i y_i + 0,15319 \sum x_i^2 y_i$$

$$= 0,44240 \cdot 81,2 - 0,56985 \cdot 108,58 + 0,15319 \cdot 191,468 = 3,380,$$

$$b_1 = -0,56985 \sum y_i + 1,05830 \sum x_i y_i - 0,32826 \sum x_i^2 y_i = 5,787 \quad \text{und}$$

$$b_2 = 0,15319 \sum y_i - 0,32826 \sum x_i y_i + 0,10942 \sum x_i^2 y_i = -2,253,$$

so daß die geschätzte Regressionsfunktion nun

$$\hat{y} = a + b_1 x + b_2 x^2 = 3,380 + 5,787x - 2,253x^2$$

ist. In Tab. 7 sind auch die geschätzten Werte \hat{y}_i angegeben, so daß sich der Schätzer für die Varianz σ^2 leicht berechnen läßt:

$$s^2 = \frac{1}{n-3} \sum (y_i - \hat{y}_i)^2 = \frac{1}{13} \cdot 2,842 = 0,219.$$

Weiter ergeben sich die Schätzer für die Varianzen sowie die Korrelationen der Schätzer a , b_1 und b_2 zu

$$s_a^2 = s^2 A = 0,219 \cdot 0,4424 = 0,097, \quad s_{b_1}^2 = 0,232, \quad s_{b_2}^2 = 0,024,$$

$$\varrho_{ab_1} = D/\sqrt{AB} = -0,833, \quad \varrho_{ab_2} = 0,696 \quad \text{und} \quad \varrho_{b_1 b_2} = -0,965.$$

Das Bestimmtheitsmaß dieser Regression, vgl. Abschnitt 1.3, ist

$$B_{Y,(X,x^2)} = 1 - \sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2 = 1 - \frac{2,842}{62,09} = 0,954$$

und somit sehr hoch, d. h. wir haben eine gute Anpassung erreicht. In Abb. 16 ist der Konfidenzstreifen zum 95% Niveau für den Erwartungswert $E(y|x_0)$ von y an der Stelle x_0 (bei Dosierung x_0) zusammen mit den Beobachtungswerten und der Regressionsfunktion eingezeichnet. Die Wertetabelle zu diesem Konfidenzstreifen findet man in Tab. 8.

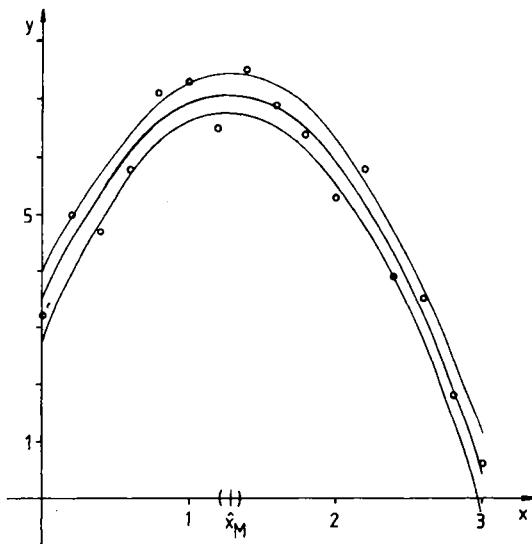


Abb. 16: Regressionsfunktion und Konfidenzstreifen für den Leistungsversuch

Tab. 8: Wertetabelle für den Konfidenzstreifen aus Abb. 16 zu den Daten aus Tab. 7, vgl. Abb. 16

x_0	$g(x_0)$	Konfidenzstreifen für $E(y x_0)$ zum Niveau 0,95	Breite
0,0	0,442	2,708 bis 4,052	1,344
0,2	0,264	3,928 bis 4,996	1,038
0,4	0,166	4,922 bis 5,746	0,824
0,6	0,122	5,688 bis 6,394	0,706
0,8	0,113	6,228 bis 6,908	0,680
1,0	0,120	6,564 bis 7,264	0,700
1,2	0,132	6,713 bis 7,447	0,734
1,4	0,140	6,688 bis 7,444	0,756
1,6	0,140	6,494 bis 7,250	0,756
1,8	0,132	6,130 bis 6,864	0,734
2,0	0,120	5,592 bis 6,292	0,700
2,2	0,113	4,867 bis 5,547	0,680
2,4	0,112	3,939 bis 4,645	0,706
2,6	0,166	2,784 bis 3,608	0,824
2,8	0,264	1,401 bis 2,439	1,038
3,0	0,442	-0,208 bis 1,136	1,344

Der Konfidenzstreifen ist, wie man in der Abb. 16 sieht, im Maximum der Regressionsfunktion ziemlich breit, verengt sich dann zunächst an beiden Seiten und wird schließlich wieder breiter.

Wir wollen nun die Hypothese

$$H_0: \beta_2 = 0 \quad \text{gegen} \quad H_1: \beta_2 \neq 0$$

zum 5% Niveau testen, um zu überprüfen, ob der quadratische Ansatz überhaupt notwendig war. Es ist

$$| (b_2 - \beta_{20})/s_{b_2} | = | (-2,253 - 0)/\sqrt{0,024} | = 14,543 > 2,16 = t_{13,0,975},$$

so daß H_0 verworfen wird, was gleichzeitig bedeutet, daß wir ein Konfidenzintervall zum 95% Niveau für das Maximum x_M der Regressionsfunktion bestimmen können.

Der Punktschätzer für x_M ist

$$\hat{x}_M = -b_1/(2b_2) = (-5,787)/(-4,506) = 1,284, \quad \text{vgl. Abb. 16,}$$

und mit

$$\begin{aligned} K_1 &= 4b_2^2 - 4 \cdot Cs^2 F_{1,n-3;1-\gamma} = 4(-2,253)^2 - 4 \cdot 0,10942 \cdot 0,219 \cdot 4,667 \\ &= 19,86, \end{aligned}$$

$$K_2 = -25,41, \quad K_3 = 32,41$$

ergibt sich das 95% Konfidenzintervall für x_M zu

$$\begin{aligned} &[(25,41 - \sqrt{25,41^2 - 19,86 \cdot 32,41})/19,86; \\ &\quad (25,41 + \sqrt{25,41^2 - 19,86 \cdot 32,41})/19,86] \\ &= [1,21; 1,35], \quad \text{vgl. Abb. 16.} \end{aligned}$$

In vielen Fällen ist es günstiger, statt des hier gewählten Ansatzes eine **Regression**

mit **Orthogonalpolynomen** durchzuführen, da sich zum einen der Arbeitsaufwand verringert und zum anderen Rechenun genauigkeiten kaum Einfluß auf die Ergebnisse haben, d.h. sie ist **numerisch günstig**, vgl. z.B. Graybill (1961). Natürlich läßt sich die Gewinnung eines allgemeinen **Regressionspolynoms k-ten Grades**

$$y = \alpha + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k$$

durch einfache Umbenennung

$$x = x_1, \quad x^2 = x_2, \dots, x^k = x_k$$

in ein Problem der **multiplen Regressionsrechnung**, vgl. Abschnitt 5, überführen.

5. Multiple Regression

Betrachtet man $k+1$ Merkmale Y, X_1, X_2, \dots, X_k und sucht einen funktionalen Zusammenhang der Gestalt

$$y(x_1, \dots, x_k) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k = \alpha + \sum_{j=1}^k \beta_j x_j,$$

so nennt man diese Funktion eine **multiple Regressionsgleichung** vom **Regressanden** Y auf die **Regressoren** X_1, \dots, X_k . Hat man die $k+1$ Merkmale an einer Stichprobe beobachtet mit den Werten bzw. Ausprägungen $y_i, x_{1i}, \dots, x_{ki}$ beim i -ten Objekt ($i = 1, \dots, n$), bzw. sind die Werte x_{ji} fest vorgegeben, so kann man aus dem **Normalengleichungssystem**

$$\begin{aligned} b_1 S Q_{X_1} + b_2 S P_{X_1 X_2} + \dots + b_k S P_{X_1 X_k} &= S P_{X_1 Y} \\ b_1 S P_{X_1 X_2} + b_2 S Q_{X_2} + \dots + b_k S P_{X_2 X_k} &= S P_{X_2 Y} \\ \vdots &\quad \vdots & \vdots \\ b_1 S P_{X_1 X_k} + b_2 S P_{X_2 X_k} + \dots + b_k S Q_{X_k} &= S P_{X_k Y} \end{aligned}$$

die **Kleinste-Quadrat-Schätzer (partiellen Regressionskoeffizienten)** b_1, \dots, b_k für β_1, \dots, β_k bestimmen; dabei ist

$$\begin{aligned} S Q_{X_j} &= \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2, \quad S P_{X_j X_{j'}} = \sum_{i=1}^n (x_{ji} - \bar{x}_j)(x_{j'i} - \bar{x}_{j'}) \quad \text{und} \\ S P_{Y X_j} &= \sum_{i=1}^n (y_i - \bar{y})(x_{ji} - \bar{x}_j) \quad \text{für } j, j' = 1, \dots, k. \end{aligned}$$

Der Schätzer a für das Absolutglied α ist dann gerade

$$a = \bar{y} - b_1 \bar{x}_1 - \dots - b_k \bar{x}_k.$$

Hierbei bezeichnet natürlich \bar{x}_j den Mittelwert der Beobachtungen des j -ten Merkmals X_j . Zu diesen Bestimmungsgleichungen für die Schätzer der unbekannten Parameter gelangt man durch Minimierung (bzgl. der Parameter) der **Residual-Quadratsumme**

$$S^2 = \sum_{i=1}^n (y_i - \alpha - \beta_1 x_{1i} - \beta_2 x_{2i} - \dots - \beta_k x_{ki})^2 = \min!_{\alpha, \beta_1, \dots, \beta_k} .$$

Das **multiple Bestimmtheitsmaß**

$$\begin{aligned} B_{Y,(x_1, \dots, x_k)} &= 1 - \sum_{i=1}^n \left(y_i - a - \sum_{j=1}^k b_j x_{ji} \right)^2 \Bigg/ \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= r_{Y,(x_1, \dots, x_k)}^2 \end{aligned}$$

gibt an, welcher Anteil der Varianz des Merkmals Y durch die Regressoren erklärt werden kann, und entspricht dem Quadrat des **multiplen Korrelationskoeffizienten** $r_{Y,(x_1, \dots, x_k)}$ zwischen Y und X_1, \dots, X_k , vgl. Abschnitt 5 in Kap. IX, der in der Literatur auch mit R bezeichnet wird ($R^2 = B_{Y,(x_1, \dots, x_k)}$). Für das Bestimmtheitsmaß gilt stets

$$B_{Y,x_i} \leqq B_{Y,(x_1, \dots, x_k)},$$

und die Regressionskoeffizienten ändern sich bei Berücksichtigung eines weiteren Merkmals, d.h. etwa: Haben wir eine Regressionsgerade der Art

$$\hat{y}(x_1) = \tilde{a} + bx_1$$

bestimmt, so gelangen wir durch Hinzunahme eines weiteren Merkmals X_2 zu einer Regressionsgleichung

$$\hat{y}(x_1, x_2) = a + b_1 x_1 + b_2 x_2,$$

für die i.a. gilt $b \neq b_1$ und natürlich $\tilde{a} \neq a$.

Setzt man, vgl. auch das Beispiel in Abschnitt 4.3.1 von Kap. XIII,

$$\hat{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & \dots & x_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix},$$

so lässt sich die allgemeine Modellgleichung in Matrizendarstellung schreiben als

$$\hat{Y} = X\beta + e$$

mit unabhängigen Fehlern e_1, \dots, e_n , für die gilt

$$Ee_i = 0 \quad \text{und} \quad \text{Var}e_i = \sigma^2.$$

Der Schätzvektor β für β nach der Methode der kleinsten Quadrate lässt sich als Lösung der Normalengleichungen

$$X^T X \beta = X^T \hat{Y} \quad (X^T \text{ Transponierte zu } X)$$

gewinnen. Ist $X^T X$ invertierbar, so ergibt sich hieraus sofort

$$\beta = (X^T X)^{-1} X^T \hat{Y}$$

als eindeutig bestimmter Schätzer für β . Setzt man

$$C = (X^T X)^{-1} = \begin{bmatrix} c_{00} & \dots & c_{0k} \\ c_{10} & \dots & c_{1k} \\ \vdots & & \vdots \\ c_{k0} & \dots & c_{kk} \end{bmatrix},$$

so gilt

$$E b_i = \beta_i, \quad \text{Var } b_i = \sigma^2 c_{ii}, \quad \text{Cov}(b_i, b_j) = \sigma^2 c_{ij},$$

wobei hier natürlich $\beta_0 = \alpha$ und $b_0 = a$, der Schätzer für α , sind. Der Streuungsparameter σ^2 kann hierbei geschätzt werden durch

$$s^2 = \frac{1}{n - k - 1} \sum_{i=1}^n \left(y_i - a - \sum_{j=1}^k x_{ji} b_j \right)^2.$$

Setzt man zudem voraus, daß die e_i einer Normalverteilung mit Mittelwert 0 und Varianz σ^2 genügen, so lassen sich folgende Aussagen gewinnen.

Für eine vorgegebene Konstante β_i^0 ist die Nullhypothese

$$H_0: \beta_i = \beta_i^0 \quad \text{zugunsten der Alternative} \quad H_1: \beta_i \neq \beta_i^0$$

zu verwerfen, falls für ein vorgegebenes Niveau γ gilt

$$\frac{|b_i - \beta_i^0|}{\sqrt{s^2 c_{ii}}} > t_{n-k-1; 1-\gamma/2}.$$

Das $1 - \gamma$ Konfidenzintervall für β_i ergibt sich als

$$[b_i - \sqrt{s^2 c_{ii}} \cdot t_{n-k-1; 1-\gamma/2}; \quad b_i + \sqrt{s^2 c_{ii}} \cdot t_{n-k-1; 1-\gamma/2}].$$

Simultane $1 - \gamma$ Konfidenzintervalle für die Regressionskoeffizienten β_i ($i = 0, \dots, k$, $\beta_0 = \alpha$) sind nach Scheffé gegeben durch

$$[b_i - \sqrt{(k+1)s^2 c_{ii} F_{k+1, n-k-1; 1-\gamma}}, \quad b_i + \sqrt{(k+1)s^2 c_{ii} F_{k+1, n-k-1; 1-\gamma}}].$$

Allgemein erhält man mit einer $(k+1) \times q$ dimensionalen Matrix

$$L = (l_1, \dots, l_q),$$

deren Spalten $l_j = (l_{j0}, \dots, l_{jk})^T$ die interessierenden Linearkombinationen

$$l_j^T \beta = \sum_{i=0}^k l_{ji} \beta_i$$

kennzeichnen, als Grenzen für simultane Konfidenzintervalle zum Niveau $1 - \gamma$ für $l_j^T \beta$, $j = 1, \dots, q$,

$$l_j^T \beta \pm \sqrt{d F_{d, n-k-1; 1-\gamma} s^2 l_j^T C l_j}.$$

Dabei ist $d = \text{Rang } L$ gleich der Anzahl der linear unabhängigen Zeilen (bzw. Spalten) von L .

Zu gewissen vorgegebenen Realisationen x_1, x_2, \dots, x_k der Regressoren schätzt man den zu erwartenden Wert \hat{y} des Regressanden (**Prognosewert**) durch

$$\hat{y} = a + \sum_{j=1}^k x_j b_j.$$

Ein Konfidenzintervall zum Niveau $1 - \gamma$ für \hat{y} hat die Grenzen

$$\hat{y} \pm \sqrt{s^2 x^T C x} t_{n-k-1; 1-\gamma/2} \quad \text{mit} \quad x = (1, x_1, \dots, x_k)^T.$$

Wollen wir feststellen, ob es überhaupt notwendig ist, alle Regressoren X_1, \dots, X_k zu betrachten oder ob vielleicht q bestimmte Regressoren zur Erklärung von Y überflüssig sind (**Modellreduktion**), so berechnen wir für die Regression auf X_1, \dots, X_k

$$SSR_k = B_{Y,(x_1, \dots, x_k)} \sum_{i=1}^n (y_i - \bar{y})^2$$

und für die Auswahl $X_{(1)}, \dots, X_{(k-q)}$ aus den X_1, \dots, X_k

$$SSR'_{k-q} = B_{Y,(x_{(1)}, \dots, x_{(k-q)})} \sum_{i=1}^n (y_i - \bar{y})^2$$

und testen dann zum Niveau γ die Hypothese, daß $X_{(1)}, \dots, X_{(k-q)}$ als Regressoren ausreichend sind. Sie muß verworfen werden, falls gilt

$$\frac{SSR_k - SSR'_{k-q}}{s^2 q} > F_{q, n-k-1; 1-\gamma}.$$

Dabei haben wir unterstellt, daß das Merkmal Y normalverteilt ist mit einem Erwartungswert an einer Stelle x_{10}, \dots, x_{k0} von

$$\alpha + \beta_1 x_{10} + \dots + \beta_k x_{k0}$$

und Varianz σ^2 , die durch

$$s^2 = \sum_{i=1}^n (y_i - a - b_1 x_{1i} - \dots - b_k x_{ki})^2 / (n - k - 1)$$

geschätzt werden kann, daß die Realisationen von Y unabhängig sind und daß die x_{ji} fest vorgegeben oder fehlerfrei gemessen sind. Die in der Regressionsanalyse erhaltenen Ergebnisse gelten unter der Bedingung, daß die x_{ji} ($j = 1, \dots, k$, $i = 1, \dots, n$) die wirklichen Verhältnisse in der interessierenden Grundgesamtheit ohne größere Fehler widerspiegeln; man sagt dann auch, die Ergebnisse sind *bedingt* durch die x_{ji} .

Lehnt man bei obigem **Reduktions-Test** die Hypothese nicht ab und schließt daraus, daß die spezifizierten Einflußgrößen $X_{(1)}, \dots, X_{(k-q)}$ zur Erklärung von Y über eine Regressionsbeziehung hinreichend sind, so sei aufgrund der i. d. R. hohen „Korrelationen zwischen den Regressoren“ (**Multikollinearität**) davor gewarnt, daraus direkt zu folgern, daß die hierbei eliminierten, restlichen Größen keinen Einfluß auf Y haben. Vielmehr kann man i. a. nur sagen, daß sie bei gleichzeitiger Berücksichtigung von $X_{(1)}, \dots, X_{(k-q)}$ keinen wesentlichen *zusätzlichen* Beitrag zur Erklärung von Y leisten.

Beispiel: Die Bearbeitungszeit Y eines Werkstücks in einer Maschine hängt vom Durchmesser X_1 und von der Länge X_2 des Werkstücks ab. Aufgrund von $n = 10$ Beobachtungen, vgl. Tab. 9, soll die Regressionsgleichung

$$\hat{y}(x_1, x_2) = a + b_1 x_1 + b_2 x_2$$

bestimmt werden.

Löst man das sich ergebende Normalengleichungssystem

$$\begin{aligned} 12,4 b_1 - 61 b_2 &= 70 \\ -61 b_1 + 954 b_2 &= -187, \end{aligned}$$

Tab. 9: Bearbeitungszeiten, Durchmesser und Länge von 10 Werkstücken, geschätzte Bearbeitungszeiten $\hat{y}_i(x_1, x_2)$ und $\hat{y}_i(x_{(1)})$ mit $x_{(1)} = x_1$

i	y_i	x_{1i}	x_{2i}	$\hat{y}_i(x_1, x_2)$	$\hat{y}_i(x_{(1)})$
1	24	4	31	22,401	19,742
2	28	6	3	29,311	31,032
3	25	5	7	23,446	25,387
4	13	3	29	15,090	14,097
5	27	5	27	28,266	25,387
6	14	3	18	12,439	14,097
7	15	3	28	14,849	14,097
8	32	6	17	32,685	31,032
9	14	4	8	16,858	19,742
10	28	5	12	24,651	25,387

so ergeben sich die Kleinst-Quadrat-Schätzer

$$b_1 = 6,829, \quad b_2 = 0,241$$

und somit

$$a = 22 - 6,829 \cdot 4,4 - 0,241 \cdot 18 = -12,386,$$

so daß sich

$$\hat{y}(x_1, x_2) = -12,386 + 6,829 x_1 + 0,241 x_2$$

ergibt; das Bestimmtheitsmaß dieser Regression ist

$$B_{Y,(x_1,x_2)} = 1 - \frac{34,974}{468} = 0,925.$$

Die Varianz σ^2 der Bearbeitungszeiten bei einer gegebenen Kombination von Durchmesser und Länge eines Werkstückes wird geschätzt durch, vgl. Tab. 9,

$$\begin{aligned} s^2 &= \sum_{i=1}^{10} (y_i + 12,386 - 6,839 x_{1i} - 0,241 x_{2i})^2 / (10 - 2 - 1) \\ &= \sum_{i=1}^{10} (y_i - \hat{y}_i(x_1, x_2))^2 / 7 = 34,974 / 7 = 4,996. \end{aligned}$$

Reicht es vielleicht zum 5% Niveau schon aus, eine Regressionsgerade

$$\hat{y}(x_{(1)}) = \tilde{a} + bx_{(1)} \quad \text{mit} \quad x_{(1)} = x_1$$

zu bestimmen, um die Bearbeitungszeiten zu erklären?

Die Regressionsgerade ergibt sich zu, vgl. Abschnitt 1,

$$\hat{y}(x_{(1)}) = -2,838 + 5,645 x_{(1)}$$

und das Bestimmtheitsmaß dieser linearen Regression ist

$$B_{Y,x_{(1)}} = 1 - \frac{72,8387}{468} = 0,844.$$

Somit ergibt sich

$$\text{SSR}_2 = B_{Y,(X_1, X_2)} \sum_{i=1}^n (y_i - \bar{y})^2 = 0,925 \cdot 468 = 432,9$$

$$\text{SSR}'_1 = B_{Y,(X_1)} \sum_{i=1}^n (y_i - \bar{y})^2 = 0,844 \cdot 468 = 394,992$$

und

$$\frac{\text{SSR}_2 - \text{SSR}'_1}{s^2 q} = \frac{432,9 - 394,992}{4,996 \cdot 1} = \frac{37,908}{4,996} = 7,588 > 5,59 = F_{1,7;0,95} \\ = (t_{7,0,975})^2$$

d. h. das Merkmal X_2 ist signifikant notwendig, um die Bearbeitungszeit Y eines Werkstücks zu erklären.

Bestimmt man eine Regression von der Bearbeitungszeit Y auf die Länge X_2 des Werkstücks, so ergibt sich die Regressionsgerade

$$\hat{y}(x_2) = 25,528 - 0,196 x_2;$$

man sieht hieraus, daß die zu einem Merkmal gehörigen Regressionskoeffizienten von Regressions- zu Regressionsfunktion erheblich abweichen, ja sogar das Vorzeichen wechseln können, was bei einer inhaltlichen Interpretation der Koeffizienten zu beachten ist.

Haben wir etwa zwischen den Variablen y, z, w eine **nichtlineare Regressionsgleichung** z. B. der folgenden Art

$$y(z, w) = \beta_0 + \beta_1 z + \beta_2 z^2 + \beta_3 z \cdot w + \beta_4 \sin w,$$

so erhalten wir wieder eine Regressionsbeziehung obiger Art mittels der Umbenennung

$$x_1 = z, \quad x_2 = z^2, \quad x_3 = z \cdot w, \quad x_4 = \sin w.$$

Sind y_i, z_i, w_i die beobachteten Werte für $i = 1, \dots, n$, so erhält man in der allgemeinen Modellgleichung die Regressorenwertematrix

$$X = \begin{pmatrix} 1 & z_1 & z_1^2 & z_1 \cdot w_1 & \sin w_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & z_n & z_n^2 & z_n \cdot w_n & \sin w_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} & x_{41} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} & x_{4n} \end{pmatrix},$$

so daß die Parameter wie oben beschrieben geschätzt und getestet werden können. Für weitere Vorgehensweisen sei etwa auch verwiesen auf Kap. XII, Abschnitt 2.5 und 5, Kap. XIV, Abschnitt 2, Elpelt/Hartung (1983a, 1984, 1985), Hartung/Werner (1983, 1984) sowie Kap. II, V und X in Hartung/Elpelt (1984).

6. Regression bei Fehlern in den Variablen

In den bisherigen Abschnitten wurde, obwohl nicht immer explizit erwähnt, vorausgesetzt, daß die Variablen x_i der Regression fehlerfrei gemessen werden. Ist dies nicht der Fall, so muß man andere Methoden der Regressionsrechnung verwenden und sogenannte Modelle mit Fehlern in den Variablen betrachten.

Hier nun sollen parametrische Verfahren vorgestellt werden, die zur Schätzung der Parameter einer Regressionsgeraden

$$y = \alpha + \beta x$$

verwandt werden können, wenn sowohl die Variable x als auch die Variable y nicht fehlerfrei gemessen werden können. Es sind dann nur die Meßergebnisse x'_1, \dots, x'_n und y'_1, \dots, y'_n von den wahren Werten x_1, \dots, x_n und y_1, \dots, y_n bekannt, d.h. es ist

$$x'_i = x_i + \varepsilon_i \quad \text{und} \quad y'_i = y_i + \delta_i.$$

Weiter muß man annehmen, daß die Erwartungswerte von ε_i und δ_i Null sind und daß ε_i für $i = 1, \dots, n$ die Varianz σ_1^2 und δ_i für $i = 1, \dots, n$ die Varianz σ_2^2 haben. Statt des üblichen Regressionsansatzes muß man hier also von einem Modell

$$(y'_i - \delta_i) = \alpha + \beta(x'_i - \varepsilon_i), \quad \text{d.h.}$$

$$y'_i = \alpha + \beta x'_i + (\delta_i - \beta \varepsilon_i) \quad \text{für } i = 1, \dots, n$$

ausgehen. Weiter verlangt man, daß die δ_i und ε_i untereinander und voneinander unabhängig sind. Die Fehlerterme $(\delta_i - \beta \varepsilon_i)$ sind nun im allgemeinen abhängig von x'_i , so daß die Parameter α und β nicht mehr nach der Methode der kleinsten Quadrate geschätzt werden können. Um überhaupt vernünftige Schätzungen durchführen zu können, muß man noch weitere Voraussetzungen machen, vgl. Abschnitte A, B, C. In Abschnitt D wird schließlich noch ein Modell vorgestellt, daß durch spezielle experimentelle Anforderungen an die x_1, \dots, x_n weiterhin die Möglichkeit der Schätzung nach der Methode der kleinsten Quadrate gibt.

A. Die Varianz σ_1^2 ist bekannt

Ist die Varianz σ_1^2 der Fehler $\varepsilon_1, \dots, \varepsilon_n$ bekannt, so kann man den Steigungsparameter β der Regressionsgeraden

$$y = \alpha + \beta x$$

durch

$$b = \frac{\sum_{i=1}^n x'_i y'_i - n \bar{x}' \bar{y}'}{\sum_{i=1}^n (x'_i)^2 - n (\bar{x}')^2 - (n-1) \sigma_1^2}$$

schätzen. Unter Verwendung dieses Schätzers für β ergibt sich dann als Schätzer für α wie üblich

$$a = \bar{y}' - b \bar{x}'.$$

Beispiel: Bei der Feststellung der Zusammensetzung eines Materials macht man sich die Tatsache zunutze, daß sich im Idealfall die Anteile der einzelnen Inhaltsstoffe in den Höhen von Absorptionspeaks bei spektrophotometrischen Analysen widerspiegeln.

Bei sechs rohen und vulkanisierten Äthylen – Prophylen Copolymeren wurden – bei bekanntem Prophylengehalt – jeweils die Höhen der Absorptionspeaks bei 892 cm^{-1} und 915 cm^{-1} gemessen (vgl. Brown/Tryon/Mandel (1963): „Determination of propylene in ethylene-propylene copolymers by infrared spectrophotometry“, Anal. Chemistry, 35, S. 2172–2176).

Aus diesen Werten berechnet man nun jeweils

$$\ln \left(100 \frac{\text{Absorptionspeak } 915 \text{ cm}^{-1}}{\text{Absorptionspeak } 892 \text{ cm}^{-1}} \right);$$

die sich ergebenden Werte sind in **Tabelle 10** zusammengestellt.

Tab. 10: Ergebnisse bei rohen und vulkanisierten Äthylen-Prophylen Copolymeren

x'_i (roh)	y'_i (vulkanisiert)
2,569	2,646
2,319	2,395
2,058	2,140
1,911	2,000
1,598	1,678
0,584	0,711

Nun möchte man aus den Daten der rohen Äthylen-Prophylen-Copolymere auf die der vulkanisierten schließen und betrachtet hierzu einen linearen Regressionsansatz. Da beide Variablen x und y nur fehlerbehaftet gemessen werden konnten, kann man die Parameter der Regressionsgeraden nicht mehr nach der Methode der kleinsten Quadrate schätzen. Geht man aber davon aus, daß die Varianz der Abweichungen ε_i von den wahren Werten x_1, \dots, x_n gerade $\sigma_1^2 = 0,007$ ist, so ergibt sich

$$b = \frac{23,674 - 6 \cdot 1,840 \cdot 1,928}{22,759 - 6 \cdot 3,386 - 5 \cdot 0,007} = \frac{2,389}{2,408} = 0,992$$

und somit

$$a = 1,928 - 0,992 \cdot 1,840 = 0,103.$$

Damit ist die gesuchte Regressionsgerade

$$\hat{y} = 0,103 + 0,992 x.$$

B. Die Varianz σ_2^2 ist bekannt

Ist nicht die Varianz der Fehler ε_i , sondern vielmehr die der Fehlerterme δ_i bekannt, so können natürlich die in Abschnitt A angegebenen Schätzer für die Parameter einer Regressionsgeraden nicht mehr verwandt werden. Vielmehr muß man den Steigungsparameter β durch

$$b = \frac{\sum_{i=1}^n (y'_i)^2 - n(\bar{y}')^2 - (n-1)\sigma_2^2}{\sum_{i=1}^n x'_i y'_i - n\bar{x}'\bar{y}'}$$

und α durch

$$a = \bar{y}' - b\bar{x}'$$

schätzen.

Beispiel: Für das Beispiel aus Abschnitt A, vgl. Tabelle 10, wollen wir nun unter der Annahme, daß

$$\sigma_2^2 = 0,007$$

bekannt ist, die Regressionsgerade

$$\hat{y} = a + bx$$

bestimmen. Es ergibt sich

$$b = \frac{24,638 - 6 \cdot 3,718 - 5 \cdot 0,007}{23,674 - 6 \cdot 1,840 \cdot 1,928} = \frac{2,295}{2,389} = 0,961$$

und

$$a = 1,928 - 0,961 \cdot 1,840 = 0,160.$$

C. Der Quotient der Varianzen σ_1^2 und σ_2^2 ist bekannt

Ist nicht eine der Varianzen σ_1^2, σ_2^2 selbst bekannt, sondern weiß man vielmehr etwas über das Verhältnis der beiden zueinander, d. h. kennt man den Quotienten $\frac{\sigma_2^2}{\sigma_1^2} = \lambda$, so ist es auch möglich, die Parameter der Regressionsgeraden zu schätzen. Und zwar ist

$$b = \frac{\sum_{i=1}^n (y_i)^2 - n(\bar{y}')^2 - \lambda \left(\sum_{i=1}^n (x_i)^2 - n(\bar{x}')^2 \right)}{2 \left(\sum_{i=1}^n x'_i y'_i - n \bar{x}' \bar{y}' \right)}$$

$$+ \frac{\sqrt{\left(\sum_{i=1}^n (y_i)^2 - n(\bar{y}')^2 - \lambda \left(\sum_{i=1}^n (x_i)^2 - n(\bar{x}')^2 \right) \right)^2 + 4\lambda \left(\sum_{i=1}^n x'_i y'_i - n \bar{x}' \bar{y}' \right)^2}}{2 \left(\sum_{i=1}^n x'_i y'_i - n \bar{x}' \bar{y}' \right)}$$

ein Schätzer für den Steigungsparameter β und

$$a = \bar{y}' - b\bar{x}'$$

ein Schätzer für den Parameter α .

Beispiel: Wir kommen noch einmal auf das Beispiel aus Abschnitt A zurück. Es ist sicherlich vertretbar davon auszugehen, daß die Varianzen der Beobachtungen an rohen und vulkanisierten Äthylen-Prophylen Copolymeren gleich sind, d.h.

$$\frac{\sigma_2^2}{\sigma_1^2} = 1$$

Ausgehend von dieser Annahme ergeben sich als Schätzer für die Parameter der Regressionsgeraden

$$\begin{aligned}
 b &= \frac{24,638 - 6 \cdot 3,718 - 1(22,759 - 6 \cdot 3,385)}{2(23,674 - 6 \cdot 1,840 \cdot 1,928)} \\
 &\quad + \frac{\sqrt{(24,638 - 6 \cdot 3,718 - 1(22,759 - 6 \cdot 3,385))^2 + 4 \cdot 1(23,674 - 6 \cdot 1,840 \cdot 1,928)^2}}{2(23,674 - 6 \cdot 1,840 \cdot 1,928)} \\
 &= \frac{-0,119 + \sqrt{(-0,119)^2 + 4 \cdot (2,389)^2}}{2 \cdot 2,389} = \frac{-0,119 + \sqrt{22,843}}{4,778} = 0,975 \quad \text{und} \\
 a &= 1,928 - 0,975 \cdot 1,840 = 0,134.
 \end{aligned}$$

Die geschätzte Regressionsgerade ist somit

$$\hat{y} = 0,134 + 0,975x.$$

D. Das Berkson-Modell

Viele Experimente sind so angelegt, daß die Variablen x_i und y_i zwar fehlerbehaftet sind, man aber trotzdem eine Regressionsgerade

$$y = \alpha + \beta x$$

durch Schätzung der Parameter α und β nach der Methode der kleinsten Quadrate schätzen kann.

Wie Berkson (1950) herausstrich, ist es häufig so, daß die x_1, \dots, x_n keine Zufallsvariablen sind, sondern im vorhinein festgelegte Werte. Die wirklich realisierten, allerdings unbekannten Werte x'_1, \dots, x'_n weichen dann lediglich in der Weise von x_1, \dots, x_n ab, daß die $\varepsilon_1, \dots, \varepsilon_n$ gerade solche Werte sind, daß gilt

$$x'_i = x_i + \varepsilon_i \quad \text{für } i = 1, \dots, n.$$

Dann sind zwar x'_i und ε_i korreliert, nicht aber ε_i und x_i , und man kann, da ja x_1, \dots, x_n vorgegebene Werte sind, ein Modell

$$(y'_i - \delta_i) = \alpha + \beta(x_i + \varepsilon_i), \quad \text{d. h.}$$

$$y'_i = \alpha + \beta x_i + (\delta_i + \beta \varepsilon_i) \quad \text{für } i = 1, \dots, n$$

betrachten. Hier sind nun die x_i und die $(\delta_i + \beta \varepsilon_i)$ für $i = 1, \dots, n$ unkorreliert, so daß man den Steigungsparameter β der Regressionsgeraden durch

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y'_i - \bar{y}')}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

und α durch

$$a = \bar{y}' - b\bar{x}$$

in üblicher Weise nach der Methode der kleinsten Quadrate schätzen kann. Lediglich die Interpretation der Fehlerterme ist hier eine andere als bei der gewöhnlichen linearen Regression. Der Fehler röhrt zum einen her aus der Meßungenauigkeit der Werte y'_i , zum anderen aus der Ungenauigkeit der y'_i , die aus dem Unterschied zwischen wahren und festgelegten Werten x'_i bzw. x_i herrühren.

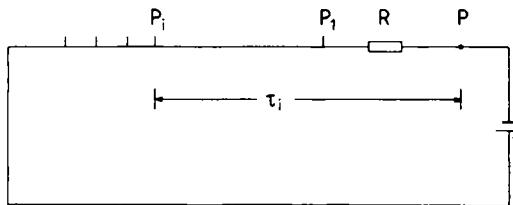


Abb. 17: Versuchsaufbau zur Bestimmung des Widerstands eines Drahtes

Beispiel: Zur Bestimmung des Widerstandes eines Drahtes wird der Versuchsaufbau aus Abb. 17 verwendet. Bei vorgegebener Standardstromstärke 1 soll der Widerstand ϱ eines Drahtes pro Längeneinheit bestimmt werden. Ferner möchte man noch den zwischengeschalteten Widerstand R bestimmen. Dazu werden die Entfernung τ_i und die Spannung y_i zwischen zwei Punkten P (fest) und P_i gemessen. Die Spannungen y_1, \dots, y_n können natürlich nicht exakt gemessen werden, man beobachtet also vielmehr Spannungen y'_1, \dots, y'_n . Auch der Abstand zwischen den Punkten P und P_1, P_2, \dots bzw. P_n kann natürlich nicht exakt bestimmt werden. Gibt man aber für die Abstände von P nach P_i feste Werte τ_i vor, die dann aber nicht so exakt realisiert werden können, so kann man ausgehend vom Berkson-Modell ϱ und R dennoch nach der Methode der kleinsten Quadrate schätzen. Die Regressionsgerade

$$y = R + \varrho\tau,$$

deren Parameter gerade die zu schätzenden Widerstände ϱ und R sind, kann dann ausgehend von der Beziehung

$$y'_i = R + \varrho\tau_i + (\delta_i + \varrho\varepsilon_i)$$

geschätzt werden. In Tabelle 11 sind für einige Längen τ_i die Spannungen y'_i zwischen den Punkten P und P_i zusammengestellt.

Es ergibt sich

$$\tilde{\varrho} = \frac{\sum_{i=1}^{11} \tau_i y'_i - 11 \bar{\tau} \bar{y}'}{\sum_{i=1}^{11} \tau_i^2 - 11 \bar{\tau}^2} = \frac{3318 - 11 \cdot 10 \cdot 28,127}{1210 - 11 \cdot 100} = \frac{224,03}{110} = 2,037$$

als Schätzer für den Drahtwiderstand pro Längeneinheit.

Der eingebaute Widerstand R hat den geschätzten Wert

$$\tilde{R} = \bar{y}' - \tilde{\varrho}\bar{\tau} = 28,127 - 2,037 \cdot 10 = 7,757.$$

7. Regressionsgerade nach Wald

Das **Waldsche Verfahren** ist wie die Verfahren aus Abschnitt 6 auch dann anwendbar, wenn im Regressionsansatz

$$y_i = \alpha + \beta_1 x_i + e_i,$$

die Größen x_i und y_i nicht fehlerfrei gemessen werden können, benötigt aber keine

Tab. 11: Versuchsergebnisse zur Bestimmung des Widerstandes eines Drahtes, vgl. Abb. 17

i	vorgegebene Länge τ_i	gemessene Spannung y'_i
1	5	17,6
2	6	20,1
3	7	21,5
4	8	25,1
5	9	25,8
6	10	28,4
7	11	30,6
8	12	31,7
9	13	33,8
10	14	36,5
11	15	38,3

Angaben über die Varianzen der Fehler. Man ordnet zunächst die n Beobachtungspaare $(x_1, y_1), \dots, (x_n, y_n)$ so, daß die x_i mit steigendem Index stets größer werden. Die geordneten Paare unterteilt man nun in zwei gleich große Gruppen. Ist die Anzahl n der Beobachtungspaare ungerade, so wird das mittlere Paar weggelassen. Bezeichnen wir also die geordneten Beobachtungspaare mit $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ so bilden wir die Gruppen

$(x_{(1)}, y_{(1)}), \dots, (x_{(n/2)}, y_{(n/2)})$ und $(x_{(n/2+1)}, y_{(n/2+1)}), \dots, (x_{(n)}, y_{(n)})$

falls n gerade und

$(x_{(1)}, y_{(1)}), \dots, (x_{((n-1)/2)}, y_{((n-1)/2)})$ und $(x_{((n+1)/2)}, y_{((n+1)/2)}), \dots, (x_{(n)}, y_{(n)})$

falls n ungerade ist.

Sodann berechnen wir die Gruppenmittelwerte in beiden Gruppen $(\bar{x}_1; \bar{y}_1)$ und (\bar{x}_2, \bar{y}_2) sowie die Gesamtmittel \bar{x} , \bar{y} .

Dann können wir die Parameter β und α schätzen durch

$$b = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1} \quad \text{und} \quad a = \bar{y} - b\bar{x}.$$

Beispiel: Man vermutet, daß bei der Herstellung eines Produktes die Lagerzeit der benötigten Rohstoffe einen Einfluß auf die Festigkeit des Endproduktes hat. Daher wurden verschiedene lange gelagerte Rohprodukte zur Fertigung verwendet und

Tab. 12: Wahre und geschätzte Festigkeiten des Produkts, vgl. Abb. 18

x_i	y_i	\hat{y}_i	x_i	y_i	\hat{y}_i
3	7,2	11,4335	11	20,0	20,0975
4	13,0	12,5165	12	21,7	21,1805
5	12,6	13,5995	13	23,3	22,2635
6	14,6	14,6825	14	23,6	23,3465
7	16,2	15,7655	15	23,4	24,4295
8	18,1	16,8485	16	25,3	25,5125
9	20,7	17,9315	17	26,7	26,5955
10	19,4	19,0145	18	27,1	27,6785

dann jeweils die Festigkeit des Endproduktes gemessen. In **Tabelle 12** sind die gemessenen $n = 16$ Wertepaare zusammengestellt.

Wir wollen die zugehörige Regressionsgerade nach dem Waldschen Verfahren schätzen. Die Beobachtungen sind schon nach der Größe der x-Komponente geordnet, so daß wir, da $n = 16$ gerade ist, neben \bar{x} und \bar{y} die Mittelwerte \bar{x}_1 und \bar{y}_1 der ersten acht und \bar{x}_2 , \bar{y}_2 der letzten acht Beobachtungen berechnen müssen.

Es ist

$$\bar{x} = \frac{168}{16} = 10,5, \quad \bar{x}_1 = \frac{1}{8} \sum_{i=1}^8 x_i = \frac{52}{8} = 6,5,$$

$$\bar{x}_2 = \frac{1}{8} \sum_{i=9}^{16} x_i = \frac{116}{8} = 14,5, \quad \bar{y} = \frac{312,9}{16} = 19,556,$$

$$\bar{y}_1 = \frac{1}{8} \sum_{i=1}^8 y_i = \frac{121,8}{8} = 15,225, \quad \bar{y}_2 = \frac{1}{8} \sum_{i=9}^{16} y_i = \frac{191,1}{8} = 23,888,$$

so daß sich

$$b = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1} = \frac{23,888 - 15,225}{14,5 - 6,5} = \frac{8,663}{8} = 1,083$$

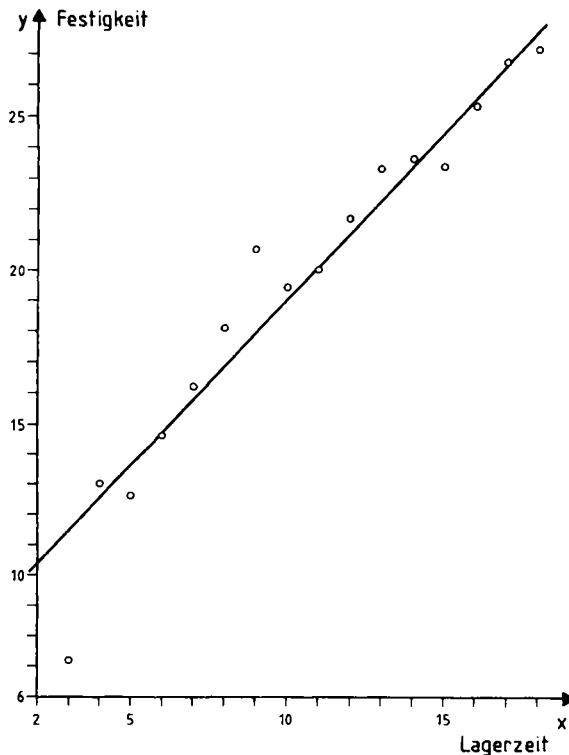


Abb. 18: Beobachtungswerte und Regressionsgerade für die Produktfestigkeiten nach dem Waldschen Verfahren

und

$$a = \bar{y} - b\bar{x} = 19,556 - 1,083\bar{x} = 8,1845$$

als Schätzer ergeben. Die geschätzte Regressionsgerade ist also

$$\hat{y} = 8,1845 + 1,083x.$$

In Tabelle 12 sind neben den wahren auch die geschätzten Produktfestigkeiten eingetragen; diese Werte sind in **Abb. 18** graphisch dargestellt.

Eine ausführliche Darstellung von im Rahmen der Regressionsanalyse auftretenden Problemen findet man etwa in den Büchern von Draper/Smith (1981), Graybill (1961, 1976), Ezekiel/Fox (1970), Seber (1977), Chatterjee/Price (1977), Krafft (1978), Schach/Schäfer (1978), Myers (1979), Trenkler (1981), Hartung/Werner (1984) sowie Hartung/Elpelt (1984).

Speziell bzgl. allgemeiner Modelle mit Fehlern in den Regressoren sei etwa verwiesen auf Schneeweß (1976) und bzgl. Modellen mit festen und zufälligen Parametern bzw. Effekten auf Elpelt/Hartung (1983 a, 1984), Hartung/Elpelt (1984).

Bisher haben wir stets vorausgesetzt, daß alle Beobachtungen gleiche Varianz haben und unkorreliert, d. h. – wegen der Normalverteilungsannahme – unabhängig sind. Ist dies nicht der Fall, so muß auf kompliziertere Verfahren zurückgegriffen werden: Sind die Varianzen und/oder Korrelationen bekannt, so läßt sich die gewichtete Methode der kleinsten Quadrate bzw. die Aitken-Schätzung verwenden; ansonsten bietet sich die iterierte Aitken-Schätzung als adäquate Methode an, vgl. Aitken (1935) und etwa Kakwani (1967), Roß (1981), Jöckel (1982), Voet (1985) sowie Abschnitt 5 in Kap. XII.

Stehen Vorinformationen über die Parameter in Form von linearen Gleichungs-Nebenbedingungen (in der Varianzanalyse Reparametrisierungsbedingungen genannt, vgl. Kap. XI) zur Verfügung, so können allgemein mittels des Konzeptes der Restringierten (Moore-Penrose-)Pseudoinversen restringierte Schätzungen (kleinsten Norm) gewonnen werden; man vgl. hierzu etwa Minamide/Nakamura (1970), Holmes (1972), Hartung (1976 b, 1977 a, 1979 a). Das Testen allgemeiner Linearer Hypothesen unter solchen Nebenbedingungen, bei dem auch das sogenannte Testbarkeitsproblem (insbesondere bei unbalancierten Versuchsdaten) auftritt, wird ausführlich in Hartung/Werner (1980, 1983, 1984) sowie Hartung/Elpelt (1984) diskutiert.

Liegen Vorinformationen in allgemeiner Form, z. B. Ungleichungen, vor, so kann das Regressionsproblem mit (Optimierungs-, Minimax-)Verfahren angegangen werden, wie sie etwa in Stoer (1971), Liew (1976), Gaul/Hartung (1979), Hartung (1977 b, 1978 a, 1978 b, 1982 a, 1982 b), Elpelt/Hartung (1983 b) dargestellt werden.

Methoden zur Bestimmung von Schätzungen kleinster Norm im Regressionsproblem mit allgemeinen konvexen Nebenbedingungen finden sich in Hartung (1978 c, 1978 d, 1980, 1985).

Kapitel XI: Varianzanalyse

Die **Varianzanalyse** ist eine recht **universelle Auswertungsmethode** für Versuche, bei denen **mehrere Einflußfaktoren** zu berücksichtigen sind. Sie wurde im wesentlichen von **Sir Ronald Aylmer Fisher** (1890–1962) in den zwanziger Jahren entwickelt.

Die verschiedenen Einflußfaktoren können dabei bereits von den Versuchsobjekten getragen oder den Objekten erst zugeordnet werden.

Untersucht wird ein Merkmal, d. h. die Merkmalsausprägungen des Merkmals werden bei den einzelnen Versuchsobjekten gemessen, auf das diese Faktoren einen Einfluß haben (können). In einem (zum Teil) kontrollierten Experiment wird also der Einfluß der Faktoren auf das betreffende Merkmal überprüft. Zum einen möchte man dabei wissen, ob überhaupt ein Einfluß vorhanden ist, zum anderen ist man daran interessiert, die Stärke des Einflusses zu untersuchen.

Die Varianzanalyse beantwortet die Frage, ob die verschiedenen Stufen eines Einflußfaktors statistisch signifikant unterschiedliche Wirkungen auf das interessierende Merkmal zeigen und quantifiziert ihre Effekte.

Ist man an der qualitativen Beurteilung der Zusammenhänge der Einflußfaktoren untereinander und zu dem spezifizierten Merkmal interessiert, so kommen Analysen mittels Kontingenztafeln (insbesondere, wenn das Merkmal nur nominal skaliert ist), die in Kap. VII beschrieben werden, sowie Methoden der Korrelationsanalyse, vgl. Kap. IX, zur Anwendung. Insbesondere wird in Abschnitt 5 des Kap. VII der Vergleich mehrerer Meßreihen aus diskreten Verteilungen ausführlich dargestellt.

Bezüglich allgemeiner Aspekte zur **Planung und Gestaltung von Versuchen** sei auf Kap. V verwiesen, wo speziell im Abschnitt 5 verschiedene **Versuchspläne** bereits vorgestellt wurden, die sich mit Methoden der Varianzanalyse auswerten lassen. Ebenfalls dort finden sich ausführliche Bemerkungen zur **Anwendbarkeit** des im folgenden verwandten **F-Tests** im Falle von Datensituationen, die den gestellten Forderungen nicht voll genügen.

Im Abschnitt 1 werden wir uns hier zunächst mit der Auswertung von p unabhängigen Meßreihen (**vollständig randomisierter Versuchsplan**) mittels **einfacher Varianzanalyse** beschäftigen.

Die Auswertung von **einfachen Blockexperimenten** erfolgt im Abschnitt 2, und die **zweifache Varianzanalyse (zweifache Kreuzklassifikation)** wird im Abschnitt 3 behandelt. Schließlich werden im Abschnitt 4 noch einfache Modelle der **eigentlichen Varianzanalyse (Modelle mit zufälligen Effekten)** vorgestellt; diese heißen auch **Modelle II der Varianzanalyse**, während die in den Abschnitten 1 bis 3 betrachteten **Modelle mit festen Effekten** auch **Modelle I der Varianzanalyse** genannt werden.

In diesem Kapitel bezeichnet $F_{n_1, n_2; \gamma}$ stets das γ -Quantil der F-Verteilung mit n_1 und n_2 Freiheitsgraden und $q_{n, v; \gamma}$ das γ -Quantil der sogenannten Studentisierten Spannweite q mit n und v Freiheitsgraden; diese Quantile sind im Anhang vertafelt.

1. Vergleich von p unabhängigen Meßreihen (Stichproben) – einfache Varianzanalyse, vollständig randomisierter Versuchsplan

Sollen p Meßreihen der Längen n_1, n_2, \dots, n_p miteinander verglichen werden, so lassen sich die $n_1 + n_2 + \dots + n_p = N > p$ Versuchsauskommen Y_{ij} für $i = 1, \dots, p$ und $j = 1, \dots, n_i$ übersichtlich in Form einer Tabelle darstellen. Haben alle Meßreihen die gleiche Länge $n = n_1 = \dots = n_p$, so ergibt sich **Tab. 1**, und sonst **Tab. 2**.

Tab. 1: Schematische Darstellung der Beobachtungsergebnisse bei p Meßreihen gleicher Länge n

		Meßreihe					
		1	2	3	...	p	
		1	Y_{11}	Y_{21}	Y_{31}	...	Y_{p1}
		2	Y_{12}	Y_{22}	Y_{32}	...	Y_{p2}
		⋮	⋮	⋮	⋮	⋮	⋮
		⋮	⋮	⋮	⋮	⋮	⋮
		⋮	⋮	⋮	⋮	⋮	⋮
		n	Y_{1n}	Y_{2n}	Y_{3n}	...	Y_{pn}

Tab. 2: Schematische Darstellung der Beobachtungsergebnisse bei p Meßreihen ungleicher Längen n_1, \dots, n_p

		Meßreihe					
		1	2	3	...	p	
		1	Y_{11}	Y_{21}	Y_{31}	...	Y_{p1}
		2	Y_{12}	Y_{22}	Y_{32}	...	Y_{p2}
		⋮	⋮	⋮	⋮	⋮	⋮
		⋮	⋮	Y_{2n_2}	⋮	⋮	⋮
		⋮	Y_{1n_1}	⋮	⋮	⋮	Y_{pn_p}
		⋮	⋮	Y_{3n_3}	⋮	⋮	⋮

Als **Beispiele** für Probleme, die den Vergleich von p Meßreihen erfordern, sind etwa zu nennen:

- (a) p Maschinen stehen für die Produktion eines bestimmten Gutes zur Auswahl. Von der i -ten Maschine sind bei der Herstellung von n_i Probeexemplaren charakteristische Werte (wie zum Beispiel der Durchmesser oder das Gewicht) Y_{i1}, \dots, Y_{in_i} gemessen worden, entweder direkt oder als Abweichung von einem Sollwert. Kann auf einen Qualitätsunterschied zwischen den Maschinen geschlossen werden?
- (b) p Medikamente stehen für eine medizinische Behandlung zur Verfügung. Das i -te Medikament wird n_i Versuchspersonen verabreicht und die Wirkung an geeigneten Kenngrößen (wie zum Beispiel Eisengehalt des Blutserums oder Blutdruck) gemessen als Y_{i1}, \dots, Y_{in_i} . Kann auf verschiedenartige Wirkung (bezüglich des untersuchten Merkmals) der Medikamente geschlossen werden?
- (c) In p verschiedenen Regionen werden die Umsätze der Firmen einer Branche erhoben. Können die Verteilungen der Umsätze in den Regionen als unterschiedlich angesehen werden?
- (d) In p vornehmlich katholischen, vornehmlich evangelischen bzw. gemischtreligiösen Ländern werden die Anzahlen der unehelichen Geburten über einige Jahre hinweg erhoben. Kann für alle Länder eine steigende Tendenz konstatiert werden?

Die Meßreihen können natürlich auch dadurch entstehen, daß man zum Vergleich von p Behandlungen N Objekte aus einer Grundgesamtheit auswählt und n_i dieser Objekte zufällig, vgl. Abschnitt 8 in Kap. III, die i -te Behandlung zuordnet (**vollständig randomisierter Versuchsplan**). Beispiele hierfür sind etwa:

- (a) Der Einfluß verschiedener Vorbereitungstherapien auf den Erfolg einer Psychotherapie soll untersucht werden. Aus der Gesamtheit der möglichen Versuchspersonen werden 20 Personen zufällig ausgewählt, und je 5 Personen wird eine der interessierenden 4 Vorbereitungstherapien zufällig zugeordnet. Bestehen Unterschiede zwischen den Vorbereitungstherapien und welcher Art sind sie?
- (b) Um den Einfluß von 5 verschiedenen Lagermethoden auf die Haltbarkeit eines bestimmten Produktes zu überprüfen, werden 25 Produkte zufällig aus der Grundgesamtheit ausgewählt, und zufällig wird je 5 Produkten eine der möglichen Lagermethoden zugeordnet. Bestehen Unterschiede zwischen den Lagermethoden? Welche Lagermethode ist welcher anderen vorzuziehen?

Die p Meßreihen werden hier als unabhängig angenommen. Ein Test, der prüft, ob vorliegende Meßreihen signifikant abhängig sind, wird in Kap. IX beschrieben.

1.1. Testen auf signifikante Lokationsunterschiede

A. Der F-Test (normalverteilte Grundgesamtheit)

Entstammen die Beobachtungen Y_{i1}, \dots, Y_{in_i} der i -ten Meßreihe ($i = 1, \dots, p$) einer $N(\mu_i, \sigma^2)$ -Verteilung mit unbekanntem Erwartungswert μ_i und unbekannter, für alle Meßreihen gleicher Varianz $\sigma^2 > 0$, so kann mit

$$N = \sum_{i=1}^p n_i > p$$

der **F-Test** zum Vergleich der Mittelwerte μ_1, \dots, μ_p verwandt werden. Man berechnet für jede der p Meßreihen den Mittelwert

$$\bar{Y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij},$$

der natürlich ein Schätzer für μ_i ist, sowie das Gesamtmittel

$$\bar{Y}_{..} = \frac{1}{N} \sum_{i=1}^p \sum_{j=1}^{n_i} Y_{ij}$$

und erstellt die **Varianzanalysetafel (ANOVA-Tafel)** aus **Tab. 3**.

Tab. 3: Tafel der einfachen Varianzanalyse

Streuungs-ursache	Freiheits-grade (FG)	Quadratsumme (SS)	Mittlere Quadratsumme (MS)
Unterschiede zwischen den Meßreihen	$p - 1$	$SST = \sum_{i=1}^p n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$	$MST = SST / (p - 1)$
Zufälliger Fehler	$N - p$	$SSE = \sum_{i=1}^p \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$	$MSE = SSE / (N - p)$
Gesamt	$N - 1$	$SSG = \sum_{i=1}^p \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$	

Die Bezeichnungsweise in der Varianzanalyse wurde von R. A. Fisher eingeführt und bezieht sich auf:

ANOVA: Analysis of variance, Varianzanalyse;

(M)(S)ST: „(Mean) sum of squares for treatments“, „(mittlere) Quadratsumme der Behandlungen“, dabei versteht man die p Meßreihen als Versuchsauskommen, die p unterschiedlichen „Behandlungen“ (in einem allgemeinen Sinne) von Versuchsobjekten entsprechen;

(M)(S)SE: „(Mean) sum of squares for errors“, „(mittlere) Fehler-Quadratsumme“;

SSG: „Grand sum of squares“, „Gesamt-Quadratsumme“.

Zu dem vorgegebenen Niveau γ verwerfen wir die Nullhypothese

$$H_0: \mu_1 = \mu_2 = \dots = \mu_p$$

und schließen auf einen **signifikanten Unterschied** (zum Niveau γ) zwischen den Reihenmitteln genau dann, wenn

$$F = \frac{MST}{MSE} > F_{p-1, N-p; 1-\gamma}$$

gilt (**F-Test**). Quantile $F_{p-1, N-p; 1-\gamma}$ der F-Verteilung sind im Anhang vertafelt.

Die Größe MSE ist ein erwartungstreuer Schätzer für die Varianz σ^2 , und es gilt die sogenannte „**einfache Streuungszerlegung**“ oder auch „**Quadratsummenzerlegung**“

$$SSG = SST + SSE.$$

Mit obigen Darstellungen läßt sich dies so interpretieren, daß die „**Gesamtvariation**“ SSG sich aufspalten läßt in die „**Variation zwischen den Meßreihen**“ SST und die „**Variation innerhalb der Meßreihen**“ SSE. Der F-Test mit der Prüfgröße $F = MST/MSE$ vergleicht mithin gerade diese „Variationskomponenten“ (unter Berücksichtigung der zugehörigen Freiheitsgrade), „analysiert“ also die verschiedenen „Varianzen“ und schließt auf einen signifikanten Unterschied zwischen den Meßreihen, wenn die Variation zwischen den Meßreihen „wesentlich größer“ ist als die innerhalb der Reihen.

MSE ist χ^2 -verteilt mit $N - p$ Freiheitsgraden, MST ist bei Gültigkeit der Nullhypothese χ^2 -verteilt mit $p - 1$ Freiheitsgraden. Weiterhin sind MSE und MST stochastisch unabhängig, so daß $F = MST/MSE$ unter der Nullhypothese einer F-Verteilung mit $p - 1$ und $N - p$ Freiheitsgraden genügt, und somit kann deren $(1 - \gamma)$ -Quantil zur Festlegung dieses „wesentlich größer“ herangezogen werden.

Beispiel: Es ist die Zugfestigkeit von $p = 3$ Drahtsorten zu überprüfen. Man kann davon ausgehen, daß die jeweils $n = n_i = 6$ zur Verfügung stehenden Proben insgesamt unabhängig und gemäß einer Normalverteilung $N(\mu_i, \sigma^2)$ mit einem unbekannten, von der Drahtsorte (eventuell) abhängenden Mittelwert μ_i und einer unbekannten, für alle Drahtsorten zutreffenden Varianz $\sigma^2 > 0$ verteilt sind.

Es ist zum Niveau von $\gamma = 0,05$ zu überprüfen, ob Unterschiede in den Zugfestigkeiten statistisch gesichert sind, wenn die Höchstbelastungen Y_{ij} , $1 \leq i \leq 3$, $1 \leq j \leq 6$ (in Newton pro mm^2) aus **Tab. 4** gemessen worden sind.

Aus diesen Daten ergibt sich die ANOVA-Tafel aus **Tab. 5**, so daß mit

$$\frac{MST}{MSE} = \frac{54,02}{14,44} = 3,74 > 3,68 = F_{2,15;0,95}$$

Tab. 4: Daten zum Zugfestigkeitsversuch und Rangzahlen

i j	Drahtsorte					
	1 Y_{1j}	2 r_{1j}	2 Y_{2j}	3 r_{2j}	3 Y_{3j}	3 r_{3j}
1	9,0	6	7,3	3,5	18,0	17
2	15,4	14	15,6	15	9,6	7
3	8,2	5	14,2	12	11,5	10
4	3,9	1	13,0	11	19,4	18
5	7,3	3,5	6,8	2	17,1	16
6	10,8	9	9,7	8	14,4	13
\bar{Y}_i	9,1	38,5	11,1	51,5	15,0	81,0

Tab. 5: ANOVA-Tafel zum Zugfestigkeitsversuch

Streuungsursache	Freiheitsgrade	Quadratsummen	Mittlere Quadratsummen
Unterschiede zwischen den Meßreihen	2	SST = 108,04	$MST = \frac{108,04}{2} = 54,02$
Zufallsfehler	15	SSE = 216,58	$MSE = \frac{216,58}{15} = 14,44$
Insgesamt	17	SSG = 324,62	

ein signifikanter Unterschied zwischen den Mittelwerten der drei Meßreihen besteht.

B. Der Test von Kruskal und Wallis

Der Kruskal-Wallis-Test prüft ohne eine Normalverteilung vorauszusetzen (lediglich eine stetige Verteilung wird gefordert), ob die Mittelwerte von p Meßreihen der Längen n_1, \dots, n_p als signifikant verschieden angesehen werden können.

Dabei geht man so vor, daß man zunächst den N Meßwerten Y_{11}, \dots, Y_{pn_p} Rangzahlen r_{ij} zuordnet, wobei der r-kleinste Meßwert Y_{ij} den Rang $r_{ij} = r$ erhält. Beim Vorliegen von Bindungen, d. h. wenn mehrere Meßwerte gleich sind, werden jeweils die mittleren der in Frage stehenden Ränge zugewiesen, vgl. Beispiel. Präziser formuliert: Treten insgesamt g verschiedene Meßwerte $z_1 < z_2 < \dots < z_g$ auf und zwar z_k genau t_k -mal, $t_k \geq 1$, $\sum_{k=1}^g t_k = N$, so entspricht der Fall $g < N$ [also $t_k > 1$ für mindestens ein k] gerade dem Vorliegen von Bindungen. Dem Meßwert Y_{ij} wird jeweils der Rang $r_{ij} = \sum_{l=0}^{k-1} t_l + \frac{t_k + 1}{2}$ zugeordnet, falls $Y_{ij} = z_k$ ist; dabei wird $t_0 = 0$ gesetzt.

Für jede der Meßreihen berechnet man dann die Summe der Ränge, die den Meßwerten der betreffenden Reihe zugeordnet sind:

$$r_{i \cdot} = \sum_{j=1}^{n_i} r_{ij}, \quad 1 \leq i \leq p,$$

und hieraus die Prüfgröße

$$H = \frac{1}{B} \left[\frac{12}{N(N+1)} \sum_{i=1}^p \frac{1}{n_i} r_{i \cdot}^2 - 3(N+1) \right],$$

$$\text{wobei } B = 1 - \frac{1}{N^3 - N} \sum_{l=1}^g (t_l^3 - t_l)$$

eine Korrekturgröße ist, die in Abwesenheit von Bindungen gerade den Wert $B = 1$ annimmt. Es kann jedoch auch beim Vorliegen von „wenigen“ Bindungen, etwa wenn die Anzahl der verschiedenen Meßwerte $g \geq \frac{3}{4}N$ ist, immer noch $B = 1$ gesetzt werden.

Zum vorgegebenen Niveau γ , $0 < \gamma < 1$, wird nun auf signifikante Unterschiede in den Reihenmitteln geschlossen, wenn gilt

$$H > h_{p, (n_1, \dots, n_p); 1-\gamma}.$$

Für den Fall $p = 3$ und $n_i \leq 5$ für $i = 1, 2, 3$ sind (zum Teil approximative) Quantile $h_{p, (n_1, \dots, n_p); 1-\gamma}$ in Tab. 6 angegeben; weitere sind für diesen Fall etwa in Owen (1962) oder Hollander/Wolfe (1973) zu finden, wo die exakten angenommenen Niveaus zu einer Auswahl vorgegebener kritischer Werte vertafelt sind.

Stehen keine kritischen Werte zur Verfügung, so schließt man approximativ auf zum Niveau γ signifikante Unterschiede der Reihenmittel, falls gilt

$$H > \chi_{p-1; 1-\gamma}^2;$$

die Quantile der χ^2 -Verteilung sind im Anhang vertafelt.

Beispiel: Für das Beispiel aus Abschnitt A wollen wir nun den Kruskal-Wallis-Test zum 5% Niveau durchführen; die Rangzahlen sind in Tab. 4 bereits angegeben. Da nur eine Bindung auftritt, ist

$$B = 1 - \frac{1}{18^3 - 18} (2^3 - 2) = 0,999$$

und somit

$$\begin{aligned} H &= \frac{1}{0,999} \left[\frac{12}{18 \cdot 19 \cdot 6} \cdot (38,5^2 + 51,5^2 + 81,0^2) - 3 \cdot 19 \right] \\ &= 5,552 < 5,991 = \chi_{2; 0,95}^2. \end{aligned}$$

Mithin erkennt der Kruskal-Wallis-Test zum Niveau von 5% keine signifikanten Unterschiede zwischen den Reihenmittelwerten.

1.2. Simultane Vergleiche von p Mittelwerten

Bisher wurden nur Tests vorgestellt, die global testen, ob die Mittelwerte von p Meßreihen signifikant verschieden sind. Will man zusätzlich herausfinden, welche Mittelwerte paarweise verschieden sind, so bedient man sich der **multiplen Vergleiche**.

Tab. 6: Kritische Werte $h_{3,(n_1, n_2, n_3), 1-\gamma}$ zum Kruskal-Wallis-Test für $n_1, n_2, n_3 \leq 5$

N	n ₁	n ₂	n ₃	$h_{3,(n_1, n_2, n_3), 1-\gamma}$	
				($\gamma \approx$)	$\gamma \approx 0,05$
4	1	1	2	2,70 (0,50)	
5	1	1	3	3,20 (0,30)	
	1	2	2	3,60 (0,20)	
6	1	1	4	3,57 (0,20)	
	1	2	3	4,29 (0,10)	
	2	2	2	4,57 (0,07)	
7	1	1	5	3,86 (0,14)	
	1	2	4	4,50 (0,08)	4,82
	1	3	3	4,57 (0,10)	5,14
	2	2	3	4,50 (0,07)	4,71
8	1	2	5	4,20 (0,10)	5,00
	1	3	4	4,06 (0,09)	5,21
	2	2	4	4,46 (0,10)	5,13
	2	3	3	4,56 (0,10)	5,14
9	1	3	5	4,02 (0,10)	4,87
	1	4	4	4,07 (0,10)	4,87
	2	2	5	4,37 (0,09)	5,04
	2	3	4	4,51 (0,10)	5,40
	3	3	3	4,62 (0,10)	5,60
10	1	4	5	3,96 (0,10)	4,86
	2	3	5	4,49 (0,10)	5,11
	2	4	4	4,55 (0,10)	5,24
	3	3	4	4,70 (0,10)	5,72
11	1	5	5	4,04 (0,11)	4,91
	2	4	5	4,52 (0,10)	5,27
	3	3	5	4,41 (0,11)	5,52
	3	4	4	4,48 (0,10)	5,58
12	2	5	5	4,51 (0,10)	5,25
	3	4	5	4,52 (0,10)	5,63
	4	4	4	4,50 (0,10)	5,65
13	3	5	5	4,55 (0,10)	5,63
	4	4	5	4,62 (0,10)	5,62
14	4	5	5	4,52 (0,10)	5,64
15	5	5	5	4,56 (0,10)	5,66

A. Die Tests von Scheffé und Tukey

Man berechnet für jedes Paar (i, i') , $i < i'$, von Meßreihen die Differenz der Reihenmittel

$$\bar{Y}_i - \bar{Y}_{i'}$$

und erkennt auf zum Niveau γ signifikante Unterschiede der Mittelwerte der Reihen i und i' , falls diese Differenz beträchtlich größer ist als eine **Grenzdifferenz**.

A1. Scheffé-Test

Entstammen die Meßreihen einer $N(\mu_i, \sigma^2)$ -verteilten Grundgesamtheit, so kann (auch bei ungleichen Reihenlängen) der **Scheffé-Test** verwandt werden. Dessen Grenzdifferenz zum Niveau γ ist

$$FSD_\gamma = \sqrt{(p-1) \cdot MSE \cdot \left(\frac{1}{n_i} + \frac{1}{n_{i'}} \right) \cdot F_{p-1, N-p; 1-\gamma}}$$

A2. Tukey-Test

Bei normalverteilten Grundgesamtheiten und gleichen Reihenlängen $n = n_i$ kann der **Tukey-Test**, der mittels der Studentisierten Spannweite $q_{p, N-p}$ die Grenzdifferenz

$$HSD_\gamma = \sqrt{\frac{MSE}{n}} \cdot q_{p, N-p; 1-\gamma}$$

zum Niveau γ benutzt, verwandt werden.

Beispiel: Beim Zugfestigkeitsversuch, vgl. Tab. 4, ergibt sich

$$|\bar{Y}_1 - \bar{Y}_2| = 2; |\bar{Y}_1 - \bar{Y}_3| = 5,9; |\bar{Y}_2 - \bar{Y}_3| = 3,9.$$

Zum 5% Niveau ist die Grenzdifferenz des Scheffé-Tests gerade

$$FSD_{0,05} = \sqrt{2 \cdot 14,44 \cdot \left(\frac{1}{6} + \frac{1}{6} \right) \cdot 3,68} = 5,952$$

und die des Tukey-Tests

$$HSD_{0,05} = \sqrt{\frac{14,44}{6}} \cdot 3,67 = 5,693,$$

d. h. der Scheffé-Test erkennt kein Meßreihenpaar mit signifikant verschiedenen Mittelwerten, wohingegen der Tukey-Test erkennt, daß sich die erste und die dritte Meßreihe signifikant zum 5% Niveau unterscheiden.

B. Der Test von Steel und Dwass

Beim **Test von Steel und Dwass**, der zwar keine Normalverteilung, wohl aber gleiche Reihenlängen $n_i = n$ erfordert (und stetige Verteilung), ordnet man den zu vergleichenden Meßreihen i und k gemeinsam Rangzahlen zu (bei Bindungen verwendet man mittlere Ränge) und schließt (approximativ) auf zum Niveau γ signifikante Unterschiede zwischen den Mittelwerten der Reihen i und k , falls das Maximum der Rangsummen der beiden Reihen größer ist als

$$\frac{n(2n+1)}{2} + \frac{1}{2} + q_{p,\infty; 1-\gamma} \cdot \sqrt{\frac{1}{24} \cdot n^2 \cdot (2n+1)}.$$

1.3. Dispersionsvergleiche

Hier sollen verschiedene Tests vorgestellt werden, die überprüfen, ob die Varianzen (Dispersionen) von p Meßreihen signifikant verschieden sind.

A. Der Bartlett-Test

Bei ungleichen Reihenlängen und normalverteilten Grundgesamtheiten läßt sich der Bartlett-Test verwenden, falls für $i = 1, \dots, p$ gilt: $n_i \geq 5$ (Faustregel).

Man berechnet dazu die Reihenmittelwerte \bar{Y}_i sowie die empirischen Varianzen

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} Y_{ij}^2 - \frac{n_i}{n_i - 1} \bar{Y}_i^2$$

der Meßreihen $i = 1, \dots, p$ und damit dann

$$\begin{aligned} s^2 &= \frac{1}{N - p} \sum_{i=1}^p (n_i - 1) s_i^2 \quad \text{sowie} \\ c &= \frac{1}{3(p - 1)} \left[\sum_{i=1}^p \frac{1}{n_i - 1} - \frac{1}{N - p} \right] + 1. \end{aligned}$$

Gilt nun mit dem vorgegebenen Niveau γ

$$\frac{1}{c} \left[(N - p) \ln s^2 - \sum_{i=1}^p (n_i - 1) \ln s_i^2 \right] > \chi_{p-1; 1-\gamma}^2,$$

so sind die Varianzen als signifikant verschieden (zum Niveau γ) zu betrachten.

B. Der Levene-Test

Der Levene-Test verzichtet auf die Normalverteilungsannahme und setzt lediglich eine stetige Verteilung der den Meßreihen zugrundeliegenden Gesamtheiten vor aus.

Man berechnet die arithmetischen Mittel \bar{Y}_i der p Meßreihen und daraus die transformierten Werte

$$X_{ij} = |Y_{ij} - \bar{Y}_i|, \text{ für } 1 \leq j \leq n_i, 1 \leq i \leq p.$$

Letzteren Werten X_{ij} ordnet man Rangzahlen r_{ij} zu, wobei dem r -kleinsten der X_{ij} der Rang $r_{ij} = r$ zugewiesen wird. Beim Vorliegen von Bindungen, d. h. mehrere der X_{ij} haben gleiche Werte, werden jeweils die mittleren der in Frage stehenden Ränge zugewiesen. Auf diese Rangzahlen wendet man den Test von Kruskal und Wallis an, wie er in Abschnitt 1.1.B beschrieben wird. Ergibt die dort definierte Prüfgröße zum gewählten Niveau γ eine Signifikanz, so darf man mit der Irrtumswahrscheinlichkeit γ schließen, daß nicht alle Varianzen gleich sind.

C. Scheffé's χ^2 -Test

Scheffé's χ^2 -Test setzt zwar keine Normalverteilung voraus, aber neben der Stetigkeit der Verteilung wird gefordert, daß das vierte Moment der Verteilungsfunktion endlich ist und daß für $i = 1, \dots, p$ gilt: $n_i \geq 10$ (Faustregel).

Man berechnet zuerst die arithmetischen Mittel \bar{Y}_i der p Meßreihen und transformiert die Meßwerte dann gemäß

$$X_{ij} = (Y_{ij} - \bar{Y}_{..})^2, \text{ für } 1 \leq j \leq n_i, 1 \leq i \leq p.$$

Hieraus berechnet man

$$z_i = \ln \left[\frac{1}{n_i - 1} \sum_{j=1}^{n_i} X_{ij} \right] \quad \text{für } i = 1, \dots, p,$$

$$z = \frac{1}{N-p} \sum_{i=1}^p (n_i - 1) z_i, \quad c = 2 + \frac{N-p}{N} \left[\frac{1}{N\bar{Y}^2} \sum_{i=1}^p \sum_{j=1}^{n_i} X_{ij}^2 - 3 \right]$$

und daraus die Prüfgröße

$$S = \frac{1}{c} \sum_{i=1}^p (n_i - 1) \cdot (z_i - z)^2.$$

Gilt dann zu dem vorgegebenen Niveau γ , daß

$$S > \chi^2_{p-1; 1-\gamma}$$

ist, so darf ein Unterschied zwischen den Varianzen als statistisch gesichert betrachtet werden.

Es handelt sich hier um approximative Verfahren, welche mit zunehmender Reihenlänge (insbesondere der kürzesten Reihe!) besser werden.

Beispiel: In den Abschnitten 1.1 und 1.2 haben wir vorausgesetzt, daß die Varianzen der Meßreihen beim Zugfestigkeitsversuch, vgl. Tab. 4, gleich sind, was wir hier nun zum 5% Niveau überprüfen wollen.

Zur Durchführung des *Bartlett-Tests* berechnen wir

$$s_1^2 = 14,736, s_2^2 = 13,672, s_3^2 = 14,908$$

und damit

$$s^2 = 14,439 \quad \text{sowie} \quad c = 1,089,$$

so daß sich die Varianzen der Meßreihen zum 5% Niveau nicht als signifikant verschieden erweisen, denn es ist

$$\begin{aligned} & \frac{1}{1,089} \{ 15 \ln 14,439 - 5(\ln 14,736 + \ln 13,672 + \ln 14,908) \} \\ & = 0,010 < 5,991 = \chi^2_{2; 0,95}. \end{aligned}$$

Um den *Levene-Test* durchführen zu können, müssen zuerst die transformierten Werte X_{ij} berechnet werden, denen dann Rangzahlen zugeordnet werden. Im Beispiel treten keine Bindungen auf, so daß $B = 1$ ist, und die Rangsummen in den Reihen der transformierten Werte sind

$$r_{1.} = 49, r_{2.} = 61 \quad \text{und} \quad r_{3.} = 61,$$

so daß auch hier wegen

$$H = \frac{12}{18 \cdot 19 \cdot 6} (49^2 + 61^2 + 61^2) - 3 \cdot 19 = 0,561 < 5,991 = \chi^2_{2; 0,95}$$

keine signifikanten Unterschiede in den Varianzen der Meßreihen entdeckt werden.

1.4. Modellbetrachtung

Bisher sind wir davon ausgegangen, daß sich ein Beobachtungswert Y_{ij} aus einem Reihenmittelwert μ_i und einem zufälligen Versuchsfehler e_{ij} in der Form

$$Y_{ij} = \mu_i + e_{ij}$$

zusammensetzt. Es ist nun auch üblich, die p Behandlungseffekte (Meßreiheneffekte) μ_i aufzuspalten in ein Gesamtmittel μ und einen differentiellen Effekt β_i der i -ten Meßreihe (Behandlung) im Vergleich zum Gesamtmittel gemäß

$$Y_{ij} = \mu_i + e_{ij} = \mu + \beta_i + e_{ij} \quad \text{mit} \quad \beta_1 + \dots + \beta_p = 0.$$

Letztere Bedingung heißt **Reparametrisierungsbedingung**.

Die Hypothese gleicher Mittelwerte

$$H_0: \mu_1 = \dots = \mu_p$$

ist dann gerade identisch mit der Hypothese

$$H_0: \beta_1 = \dots = \beta_p.$$

Die Nützlichkeit dieser Zerlegung wird allerdings erst in den nachfolgend beschriebenen komplizierteren Versuchsplänen vollauf deutlich.

2. Das einfache Blockexperiment

Der Hauptnachteil der vollständig randomisierten Versuchspläne besteht darin, daß die Variabilität innerhalb der verschiedenen Gruppen gleicher Behandlung sehr groß sein kann, wodurch die Unterschiede zwischen den Behandlungseffekten teilweise verwischt werden, so daß die einschlägigen Testverfahren keinen Unterschied zwischen den Behandlungen mehr feststellen können bzw. Gruppen- und Behandlungseffekte sich überlagern. Dieser unerwünschte Umstand kann weitgehend behoben werden, wenn eine Hauptstörgröße bekannt ist, wie zum Beispiel der Intelligenzquotient bei Untersuchungen des Lernvermögens.

Liegt die Störvariable in r Ausprägungen (**Blöcke** genannt) vor und werden die Effekte von p verschiedenen Behandlungen auf ein interessierendes Merkmal untersucht, so wendet man im **einfachen Blockexperiment** jede Behandlung genau einmal in jedem Block an, und das Versuchsergebnis im i -ten Block bei Anwendung der j -ten Behandlung wird beschrieben durch die Zufallsvariable Y_{ij} , mit $i = 1, \dots, r$ und $j = 1, \dots, p$.

Beispiel: (a) In einem psychologischen Experiment soll der Einfluß von 3 verschiedenen Rhythmen eines Metronoms auf das Sprachverhalten von Stotterern untersucht werden. Da zwischen den Personen der Grundgesamtheit der Stotterer große Unterschiede bestehen, ist die natürliche Klassengröße in der Grundgesamtheit gerade 1. Wenn die Unterschiede zwischen allen Stotterern in der Grundgesamtheit nicht interessieren, kann man ein Modell mit festen Effekten wählen; vgl. auch die Einführung zu Abschnitt 4. Es werden zufällig 10 Stotterer (Blöcke) ausgewählt, und jeder von ihnen muß je zwei Minuten bei jedem Metronomrhythmus sprechen. Haben die Rhythmen einen Einfluß auf das Sprachverhalten?

(b) Es soll untersucht werden, ob zwischen vier verschiedenen Reifensorten ein Unterschied bzgl. der Profilabnutzung besteht. Da man weiß, daß die Abnutzung sehr stark vom benutzten Auto (bzw. von dessen Fahrer) abhängig ist, liegt hier eine

natürliche Blockgröße von 4 (Reifen pro Auto) vor. Bei jedem Wagen (Block) im Experiment wird dann je ein Reifen jeder Sorte verwandt.

2.1. Verfahren bei normalverteilter Grundgesamtheit

Wir gehen davon aus, daß sich das Versuchsauskommen Y_{ij} bei der j -ten Behandlung im i -ten Block aus einem Gesamtmittel μ , dem differentiellen Blockeffekt α_i , dem differentiellen Behandlungseffekt β_j sowie einem zufälligen Fehler e_{ij} zusammensetzt:

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad \text{für } i = 1, \dots, r \quad \text{und } j = 1, \dots, p,$$

wobei die zufälligen Fehler unabhängig identisch $N(0, \sigma^2)$ -verteilt sind, und stellen die Reparametrisierungsbedingungen

$$\alpha_1 + \dots + \alpha_r = 0, \quad \beta_1 + \dots + \beta_p = 0.$$

Die Schätzer für die unbekannten Parameter μ , α_i und β_j sind

$$m = \bar{Y}_{..} \quad \text{bzw. } a_i = \bar{Y}_{i.} - \bar{Y}_{..} \quad \text{bzw. } b_j = \bar{Y}_{.j} - \bar{Y}_{..},$$

$$\text{mit } \bar{Y}_{i.} = \frac{1}{p} \sum_{j=1}^p Y_{ij}, \quad \bar{Y}_{.j} = \frac{1}{r} \sum_{i=1}^r Y_{ij} \quad \text{und} \quad \bar{Y}_{..} = \frac{1}{rp} \sum_{i=1}^r \sum_{j=1}^p Y_{ij},$$

und es ergibt sich die in Tab. 7 angegebene Varianzanalysetafel, bei der wir die sogenannte **doppelte Quadratsummenzerlegung**

$$SSG = SSA + SSB + SSE$$

ausnutzen, die besagt, daß die Gesamtvariation SSG zurückgeführt werden kann auf die „Variation zwischen den Blöcken“ SSA, die „Variation zwischen den Behandlungen“ SSB und die „Zufallsvariation“ SSE der von den Block- und Behandlungseffekten bereinigten Versuchsergebnisse. *

Außerdem sind in Tab. 7 die Prüfgrößen und die kritischen Werte für die F-Tests zum Niveau γ auf signifikante Unterschiede in den Blockeffekten α_i bzw. den Behandlungseffekten β_j angegeben. Die Block- bzw. Behandlungseffekte sind signifikant verschieden zum Niveau γ , falls der Wert der Prüfgröße den kritischen Wert übersteigt.

Multiple Vergleiche der Block- und Behandlungseffekte lassen sich analog zu den in Abschnitt 1 angegebenen durchführen, indem man die Differenzen $|a_i - a_j|$ bzw. $|b_j - b_{j'}|$ berechnet und z. B. die Behandlungen j und j' als signifikant verschieden zum Niveau γ erklärt, falls $|b_j - b_{j'}|$ größer ist als die zugehörige Grenzfrequenz. Diese ist etwa für den Tukey-Test auf Behandlungsunterschiede

$$HSD_\gamma = \sqrt{\frac{MSE}{r}} \cdot q_{p,(r-1)(p-1); 1-\gamma}.$$

Beispiel: Es soll untersucht werden, ob sich vier Zigarettenarten hinsichtlich ihres Gehaltes an einer gewissen, vermutlich Krebs erzeugenden Substanz unterscheiden, wobei für die chemische Analyse sechs Laboratorien in Frage kommen. Das Problem wird als so wichtig erachtet, daß jede Zigarettenart in jedem Labor untersucht werden soll, wobei aus Kostengründen jeweils nur eine Untersuchung ange stellt werden kann. Da sich die Laboratorien in ihrer technischen und personellen Ausstattung unterscheiden, sollen die vier Meßwerte jedes einzelnen Labors zu

Tab. 7: Varianzanalysetafel für das einfache Blockexperiment

Streuungursache	Freiheitsgrade	Quadratsumme	Mittlere Quadratsumme	Prüfgröße	Kritischer Wert
Unterschiede zwischen den Blöcken	$r - 1$	$SSA = p \sum_i a_i^2$	$MSA = \frac{SSA}{r - 1}$	$F_A = \frac{MSA}{MSE}$	$F_{r-1, (r-1)(p-1), 1-\gamma}$
Unterschiede zwischen den Behandlungen	$p - 1$	$SSB = r \sum_j b_j^2$	$MSB = \frac{SSB}{p - 1}$	$F_B = \frac{MSB}{MSE}$	$F_{p-1, (r-1)(p-1), 1-\gamma}$
Fehler	$(r-1)(p-1)$	$SSE = SSG - SSA - SSB$	$MSE = \frac{SSE}{(r-1)(p-1)}$		
Gesamt	$rp - 1$	$SSG = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$			

einem Block zusammengefaßt werden, so daß eine einfache Blockanalyse mit $p = 4$ und $r = 6$ adäquat ist.

In Tab. 8 sind die Versuchsergebnisse sowie alle Schätzer m , a_i und b_j angegeben. Wir erhalten somit die Varianzanalysetafel aus Tab. 9, die auch die Prüfgrößen und die kritischen Werte zum 5% Niveau enthält. Wir sehen, daß nicht zwischen den Laboratorien, wohl aber zwischen den Zigarettenarten signifikante Unterschiede zum 5% Niveau bestehen.

Tab. 8: Versuchsergebnisse, Effektschätzer und Rangzahlen zum Laborversuch

Laboratorium i	Zigartensorte j								Schätzer a_i	
	1		2		3		4			
	Y_{i1}	r_{i1}	Y_{i2}	r_{i2}	Y_{i3}	r_{i3}	Y_{i4}	r_{i4}		
1	3	2	6	4	5	3	1	1	-4,75	
2	7	2	21	4	12	3	5	1	2,75	
3	5	2	17	4	8	3	3	1	-0,25	
4	24	4	14	3	4	2	1	1	2,25	
5	6	1	19	4	13	3	7	2	2,75	
6	3	2	13	4	6	3	1	1	-2,75	
Rangsumme R_j	13		23		17		7			
Schätzer b_j	-0,5		6,5		-0,5		-5,5		$m = 8,5$	

Tab. 9: Varianzanalysetafel zum Laborversuch

Streuungs-ursache	Freiheits-grade	Quadrat-summe	Mittlere Quadrat-summe	Prüfgröße	Kritischer Wert
Laboratorien	5	201,5	40,3	1,665	2,90
Zigarettenarten	3	438,0	146,0	6,033	3,29
Fehler	15	362,5	24,2		
Gesamt	23	1002,0			

2.2. Verteilungsfreie Verfahren

Wir gehen vom gleichen Modell aus wie in Abschnitt 2.1, fordern jedoch nur noch, daß die Fehler e_{ij} unabhängig sind und derselben stetigen Verteilung genügen. Dann läßt sich der Friedman-Test auf signifikante Unterschiede in den p Behandlungen durchführen. Innerhalb jedes Blockes werden den Meßwerten Y_{ij} dazu Ränge r_{ij} zugeordnet: Der kleinste Wert in jedem Block erhält den Rang 1, ..., der größte den Rang p ; treten Bindungen auf, so werden mittlere Ränge zugeordnet. Bezeichnet k_i die Anzahl unterschiedlicher Werte (Bindungsgruppen) im i -ten Block und t_{il} die Anzahl der Meßwerte im i -ten Block, die zur l -ten Bindungsgruppe gehören ($l = 1, \dots, k_i$), so gilt stets:

$$\sum_{l=1}^{k_i} t_{il} = p \quad \text{für } i = 1, \dots, r.$$

Bezeichnet man weiter mit

$$R_j = \sum_{i=1}^r r_{ij}$$

die resultierende Rangsumme für die j-te Behandlung ($j = 1, \dots, p$), so berechnet sich die Prüfgröße des Friedman-Tests als

$$V = \frac{12 \cdot (p-1) \cdot \sum_{j=1}^p \left[R_j - r \cdot \frac{p+1}{2} \right]^2}{r \cdot p^3 \cdot \sum_{i=1}^r \sum_{l=1}^{k_i} t_{il}^3}.$$

Treten in keinem Block Bindungen auf, d. h. ist der Umfang jeder Bindungsgruppe 1, so vereinfacht sich diese Prüfgröße zu

$$V = \frac{12}{r \cdot p \cdot (p+1)} \cdot \sum_{j=1}^p R_j^2 - 3 \cdot r \cdot (p+1).$$

Gilt dann zu dem vorgegebenen Niveau γ

$$V \geq v_{p,r;1-\gamma},$$

so sind die Unterschiede zwischen den Behandlungen als signifikant zu betrachten. Einige kritische Werte $v_{p,r;1-\gamma}$ sind in Tab. 10 zusammengestellt; einige weitere findet man etwa bei Hollander/Wolfe (1973).

Tab. 10: Kritische Werte $v_{p,r;1-\gamma}$ des Friedman-Tests auf signifikante Behandlungsunterschiede für $\gamma = 0,05$

$p \backslash r$	2	3	4	5	6	7	8
3	—	4,67	6,00	5,20	6,33	6,00	5,25
4	5,40	7,00	7,50	7,32	7,40	7,63	7,50
5	—	8,27	8,60	8,80	—	—	—

Stehen keine kritischen Werte zur Verfügung, so kann man approximativ vorgehen und auf signifikante Unterschiede schließen, falls gilt

$$V \geq \chi^2_{p-1;1-\gamma}.$$

Multiple Vergleiche zwischen den Behandlungen können mittels des **Friedman-Rangsummentests** durchgeführt werden: Die j-te und j'-te Behandlung sind signifikant verschieden zum Niveau γ , falls

$$|R_j - R_{j'}| \geq r'_{p,r;1-\gamma} \simeq q_{p,\infty;1-\gamma} \cdot \sqrt{\frac{r \cdot p \cdot (p+1)}{12}}$$

ist. Kritische Werte $r'_{p,r;1-\gamma}$ findet man zum Beispiel bei Hollander/Wolfe (1973).

Beispiel: Wir kommen auf das Beispiel (Laborversuch) aus Abschnitt 2.1 zurück; in der dortigen Tab. 8 sind bereits die Rangzahlen r_{ij} und Rangsummen R_j angegeben. Mit

$$V = \frac{12}{6 \cdot 4 \cdot 5} \cdot (13^2 + 23^2 + 17^2 + 7^2) - 3 \cdot 6 \cdot 5 \\ = 13,6 > 7,40 = v_{4,6;0,95}$$

erweisen sich die Zigarettenarten als signifikant verschieden zum 5% Niveau. Speziell sind mit

$$r'_{4,6;0,95} \approx q_{4,\infty;0,95} \cdot \sqrt{\frac{6 \cdot 4 \cdot 5}{12}} = 3,63 \cdot \sqrt{10} = 11,48$$

die Sorten 2 und 4 signifikant verschieden zum 5% Niveau.

3. Zweifache Varianzanalyse mit mehreren Beobachtungen pro Faktorstufenkombination (pro Zelle)

Falls möglich sollte man jede Behandlung mehrmals in jedem Block anwenden, vgl. auch die graphische Darstellung eines solchen Experiments in der Abb. 19 von Kap. V (Abschnitt 5, S. 356); man spricht dann von **Modellen der zweifachen Varianzanalyse**.

In diesem Abschnitt wird nun zunächst das Modell der zweifachen Varianzanalyse mit Wechselwirkungen betrachtet. Von einem solchen Modell sollte vor allem dann ausgegangen werden, wenn zwei Einflußfaktoren A und B auf r bzw. p Stufen untersucht werden, von denen keiner einen echten Blockfaktor darstellt und deren Stufen beliebig kombinierbar sind (**zweifache Kreuzklassifikation**). Später wird dann noch ein Modell mit mehreren Beobachtungen pro Zelle ohne Wechselwirkungen behandelt. Dieses kommt zur Anwendung, wenn von vornherein keine Wechselwirkungen (vgl. Kap. V, Abschnitt 5, insbesondere die dortigen Abb. 20 und 21) zwischen den Faktoren A und B zu erwarten sind, was vor allem dann der Fall sein wird, wenn einer der Einflußfaktoren als echter **Blockfaktor** nur einen **Wiederholungseffekt** bewirkt.

Beispiel: (a) Der Einfluß der Faktoren Düngung auf 3 Stufen und chemische Behandlung auf 4 Stufen auf den Ertrag einer Weinsorte soll untersucht werden. Dazu wird ein Feld, auf dem überall etwa gleiche Bodengüte zu erwarten ist, in 36 Parzellen aufgeteilt, und je 3 Parzellen wird zufällig eine der 12 möglichen Stufenkombinationen der Faktoren „Düngung“ und „chemische Behandlung“ zugeordnet. Hier wäre es wohl angebracht, ein Modell mit Wechselwirkungen zu verwenden, da Interaktionen zwischen den beiden Faktoren nicht auszuschließen sind.

(b) Um den Einfluß des Geschlechts und des Alters auf die mittlere Oberfläche von Erythrozyten zu untersuchen, wird die Gesamtheit der interessierenden Versuchspersonen in 10 Klassen eingeteilt; zum einen werden zwei Geschlechter, zum anderen 5 Altersgruppen unterschieden. Aus jeder Klasse werden dann 3 Personen zufällig ausgewählt und die mittlere Oberfläche ihrer Erythrozyten bestimmt.

(c) Der Einfluß der Anzahl der Umdrehungen eines Bohrers und des Drucks auf den Bohrer auf die Bohrgeschwindigkeit (Meter pro Stunde) beim Bohren eines Schachtes bei einem bestimmten Untergrund soll untersucht werden. Bei 4 verschiedenen Drehzahlen und 5 Druckstärken werden 40 Bohrungen gemacht, je 2 bei jeder Kombination von Drehzahl und Druck.

(d) Ein Statistiker will den Einfluß verschiedener Lehrmethoden und verschiedenen Alters der Schüler auf die Erfolge im Mathematikunterricht untersuchen. Es

stehen 3 Lehrmethoden und Schüler dreier Altersgruppen für den Versuch zur Verfügung. Aus jeder Altersgruppe werden 12 Schüler zufällig ausgewählt und je 4 einer Lehrmethode zugeordnet. Gemessen werden dann die Punktzahlen in einem Test am Ende des Versuches.

3.1. Modell mit Wechselwirkungen zwischen den Faktoren A und B

Wir betrachten das Modell

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

für $i = 1, \dots, r, j = 1, \dots, p$ und $k = 1, \dots, n$

mit den Reparametrisierungsbedingungen

$$\begin{aligned} \alpha_1 + \dots + \alpha_r &= 0, & \beta_1 + \dots + \beta_p &= 0, \\ (\alpha\beta)_{i1} + \dots + (\alpha\beta)_{ip} &= 0 \quad \text{für } i = 1, \dots, r \\ \text{und} \quad (\alpha\beta)_{1j} + \dots + (\alpha\beta)_{rj} &= 0 \quad \text{für } j = 1, \dots, p. \end{aligned}$$

Dabei bezeichnet Y_{ijk} das Versuchsauskommen bei der k -ten Wiederholung auf Stufe i des Faktors A und Stufe j des Faktors B, μ das Gesamtmittel, α_i den Effekt der i -ten Stufe des Faktors A, β_j den Effekt der j -ten Stufe des Faktors B, $(\alpha\beta)_{ij}$ den **Wechselwirkungseffekt** von i -ter Stufe des Faktors A und j -ter Stufe des Faktors B und e_{ijk} den Versuchsfehler. Weiterhin gehen wir davon aus, daß die Fehler e_{ijk} unabhängig identisch $N(0, \sigma^2)$ -verteilt sind und erhalten als **Schätzer** für das Gesamtmittel μ und die differentiellen Effekte $\alpha_i, \beta_j, (\alpha\beta)_{ij}$ die Größen

$$\begin{aligned} m &= \bar{Y}_{...}, \quad a_i = \bar{Y}_{i..} - m, \quad b_j = \bar{Y}_{..j} - m, \quad \text{bzw. } (ab)_{ij} = \bar{Y}_{ij} - a_i - b_j - m, \\ \text{mit } \bar{Y}_{...} &= \frac{1}{rpn} \sum_{i=1}^r \sum_{j=1}^p \sum_{k=1}^n Y_{ijk}, \quad \bar{Y}_{i..} = \frac{1}{pn} \sum_{j=1}^p \sum_{k=1}^n Y_{ijk}, \quad \bar{Y}_{..j} = \frac{1}{rn} \sum_{i=1}^r \sum_{k=1}^n Y_{ijk} \text{ und} \\ \bar{Y}_{ij.} &= \frac{1}{n} \sum_{k=1}^n Y_{ijk}, \end{aligned}$$

woraus sich mit der Quadratsummenzerlegung

$$SSG = SSA + SSB + SS(AB) + SSE$$

die **Varianzanalysetafel** aus Tab. 11 ergibt; dort sind auch die Prüfgrößen und die kritischen Werte für die F-Tests auf Vorliegen zum Niveau γ signifikanter Wirkungen der Faktoren A bzw. B und auf Vorliegen zum Niveau γ signifikanter Wechselwirkungen angegeben. Auf Signifikanz wird geschlossen, wenn die jeweilige Prüfgröße den zugehörigen kritischen Wert übersteigt. Die Effekte β_j der Stufen des Faktors B sind z. B. signifikant verschieden zum Niveau γ , falls der Wert der Prüfgröße F_B größer ist als der zugehörige kritische Wert. **Multiple Vergleiche** können analog zu den in Abschnitt 1 angegebenen unter Verwendung der betraglichen Differenzen der Effektschätzer durchgeführt werden.

Beispiel: In einem Experiment werden die Wirkungen dreier verschiedener Antibiotika (B-Behandlungen, Faktor B) bei Patienten mit Lungeninfektionen untersucht, wobei jeweils einer von drei Erregern (A-Behandlungen, Faktor A) vorliegt. Jedes der drei Antibiotika wird bei jedem Erreger bei zwei Patienten angewandt, so daß ein Blockexperiment mit mehrfacher Besetzung ($n = 2$) entsteht. Da Wechselwirkungen zwischen Antibiotika und Erregern nicht auszuschließen sind, wird der

Tab. 11: Varianzanalysetafel der zweifachen Varianzanalyse mit Wechselwirkungen bei n Beobachtungen pro Zelle

Streuungursache	Freiheitsgrade	Quadratsumme	Mittlere Quadratsumme	Prüfgröße	Kritischer Wert
Faktor A	r - 1	$SSA = pn \sum_i a_i^2$	$MSA = \frac{SSA}{r-1}$	$F_A = \frac{MSA}{MSE}$	$F_{r-1, pr(n-1); 1-\gamma}$
Faktor B	p - 1	$SSB = mn \sum_j b_j^2$	$MSB = \frac{SSB}{p-1}$	$F_B = \frac{MSB}{MSE}$	$F_{p-1, pr(n-1); 1-\gamma}$
Wechselwirkung	(p - 1)(r - 1)	$SS(AB) = n \sum_i \sum_j (ab)_{ij}^2$	$MS(AB) = \frac{SS(AB)}{(p-1)(r-1)}$	$F_{AB} = \frac{MS(AB)}{MSE}$	$F_{(p-1)(r-1), pr(n-1); 1-\gamma}$
Fehler	pr(n - 1)	$SSE = SSG - SSA - SSB$ - $SS(AB)$	$MSE = \frac{SSE}{pr(n-1)}$		
Gesamt	rpn - 1	$SSG = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2$			

Versuch mittels zweifacher Varianzanalyse mit Wechselwirkungen und mehrfacher Zellenbesetzung ausgewertet. Alle Aussagen sollen zum 5% Niveau gesichert werden.

Die Versuchsergebnisse sind in **Tab. 12** zusammengestellt, wo auch alle Effektschätzer bereits angegeben sind. Hieraus ergibt sich dann die Varianzanalysetafel aus **Tab. 13**, die auch Prüfgrößen und kritische Werte für die F-Tests zum 5% Niveau enthält.

Tab. 12: Versuchsergebnisse und Effektschätzer zum Beispiel

Erreger i	Antibiotikum j						Schätzer a_i	
	k	1		2		3		
		Y_{i1k}	$(ab)_{i1}$	Y_{i2k}	$(ab)_{i2}$	Y_{i3k}	$(ab)_{i3}$	
1	1	38	-4,000	40	1,667	38	2,333	0,667
	2	35		41		39		
2	1	42	3,833	39	-2,000	33	-1,833	-0,167
	2	45		33		34		
3	1	38	0,167	38	0,333	33	-0,500	-0,500
	2	41		38		36		
Schätzer b_j		2,000		0,333		-2,333		$m = 37,833$

Tab. 13: Varianzanalysetafel zum Beispiel

Streuungs-ursache	Freiheits-grade	Quadrat-summe	Mittlere Quadrat-summe	Prüfgröße	Kritischer Wert
Erreger	2	4,33	2,165	0,52	4,26
Antibiotika	2	57,33	28,665	6,88	4,26
Interaktion	4	93,33	23,333	5,60	3,63
Fehler	9	37,51	4,167		
Gesamt	17	192,50			

Zwischen den Effekten der Erreger bestehen also keine signifikanten Unterschiede, wohl aber zwischen den Antibiotika, und auch die Wechselwirkungen zwischen Antibiotika und Erregern sind signifikant (da $5,60 > 3,63$ ist), was natürlich impliziert, daß auch der Faktor „Erreger“ eine Wirkung hat.

3.2. Modell ohne Wechselwirkungen zwischen den Faktoren

Sind keine Wechselwirkungen zwischen den Faktoren A und B zu erwarten, so betrachtet man nur noch das Modell

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \quad \text{für } i = 1, \dots, r, j = 1, \dots, p, k = 1, \dots, n.$$

Die **Effektschätzer** für μ , α_i und β_j sind hier die gleichen wie im Abschnitt 3.1, und

auch die **Quadratsummen** SSA, SSB und SSG bleiben die gleichen. Lediglich gilt jetzt

$$SSE = SSG - SSA - SSB,$$

und es ergibt sich die **Varianzanalysetafel** aus **Tab. 14**, in der auch Prüfgrößen und kritische Werte für F-Tests zum Niveau γ angegeben sind.

Tab. 14: Varianzanalysetafel des Modells ohne Wechselwirkungen

Streuungs-ursache	Freiheits-grade	Quadrat summe	Mittlere Quadratsumme	Prüf-größe	Kritischer Wert
Faktor A	$r - 1$	SSA	$MSA = \frac{SSA}{r - 1}$	$F_A = \frac{MSA}{MSE}$	$F_{r-1, v, 1-\gamma}$
Faktor B	$p - 1$	SSB	$MSB = \frac{SSB}{p - 1}$	$F_B = \frac{MSB}{MSE}$	$F_{p-1, v, 1-\gamma}$
Fehler	$v = rpn - r - p + 1$	SSE	$MSE = \frac{SSE}{v}$		
Gesamt	$rpn - 1$	SSG			

Sind die Fehler e_{ijk} nicht normalverteilt, so kann man, falls sie als unabhängig mit stetiger und für alle Fehler identischer Verteilung angenommen werden können, eine Modifikation des in Abschnitt 2 vorgestellten **Friedman-Tests** auf Unterschiede in den Behandlungseffekten durchführen. Dazu werden den Beobachtungen des i-ten Blocks (der i-ten Stufe des Faktors A) Rangzahlen r_{ijk} von 1 bis np zugeordnet; der kleinste Meßwert Y_{ijk} erhält den Rang $r_{ijk} = 1, \dots$, der größte den Rang np. Bezeichnet für $j = 1, \dots, p$

$$R_j = r_{1j1} + \dots + r_{1jn} + r_{2j1} + \dots + r_{rjn}$$

die Summe der Ränge auf der j-ten Stufe des Faktors B, so wird auf zum Niveau γ signifikante Unterschiede zwischen den Stufen von B geschlossen, falls gilt

$$\nabla = \frac{12}{r \cdot p \cdot n^2 \cdot (np + 1)} \cdot \sum_{j=1}^p R_j^2 - 3 \cdot r \cdot (np + 1) > \chi_{p-1, 1-\gamma}^2.$$

Viele weitere Modelle der Varianzanalyse wie zum Beispiel **lateinische Quadrate**, **griechisch-lateinische Quadrate**, faktorielle Versuchspläne, Split-Plot-Pläne sowie auch weitere Verfahren der multiplen Vergleiche (z. B. nach **Newman-Keuls**, **Duncan**, **Bonferroni-Fisher**, **Bonferroni-Holm**) können hier nicht vorgestellt werden. Werden die Ausprägungen von Einflußfaktoren selbst erst im Experiment gemessen, so kommt man zu Methoden der **Kovarianzanalyse** oder auch zu allgemeineren Methoden, die insbesondere auch dann verwandt werden müssen, wenn die speziellen Modellannahmen dieses Kapitels nicht erfüllt sind, wie z. B. die Balanciertheitsannahmen, die durch Versuchsausfälle oder fehlende Daten bei Erhebungen verletzt werden können. Es sei auf die bereits am Ende des Abschnitts 5 in Kap. V zitierte Literatur sowie auf Scheffé (1959), Miller (1966), Graybill (1976), Holm (1979), Hartung/Werner (1980, 1983, 1984) und Hartung/Elpelt (1984) verwiesen.

Werden mehrere Merkmale bei verschiedenen Einflußfaktoren gemessen und gemeinsam analysiert, so geschieht dies mit Methoden der **Multivariaten Analysis**, vgl. z. B. Morrison (1967), Hartung/Elpelt (1984).

4. Die Komponenten der Streuung – Modelle der Varianzanalyse mit zufälligen Effekten (Modell II)

Liegen die Einflußfaktoren nur in endlich vielen Ausprägungen (Stufen des Faktors) vor und hat man all diese erfaßt bzw. bezieht sich eine gewünschte Aussage nur auf die im Experiment vorkommenden Stufen der Faktoren, so sind die sogenannten **varianzanalytischen Modelle I** (Klassifizierung nach Eisenhart (1947)) anwendbar, von denen einige in den Abschnitten 1 bis 3 beschrieben wurden. Kann jedoch nur eine Auswahl aller möglichen Faktorstufen eines Einflußfaktors im Experiment oder in der Erhebung berücksichtigt werden und möchte man trotzdem Aussagen über die gesamten Faktorstufen in der Grundgesamtheit machen, so gelangt man zu den sogenannten **Modellen II der Varianzanalyse**, die hier kurz eingeführt werden.

Modelle I heißen auch **Modelle mit festen Effekten**, da man nur an den bestimmten fest vorgegebenen Stufen der Faktoren interessiert ist; Modelle II heißen auch **Modelle mit zufälligen Effekten**, da aus allen Stufen der Einflußfaktoren nur einige Stufen zufällig für das Experiment ausgewählt werden. Treten feste und zufällige Effekte gemeinsam auf, so spricht man von **gemischten Modellen (Modellen III)**, die aus Platzgründen hier nicht behandelt werden können.

Beispiel: Will man die Unterschiede in der Leistungsfähigkeit verschiedener Maschinen untersuchen, so wird man den Arbeiter, der die Maschine bedient, als einen Einflußfaktor in das Experiment miteinbeziehen müssen. Da aber nicht jeder einzelne Arbeiter im Experiment berücksichtigt werden kann, müssen einige Arbeiter aus der Grundgesamtheit zufällig ausgewählt werden. Erfolgt die Auswahl der verschiedenen Arbeiter nur, um den Störfaktor „Arbeiter“ auszuschalten und will man keine Schlüsse über die Unterschiede zwischen den Arbeitern der Gesamtheit ziehen, so kommt als Versuchstyp ein Modell I mit festen Effekten und als Versuchsplan ein einfacher Blockplan in Frage.

Möchte man aber Unterschiede zwischen den Arbeitern in der Grundgesamtheit erkennen, so muß als Versuchstyp ein Modell III mit gemischten Effekten gewählt werden: Der Effekt der Arbeiter ist zufällig, der Effekt der Maschinen ist fest. Sind zudem die im Versuch berücksichtigten Maschinen nur zufällig aus einer Grundgesamtheit ausgewählt (z. B. zufällig ausgewählte Autos aus einer Serie), so muß, wenn Aussagen über die Grundgesamtheit aller Maschinen gemacht werden sollen, als Versuchstyp ein Modell II mit zufälligen Effekten gewählt werden.

Die Analyse von Modellen II heißt auch **eigentliche Analyse der Varianz**, da hier Anteile der Gesamtvarianz der Versuchsauskommen geschätzt werden, die auf einzelne Einflußfaktoren zurückzuführen sind (**Varianzkomponenten**). Die relative Größe der auf einen Einflußfaktor zurückzuführenden Varianz beschreibt dann die Stärke des Einflusses dieses Faktors auf das beobachtete Merkmal: große Varianz – großer Einfluß, kleine Varianz – kleiner Einfluß.

Begründer der Schätzung von Varianzkomponenten sind Friedrich Robert Helmert (1876), der ein Verfahren zur Schätzung von Varianzkomponenten in Zusammenhang mit der Vermessung der Gotthard-Tunnel-Achse angibt, und unabhängig davon Ronald Aylmer Fisher, der allgemeine Modelle zur Schätzung von Varianzkomponenten in den zwanziger Jahren entwickelte.

4.1. Die einfache hierarchische Klassifikation

Will man etwa die Variation des Metallgehaltes in einem Erzlager untersuchen, so kann man aus p Bohrlöchern (**Klassen des Faktors A**) jeweils q Gesteinsproben (**Versuchseinheiten**) entnehmen und ihren Metallgehalt bestimmen. Die Varianz σ^2 des Metallgehalts zerfällt dann in die Variation σ_A^2 zwischen den Bohrlöchern, von denen p Stück zufällig ausgewählt wurden, und die Variation σ_e^2 innerhalb der Bohrlöcher (Versuchsfehler). Diese Variationen kann man dann durch Analyse in einem **Modell der einfach hierarchischen Klassifikation** schätzen:

$$Y_{ij} = \mu + A_i + e_{ij} \quad \text{für } i = 1, \dots, p \quad \text{und } j = 1, \dots, q,$$

wobei Y_{ij} das Versuchsauskommen bei der j -ten Wiederholungsmessung auf der i -ten Stufe des Faktors A, μ das Gesamtmittel, A_i die zufällige Abweichung der i -ten Stufe des Faktors A vom Gesamtmittel ($E A_i = 0$) und e_{ij} gerade den zufälligen Versuchsfehler ($E e_{ij} = 0$) bezeichnet und die A_i und e_{ij} als voneinander und untereinander unabhängig angenommen sind. Es gilt dann

$$E Y_{ij} = \mu \quad \text{für alle } i, j,$$

$$\text{Var } Y_{ij} = \text{Var } A_i + \text{Var } e_{ij} = \sigma_A^2 + \sigma_e^2 \quad \text{für alle } i, j$$

und

$$\text{Cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma_A^2 + \sigma_e^2, & \text{falls } i = i' \text{ und } j = j' \\ \sigma_A^2, & \text{falls } i = i' \text{ und } j \neq j' \\ 0, & \text{sonst} \end{cases}$$

die Versuchsauskommen sind also nur dann korreliert, wenn sie innerhalb einer Stufe des Faktors A liegen.

Stellt man nun eine Varianzanalysetafel wie in **Tab. 15** auf, vgl. auch Abschnitt 1, so lassen sich leicht Schätzer für die Variationsanteile σ_A^2 und σ_e^2 angeben. Die üblichen erwartungstreuen ANOVA-Schätzer für σ_A^2 und σ_e^2 lassen sich direkt aus den beiden letzten Spalten der Varianzanalysetafel bestimmen zu

$$s_e^2 = \text{MSE} \quad \text{und} \quad s_A^2 = \frac{1}{q} (\text{MSA} - \text{MSE}).$$

Diese klassischen Schätzer sind zwar erwartungstreue, jedoch hat s_A^2 den großen Nachteil, auch negative Werte annehmen zu können, obwohl Varianzen stets positi-

Tab. 15: Varianzanalysetafel der einfache hierarchischen Klassifikation

Variations-ursache	Frei-hefts-grade	Quadratsumme	Mittlere Quadratsumme (MS)	$E(\text{MS})$
zwischen den Klassen	$p - 1$	$\text{SSA} = q \sum_i (\bar{Y}_i - \bar{Y}_{..})^2$	$\text{MSA} = \frac{\text{SSA}}{p - 1}$	$\sigma_e^2 + q\sigma_A^2$
innerhalb der Klassen	$p(q - 1)$	$\text{SSE} = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2$	$\text{MSE} = \frac{\text{SSE}}{p(q - 1)}$	σ_e^2
Gesamt	$pq - 1$	$\text{SSG} = \text{SSA} + \text{SSE}$		

tiv sind. Daher wollen wir hier noch den entsprechenden **positiven Schätzer** nach Hartung (1981) angeben, dem ein anderes Gütekriterium zugrundeliegt:

$$S_A^2 = \frac{q}{1+q^2} MSA.$$

Der Intraklass-Korrelationskoeffizient

$$\rho = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2}$$

ist ein Maß für die Abhängigkeit zweier Beobachtungen Y_{ij} und $Y_{ij'}$ innerhalb der selben Klasse (Stufe) und liegt theoretisch zwischen 0 und 1. **Schätzer für ρ** sind

$$r_1 = \frac{s_A^2}{s_A^2 + s_e^2}, \quad r_2 = \frac{S_A^2}{S_A^2 + s_e^2} \quad \text{oder} \quad r_3 = \frac{S_A^2}{S_A^2 + s_e^2},$$

wobei r_1 auch negative Werte annehmen kann. In der Genetik wird ρ auch **Heritabilitätskoeffizient** (h_r) genannt und ist ein Maß für die Erblichkeit eines beobachteten Merkmals.

Sind die zufälligen Fehler e_{ij} und die Effekte A_i normalverteilt, so können auch Konfidenzintervalle für die Varianzkomponenten bestimmt und Hypothesen über sie getestet werden; vgl. etwa Graybill (1976), Hartung/Voet (1985).

Beispiel: (a) Um die Variation der Milchfettmengenleistung von Bullen zu untersuchen, werden zufällig $p = 2$ Bullen ausgewählt und je $q = 3$ ihrer Töchter untersucht. Die Daten zu diesem illustrativen Versuch, vgl. Searle (1971, S. 406 f.), sind in **Tab. 16** angegeben; der Versuch ist in **Abb. 1** veranschaulicht.

Tab. 16: Daten des Bullenversuches

Bulle i	1	2
Y_{ij}	19	25
(Versuchseinheiten: Töchter)	17	5
15	15	
$\sum_j Y_{ij}$	51	45

Hieraus ergibt sich die Varianzanalysetafel aus **Tab. 17**. Der Schätzer für die Variation σ_e^2 zwischen den Töchtern der Bullen (Versuchsfehler) ist also

$$s_e^2 = MSE = 52.$$

Für die Variation σ_A^2 zwischen den Bullen ergibt sich der (hier negative) ANOVA-Schätzer

$$s_A^2 = \frac{1}{3} \cdot (6 - 52) = -15,333,$$

wohingegen der angegebene positive Schätzer hier den Wert

$$S_A^2 = \frac{3}{10} \cdot 6 = 1,800$$

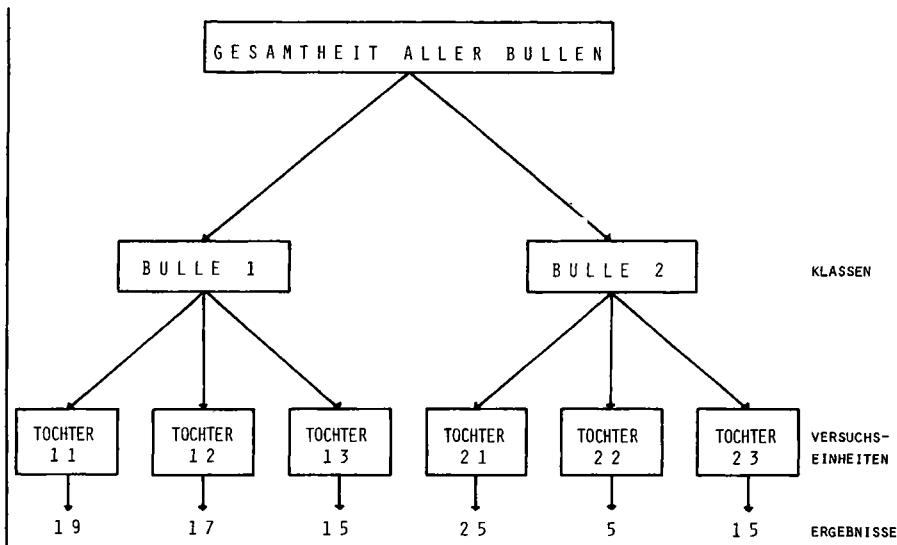


Abb. 1: Graphische Darstellung des Bullenversuchs

Tab. 17: Varianzanalysetafel des Bullenversuchs

Streuung	Freiheits-grade	Quadrat-summe	MS	E(MS)
zwischen den Bullen	1	6	6	$\sigma_e^2 + 3\sigma_A^2$
innerhalb der Bullen (d. h. zwischen den Töchtern der Bullen)	4	208	52	σ_e^2
Gesamt	5	214		

liefert. Ein Schätzer für die Abhängigkeit zwischen den Milchfettmengenleistungen zweier Töchter eines Bullen ist zum Beispiel

$$r_3 = \frac{1,800}{1,800 + 52} = 0,033.$$

(b) Um die Variation des Metallgehaltes in einem Erzlager zu untersuchen, werden aus $p = 6$ Bohrlöchern jeweils $q = 5$ Gesteinsproben entnommen und auf ihren Metallgehalt hin analysiert. In Tab. 18 sind die Analyseergebnisse Y_{ij} zusammengestellt.

Da der mittlere Metallgehalt aller 30 Proben gerade

$$\bar{Y}_{..} = 30,78$$

ist, ergibt sich die Gesamtsumme der quadratischen Abweichungen vom Mittelwert zu

$$SSG = \sum_{i=1}^p \sum_{j=1}^q (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^6 \sum_{j=1}^5 (Y_{ij} - 30,78)^2 = 411,488.$$

Tab. 18: Metallgehalte (in %) in je 5 Proben aus 6 Bohrlöchern in einem Erzlager

Bohrloch i \ Probe j	1	2	3	4	5	6
1	33,5	26,5	31,7	29,4	24,1	32,4
2	29,4	31,3	35,4	36,7	26,7	38,6
3	32,1	29,7	34,8	28,6	28,3	27,8
4	34,2	28,4	33,2	30,0	23,2	30,3
5	28,7	30,6	36,1	35,2	25,6	30,9
\bar{Y}_i	31,58	29,30	34,24	31,98	25,58	32,00

Diese Gesamtsumme lässt sich aufspalten in die Variation zwischen den Bohrlöchern

$$SSA = q \sum_{i=1}^p (\bar{Y}_i - \bar{Y}_{..})^2 = 5 \sum_{i=1}^6 (\bar{Y}_i - 30,78)^2 = 5 \cdot 44,7704 = 223,852$$

und die Restvariation innerhalb der Bohrlöcher

$$SSE = SSG - SSA = \sum_{i=1}^6 \sum_{j=1}^5 (Y_{ij} - \bar{Y}_{i.})^2 = 187,636.$$

Die Varianzanalysetafel des Versuchs ergibt sich somit, wie in **Tab. 19** dargestellt.

Tab. 19: Varianzanalysetafel des Erzlagerversuchs

Streuung	Freiheitsgrade	Quadratsumme	MS	E(MS)
zwischen den Bohrlöchern	5	223,852	44,770	$\sigma_e^2 + 5\sigma_A^2$
innerhalb der Bohrlöcher (zwischen den Proben der Bohrlöcher)	24	187,636	7,818	σ_e^2
Gesamt	29	411,488		

Aus der Tab. 19 ergibt sich nun als Schätzer für die Varianz σ_e^2 innerhalb der Bohrlöcher

$$s_e^2 = MSE = 7,818.$$

Der ANOVA-Schätzer für die Streuung σ_A^2 zwischen den Bohrlöchern ist

$$s_A^2 = \frac{MSA - MSE}{q} = \frac{44,770 - 7,818}{5} = 7,390;$$

dagegen liefert der angegebene stets positive Schätzer hier den Wert

$$S_A^2 = \frac{q}{1+q^2} MSA = \frac{5}{26} 44,770 = 8,610.$$

Die drei möglichen Schätzwerte für das Verhältnis der Variation zwischen den Bohrlöchern zur Gesamtvariation des Metallgehalts im Erzlager sind

$$r_1 = \frac{7,390}{7,390 + 7,818} = 0,486, \quad r_2 = \frac{8,610}{7,390 + 7,818} = 0,566$$

und $r_3 = \frac{8,610}{8,610 + 7,818} = 0,524;$

alle drei Schätzer nehmen hier also Werte zwischen 0 und 1 an.

Eine ausführlichere Darstellung der Analyse von Modellen mit zufälligen Effekten findet man etwa bei Scheffé (1959), Searle (1971) oder Graybill (1976). Weitere Varianzanalysemmodelle mit zufälligen Effekten unter Berücksichtigung der positiven Schätzer nach Hartung (1981) werden in Elpelt/Hartung (1982), Hartung/Elpelt (1984) und in Hartung/Voet (1985) betrachtet.

4.2. Ein nicht-klassisches Varianzanalysemmodell aus der Geodäsie

Eine wesentliche Aufgabe der Geodäsie ist die Schätzung von Gewichtsverhältnissen bei der Ausgleichung verschiedenartiger Beobachtungen. Ein Beispiel aus diesem Problemkreis ist die Ausgleichung von Winkel- und Streckenmessungen. Wohl kaum weniger wichtig ist aber die Schätzung der Variation von Messungen. Bei elektronischen Messungen etwa setzt sich die Varianz der Meßwerte aus einem konstanten Anteil und einem Anteil, der abhängt von dem, was gemessen wird, zusammen: Der konstante Anteil ist abhängig von der Präzision des Meßgerätes, die Restvarianz ist z. B. bei Streckenmessungen abhängig von der Länge der Strecke.

Ein Verfahren zur Schätzung solcher Varianzkomponenten bei geodätischen Messungen wurde, basierend auf heuristischen Überlegungen, von Helmert (1876) entwickelt. Er interessierte sich damals für die Beobachtungsfehler bei der Vermessung für die Gotthard-Tunnel-Achse. Dieses Verfahren wurde später verbessert; vgl. Helmert (1907). Das strengere Verfahren von Helmert führt zu statistisch optimalen Schätzern, wie sie erst von Rao (1970) allgemein vorgeschlagen wurden. Die Geodäsie war also aufgrund heuristischer Überlegungen der statistischen Theorie der Varianzkomponentenschätzung weit voraus. Dieses Schätzverfahren wird etwa in Grafarend/Schaffrin (1979) beschrieben.

Für einen speziellen Fall wollen wir hier die minimalverzerrten Varianzkomponentenschätzer nach Helmert sowie einen weiteren Schätzer mit dieser Eigenschaft angeben. (Bei Normalverteilungsannahme ist ersterer der Schätzer gleichmäßig kleinster Varianz, dagegen liefert letzterer stets eine „gleichmäßig kleinste“ Schätzung; vgl. Hartung (1981).)

Wir betrachten nun einen speziellen Fall der Varianzkomponentenschätzung in einem gerichteten geodätischen Netz. Die Höhendifferenzen zwischen einem Ort 1 und einem Ort 2 sowie zwischen den Orten 2 und 3 werden je zweimal gemessen. Die unbekannten absoluten Höhen der Orte 1, 2 und 3 seien mit h_1 , h_2 und h_3 bezeichnet. Gemessen werden zum Zeitpunkt t_1 die Höhendifferenzen Y_{12} zwischen den Orten 1 und 2, Y_{23} zwischen den Orten 2 und 3 und zum Zeitpunkt t_2 die Höhendifferenzen Y_{21} zwischen 2 und 1 sowie Y_{32} zwischen 3 und 2, vgl. Abb. 2. Wären die Messungen exakt, so würde natürlich gelten

$$Y_{12} = -Y_{21} \quad \text{und} \quad Y_{23} = -Y_{32}.$$

Die Höhendifferenzen zwischen den Orten 1 und 2 können dann durch

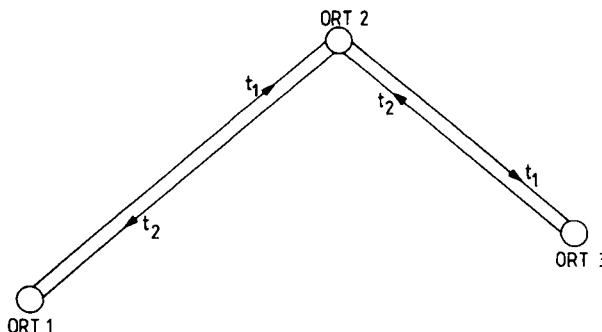


Abb. 2: Höhenmessungen in einem gerichteten geodätischen Netz

$$\widehat{h_2 - h_1} = \frac{1}{2} \cdot (Y_{12} - Y_{21})$$

und die zwischen den Orten 3 und 2 durch

$$\widehat{h_2 - h_3} = \frac{1}{2} \cdot (Y_{32} - Y_{23})$$

geschätzt werden.

Nehmen wir nun einmal an, daß die Ortshöhen selbst unkorreliert und die Varianzen der Höhen zum gleichen Zeitpunkt t_i ($i = 1, 2$) identisch sind, nämlich gleich σ_i^2 bzw. σ_2^2 .

Basierend auf dem Verfahren von Helmert ergibt sich dann als Schätzung für σ_i^2 ($i = 1, 2$), vgl. Grafarend/Schaffrin (1979),

$$\begin{aligned} s_i^2 = \frac{1}{6} \cdot & (Y_{12}^2 + Y_{23}^2 + Y_{21}^2 + Y_{32}^2 + Y_{12} \cdot Y_{23} + 2 \cdot Y_{12} \cdot Y_{21} \\ & + Y_{12} \cdot Y_{32} + Y_{23} \cdot Y_{21} + 2 \cdot Y_{23} \cdot Y_{32} + Y_{21} \cdot Y_{32}). \end{aligned}$$

Der andere erwähnte Schätzer für σ_i^2 ($i = 1, 2$) hingegen ergibt sich zu

$$\begin{aligned} S_i^2 = \frac{1}{10} \cdot & (Y_{12}^2 + Y_{23}^2 + Y_{21}^2 + Y_{32}^2 + 2 \cdot Y_{12} \cdot Y_{21} + 2 \cdot Y_{23} \cdot Y_{32} \\ & - Y_{12} \cdot Y_{23} - Y_{12} \cdot Y_{32} - Y_{23} \cdot Y_{21} - Y_{32} \cdot Y_{21}). \end{aligned}$$

Eine beliebige Linearkombination $p_1 \cdot \sigma_1^2 + p_2 \cdot \sigma_2^2$ der Varianzen wird nach Helmert durch

$$\begin{aligned} \frac{1}{6} \cdot & (Y_{12}^2 + Y_{23}^2 + Y_{21}^2 + Y_{32}^2 + Y_{12} \cdot Y_{23} + 2 \cdot Y_{12} \cdot Y_{21} + Y_{12} \cdot Y_{32} \\ & + Y_{23} \cdot Y_{21} + 2 \cdot Y_{23} \cdot Y_{32} + Y_{21} \cdot Y_{32}) \cdot (p_1 + p_2) \end{aligned}$$

geschätzt, wohingegen der andere Schätzer

$$\begin{aligned} \frac{1}{10} \cdot & (Y_{12}^2 + Y_{23}^2 + Y_{21}^2 + Y_{32}^2 + 2 \cdot Y_{12} \cdot Y_{21} + 2 \cdot Y_{23} \cdot Y_{32} \\ & - Y_{12} \cdot Y_{23} - Y_{12} \cdot Y_{32} - Y_{23} \cdot Y_{21} - Y_{32} \cdot Y_{21}) \cdot (p_1 + p_2) \end{aligned}$$

liefert.

Beispiel: Bei einer Höhenmessung wurde zwischen den Orten 1 und 2 eine Differenz $Y_{12} = 223,4$ und zwischen den Orten 2 und 3 eine Differenz $Y_{23} = -345,8$ der absoluten Höhen zum Zeitpunkt t_1 beobachtet. Zur Zeit t_2 waren die beobachteten Höhendifferenzen zwischen den Orten 2 und 1 bzw. 3 und 2 dann gerade $Y_{21} = -218,7$ und $Y_{32} = 347,1$.

Die Varianz der absoluten Höhen zum Zeitpunkt t_1 bzw. t_2 wird dann nach Helmert durch

$$s_1^2 = s_2^2 = \frac{1}{6} \cdot 29,89 = 4,982$$

oder durch

$$S_1^2 = S_2^2 = \frac{1}{10} \cdot 17,67 = 1,767$$

geschätzt. Jede Linearkombination $p_1 \cdot \sigma_1^2 + p_2 \cdot \sigma_2^2$ wird dann durch

$$(p_1 + p_2) \cdot 4,982 \text{ bzw. durch } (p_1 + p_2) \cdot 1,767$$

geschätzt. Damit ist z. B. die Varianz der Höhendifferenzen gemäß

$$(1+1) \cdot 4,982 = 9,964 \text{ bzw. } (1+1) \cdot 1,767 = 3,534$$

schätzbar.

Außer auf die bereits zitierte Literatur zum Gebiet der Varianzkomponentenmodelle sei hier auf Elpelt (1983, 1984), Hartung/Heine (1984), Hartung/Klössener (1985), Hartung/Voet (1984), Klössener (1983), Pukelsheim (1981), Voet (1983) und insbesondere auf den „Modellkatalog Varianzkomponenten“ von Hartung/Voet (1985) hingewiesen, in dem auf ca. 500 Seiten neben einer allgemeinverständlichen Darstellung der wesentlichen Ergebnisse zum Schätzen und Testen von Varianzkomponenten die gebräuchlichsten Modelle (22 Stück) speziell für die Praxis aufbereitet sind, u. a. mit jeweils expliziter Angabe der Varianzkomponentenschätzer (ANOVA-Schätzer bzw. MINQUE und „positiver Schätzer“), der Schätzer für zugehörige Varianz bzw. MSE und Kovarianzen sowie aller möglichen Tests über die Varianzkomponenten; für jedes Modell wird mindestens ein konkretes, reales Beispiel in allen Variationen durchgerechnet (zu beziehen bei: Prof. Dr. J. Hartung, Fachbereich Statistik, Universität Dortmund, Postfach 500500, 4600 Dortmund 50).

Kapitel XII: Zeitreihenanalyse

In vielen Bereichen ist es üblich, den zeitlichen Ablauf gewisser Phänomene festzuhalten und zu analysieren.

Beispiel: (a) Das wohl bekannteste Beispiel für eine derartige Vorgehensweise in der Medizin dürfte die Fieberkurve sein. Zu verschiedenen, festgelegten Zeiten wird die jeweils gemessene Körpertemperatur des betreffenden Patienten in ein Fieberblatt eingetragen und der so erhaltene Punkt mit dem vorhergehenden geradlinig verbunden. Aus dem Verlauf der Fieberkurve kann man dann etwa Rückschlüsse auf den Verlauf der Erkrankung und auf Komplikationen ziehen.

(b) Aus dem Bereich der (Land-)Wirtschaft ist der sogenannte Hanauische Schweinezyklus wohlbekannt, vgl. Hanau (1930). Hanau stellt dabei für die Schweinepreise der Jahre 1896–1914 einen Vierjahresrhythmus fest: Auf zweijährige Preissteigerungen folgen zwei Jahre des Preisverfalls ehe die Preise wieder steigen.

(c) Meteorologische Institute halten in regelmäßigen (wenn auch unterschiedlichen) Abständen Niederschlagsmenge, Lufttemperatur, Windrichtung und -stärke fest.

(d) Seismologen registrieren ständig mit ihren hochempfindlichen Geräten die Erschütterungen der Erdkruste, um so (unter anderem) Aufschlüsse über eventuelle Besonderheiten der seismologischen Wellen zu erhalten, die als Vorboten für Erdbeben dienen und so einen wichtigen Beitrag zur Frühwarnung der betroffenen Bevölkerung leisten können.

(e) Insbesondere für eine Verbesserung der Lebensumstände von Schwerbehinderten werden Maschinen konstruiert, welche verbale Anweisungen „verstehen“ können. Um eine derartige automatische Spracherkennung gewährleisten zu können, ist für jeden Laut (und im Grunde sogar für jede Lautkombination) der Schwingungsverlauf in der Zeit zu untersuchen, der nicht nur von Mensch zu Mensch, sondern auch individuell (etwa durch Erkältungskrankheiten) variiert.

(f) Um Aufschlüsse über die Eigenschaften weit entfernter Fixsterne, Spiralnebel usw. zu erhalten, haben Astronomen umfassende Aufzeichnungen des zeitlichen Verlaufs der elektromagnetischen Strahlung und ihrer Intensitätsverteilung für verschiedene Wellenlängen vorgenommen. Aufgrund solcher Information konnte die Existenz eines Neutronensterns im Krebsnebel (crab nebula, PSR $\varnothing 531 + 21$), im Sternbild des Stiers, nachgewiesen werden.

Die Liste der Beispiele ließe sich noch beliebig verlängern. Prinzipiell nennt man jede zeitliche Folge quantitativer Beobachtungswerte zu einem bestimmten Vorgang eine **Zeitreihe**. Die Betrachtung von Zeitreihen verfolgt einen oder mehrere der folgenden Zwecke:

- wie in den Beispielen (c), (e) und (f) ist eine Beschreibung des zeitlichen Vorgangs selbst von Interesse;
- in den Beispielen (a), (c) und (d) dient die zeitliche Beobachtung der Kontrolle, d. h. man möchte Veränderungen mit der Zeit möglichst frühzeitig erkennen, um gegebenenfalls rechtzeitig Gegenmaßnahmen einzuleiten;
- man ist, vgl. die Beispiele (a)–(d), an einer Prognose der zukünftigen Entwicklung interessiert.

In jedem Fall können Beobachtungswerte nur über einen endlichen Zeitraum hinweg erhoben werden, wohl aber können im Einzelfall unendlich viele Einzeldaten zur Verfügung stehen, wie dies in den Beispielen (d), (e) und (f) der Fall ist.

Bei jeder numerischen Auswertung wird man sich zudem auf endlich viele dieser Beobachtungen beschränken müssen, da die Rechenkapazität der Menschen wie der Computer begrenzt ist. Weil auch der unvoreingenommenste Analytiker Gefahr läuft, bei willkürlicher Auswahl durch zu starke Berücksichtigung ihm besonders interessant erscheinender Zeitpunkte (für die z. B. Minima oder Maxima auftreten) den „typischen“ Verlauf der Zeitreihe systematisch zu verfälschen, werden in praxi **äquidistante Zeitpunkte** ausgewählt. Dabei sollte das jeweils übersprungene Zeintervall so klein sein, daß auch alle wesentlichen Phänomene erfaßt werden können. Ändert sich eine Zeitreihe nur sehr langsam, so wird man mit weniger Beobachtungen pro Zeiteinheit auskommen als bei starker Oszillation.

Allerdings darf man nicht annehmen, daß für jede Zeitreihe eine bessere Erfassung durch z. B. eine Verdoppelung der Zahl der Beobachtungen pro Zeiteinheit möglich ist. Gerade bei wirtschaftlichen Zeitreihen scheitert ein solches Vorgehen oft daran, daß manche Größen etwa nur monatlich ermittelt oder veröffentlicht werden; man denke etwa an die Arbeitslosenzahlen.

In diesem Zusammenhang ist auch daran zu denken, daß einer theoretisch möglichen Verbreiterung der Datenbasis nicht nur finanzielle und organisatorische Schwierigkeiten gegenüberstehen. Da jede Informationsvermittlung fehlerbehaftet ist, kann gerade bei kurzen Berichtszeiträumen die gewünschte Information durch (zufällige oder systematische) Fehler völlig überlagert werden, wohingegen eine längere Zwischenzeit ein – wenn auch sicher nur recht unvollkommenes – „Ausmitten“ der Fehler bewirken kann.

Beispiel: Die offiziellen Arbeitslosenzahlen beziehen sich auf die bei den Arbeitsämtern als arbeitslos Gemeldeten. Die eigentlich interessierende Zahl der unbeschäftigten Arbeitsfähigen und zugleich -willigen dürfte so auch im Normalfalle recht gut approximiert werden. Bei wöchentlicher Erfassung könnte jedoch eine sehr heiße Sommerwoche zu einer merklichen Überschätzung letzterer Zahl durch die Anzahl der in den Listen geführten Arbeitslosen führen.

Besonders bei ökonomischen Zeitreihen mit monatlichen oder vierteljährlichen Berichtszeiträumen tritt das Problem der **Kalendereffekte** auf: Durch die unterschiedliche Länge der Monate und durch wichtige Festtage wie z. B. Weihnachten treten zum Teil extreme Verzerrungen auf. An dieser Stelle sei nur darauf hingewiesen, daß eine Verwendung der Arbeitstage zur Korrektur der Daten nicht unproblematisch ist, da sie nur ein sehr unvollkommenes Substitut für die tatsächlich geleistete Arbeitszeit darstellen. Für eine genauere Besprechung dieser Problematik sei verwiesen auf Brubacher/Wilson (1976) sowie Cleveland/Devlin (1980).

Selbstverständlich ist auch darauf zu achten, daß die betrachtete Zeitreihe in der Tat vergleichbare Werte enthält. Wirtschaftliche Zeitreihen sollten also etwa nur inflationsbereinigte Daten enthalten, gegebenenfalls auch veränderten Flächen- und Einwohnerstand berücksichtigen (Gebietsreform).

Wir wollen an dieser Stelle zunächst einen kurzen **Überblick** über die Inhalte dieses Kapitels geben.

Im ersten Abschnitt beschäftigen wir uns mit deskriptiven Methoden der Zeitreihenanalyse. Wir behandeln dabei im wesentlichen den Fall, daß sich eine Zeitreihe

y_1, \dots, y_n der Länge n additiv aus einer glatten Komponente (i. w. eine Trendkomponente), einer saisonalen Komponente und einer irregulären Komponente zusammensetzt, vgl. Abschnitt 1.1. Im Abschnitt 1.2 werden einige spezielle, klassische Trendmodelle für Zeitreihen ohne saisonale Komponente behandelt. Im Abschnitt 1.3 beschäftigen wir uns mit der Schätzung und Elimination von glatten Komponenten und der Bereinigung einer Zeitreihe von saisonalen Komponenten (bekannte Verfahren sind z. B. das Berliner Verfahren, Census X-11 oder ASA-II), wobei in der Regel eine konstante Saisonfigur mit bekannter Periode vorausgesetzt wird, d. h. die saisonale Komponente wiederholt sich nach einer gewissen Zeit gleichförmig. Auf die in Abschnitt 1.2 und 1.3 angesprochene Problematik kommen wir in Abschnitt 3.5 und 3.6 noch einmal zurück, wo die harmonische Analyse und das Berliner Verfahren, ein spezielles Saisonbereinigungsverfahren dargestellt werden. In Abschnitt 1.4 beschäftigen wir uns dann noch mit Kenngrößen für den Zusammenhang zwischen den Beobachtungswerten einer Zeitreihe.

Im **zweiten Abschnitt** werden dann Modelle vorgestellt, bei denen eine Zeitreihe y_1, \dots, y_n als endliche Realisation stochastischer Prozesse aufgefaßt wird. Die in den Abschnitten 2.1–2.3 dargestellten autoregressiven, moving average und gemischten Prozesse erklären sich aus sich selbst heraus und/oder einem Fehlerprozeß, der ein sogenanntes weißes Rauschen ist. Im Abschnitt 2.4 werden dann Verallgemeinerungen solcher Prozesse bzw. Transformationen, die derartige Prozesse liefern, behandelt. Möchte man an eine konkrete Zeitreihe einen solchen Prozeß anpassen, d. h. möchte man sie mittels eines solchen Prozesses erklären, so muß, bevor die Prozeßparameter geschätzt werden können, das adäquate Modell (d. h. der adäquate Prozeß) identifiziert werden. Hierzu kann die im Abschnitt 2.5 dargestellte und im Abschnitt 2.6 an einem Beispiel demonstrierte Methode von Box und Jenkins herangezogen werden. Neben Modellidentifikation und Parameterschätzung dient dieses Verfahren der Güteprüfung des Modells und der Prognose zukünftiger Werte der Zeitreihe.

In den Abschnitten 1 und 2 haben wir Zeitreihen bzw. stochastiche Prozesse im Zeitbereich behandelt. Die dort in einer Zeitreihe enthaltene Information läßt sich mittels der Spektralanalyse, die im **Abschnitt 3** dargestellt wird, in den Frequenzbereich transferieren. Zur Beurteilung einer Zeitreihe oder eines Prozesses im Frequenzbereich dienen das Spektrum und die Spektraldichte, die in Abschnitt 3.2 im Anschluß an einen kurzen Abriß über komplexe Zahlen in Abschnitt 3.1 vorgestellt werden. Im Abschnitt 3.3 gehen wir dann auf die Auswirkung von Glättungs- und Saisonbereinigungsverfahren, vgl. Abschnitt 1.3, auf die Spektraldichte ein. Hierüber lassen sich dann auch die Auswirkungen im Zeitbereich charakterisieren. Schließlich beschäftigen wir uns in Abschnitt 3.4 mit der Schätzung von Spektraldichten, und im Anschluß daran werden in Abschnitt 3.5 und 3.6 die harmonische Analyse und das Berliner Verfahren geschildert, die inhaltlich eigentlich mehr zu Abschnitt 1 gehören, jedoch aus technischen Gründen erst im Abschnitt 3 abgedeckt werden.

Im **Abschnitt 4** beschäftigen wir uns dann mit der Analyse des Zusammenhangs zweier Zeitreihen. Zunächst betrachten wir in Abschnitt 4.1 den Zusammenhang im Zeitbereich und dann in Abschnitt 4.2 den im Frequenzbereich. Bei letzterer Problemstellung spricht man auch von Kreuzspektralanalyse.

Schließlich werden im **Abschnitt 5** noch Mischformen von Regressions- und Zeitreihenmodellen behandelt. Zunächst betrachten wir in Abschnitt 5.1 Regressionsmodelle, bei denen – im Gegensatz zur Regressionsanalyse in Kap. X – die Fehler

nicht mehr unabhängig sind, sondern z. B. einem der in Abschnitt 2 dargestellten Prozesse folgen. Im Abschnitt 5.2 wird zudem zugelassen, daß einige der Regressoren im Modell keine echten, exogenen Variablen sondern vielmehr zeitverzögerte, endogene Variablen sind, d. h. der Regressand taucht zeitverzögert auch als Regressor auf.

Es sei an dieser Stelle noch kurz der Begriff des **Kalman-Filters** erwähnt. Dabei geht man von einem sogenannten **Zustandsraummodell** aus, das sich aus einer Systemgleichung für die eigentlich interessierende, u. U. nichtbeobachtbare, zeitabhängige Größe x_t und einer Beobachtungsgleichung für y_t zusammensetzt, wobei y_t durch x_t und einen zufälligen Rest erklärt wird. Ziel des Kalman-Filter-Ansatzes, vgl. etwa Kalman (1960), ist es nun, x_t rekursiv zu schätzen aus der Beobachtung zu y_t und der Schätzung für x_{t-1} zum direkt vorausgehenden Zeitpunkt. Diesem allgemeinen Ansatz, den wir hier nicht näher erläutern wollen, man vgl. aber etwa Schlittgen/Streitberg (1984), lassen sich die meisten der in Abschnitt 1 und 2 dargestellten Methoden theoretisch unterordnen.

Wir können in diesem Kapitel nur jeweils getrennt einzelne Aspekte der Zeitreihenanalyse behandeln. **Wie geht man nun aber konkret vor**, wenn eine bestimmte Zeitreihe y_1, \dots, y_n analysiert werden soll. In den **meisten** Fällen wird man zunächst versuchen, die Zeitreihe durch ein Regressionsmodell nach der gewöhnlichen Methode der kleinsten Quadrate zu approximieren. Als Regressoren kommen hier neben exogenen, zusätzlich erhobenen Variablen und verzögerten Variablen auch explizite Funktionen der Zeit (vgl. z. B. Abschnitt 1.2, 3.5 und 3.6) in Frage. Aus den daraus resultierenden Schätzungen \hat{y}_t werden geschätzte Residuen $\hat{\epsilon}_t = y_t - \hat{y}_t$ für $t = 1, \dots, n$ bestimmt, an diese wird man dann einen stochastischen Prozeß nach der Methode von Box und Jenkins anpassen und hieraus die Kovarianzmatrix von y_1, \dots, y_n schätzen. Unter Verwendung dieser Schätzmatrix kann dann mittels Aitken-Schätzung, vgl. Abschnitt 5, eine neue Anpassung durch das Regressionsmodell berechnet werden; die neuen, daraus resultierenden Residuen-Schätzer werden wieder dem Box-Jenkins-Verfahren unterzogen, dann wird ein neuer Schätzer für die Kovarianzmatrix von y_1, \dots, y_n bestimmt usw., bis sich die Ergebnisse (hoffentlich) stabilisieren.

Der weitergehend an Methoden der Zeitreihenanalyse interessierte Leser sei an dieser Stelle auch verwiesen auf die Bücher von Grenander/Rosenblatt (1957), Hannan (1970), Anderson (1971), Kendall (1973), Brillinger (1975), Chatfield (1975), Box/Jenkins (1976), Fuller (1976), Schlittgen/Streitberg (1984) sowie auf den Übersichtsartikel von Heiler (1981). Speziell bzgl. Saisonbereinigungsverfahren sei z. B. auf Stier (1980) und bzgl. spektralanalytischer Methoden auch auf Naeve (1969) sowie König/Wolter (1972) hingewiesen.

1. Deskriptive Methoden der Zeitreihenanalyse

1.1. Die Komponenten einer Zeitreihe

Häufig setzen sich die Beobachtungswerte y_t ($t = 1, \dots, n$) einer Zeitreihe additiv aus drei Komponenten zusammen, nämlich aus der **glatten Komponente** G_t , der **saisonalen Komponente** S_t und der **irregulären Komponente** R_t . Wir werden im folgenden solche **additiven Zeitreihenmodelle**

$$y_t = G_t + S_t + R_t \quad \text{für } t = 1, \dots, n$$

betrachten.

Die glatte Komponente G_t gibt den fiktiven Verlauf der Zeitreihe bei Fehlen saisonaler und irregulärer Schwankungen an. Mitunter wird die glatte Komponente noch in eine Trendkomponente und eine zyklische Komponente aufgeteilt: Der Verlauf der **Trendkomponente** ist bedingt durch langfristige Ursachen wachsend oder fallend (mit der Zeit t), wohingegen die **zyklische Komponente** einen wellenförmigen Verlauf aufgrund sich stetig, aber langsam ändernder Einflüsse nimmt. Im Falle wirtschaftlicher Zeitreihen etwa entsteht eine zyklische Komponente oft durch den Konjunkturzyklus. Ob eine signifikante Trendkomponente in einer Zeitreihe vorhanden ist, d. h. ob der Verlauf signifikant steigt oder fällt, lässt sich mit Hilfe der in Abschnitt 5.4.3 des Kap. IV dargestellten Tests auf Trend überprüfen, vgl. auch das Beispiel in Abschnitt 1.3.3.

Ebenfalls wellenförmig ist der Verlauf der saisonalen Komponente, die durch einen periodischen Zeit-Einfluß auf die Beobachtungswerte entsteht. Ein solcher periodischer Einfluß, wie z. B. der Einfluß der Jahreszeiten auf die Auftragslage im Baugewerbe, ist natürlich nur bei manchen Zeitreihen vorhanden.

Schließlich ist die irreguläre Komponente R_t , das Ergebnis kurzfristiger, sich unregelmäßig verändernder Ursachen. Wir gehen davon aus, daß die irreguläre Komponente zufällig um Null schwankt.

In den weiteren Ausführungen im Abschnitt 1 werden wir davon ausgehen, daß sich der Verlauf der saisonalen Komponente, falls sie überhaupt vorhanden ist, nach einer festen Anzahl p von Zeitpunkten wiederholt, d. h. es gilt $S_t = S_{t+p}$ für $t = 1, \dots, n-p$. Man spricht auch von Zeitreihen mit **konstanter Saisonfigur**.

Es gibt aber auch Zeitreihen, deren Saisonfigur sich im Laufe der Zeit ändert. Oft bleibt dann nur die Form der Schwankungen gleich, nicht aber ihre Stärke. Ein solcher Verlauf der saisonalen Komponente entspricht einer **Saisonfigur mit variabler Amplitude**. Die Amplitude kann sich entweder stetig oder auch sprunghaft verändern; in vielen Fällen verhält sie sich proportional zur glatten Komponente. Ist die Veränderung auch proportional zur irregulären Komponente, so lassen sich die Beobachtungswerte durch das **multiplikative Zeitreihenmodell**

$$y_t = G_t \cdot S_t \cdot R_t \quad \text{für } t = 1, \dots, n$$

beschreiben. Geht man zu den Logarithmen der Zeitreihe über,

$$\ln y_t = \ln(G_t \cdot S_t \cdot R_t) = \ln G_t + \ln S_t + \ln R_t \quad \text{für } t = 1, \dots, n,$$

so trifft für diese wiederum unser additives Zeitreihenmodell mit konstanter Saisonfigur zu.

Wir werden uns in den Abschnitten 1.2 und 1.3 mit der Schätzung bzw. Elimination von glatter und saisonaler Komponente beschäftigen. Im Abschnitt 1.2 betrachten wir zunächst nichtlineare Trendmodelle für Zeitreihen ohne saisonale Komponente und im Abschnitt 1.3 werden wir sowohl für den Fall von Zeitreihen ohne saisonale Komponente als auch für den von Zeitreihen mit konstanter Saisonfigur Methoden der Trend- bzw. Saison-Schätzung und -Elimination behandeln.

Mitunter weiß man auch, daß eine Zeitreihe i. w. durch Überlagerung von harmonischen Schwingungen (Sinus- und Kosinus-Schwingungen) bestimmter Perioden festgelegt ist. Die Approximation einer Zeitreihe durch solche Schwingungen nennt man **harmonische Analyse**. Auf diese Vorgehensweise werden wir erst in Abschnitt 3.5 eingehen. Noch allgemeinere Schätzmethoden zur Approximation von Zeitreihen werden etwa in Förster (1969) behandelt.

1.2. Nichtlineare Trendmodelle – Trendschätzung mittels nichtlinearer Regression

Eine Zeitreihe y_1, \dots, y_n , deren Beobachtungsdaten lediglich aus einer Trendkomponente und einer irregulären Komponente zusammengesetzt sind, läßt sich vielfach durch eine Funktion approximieren, d. h. an die Trendkomponente kann eine Funktion der Zeit angepaßt werden, die von unbekannten, zu schätzenden Parametern β_1, \dots, β_k abhängt.

Die Parameter einer solchen Funktion

$$y(t) = y(t; \beta_1, \dots, \beta_k)$$

können – wie in Regressionsmodellen, vgl. Kap. X – nach der **Methode der kleinsten Quadrate** geschätzt werden, falls die Abstände zwischen den Beobachtungszeitpunkten $t = 1, \dots, n$ äquidistant sind: Zu gegebenen Beobachtungsdaten $(y_1, 1), (y_2, 2), \dots, (y_n, n)$ muß die **Summe der quadratischen Abweichungen**

$$S^2 = S^2(\beta_1, \dots, \beta_k) = \sum_{t=1}^n (y_t - y(t; \beta_1, \dots, \beta_k))^2$$

minimiert werden.

Ist die Funktion $y(t)$ linear in den Parametern (z. B. linearer oder polomialer Trend), so können die in Kap. X dargestellten Methoden der linearen Regression eingesetzt werden, ebenso natürlich bei (bzgl. der Parameter) linearisierbaren Funktionen, vgl. Abschnitt 3 in Kap. X; es sei aber auch auf Abschnitt 2 in Kap. XIV hingewiesen.

Ist die Funktion $y(t)$ nichtlinear in den Parametern, so kann diese Minimierung entweder direkt mittels Gradientenverfahren oder z. B. Verfahren vom Davidon-Fletcher-Powell-Typ oder z. B. auch durch Nullsetzen der partiellen Ableitungen von S^2 nach den Parametern β_1, \dots, β_k mit Lösen des dadurch erzeugten Gleichungssystems, etwa mit Newton-Verfahren, erfolgen.

Wir wollen auf solche numerischen Verfahren hier nicht eingehen und verweisen etwa auf Himmelblau (1972), Blum/Oettli (1975) und Stoer (1979). Es sei lediglich noch darauf hingewiesen, daß man i. a. (z. B. anhand der zweiten Ableitungen und/oder mehrerer Funktionswerte) überprüfen sollte, ob der erhaltene Extrempunkt wirklich ein globales Minimum der Funktion S^2 liefert.

Wir wollen hier die spezielle Vorgehensweise bei einigen klassischen nichtlinearen Trendfunktionen, nämlich der **logistischen Funktion**, der **Mitscherlich-Funktion** oder **Gooperitz-Kurve** und der **allometrischen Funktion**, darstellen. Schon an diesen Beispielen zeigt sich, daß eine Minimierung der Summe der Abweichungsquadrate recht aufwendig sein kann.

An dieser Stelle sei noch vermerkt, daß jede Trendfunktion auch allgemein als lineare oder nichtlineare Regressionsfunktion in Abhängigkeit von einem Regressor oder einer Funktion der Zeit $x(t)$, vgl. Kap. X, verwandt werden kann. An die Stelle der Beobachtungszeitpunkte $t = 1, \dots, n$ treten dann die Beobachtungswerte x_1, \dots, x_n des Regressors bzw. die Funktionswerte $x(1), \dots, x(n)$; diese Vorgehensweise wird in einem Beispiel zur allometrischen Funktion in Abschnitt 1.2.3 demonstriert.

1.2.1. Die logistische Funktion

Die **logistische Funktion** oder **logistische Kurve** eignet sich mitunter zur Repräsentation von **Wachstumskurven**. Sie hat gegenüber anderen Wachstumsfunktionen den

Vorteil, daß sie nicht mit der Zeit unendlich wird, sondern einer Asymptote – der sogenannten **Sättigungsgrenze** – zustrebt. Daher kann sie auch zur Extrapolation, d. h. zur Prognose, von Werten in einen größeren zukünftigen Zeitbereich hinein verwandt werden.

Der Entdecker der logistischen Funktion

$$y(t) = y(t; \alpha, \beta, \gamma) = \frac{\gamma}{1 + \beta e^{-\alpha t}},$$

deren prinzipieller Verlauf für $\alpha > 0$ in Abb. 1 dargestellt ist, war Verhulst (1838) und wiederentdeckt wurde die Funktion erst von Pearl/Reed (1920), man vgl. aber auch Kuznets (1930), Mitchell (1941) und Winkler (1951).

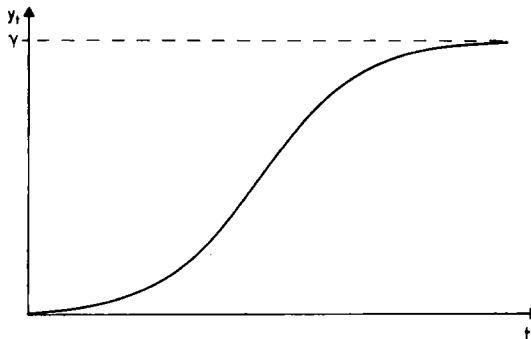


Abb. 1: Prinzipielle Gestalt einer logistischen Funktion $y_t = \gamma/(1 + \beta e^{-\alpha t})$ für $\alpha > 0$

Wie man in der Abb. 1 sieht, strebt die logistische Funktion im Falle $\alpha > 0$ mit wachsendem t gegen die **Sättigungsgrenze** γ . Ist umgekehrt $\alpha < 0$, so läuft sie für $t \rightarrow \infty$ gegen Null und für $t \rightarrow 0$ gegen $\gamma/(1 + \beta)$.

Die logistische Funktion $y(t)$ ist nun bzgl. der Parameter α und β nicht linearisierbar, so daß die in Kap. X dargestellte allgemeine Vorgehensweise der linearen Regressionsrechnung nicht direkt zur Schätzung der unbekannten Parameter herangezogen werden kann. Jedoch entwickelte Tintner (1958) eine Methode, die doch wieder eine Parameterschätzung mittels linearer Regression erlaubt. Sei im Folgenden $y_t > 0$ für $t = 1, \dots, n$.

Geht man von der Funktion $y(t)$ zur Umkehrfunktion über, so erhält man

$$\begin{aligned} \frac{1}{y(t)} &= \frac{1 + \beta e^{-\alpha t}}{\gamma} = \frac{1 - e^{-\alpha}}{\gamma} + e^{-\alpha} \cdot \frac{1 + \beta e^{-\alpha(t-1)}}{\gamma} \\ &= \frac{1 - e^{-\alpha}}{\gamma} + e^{-\alpha} \cdot \frac{1}{y(t-1)}. \end{aligned}$$

Hat man an den Stellen $t = 1, 2, \dots, n$ die (von Null verschiedenen) Größen y_1, \dots, y_n beobachtet, die auch als Realisationen von gleichbezeichneten Zufallsvariablen y_1, \dots, y_n aufgefaßt werden können, so kann aufgrund obiger Beziehung eine lineare Regression von $1/y_t$ auf $1/y_{t-1}$ durchgeführt werden:

$$\frac{1}{y_t} = \frac{1 - e^{-\alpha}}{\gamma} + e^{-\alpha} \cdot \frac{1}{y_{t-1}} + \varepsilon_t \quad \text{für } t = 2, \dots, n;$$

hier bezeichnen $\varepsilon_2, \dots, \varepsilon_n$ unabhängige Zufallsvariablen mit Erwartungswert 0 und Varianz σ^2 . $(1 - e^{-\alpha})/\gamma$ ist dann gerade das Absolutglied und $e^{-\alpha}$ der Steigungsparameter der Regression von $1/y_t$ auf $1/y_{t-1}$. (Dieser Ansatz entspricht der Annahme, daß der zu $1/y_t$ gehörige Fehler-Prozeß ein AR(1)-Prozeß mit Parameter $-e^{-\alpha}$ ist; für $\alpha > 0$ ist dieser Prozeß auch stationär, vgl. hierzu Abschnitt 2.1 und Abschnitt 5.1 dieses Kapitels.)

Wir können nun zunächst Absolutglied und Steigungsparameter, wie in Abschnitt 1.1 von Kap. X angegeben, nach der Methode der kleinsten Quadrate schätzen und erhalten dann als Schätzer für die unbekannten Parameter α und γ zunächst die Größen

$$\hat{\alpha} = -\ln \frac{\sum_{t=2}^n \frac{1}{y_t y_{t-1}} - \frac{1}{n-1} \sum_{t=2}^n \frac{1}{y_{t-1}} \sum_{t=2}^n \frac{1}{y_t}}{\sum_{t=2}^n \left(\frac{1}{y_{t-1}}\right)^2 - \frac{1}{n-1} \left(\sum_{t=2}^n \frac{1}{y_{t-1}}\right)^2},$$

$$\hat{\gamma} = (n-1)(1 - e^{-\hat{\alpha}}) \left/ \left(\sum_{t=2}^n \frac{1}{y_t} - e^{-\hat{\alpha}} \sum_{t=2}^n \frac{1}{y_{t-1}} \right) \right..$$

Tinter (1958) schlägt vor, den verbleibenden, unbekannten Parameter β nach der Formel von Rhodes (1940) durch

$$\hat{\beta} = e^{(n+1)\hat{\alpha}/2 + [\sum_{t=1}^n \ln(\hat{y}/y_t - 1)]/n}$$

zu schätzen. Allerdings eignet sich dieser Schätzer nur dann, wenn alle Beobachtungswerte y_1, \dots, y_n unterhalb der geschätzten Sättigungsgrenze \hat{y} liegen. Ist dies nicht der Fall, so ist

$$\hat{\beta} = \left(\sum_{t=1}^n \frac{\hat{y} - y_t}{y_t} e^{-\hat{\alpha}t} \right) \left/ \sum_{t=1}^n e^{-2\hat{\alpha}t} \right..$$

ein möglicher Schätzer für β .

Als Schätzung bzw. Prognose für den Wert von $y(t)$ an einer Stelle t verwendet man dann natürlich

$$\hat{y}(t) = \frac{\hat{y}}{1 + \hat{\beta} e^{-\hat{\alpha}t}},$$

wobei für $t = 1, \dots, n, n+1, n+2, \dots$ oft auch geschrieben wird

$$\hat{y}(t) = \hat{y}_t.$$

Eine andere Möglichkeit, die Parameter der logistischen Funktion (mit entsprechend anderem Fehleransatz) zu schätzen, wird am Ende des nachfolgenden Abschnitts 1.2.2 beschrieben.

Beispiel: Mitunter läßt sich das Bevölkerungswachstum in einem Land durch die logistische Funktion repräsentieren, wobei allerdings die Schätzung stark von der gewählten Datenbasis abhängt. Um dies zu demonstrieren, wollen wir hier die

Parameter der logistischen Funktion für die Bevölkerungsentwicklung in Nordrhein-Westfalen schätzen und die Bevölkerungsentwicklung auch in die Zukunft prognostizieren. Als Datenbasis wählen wir zum einen die Bevölkerungszahlen der Jahre 1935–1980 in Fünfjahresabständen und zum anderen die der Jahre 1961–1979 in Zweijahresabständen (vgl. Statistisches Jahrbuch Nordrhein-Westfalen 1983).

Zunächst wollen wir aufgrund der in Tab. 1 angegebenen fünfjährigen Bevölkerungszahlen y_1, \dots, y_{10} von 1935–1980 die Parameter der logistischen Funktion

$$y(t) = \frac{\gamma}{1 + \beta e^{-\alpha t}}$$

für die Bevölkerungsentwicklung Nordrhein-Westfalens mittels der zuvor beschriebenen Methode schätzen.

Tab. 1: Bevölkerungsentwicklung in Nordrhein-Westfalen 1935–1980 und mittels logistischer Funktion (bis 1995) prognostizierte Bevölkerungszahlen

Jahr	t	Bevölkerung y_t (in Millionen)	Schätzung \hat{y}_t (in Millionen)
1935	1	11,772	12,051
1940	2	12,059	12,723
1945	3	11,200	12,373
1950	4	12,926	13,995
1955	5	14,442	14,585
1960	6	15,694	15,142
1965	7	16,661	15,663
1970	8	16,914	16,146
1975	9	17,176	16,592
1980	10	17,044	17,001
1985	11	—	17,374
1990	12	—	17,713
1995	13	—	18,018

Aus den Daten y_1, \dots, y_{10} in Tab. 1 ergibt sich dann

$$\begin{aligned} \alpha &= -\ln \left(\frac{0,045355763 - 1/9 \cdot 0,644847022 \cdot 0,618571363}{0,047391926 - 1/9 \cdot (0,644847022)^2} \right) \\ &= -\ln 0,870866225 \\ &= 0,138266902, \end{aligned}$$

und somit erhalten wir als Schätzung für die Sättigungsgrenze γ den Wert

$$\hat{\gamma} = \frac{9 \cdot (1 - 0,870866225)}{0,618571363 - 0,870866225 \cdot 0,644847022} = 20,39102043,$$

d. h. die Sättigungsgrenze für die nordrhein-westfälische Bevölkerung wird auf ca. 20 Millionen Einwohner geschätzt. Da y_1, \dots, y_{10} alle kleiner als $\hat{\gamma}$ sind, können wir den Parameter β der logistischen Funktion hier mittels der Formel von Rhodes schätzen. Es ergibt sich

$$\hat{\beta} = e^{1/2 \cdot 11 \cdot 0,138266902 + 1/10 \cdot (-9,903451248)} = 0,794631206,$$

so daß die Bevölkerung Nordrhein-Westfalens zum Zeitpunkt t durch

$$\hat{y}(t) = \frac{20,39102043}{1 + 0,794631206 \cdot e^{-0,138266902 \cdot t}}$$

geschätzt bzw. prognostiziert wird. Als Schätzung für die Einwohnerzahl im Jahre 1979 ergibt sich wegen $t = 9$ für 1975 und $t = 10$ für 1980

$$\hat{y}(9,8) = 16,922;$$

die tatsächliche Einwohnerzahl belief sich auf 17,002 Millionen, so daß sie also um 0,08 Millionen oder 80000 Einwohner unterschätzt wird. In Tab. 1 sind die Prognosen $\hat{y}(t) = \hat{y}$, für $t = 1, \dots, 13$, also für die Jahre 1935–1995 in Fünfjahresabständen bereits angegeben, und in Abb. 2 ist die Funktion $\hat{y}(t)$ auch graphisch dargestellt.

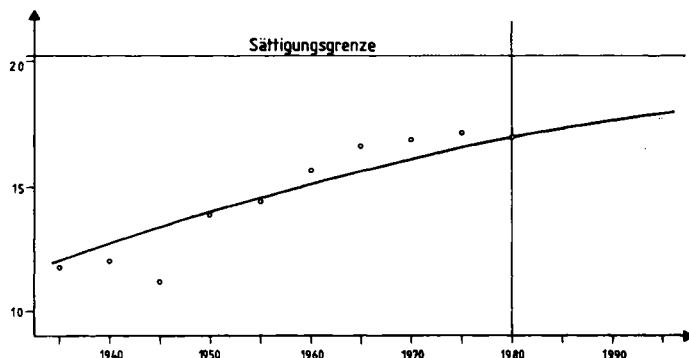


Abb. 2: Schätzung der Bevölkerungszahlen Nordrhein-Westfalens mit Hilfe der logistischen Funktion auf der Basis von 1935–1980

Als Maß für die Anpassungsgüte der geschätzten logistischen Funktion an die Daten kann hier – wie bei linearen Regressionsfunktionen – das Bestimmtheitsmaß B verwendet werden. Es ist mit

$$B = 1 - \frac{\sum_{t=1}^{10} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{10} (y_t - \bar{y})^2} = 1 - \frac{8,6373}{52,2483} = 0,8347, \quad \bar{y} = \frac{1}{10} \sum_{t=1}^{10} y_t,$$

hier recht hoch.

Um die Abhängigkeit der Schätzung von der gewählten Datenbasis aufzuzeigen, wollen wir die Bevölkerungsentwicklung Nordrhein-Westfalens auch noch anhand der zweijährigen Einwohnerzahlen von 1961 bis 1979, vgl. Tab. 2, durch eine logistische Funktion schätzen.

Als Schätzung für den Parameter α der logistischen Funktion ergibt sich aus den Daten y_1, \dots, y_{10} der Tab. 2

Tab. 2: Einwohnerzahlen y_t Nordrhein-Westfalens 1961–1979 und geschätzte bzw. prognostizierte Bevölkerungszahlen \hat{y}_t für 1961–1989

Jahr	t	Bevölkerung y_t (in Millionen)	Schätzung \hat{y}_t (in Millionen)
1961	1	15,920	15,920
1963	2	16,280	16,351
1965	3	16,661	16,636
1967	4	16,835	16,820
1969	5	17,044	16,939
1971	6	17,091	17,014
1973	7	17,223	17,062
1975	8	17,176	17,092
1977	9	17,052	17,112
1979	10	17,002	17,124
1981	11	—	17,131
1983	12	—	17,136
1985	13	—	17,139
1987	14	—	17,141
1989	15	—	17,142

$$\begin{aligned}\hat{\alpha} &= -\ln \frac{0,0316695 - 1/9 \cdot 0,531771 \cdot 0,5357684}{0,0319153 - 1/9 \cdot (0,5357684)^2} \\ &= -\ln 0,6303318 = 0,461509,\end{aligned}$$

so daß die Sättigungsgrenze γ geschätzt wird als

$$\hat{\gamma} = \frac{9 \cdot (1 - 0,6303318)}{0,531771 - 0,6303318 \cdot 0,5357684} = 17,144329.$$

Im Gegensatz zur geschätzten Sättigungsgrenze von ca. 20 Millionen Einwohnern bei der Datenbasis aus Tab. 1, wird hier ein Maximum von ca. 17 Millionen Einwohnern geschätzt. Diese Zahl ist 1979 schon nahezu erreicht und wurde in den Jahren 1973 und 1975 sogar schon überschritten, vgl. Tab. 2. Zur Schätzung des Parameters β können wir daher auch nicht die Formel von Rhodes benutzen, sondern verwenden

$$\hat{\beta} = \left(\sum_{t=1}^{10} \frac{\hat{y}_t - y_t}{y_t} e^{-\hat{\alpha} t} \right) / \sum_{t=1}^{10} e^{-2\hat{\alpha} t} = \frac{0,0804586}{0,6591855} = 0,1220576.$$

In der Tab. 2 sind die mittels der hier geschätzten logistischen Funktion

$$\hat{y}(t) = \frac{17,144329}{1 + 0,1220576 \cdot e^{-0,461509 \cdot t}}$$

prognostizierten Einwohnerzahlen $\hat{y}(t) = \hat{y}_t$ in den Jahren 1961–1989 ($t = 1, \dots, 15$) bereits angegeben; in Abb. 3 ist die Funktion $\hat{y}(t)$ auch graphisch dargestellt.

Das Bestimmtheitsmaß als Maß für die Güte der Anpassung ist hier noch be-

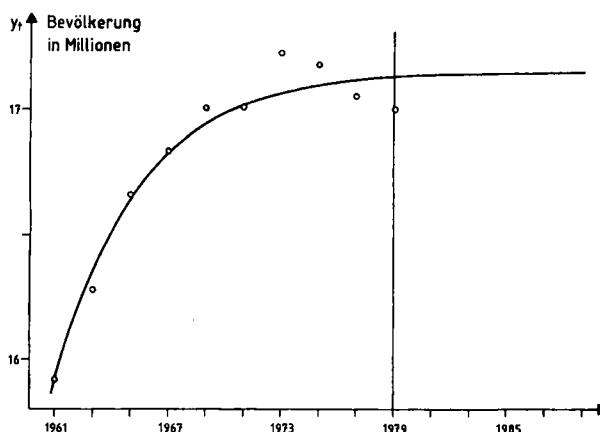


Abb. 3: Schätzung der Bevölkerungszahlen Nordrhein-Westfalens mit Hilfe der logistischen Funktion auf der Basis von 1961–1979

trächtlich höher als im ersten Fall:

$$B = 1 - \frac{\sum_{t=1}^{10} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{10} (y_t - \bar{y})^2} = 1 - \frac{0,0743}{1,6261} = 0,9543, \quad \bar{y} = \frac{1}{10} \sum_{t=1}^{10} y_t.$$

Auch die Prognose zukünftiger Werte mittels dieser logistischen Funktion scheint recht gut zu sein. Für das Jahr 1981 ($t = 11$) beispielsweise beträgt die Differenz zwischen prognostizierter Einwohnerzahl

$$\hat{y}(11) = \hat{y}_{11} = 17,131$$

und tatsächlicher Einwohnerzahl ($y_{11} = 17,049$ lt. Jahrbuch 1983) nur 8200 Einwohner.

1.2.2. Die Mitscherlich-Funktion und die Gompertz-Kurve

Bei der längerfristigen Beschreibung von Ertrags- und Wachstumskurven findet die **Mitscherlich-Funktion**

$$y(t) = \alpha + \beta e^{\gamma t} \quad \text{mit } \gamma < 0$$

ihre Anwendung; sie strebt mit wachsendem t gegen die Sättigungsgrenze α und nimmt an der Stelle $t = 0$ den Wert $\alpha + \beta$ an. Da die Mitscherlich-Funktion bzgl. der unbekannten Parameter β und γ nicht linearisiert werden kann, muß zur Bestimmung der Kleinst-Quadrat-Schätzer für α , β und γ ein iteratives Verfahren verwendet werden.

Liegt eine Zeitreihe y_0, \dots, y_n vor und betrachtet man den Ansatz

$$y_t = \alpha + \beta e^{\gamma t} + \varepsilon_t \quad \text{für } t = 0, \dots, n,$$

wobei ε_t die irreguläre Komponente (den Fehler) bezeichnet, so sind also Schätzer $\hat{\alpha}$,

$\hat{\beta}$ und $\hat{\gamma}$ für α , β und γ gesucht, die die Funktion

$$S^2 = \sum_{t=0}^n (y_t - \alpha - \beta e^{\gamma t})^2$$

minimieren, d. h. die

$$\begin{aligned} \sum_{t=0}^n y_t - (n+1)\alpha - \hat{\beta} \sum_{t=0}^n e^{\hat{\gamma}t} &= 0, \\ \sum_{t=0}^n y_t e^{\hat{\gamma}t} - \alpha \sum_{t=0}^n e^{\hat{\gamma}t} - \hat{\beta} \sum_{t=0}^n e^{2\hat{\gamma}t} &= 0, \\ \sum_{t=0}^n y_t \cdot te^{\hat{\gamma}t} - \alpha \sum_{t=0}^n te^{\hat{\gamma}t} - \hat{\beta} \sum_{t=0}^n te^{2\hat{\gamma}t} &= 0 \end{aligned}$$

liefern (Nullstellen der partiellen Ableitungen).

Bestimmt man den Schätzer $\hat{\gamma}$ für γ iterativ aus der Gleichung

$$\begin{aligned} g(\hat{\gamma}) &= \left[\sum_{t=0}^n y_t e^{\hat{\gamma}t} - \frac{1}{n+1} \sum_{t=0}^n y_t \sum_{t=0}^n e^{\hat{\gamma}t} \right] \left[\sum_{t=0}^n te^{2\hat{\gamma}t} - \frac{1}{n+1} \sum_{t=0}^n e^{\hat{\gamma}t} \sum_{t=0}^n te^{\hat{\gamma}t} \right] \\ &\quad - \left[\sum_{t=0}^n e^{2\hat{\gamma}t} - \frac{1}{n+1} \left(\sum_{t=0}^n e^{\hat{\gamma}t} \right)^2 \right] \left[\sum_{t=0}^n y_t \cdot te^{\hat{\gamma}t} - \frac{1}{n+1} \sum_{t=0}^n y_t \sum_{t=0}^n te^{\hat{\gamma}t} \right] = 0, \end{aligned}$$

so ergeben sich die Schätzer für α und β zu

$$\begin{aligned} \hat{\alpha} &= \frac{1}{n+1} \sum_{t=0}^n y_t - \frac{\hat{\beta}}{n+1} \sum_{t=0}^n e^{\hat{\gamma}t} \quad \text{bzw.} \\ \hat{\beta} &= \frac{\sum_{t=0}^n y_t \cdot te^{\hat{\gamma}t} - \frac{1}{n+1} \sum_{t=0}^n y_t \sum_{t=0}^n te^{\hat{\gamma}t}}{\sum_{t=0}^n te^{2\hat{\gamma}t} - \frac{1}{n+1} \sum_{t=0}^n e^{\hat{\gamma}t} \sum_{t=0}^n te^{\hat{\gamma}t}}. \end{aligned}$$

Als Startwert für das iterative Verfahren zur Bestimmung von $\hat{\gamma}$ kann man nach der Methode von Verhagen

$$\gamma_0 = \frac{\sum_{t=1}^n (y_t - \bar{y})(z_t - \bar{z}) \sum_{t=1}^n (t - \bar{t})^2 - \sum_{t=1}^n (t - \bar{t})(z_t - \bar{z}) \sum_{t=1}^n (t - \bar{t})(y_t - \bar{y})}{\sum_{t=1}^n (z_t - \bar{z})^2 \sum_{t=1}^n (t - \bar{t})^2 - \left(\sum_{t=1}^n (t - \bar{t})(z_t - \bar{z}) \right)^2}$$

wählen, wobei

$$\bar{y} = \frac{1}{n+1} \sum_{t=0}^n y_t, \quad \bar{t} = \frac{1}{n+1} \sum_{t=0}^n t,$$

$$z_i = \frac{1}{2} \sum_{t=1}^i (y_{t-1} + y_t)(t - (t-1)) = \frac{1}{2} \sum_{t=1}^i (y_{t-1} + y_t) \quad \text{für } i = 1, \dots, n$$

$$\text{und } \bar{z} = \frac{1}{n} \sum_{t=1}^n z_t.$$

Wird anstelle der Zeit t hier ein beliebiger Regressor x_t verwandt, so muß zum Zeitpunkt $t = 0$ auch $x_t = 0$ gelten.

Als recht einfaches iteratives Verfahren zur Bestimmung von γ soll hier die „regula falsi“, vgl. etwa Stoer (1979), vorgestellt werden.

Zu einem Startwert w_0 sucht man ein v_0 , so daß gilt

$$g(w_0)g(v_0) < 0.$$

Beginnend mit $i = 0$ berechnet man dann im i -ten Schritt

$$\mu_i = w_i - g(w_i) \frac{w_i - v_i}{g(w_i) - g(v_i)}.$$

Ist $g(\mu_i) = 0$, so ist μ_i die gesuchte Nullstelle der Funktion g , ist $g(\mu_i) \neq 0$, so setzt man

$$(w_{i+1}, v_{i+1}) = \begin{cases} (\mu_i, v_i), & \text{falls } g(\mu_i)g(w_i) > 0 \\ (\mu_i, w_i), & \text{falls } g(\mu_i)g(w_i) < 0. \end{cases}$$

Beispiel: In Abschnitt 1.2.1 wurde das Wachstum der nordrhein-westfälischen Bevölkerung aufgrund der zweijährigen Bevölkerungszahlen von 1961–1979 mittels der logistischen Funktion geschätzt, wohingegen wir hier einmal bei gleicher Datenbasis das Bevölkerungswachstum mit Hilfe der Mitscherlich-Funktion

$$y(t) = \alpha + \beta e^{\gamma t}$$

schätzen wollen. Die iterative Bestimmung des Schätzers $\hat{\gamma}$ für γ soll mittels der regula falsi erfolgen, wobei der Startwert γ_0 nach der Methode von Verhagen verwandt wird, vgl. auch Tab. 3:

$$\begin{aligned} \gamma_0 &= \frac{83,326446 \cdot 60 - 1020,52 \cdot 4,922015}{17358 \cdot 311 \cdot 60 - (1020,52)^2} \\ &= \frac{-23,427988}{37,5896} = -0,6232572. \end{aligned}$$

Tab. 3: Schätzung der Bevölkerungsentwicklung in Nordrhein-Westfalen mit der Mitscherlich-Funktion (Daten y_t entnommen aus: Statistisches Jahrbuch Nordrhein-Westfalen 1983)

Jahr	t	Bevölkerung y_t (in Millionen)	Hilfsgrößen z_t	Schätzung \hat{y}_t (in Millionen)	$y_t - \hat{y}_t$
1961	0	15,920	–	15,8721	0,0479
1963	1	16,280	16,1000	16,3701	-0,0901
1965	2	16,661	32,5705	16,6752	-0,0142
1967	3	16,835	49,3185	16,8622	-0,0272
1969	4	17,044	66,2580	16,9767	0,0673
1971	5	17,091	83,3255	17,0469	0,0441
1973	6	17,223	100,4825	17,0899	0,1331
1975	7	17,176	117,6820	17,1162	0,0598
1977	8	17,052	134,7960	17,1324	-0,0804
1979	9	17,002	151,8230	17,1423	-0,1403
1981	10	–	–	17,1483	–
1983	11	–	–	17,1520	–
1985	12	–	–	17,1543	–
1987	13	–	–	17,1557	–

Weiter ist

$$\begin{aligned} g(\gamma_0) &= (35,007955 - 36,212084)(0,5662902 - 0,56262005) \\ &\quad - (1,4035136 - 0,4630431)(40,998725 - 41,15128) \\ &= 0,0952003 > 0, \end{aligned}$$

so daß man ein v_0 mit $g(v_0) < 0$ suchen muß. Für

$$v_0 = -0,5 \text{ ergibt sich } g(v_0) = 0,0094857$$

und für

$$v_0 = -0,4 \text{ erhalten wir } g(v_0) = -0,1008734.$$

Wir wählen daher $v_0 = -0,4$ und wissen auch bereits, daß $\hat{\gamma}$ zwischen $-0,4$ und $-0,5$ liegt. Daher wählen wir anstelle des Startwerts γ_0 direkt den Startwert

$$\tilde{\gamma}_0 = -0,5.$$

Es ist dann

$$\mu_0 = -0,4914047 \text{ und } g(\mu_0) = 0,0014408,$$

so daß wir wegen $g(\mu_0)g(\tilde{\gamma}_0) > 0$ nun

$$\gamma_1 = \mu_0 = -0,4914047 \text{ und } v_1 = v_0 = -0,4$$

setzen. Dann berechnen wir μ_1 usw., bis $g(\mu_i) = 0$ gilt.

In diesem Beispiel ergibt sich schon im zweiten Schritt, die Iterationsschritte 0–2 sind in Tab. 4 zusammengestellt, ein Wert $g(\mu_2) \simeq 0$, so daß bereits hier das Verfahren abgebrochen werden kann; in Abb. 4 ist die Funktion $g(w)$ für $w < 0$ auch graphisch dargestellt.

Wegen $g(\mu_2) \simeq 0$, vgl. Tab. 4, gilt ungefähr

$$\hat{\gamma} \simeq \mu_2 \simeq -0,4899,$$

und wir erhalten als Schätzwerte für die Parameter β und α gerade

$$\hat{\beta} = \frac{65,017396 - 16,8284 \cdot 3,8613892}{0,9612342 - 0,256265 \cdot 3,8613892} = -\frac{0,036394}{0,028305} = -1,2858$$

Tab. 4: Iterationsschritte zur Bestimmung des Schätzers $\hat{\gamma}$ für γ im Beispiel

Schritt i

0	$\tilde{\gamma}_0 = -0,5$ $v_0 = -0,4$ $\mu_0 = -0,4914047$	$g(\tilde{\gamma}_0) = 0,0094857$ $g(v_0) = -0,1008734$ $g(\mu_0) = 0,0014408$
1	$\gamma_1 = -0,4914047$ $v_1 = -0,4$ $\mu_1 = -0,4901175$	$g(\gamma_1) = 0,0014408$ $g(v_1) = -0,1008734$ $g(\mu_1) = 0,0002114$
2	$\gamma_2 = -0,4901175$ $v_2 = -0,4$ $\mu_2 = -0,4899291$	$g(\gamma_2) = 0,0002114$ $g(v_2) = -0,1008734$ $g(\mu_2) = 0,000003$

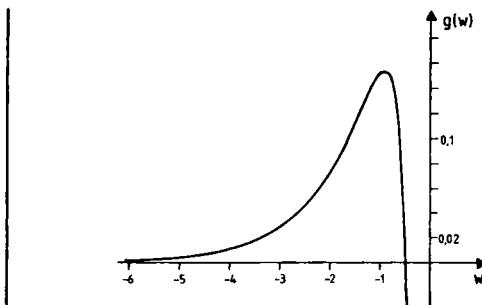


Abb. 4: Verlauf der Funktion $g(w)$ für $w < 0$

und

$$\alpha = 16,8284 - \frac{-1,2858}{10} \cdot 2,56265 = 17,1579.$$

Die Bevölkerungsentwicklung Nordrhein-Westfalens wird also durch die Mitscherlich-Funktion

$$\hat{y}(t) = \alpha + \beta e^{\gamma t} = 17,1579 - 1,2858 e^{-0,4899t}$$

geschätzt. Die Schätzwerte $\hat{y}(t) = \hat{y}_t$ für die Jahre 1961–1987 ($t = 0, \dots, 13$) wurden bereits in Tab. 3 mitangegeben; mit ihnen berechnet man das recht hohe Bestimmtheitsmaß

$$B = 1 - \frac{\sum_{t=0}^9 (y_t - \hat{y}_t)^2}{\sum_{t=0}^9 (y_t - \bar{y})^2} = 1 - \frac{0,0653}{1,6261} = 0,9598, \quad \bar{y} = \frac{1}{10} \sum_{t=0}^9 y_t.$$

In Abb. 5 ist der Verlauf der geschätzten Mitscherlich-Funktion auch graphisch veranschaulicht.

Neben logistischer Funktion und Mitscherlich-Funktion wird oft auch die **Gompertz-Kurve**

$$y(t) = e^{\alpha + \beta r^t}$$

mit den Parametern α , β und r ($0 < r < 1$) zur Beschreibung von Wachstumsverläufen verwandt. Dabei muß natürlich vorausgesetzt werden, daß man $y(t) > 0$ für alle t annehmen kann. Die Gompertz-Kurve wird oftmals auch für den Logarithmus von $y(t)$, also in der Form

$$\ln y(t) = \alpha + \beta r^t$$

angegeben. Man sieht, daß sie in dieser Darstellung gerade der Mitscherlich-Funktion mit

$$e^\gamma = r, \quad \gamma < 0, \quad 0 < r < 1$$

entspricht. Man kann die Parameter der Gompertz-Kurve also wie die der Mitscherlich-Funktion schätzen, wenn man anstelle von y_1, \dots, y_n die Werte

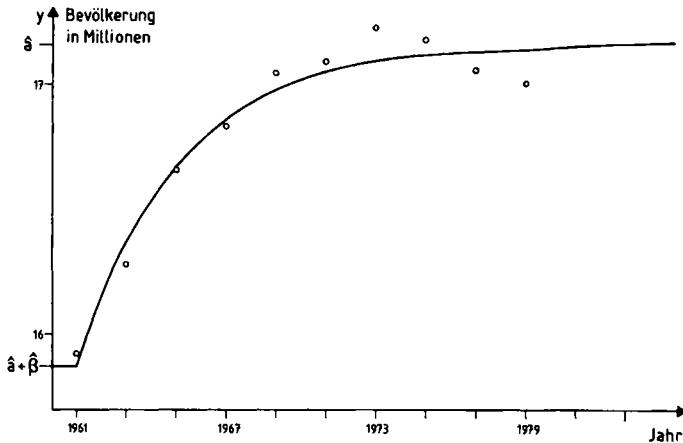


Abb. 5: Schätzung der Bevölkerungszahlen Nordrhein-Westfalens mit Hilfe der Mitscherlich-Funktion auf der Basis von 1961–1979

$\ln y_1, \dots, \ln y_n$ betrachtet. Die Parameterschätzungen $\hat{\alpha}$ und $\hat{\beta}$ sind natürlich direkt diejenigen für α bzw. β und der Parameter r wird geschätzt durch

$$\hat{r} = e^{\hat{\gamma}}.$$

Die Parameter der Mitscherlich-Funktion können auch mittels der in Abschnitt 1.2.1 beschriebenen Methoden geschätzt werden. Geht man nämlich bei der Mitscherlich-Funktion von

$$y(t) = \alpha + \beta e^{\gamma t}$$

über zu

$$y^*(t) = \frac{1}{y(t)} = \frac{1}{\alpha + \beta e^{\gamma t}} = \frac{1/\alpha}{1 + \beta/\alpha \cdot e^{\gamma t}} = \frac{\gamma^*}{1 + \beta^* e^{-\alpha^* t}},$$

so erhält man gerade eine logistische Funktion mit den Parametern

$$\alpha^* = -\gamma, \quad \beta^* = \beta/\alpha \quad \text{und} \quad \gamma^* = 1/\alpha.$$

Natürlich ist hier dann der Fehleransatz ein anderer. Bei der Schätzung der Parameter geht man ja über zu

$$\frac{1}{y^*(t)} = y(t) = \frac{1 - e^{-\alpha^*}}{\gamma^*} + e^{-\alpha^*} \frac{1}{y^*(t-1)} = \frac{1 - e^{-\alpha^*}}{\gamma^*} + e^{-\alpha^*} y(t-1)$$

und unterstellt für $y(t)$ einen AR(1)-Fehlerprozeß (ε_t), vgl. Abschnitt 2.1, mit Parameter $-e^{-\alpha^*}$ anstelle eines weißen Rauschens bei Verwendung der in diesem Abschnitt vorgestellten Schätzmethoden.

Umgekehrt können für $\alpha > 0$ die Parameter der logistischen Funktion

$$y(t) = \gamma/(1 + \beta e^{-\alpha t})$$

mit den hier beschriebenen Methoden geschätzt werden. Geht man über zu

$$y^*(t) = \frac{1}{y(t)} = \frac{1 + \beta e^{-\alpha t}}{\gamma} = \frac{1}{\gamma} + \frac{\beta}{\gamma} e^{-\alpha t} = \alpha^* + \beta^* e^{\gamma^* t},$$

so hat man für $y^*(t)$ gerade eine Mitscherlich-Funktion mit Parametern

$$\alpha^* = 1/\gamma, \quad \beta^* = \beta/\gamma \quad \text{und} \quad \gamma^* = -\alpha.$$

An die Stelle des AR(1)-Fehlerprozesses für $y^*(t) = 1/y(t)$ in Abschnitt 1.2.1 bei der Schätzung der Parameter der logistischen Funktion tritt dann natürlich ein weißes Rauschen.

1.2.3. Die allometrische Funktion

Sowohl im biologischen als auch im ökonomischen Bereich wird die **allometrische Funktion**

$$y(t) = \beta t^\alpha$$

zur Beschreibung von Wachstumsverläufen verwandt. Tritt eine Funktion $x(t)$ der Zeit an die Stelle von t , also

$$y(t) = \beta x(t)^\alpha,$$

so ist die allometrische Funktion gerade eine spezielle **Cobb-Douglas-Produktionsfunktion**, die das Wachstum von $y(t)$ und $x(t)$ relativ zueinander beschreibt; man vgl. hierzu auch das nachfolgende Beispiel.

Die allometrische Funktion ist zwar durch den Übergang zu Logarithmen linearisierbar in der Form (für $y(t) > 0$)

$$\ln y(t) = \ln \beta + \alpha \ln t,$$

vgl. Abschnitt 3 in Kap. X, jedoch liefert bei gegebener, beobachteter Zeitreihe y_1, \dots, y_n die direkte Minimierung von

$$S^2 = \sum_{t=1}^n (y_t - \beta t^\alpha)^2$$

oft bessere Resultate. Beim Linearisieren geht man von einer multiplikativen irrekulären Komponente ε_t aus, d.h.

$$y_t = \beta t^\alpha \cdot \varepsilon_t \quad \text{für } t = 1, \dots, n,$$

wohingegen bei der direkten Minimierung der Fehlerterm additiv angenommen wird:

$$y_t = \beta t^\alpha + \varepsilon_t \quad \text{für } t = 1, \dots, n.$$

Beim multiplikativen Fehleransatz führt man natürlich, wie im Abschnitt 1 des Kap. X beschrieben, eine lineare Regression von $\ln y(t)$ auf $\ln t$ mit Steigungsparameter α und Absolutglied $\ln \beta$ durch. Dagegen ist beim additiven Fehleransatz eine numerische Minimierung von S^2 vonnöten. Hierbei kann man z.B. so vorgehen, daß man den Schätzer $\hat{\alpha}$ für α iterativ als Lösung der Gleichung

$$g(\hat{\alpha}) = \sum_{t=1}^n y_t t^{\hat{\alpha}} \ln t \cdot \sum_{t=1}^n t^{2\hat{\alpha}} - \sum_{t=1}^n t^{2\hat{\alpha}} \ln t \cdot \sum_{t=1}^n y_t t^{\hat{\alpha}} = 0$$

bestimmt und dann

$$\hat{\beta} = \frac{\sum_{t=1}^n y_t t^{\hat{\alpha}}}{\sum_{t=1}^n t^{2\hat{\alpha}}}$$

als Schätzer für den Parameter β berechnet. Als Startwert für das Iterationsverfah-

ren kann man den Startwert α_0 für α wählen, der sich bei Verwendung der linearisierten Funktion ergibt.

Beispiel: In der Abb. 3 des Kapitels X ist eine Regressionsgerade für die Regression vom durchschnittlichen Nettoeinkommen auf das durchschnittliche Bruttoeinkommen in der Bundesrepublik Deutschland auf der Datenbasis der Jahre 1960–1970 (Daten: Statistisches Bundesamt) dargestellt. Man erkennt dort, daß eigentlich eine nichtlineare Regressionsfunktion angebracht wäre.

Anhand dieses Beispiels wollen wir einmal demonstrieren, daß sich (nichtlineare) Trendfunktionen auch als gewöhnliche Regressionsfunktionen eignen, und mittels allometrischer Funktion eine Regression vom durchschnittlichen Nettoeinkommen auf das durchschnittliche Bruttoeinkommen durchführen.

In der Tab. 5 sind für die Jahre 1960–1970 ($t = 0, \dots, 10$) Nettoeinkommen $y(t) = y_t$ und Bruttoeinkommen $x(t) = x_t$ angegeben. Ausgehend von diesen Daten wollen wir zum einen die Parameter bei Linearisierung und zum anderen bei direkter Minimierung schätzen. Dabei wollen wir die Beobachtungen y_0 und x_0 im Jahr 1960 als Standard ansehen, d. h. wir betrachten anstelle von y_t und x_t die Größen

$$\tilde{y}_t = y_t - y_0 \quad \text{und} \quad \tilde{x}_t = x_t - x_0 \quad \text{für } t = 1, \dots, 10,$$

die ebenfalls in Tab. 5 angegeben sind, bzw. die Funktion

$$\tilde{y}(t) = \beta \tilde{x}(t)^\alpha \quad \text{mit} \quad \tilde{y}(t) = y_t - y_0, \quad \tilde{x}(t) = x_t - x_0 \quad \text{für alle } t.$$

Tab. 5: Durchschnittliche Brutto- und Nettoeinkommen in der Bundesrepublik Deutschland von 1960 bis 1970 (in Tausend DM)

Jahr	t	Brutto x_t	Netto y_t	$\tilde{x}_t = x_t - x_0$	$\tilde{y}_t = y_t - y_0$
1960	0	6,148	5,178	—	—
1961	1	6,775	5,664	0,627	0,486
1962	2	7,395	6,151	1,247	0,973
1963	3	7,850	6,501	1,702	1,323
1964	4	8,556	7,045	2,408	1,867
1965	5	9,336	7,746	3,188	2,568
1966	6	10,014	8,200	3,866	3,022
1967	7	10,349	8,437	4,201	3,259
1968	8	10,988	8,841	4,840	3,663
1969	9	12,003	9,499	5,855	4,321
1970	10	13,773	10,660	7,625	5,482

Mittels linearer Regression von $\ln \tilde{y}(t)$ auf $\ln \tilde{x}(t)$ ergibt sich als Schätzer für α zunächst mit

$$\overline{\ln \tilde{x}} = \frac{1}{10} \sum_{t=1}^{10} \ln \tilde{x}_t, \quad \overline{\ln \tilde{y}} = \frac{1}{10} \sum_{t=1}^{10} \ln \tilde{y}_t$$

der Wert

$$\hat{\alpha} = \frac{\sum_{t=1}^{10} (\ln \tilde{x}_t - \overline{\ln \tilde{x}})(\ln \tilde{y}_t - \overline{\ln \tilde{y}})}{\sum_{t=1}^{10} (\ln \tilde{x}_t - \overline{\ln \tilde{x}})^2} = 0,9770,$$

und somit erhalten wir als Schätzer für $\ln \beta$

$$\ln \hat{\beta} = \overline{\ln \tilde{y}} - \alpha \overline{\ln \tilde{x}} = -0,2397,$$

d.h. der Schätzer für den Parameter β selbst ist

$$\hat{\beta} = e^{-0,2397} = 0,7869.$$

Bei einem auf das Jahr 1960 standardisierten Bruttolohn $\tilde{x}(t)$ erhält man also als Schätzung für den ebenso standardisierten Nettolohn

$$\hat{y}(t) = 0,7869 \tilde{x}(t)^{0,9770},$$

und die Schätzung für den tatsächlichen Nettolohn $\tilde{y}(t)$ ergibt sich zu

$$\hat{y}(t) = \hat{y}(t) + y_0 = \hat{y}(t) + 5,178.$$

In der Rubrik „Linearisieren der allometrischen Funktion“ in Tab. 6 sind die Schätzungen $\hat{y}(t) = \hat{y}_t$ und $\hat{y}(t) = \hat{y}_t$ für die Jahre 1961–1970 ($t = 1, \dots, 10$) sowie die Differenzen

$$d_t = \tilde{y}_t - \hat{y}_t = y_t - \hat{y}_t$$

angegeben.

Zum Vergleich wollen wir die Parameter α und β nun direkt in einem Regressionsmodell mit additivem Fehleransatz schätzen. Zunächst schätzen wir hier den Parameter α durch iterative Lösung der Gleichung

$$g(\alpha) = \sum_{t=1}^{10} \tilde{y}_t \tilde{x}_t^{\hat{\alpha}} \ln \tilde{x}_t \cdot \sum_{t=1}^{10} \tilde{x}_t^{2\hat{\alpha}} - \sum_{t=1}^{10} \tilde{x}_t^{2\hat{\alpha}} \ln \tilde{x}_t \cdot \sum_{t=1}^{10} \tilde{y}_t \tilde{x}_t^{\hat{\alpha}} = 0$$

mittels regula falsi, vgl. Abschnitt 1.2.2, wobei wir als Startwerte

Tab. 6: Regression vom durchschnittlichen Nettoeinkommen auf das durchschnittliche Bruttoeinkommen in der Bundesrepublik Deutschland mittels allometrischer Funktion (Angaben in Tausend DM)

Jahr	t	Linearisieren der allometrischen Funktion			direkte Minimierung der quadratischen Abweichungen		
		\hat{y}_t	$\hat{y}_t = \hat{y}_t + y_0$	$d_t = y_t - \hat{y}_t$	\hat{y}_t	$\hat{y}_t = \hat{y}_t + y_0$	$d_t = y_t - \hat{y}_t$
1961	1	0,499	5,677	-0,013	0,552	5,730	-0,066
1962	2	0,976	6,154	-0,003	1,042	6,220	-0,069
1963	3	1,323	6,501	0,000	1,388	6,566	-0,065
1964	4	1,857	7,035	0,010	1,912	7,090	-0,045
1965	5	2,443	7,621	0,125	2,477	7,655	0,091
1966	6	2,949	8,127	0,073	2,960	8,138	0,062
1967	7	3,198	8,376	0,061	3,196	8,374	0,063
1968	8	3,673	8,851	-0,010	3,642	8,820	0,021
1969	9	4,424	9,602	-0,103	4,341	9,519	-0,020
1970	10	5,726	10,904	-0,244	5,540	10,718	-0,058

$\alpha_0 = 0,93$ und $v_0 = 0,92$ wählen. Da gilt

$$\begin{aligned} g(\alpha_0) &= g(0,93) \\ &= 184,0840930 \cdot 134,4888652 - 219,3352772 \cdot 112,9564904 \\ &= -18,08236000, \end{aligned}$$

$$\begin{aligned} g(v_0) &= g(0,92) \\ &= 180,9319982 \cdot 130,1772045 - 211,8752269 \cdot 111,1314568 \\ &= 7,21910000 \end{aligned}$$

und somit

$$g(\alpha_0)g(v_0) < 0,$$

sind α_0 und v_0 zulässige Startwerte, und wir können μ_0 berechnen:

$$\mu_0 = \alpha_0 - g(\alpha_0) \frac{\alpha_0 - v_0}{g(\alpha_0) - g(v_0)} = 0,922853235,$$

d.h.

$$g(\mu_0) = 0,207270000.$$

Damit ist dann

$$\begin{aligned} \alpha_1 &= \mu_0, \quad g(\alpha_1) > 0 \quad \text{und wegen} \quad g(\alpha_0)g(\mu_0) < 0 \\ v_1 &= \alpha_0, \quad g(v_1) = g(\alpha_0) < 0, \end{aligned}$$

so daß sich

$$\mu_1 = 0,922934226$$

ergibt. Diese Resultate sowie die weiteren Iterationsschritte sind in Tab. 7 zusammengestellt.

Tab. 7: Iterative Bestimmung des Schätzwertes α mit der regula falsi

Schritt		
0	$\alpha_0 = 0,93$ $v_0 = 0,92$ $\mu_0 = 0,922853235$	$g(\alpha_0) = -18,08236000$ $g(v_0) = 7,21910000$ $g(\mu_0) = 0,20727000$
1	$\alpha_1 = \mu_0$ $v_1 = \alpha_0$ $\mu_1 = 0,922934226$	$g(\alpha_1) = g(\mu_0)$ $g(v_1) = g(\alpha_0)$ $g(\mu_1) = 0,00583000$
2	$\alpha_2 = \mu_1$ $v_2 = \alpha_0$ $\mu_2 = 0,922936504$	$g(\alpha_2) = g(\mu_1)$ $g(v_2) = g(\alpha_0)$ $g(\mu_2) = 0,00017000$
3	$\alpha_3 = \mu_2$ $v_3 = \alpha_0$ $\mu_3 = 0,922936570$	$g(\alpha_3) = g(\mu_2)$ $g(v_3) = g(\alpha_0)$ $g(\mu_3) = 0$

Im dritten Schritt erhalten wir den Wert μ_3 mit $g(\mu_3) = 0$, so daß sich als Schätzer für den Parameter α der Wert

$$\alpha = \mu_3 = 0,922936570$$

und damit als Schätzer für den Parameter β der allometrischen Funktion

$$\hat{\beta} = \frac{\sum_{t=1}^{10} \hat{y}_t \hat{x}_t^2}{\sum_{t=1}^{10} \hat{x}_t^{2\hat{\alpha}}} = \frac{111,6641257}{131,4279186} = 0,849622568$$

ergibt.

Zu einem vorgegebenen standardisierten Bruttoeinkommen $\tilde{x}(t)$ ergibt sich somit als Schätzung für das standardisierte Nettoeinkommen

$$\hat{y}(t) = 0,849622568 \cdot \tilde{x}(t)^{0,922936570};$$

das entsprechende nichtstandardisierte Nettoeinkommen wird natürlich durch

$$\hat{y}(t) = \hat{y}(t) + y_0 = \hat{y}(t) + 5,178$$

geschätzt. In der Tab. 6 wurden bereits für die Jahre 1961–1970 ($t = 1, \dots, 10$) die resultierenden Größen \hat{y}_t , \hat{y}_t sowie die Differenzen

$$d_t = \tilde{y}_t - \hat{y}_t = y_t - \hat{y}_t \quad \text{für } t = 1, \dots, 10$$

angegeben.

Berechnet man ausgehend von Tab. 5 und Tab. 6 die Bestimmtheitsmaße der allometrischen Funktionen, so ergibt sich bei Linearisierung

$$B_L = 0,995712253$$

und bei direkter Minimierung

$$B_D = 0,998393593.$$

Wir sehen, daß beide Anpassungen sehr gut sind, sich der hohe Wert von ca. 0,9957 bei Linearisierung jedoch bei direkter Minimierung noch auf ca. 0,9984 erhöhen lässt.

Betrachten wir nun einmal die allometrische Funktion

$$\hat{y}(t) = 5,178 + 0,849622568 \cdot \tilde{x}(t)^{0,922936570},$$

die in Abb. 6 auch graphisch dargestellt ist. Trotz der wenigen Daten und unter

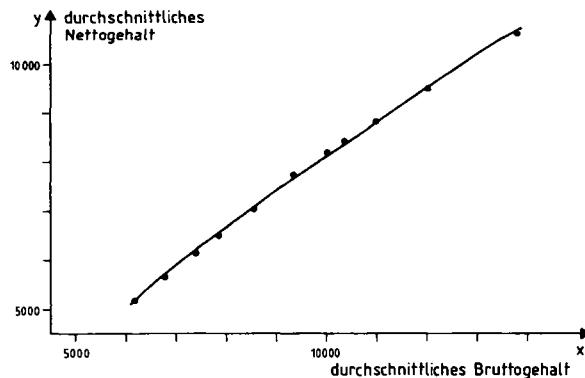


Abb. 6: Regression vom durchschnittlichen Nettogehalt auf das durchschnittliche Bruttogehalt in der Bundesrepublik Deutschland mittels allometrischer Funktion auf der Basis von 1960–1970

Ausschluß der Berücksichtigung von Steuergesetzen kann man aus dieser Regressionsfunktion z. B. folgende Schlüsse ziehen.

Steigt das auf das Jahr 1960 standardisierte Bruttogehalt um das y -fache, so steigt das standardisierte Nettogehalt nur um das $y^{0.922936570}$ -fache, d. h. z. B. bei zehnprozentiger Steigerung des standardisierten Bruttogehaltes steigt das standardisierte Nettogehalt nur um das

$$1,1^{0.922936570} = 1,091950177\text{-fache}$$

oder um ca. 9,2%.

Außerdem läßt sich eine grobe „Grenznutzen-Betrachtung“ folgender Art durchführen. Hätte man z. B. Brutto 1000 DM mehr verdient, so wäre dies nach der Funktion ein Gewinn von Netto

$$[\hat{\beta}(\tilde{x}_t + 1)^{\tilde{a}} - \hat{y}_t] \cdot 1000 \text{ DM im Jahre 1960} + t,$$

also von 779 DM im Jahre 1961, von 752 DM im Jahre 1962, von 738 DM im Jahre 1963, von 722 DM im Jahre 1964, von 709 DM im Jahre 1965, von 700 DM im Jahre 1966, von 696 DM im Jahre 1967, von 689 DM im Jahre 1968, von 680 DM im Jahre 1969 und von 667 DM im Jahre 1970. In diesen 10 Jahren ist nach der Funktion bei den 1000 DM Brutto also ein „Gewinnverlust“ von 779 DM – 667 DM = 112 DM eingetreten.

Zur Überprüfung der kurzfristigen Prognosefähigkeit sollen nun noch geschätzte und wahre durchschnittliche Nettogehälter in den Jahren 1971 bis 1974 gegenübergestellt werden. In Tab. 8 sind wahre durchschnittliche Brutto- und Nettogehälter y_t , x_t und geschätzte Bruttogehälter \hat{y}_t für $t = 11, \dots, 14$ bei linearer Regressionsrechnung

$$\hat{y}_t = 0,806985 + 0,727 \cdot x_t,$$

vgl. Abb. 3 in der Einleitung von Kap. X, bei Linearisierung der allometrischen Funktion

$$\hat{y}_t = 0,7869 \cdot (x_t - 6,148)^{0.9770} + 5,178$$

sowie bei direkter Minimierung mit allometrischer Funktion

$$\hat{y}_t = 0,849622568 \cdot (x_t - 6,148)^{0.922936570} + 5,178$$

angegeben.

Tab. 8: Wahre durchschnittliche Brutto- und Nettogehälter in den Jahren 1971–1974 sowie Prognosen der durchschnittlichen Bruttogehälter (in Tausend DM)

Jahr	t	Brutto x_t	Netto y_t	\hat{y}_t bei linearer Regression	\hat{y}_t bei Linearisierung der allometrischen Funktion	\hat{y}_t bei direkter Minimierung mittels allometrischer Funktion
1971	11	15,392	11,698	11,997	12,089	11,795
1972	12	16,770	12,750	12,999	13,094	12,700
1973	13	18,782	13,804	14,461	14,556	14,006
1974	14	20,928	15,170	16,022	16,110	15,382

Man erkennt deutlich, daß bei allen Regressionsfunktionen die Entwicklung der Nettogehälter zu optimistisch prognostiziert wird. Die besten Prognosen ergeben sich bei direkter Minimierung mit allometrischer Funktion und auch die Prognosen bei einfacher linearer Regressionsrechnung, bei der ein Bestimmtheitsmaß von 0,96681 erreicht wird, sind in diesem Zeitraum besser als die bei Linearisierung der allometrischen Funktion.

1.3. Trend- und Saison-Schätzung bzw. -Elimination durch Glättung bzw. Filterung

Wir betrachten hier wiederum das additive Zeitreihenmodell aus Abschnitt 1.1 und lassen im Gegensatz zu den Trendmodellen in Abschnitt 1.2 zu, daß eine Saisonkomponente vorhanden ist; allerdings unterstellen wir, wie bereits in Abschnitt 1.1 erwähnt, daß eine beobachtete Zeitreihe y_1, \dots, y_n eine konstante Saisonfigur hat. Wir werden hier verschiedene spezielle **Glättungsverfahren** für solche Zeitreihen behandeln, die es ermöglichen, Trend und Saison zu schätzen bzw. zu eliminieren und zukünftige Beobachtungswerte zu prognostizieren. Außerdem werden wir eine allgemeine Klasse von Glättungsverfahren – die **linearen Filter** – behandeln, in die sich die speziellen Glättungsverfahren einordnen lassen.

1.3.1. Gleitende Durchschnitte

Mit Hilfe **gleitender Durchschnitte (moving averages)** kann bei Zeitreihen y_1, \dots, y_n ohne saisonale Komponente bzw. mit konstanter Saisonfigur die glatte Komponente G_t ermittelt werden.

Unter einem **gleitenden Durchschnitt der (ungeraden) Ordnung $2k + 1$** für den Wert y_t einer Zeitreihe versteht man das arithmetische Mittel aus $y_{t-k}, \dots, y_t, \dots, y_{t+k}$, mithin den Ausdruck

$$y_t^* = \frac{1}{2k+1} \sum_{j=-k}^k y_{t+j},$$

der für $t = k+1, k+2, \dots, n-k$ berechnet werden kann. Somit hat die zu einer Zeitreihe y_1, \dots, y_n der Länge n gehörige Reihe der gleitenden Durchschnitte $y_{k+1}^*, \dots, y_{n-k}^*$ die Länge $n-2k$.

Ein **gleitender Durchschnitt der (geraden) Ordnung $2k$** zum Wert y_t einer Zeitreihe hingegen ist gegeben durch

$$y_t^* = \frac{1}{2k} \left(\frac{1}{2} y_{t-k} + \sum_{j=-(k-1)}^{k-1} y_{t+j} + \frac{1}{2} y_{t+k} \right).$$

Auch dieser gleitende Durchschnitt kann für $t = k+1, \dots, n-k$ berechnet werden, denn wie beim gleitenden Durchschnitt der Ordnung $2k+1$ werden die Werte $y_{t-k}, \dots, y_t, \dots, y_{t+k}$ bei der Bildung von y_t^* berücksichtigt, der erste und letzte Wert allerdings jeweils mit dem Gewicht $\frac{1}{4k}$, die übrigen mit dem Gewicht $\frac{1}{2k}$.

Bei der Berechnung gleitender Durchschnitte lassen sich im Falle einer ungeraden Ordnung $2k+1$ bzw. im Falle einer geraden Ordnung $2k$ folgende **Rekursionsformeln** anwenden:

$$y_{t+1}^* = y_t^* + \frac{1}{2k+1} (y_{t+k+1} - y_{t-k})$$

bzw.

$$y_{t+1}^* = y_t^* + \frac{1}{4k} (y_{t+k+1} + y_{t+k} - y_{t-k} - y_{t-k+1}).$$

Bei einer **Zeitreihe ohne saisonale Komponente**, d. h.

$$y_t = G_t + R_t,$$

gilt bei Übergang zu gleitenden Durchschnitten

$$y_t^* = G_t^* + R_t^*,$$

und da die Werte der irregulären Komponente R_t um Null schwanken, kann man annehmen, daß das Mittel R_t^* solcher Werte in der Nähe von Null liegt.

So gilt

$$y_t^* \approx G_t^*$$

und die glatte Komponente G_t kann durch G_t^* und somit durch die gleitenden Durchschnitte y_t^* approximiert werden. Die Approximation ist umso besser, je mehr sich G_t im Intervall $[t-k, t+k]$ einer linearen Form nähert.

Bei einer **Zeitreihe mit saisonaler Komponente**

$$y_t = G_t + S_t + R_t$$

gilt für die gleitenden Durchschnitte

$$y_t^* = G_t^* + S_t^* + R_t^*,$$

d. h. y_t^* kann dann zur Schätzung der glatten Komponente G_t verwendet werden, wenn das Mittel S_t^* der Werte der Saisonkomponente S_t verschwindet. Liegt eine **konstante Saisonalfigur** mit Periode p vor, d. h. gilt $S_t = S_{t+p}$ für $t = 1, \dots, n-p$, so läßt sich S_t für $t = 1, \dots, p$ so normieren, daß gilt

$$S_t^* = \frac{1}{p+1} \sum_{j=-(p-1)/2}^{(p-1)/2} S_{t+j} = 0, \quad \text{falls } p \text{ ungerade, bzw.}$$

$$S_t^* = \frac{1}{p} \left(\frac{1}{2} S_{t-p/2} + \sum_{j=-(p/2-1)}^{p/2-1} S_{t+j} + \frac{1}{2} S_{t+p/2} \right) = 0, \quad \text{falls } p \text{ gerade.}$$

Bildet man also gleitende Durchschnitte y_t^* der Ordnung p (oder der Ordnung $h \cdot p$, wobei h eine ganze Zahl ist), so verschwindet S_t^* , und die glatte Komponente G_t kann wiederum durch y_t^* approximiert werden.

Bei Zeitreihen ohne saisonale Schwankungen stellt sich die Frage, wie groß man die **Ordnung der gleitenden Durchschnitte** wählen soll. Da durch gleitende Durchschnitte starke Krümmungen der glatten Komponente abgeschliffen werden, und zwar umso stärker, je höher die Ordnung, sollte die Ordnung bei stark gekrümmtem Verlauf der glatten Komponente nicht allzu groß sein. Andererseits ist darauf zu achten, daß bei kleiner Ordnung der Wert R_t^* weit von Null abweichen kann und zwar insbesondere bei starken irregulären Schwankungen.

Beispiel: Für die Zeitreihe y_1, \dots, y_{51} , „Anzahl der Arbeitslosen im Baugewerbe am Monatsende von Juli 1975 bis September 1979“ (vgl. „Wirtschaft und Statistik“, Hrsg. Statistisches Bundesamt Wiesbaden (1976–1980), Kohlhammer Verlag, Stuttgart/Mainz), vgl. Tab. 9, soll die glatte Komponente geschätzt werden. Aufgrund der Zeitreihe, die in Abb. 7 auch graphisch dargestellt ist, können wir von

Tab. 9: Zahl der Arbeitslosen im Baugewerbe y_t , Schätzer y_t^* für die glatte Komponente G_t , saisonbereinigte Werte $y_t - \hat{s}_t$ und Hilfsgrößen Σ_t für den Test auf Trend

Monat		t	y_t	y_t^*	$y_t - \hat{s}_t$	$\Sigma_t = \sum_{j=t+1}^{51} \text{sgn}(y_j - y_t)$
Juli	1975	1	60572	—	—	-22
Aug.	1975	2	52461	—	—	-15
Sept.	1975	3	47357	—	—	-10
Okt.	1975	4	48320	—	—	-13
Nov.	1975	5	60219	—	—	-18
Dez.	1975	6	84418	—	—	-27
Jan.	1976	7	119916	66715	66780	-42
Febr.	1976	8	124350	64421	68855	-43
März	1976	9	87309	62541	67343	-30
April	1976	10	57035	60883	59791	-21
Mai	1976	11	39903	59203	54637	- 8
Juni	1976	12	34053	57508	54093	- 3
Juli	1976	13	29905	56318	52275	+ 6
Aug.	1976	14	28068	55293	51175	+15
Sept.	1976	15	26634	53992	50473	+18
Okt.	1976	16	29259	53226	49872	+ 9
Nov.	1976	17	38942	53242	50965	- 2
Dez.	1976	18	65036	53496	54155	-17
Jan.	1977	19	110728	53754	57592	-32
Febr.	1977	20	108931	53997	53436	-31
März	1977	21	71517	54197	51551	-20
April	1977	22	54428	54386	57184	-17
Mai	1977	23	42911	54591	57645	-10
Juni	1977	24	37123	54639	57163	- 7
Juli	1977	25	33044	54102	55414	- 4
Aug.	1977	26	30755	53425	53862	- 1
Sept.	1977	27	28742	53388	52581	0
Okt.	1977	28	31698	53095	52311	- 1
Nov.	1977	29	41427	52273	53450	- 4
Dez.	1977	30	63685	51472	52804	-11
Jan.	1978	31	99189	50720	46053	-18
Febr.	1978	32	104240	50138	48745	-19
März	1978	33	75304	49626	55338	-14
April	1978	34	43622	49051	46378	- 9
Mai	1978	35	33990	48179	48724	- 8
Juni	1978	36	26819	46935	46859	- 3
Juli	1978	37	25291	45896	47661	- 2
Aug.	1978	38	24538	44931	47645	- 1
Sept.	1978	39	22685	43163	46524	+ 2
Okt.	1978	40	23945	41385	44558	+ 1
Nov.	1978	41	28245	40134	40268	0
Dez.	1978	42	47017	39094	36136	- 3
Jan.	1979	43	90920	38309	37784	- 8
Febr.	1979	44	89340	37614	33845	- 7
März	1979	45	47792	36984	27826	- 6
April	1979	46	28448	—	—	- 5
Mai	1979	47	19139	—	—	- 4
Juni	1979	48	16728	—	—	- 3
Juli	1979	49	16523	—	—	0
Aug.	1979	50	16622	—	—	- 1
Sept.	1979	51	15499	—	—	-

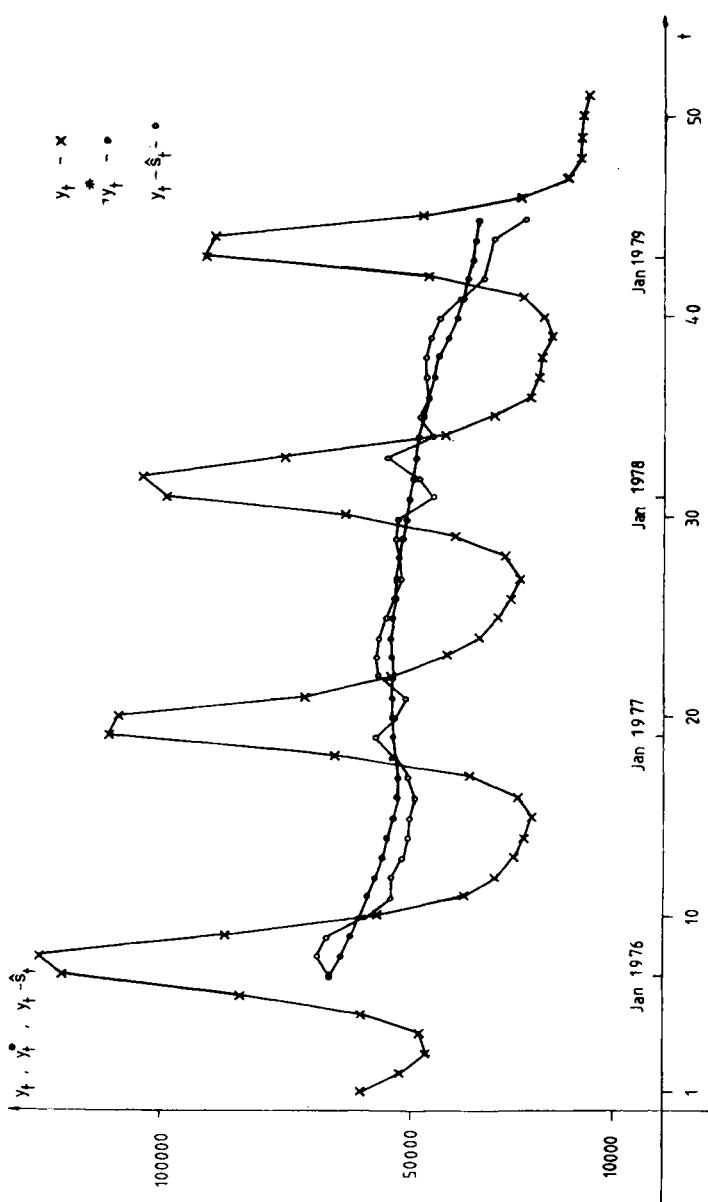


Abb. 7: Zeitreihe der Arbeitslosen im Baugewerbe y_t am Monatsende in den Monaten Mai 1975 bis September 1979, gleitende Durchschnitte y_t^* der Ordnung 12 und saisonbereinigte Zeitreihe $y_t - \hat{s}_t$

einer konstanten Saisonfigur mit 12-monatiger Periode ausgehen, so daß wir die glatte Komponente G_t durch gleitende Durchschnitte

$$y_t^* = \frac{1}{12} \left(\frac{1}{2} y_{t-6} + \sum_{j=-5}^5 y_{t+j} + \frac{1}{2} y_{t+6} \right) \quad \text{für } t = 7, \dots, 45$$

der geraden Ordnung 12 schätzen können; diese Werte y_t^* sind ebenfalls in Tab. 9 angegeben und in Abb. 7 veranschaulicht. Auf die weiteren in Tab. 9 bzw. Abb. 7 angegebenen Größen werden wir später eingehen.

Ein Nachteil bei der Berechnung gleitender Durchschnitte ist, daß die letzten Werte der glatten Komponente nicht geschätzt werden können. Daher soll nun eine Methode vorgestellt werden, mittels derer die **Reihe der gleitenden Durchschnitte y_t^* aufgefüllt** werden kann; man vgl. aber auch Abschnitt 3.6.

Im Falle gleitender Durchschnitte der ungeraden Ordnung $2k+1$ werden bei dieser Methode zur Bestimmung von $y_{n-k+1}^*, \dots, y_n^*$ stets die letzten $2k+1$ Glieder y_{n-2k}, \dots, y_n der Zeitreihe bei verschiedener Gewichtung gemittelt:

$$y_{n-k+r}^* = \sum_{j=-k}^k \left(\frac{1}{2k+1} + \frac{12rj}{2k(2k+1)(2k+2)} \right) y_{n-k+j} \quad \text{für } r = 1, \dots, k.$$

Mit zunehmendem r werden die Gewichte

$$\frac{1}{2k+1} + \frac{12rj}{2k(2k+1)(2k+2)}, \quad j = -k, \dots, k,$$

der letzten Glieder immer größer, während die ersten Glieder immer kleinere und schließlich negative Gewichte erhalten.

Beispiel: Für $2k+1 = 5$, d.h. $k = 2$, ergibt sich etwa die folgende Gewichtung:

$$\begin{aligned} r = 1: \quad & \text{Gewicht für } y_{n-4} = \frac{1}{5} - \frac{2}{10} = 0 \\ & \text{Gewicht für } y_{n-3} = \frac{1}{5} - \frac{1}{10} = \frac{1}{10} = 0.1 \\ & \text{Gewicht für } y_{n-2} = \frac{1}{5} + 0 = \frac{1}{5} = 0.2 \\ & \text{Gewicht für } y_{n-1} = \frac{1}{5} + \frac{1}{10} = \frac{3}{10} = 0.3 \\ & \text{Gewicht für } y_n = \frac{1}{5} + \frac{2}{10} = \frac{2}{5} = 0.4 \\ r = 2: \quad & \text{Gewicht für } y_{n-4} = \frac{1}{5} - \frac{2}{5} = -\frac{1}{5} = -0.2 \\ & \text{Gewicht für } y_{n-3} = \frac{1}{5} - \frac{1}{5} = 0 \\ & \text{Gewicht für } y_{n-2} = \frac{1}{5} + 0 = \frac{1}{5} = 0.2 \\ & \text{Gewicht für } y_{n-1} = \frac{1}{5} + \frac{1}{5} = \frac{2}{5} = 0.4 \\ & \text{Gewicht für } y_n = \frac{1}{5} + \frac{2}{5} = \frac{3}{5} = 0.6, \end{aligned}$$

d.h. es ist

$$\begin{aligned} y_{n-1}^* &= 0 \cdot y_{n-4} + 0.1 y_{n-3} + 0.2 y_{n-2} + 0.3 y_{n-1} + 0.4 y_n \\ y_n^* &= -0.2 y_{n-4} + 0 \cdot y_{n-3} + 0.2 y_{n-2} + 0.4 y_{n-1} + 0.6 y_n. \end{aligned}$$

Ein weiteres Problem, auf das man bei Übergang zu gleitenden Durchschnitten stößt, ist der sogenannte **Slutzky-Yule-Effekt**, vgl. Slutsky (1937). Selbst wenn die irregulären Komponenten R_t einer Zeitreihe unabhängig identisch normalverteilt sind, geht bei Übergang zu R_t^* nicht nur die Unabhängigkeit verloren; es können darüberhinaus vielmehr starke zyklische Effekte auftreten, die eine Analyse der gleitenden Durchschnitte y_t^* unmöglich machen.

Wie wir zuvor gesehen haben, kann die glatte Komponente G_t einer Zeitreihe y_1, \dots, y_n mit konstanter Saisonfigur der Periode p mittels gleitender Durchschnitte der Ordnung p geschätzt werden. Die Differenzen

$$d_t = y_t - y_t^*$$

setzen sich dann näherungsweise aus saisonaler und irregulärer Komponente zusammen, d.h.

$$d_t \approx S_t + R_t.$$

Wegen der Konstanz der Saisonfigur wissen wir weiterhin, daß für $t = 1, \dots, p$ die Größen $d_t, d_{t+p}, d_{t+2p}, \dots$ jeweils um einen festen Wert schwanken, der näherungsweise durch das arithmetische Mittel

$$\bar{d}_t = \frac{1}{n_t} \sum_{j=0}^{n_t-1} d_{t+jp} \quad \text{für } t = 1, \dots, p$$

gegeben ist; hierbei bezeichnet n_t die Anzahl der Perioden, die zur Berechnung von d_t zur Verfügung stehen.

Normiert man die Größen $\bar{d}_1, \dots, \bar{d}_p$ noch wie folgt:

$$\hat{s}_t = \bar{d}_t - \frac{1}{p} \sum_{j=1}^p \bar{d}_j \quad \text{für } t = 1, \dots, p,$$

so kann $\hat{s}_t = \hat{s}_{t+p} = \hat{s}_{t+2p} = \dots$ als **Schätzer für die Saisonkomponenten** $S_t = S_{t+p} = S_{t+2p} = \dots$ verwandt werden, da die irreguläre Komponente um Null schwankt.

Die Differenz zwischen der ursprünglichen Zeitreihe und der geschätzten Saisonkomponente $y_t - \hat{s}_t$ nennt man auch **saisonbereinigte Zeitreihe**.

Beispiel: Für die in Tab. 9 angegebene Zeitreihe mit Periode $p = 12$ soll die Saisonkomponente S_t geschätzt und die saisonbereinigte Zeitreihe bestimmt werden. Die benötigten Hilfsgrößen d_t und \bar{d}_t sowie die Schätzer \hat{s}_t für die Saisonkomponente,

Tab. 10: Hilfsgrößen d_t und \bar{d}_t sowie Schätzer \hat{s}_t für die Saisonkomponente zur Zeitreihe y_1, \dots, y_{51} bzw. zu den gleitenden Durchschnitten y_7^*, \dots, y_{45}^* aus Tab. 9

Monat	d _t (gerundete Werte)				d _t (gerundet)	s _t (gerundet)
	1976	1977	1978	1979		
Januar	53201	56974	48469	52611	52814	53136
Februar	59929	54934	54102	51727	59173	55495
März	24768	17320	25678	10808	19644	19966
April	-3848	42	-5429	-	-3078	-2756
Mai	-19300	-11680	-14189	-	-15056	-14734
Juni	-23455	-17516	-20116	-	-20362	-20040
Juli	-26413	-21058	-20605	-	-22692	-22370
August	-27225	-22670	-20393	-	-23429	-23107
September	-27358	-24646	-20478	-	-24161	-23839
Oktober	-23967	-21397	-17439	-	-20934	-20612
November	-14300	-10846	-11889	-	-12345	-12023
Dezember	11540	12213	7923	-	10559	10881

die sich durch

$$\hat{s}_t = \bar{d}_t - \frac{1}{12} \sum_{i=1}^{12} \bar{d}_i = \bar{d}_t + \frac{1}{12} \cdot 3867 \approx \bar{d}_t + 322$$

ergeben, sind in Tab. 10 zusammengestellt. Die resultierende saisonbereinigte Zeitreihe wurde bereits in Tab. 9 bzw. Abb. 7 angegeben.

Speziellere und z. T. auch aufwendigere Verfahren zur Saisonbereinigung sind etwa das **Berliner Verfahren**, das vom Statistischen Bundesamt eingesetzt wird, das **Census X-11** Verfahren des amerikanischen „National Bureau of Census“ oder **ASA-II**, ein Verfahren, das von verschiedenen Wirtschaftsforschungsinstituten entwickelt wurde. Recht ausführlich sind diese Verfahren in Stier (1980) dargestellt; verwiesen sei aber z. B. auch auf die Beiträge in Schäffer (1976).

Das Berliner Verfahren geht von einer konstanten Saisonfigur mit Periode p aus. In seiner Grundform basiert es auf einem Regressionsansatz ähnlich dem bei der harmonischen Analyse. Da wir auf diese erst in Abschnitt 3.5 eingehen, werden wir auch erst im Anschluß daran die Grundzüge des Berliner Verfahrens erläutern, vgl. Abschnitt 3.6.

Das Census X-11-Verfahren, das in Shiskin et al. (1967) dargestellt wird, geht von einer variablen Saisonfigur aus. Es ist ein iteratives Verfahren, bei dem zunächst – wie im Falle einer konstanten Saisonfigur – die „saisonbereinigte“ Zeitreihe bestimmt wird. Für diese „saisonbereinigte“ Zeitreihe wird dann wiederum die glatte Komponente und damit dann die Saisonkomponente und die zweite „saisonbereinigte“ Zeitreihe bestimmt, usw.

ASA-II schließlich ist eine Mixtur der Ansätze des Berliner Verfahrens und des Census X-11-Verfahrens.

1.3.2. Polynome und Splines

Wie in Abschnitt 1.3.1 bereits erwähnt, werden durch den Übergang zu gleitenden Durchschnitten starke Krümmungen der glatten Komponente abgeschliffen. Dies läßt sich dadurch vermeiden, daß man zu gleitenden Durchschnitten mit unterschiedlich stark gewichteten Gliedern übergeht.

Paßt man jeweils $2k + 1$ Glieder einer Zeitreihe nach der Methode der kleinsten Quadrate durch eine Gerade an, vgl. Abschnitt 1 in Kap. X, und bestimmt den Wert der geschätzten Geraden an der Stelle des $(k + 1)$ -ten, mittleren Gliedes, so erhält man gerade den einfachen gleitenden Durchschnitt der Ordnung $2k + 1$ an dieser Stelle. Bei starken Krümmungen der glatten Komponente liegt es daher nahe, jeweils $2k + 1$ Glieder einer Zeitreihe durch ein **Polynom** oder einen **Spline**, vgl. etwa Hebbel (1982) oder Schäffer (1978), höherer Ordnung zu approximieren und den Wert des geschätzten Polynoms oder Splines an der Stelle des $(k + 1)$ -ten, mittleren Gliedes als **gewogenen gleitenden Durchschnitt der Ordnung $2k + 1$** an dieser Stelle zu verwenden.

Wir wollen hier speziell die Polynomanpassung behandeln. Dabei ist es zweckmäßig, die jeweils zu berücksichtigenden $2k + 1$ Werte einer Zeitreihe mit $t = -k, -k + 1, \dots, 0, \dots, k - 1, k$ zu indizieren. Wollen wir dann an die Reihe y_{-k}, \dots, y_k ein Polynom

$$P(t) = \alpha_0 + \alpha_1 t + \dots + \alpha_q t^q$$

vom Grad q mittels der Methode der kleinsten Quadrate anpassen, so müssen die Ableitungen der Funktion

$$S^2 = S^2(\alpha_0, \dots, \alpha_q) = \sum_{t=-k}^k (y_t - \alpha_0 - \alpha_1 t - \dots - \alpha_q t^q)^2$$

nach den unbekannten, zu schätzenden Parametern $\alpha_0, \dots, \alpha_q$ verschwinden, d. h. es muß

$$\sum_{t=-k}^k t^j y_t - \alpha_0 \sum_{t=-k}^k t^j - \alpha_1 \sum_{t=-k}^k t^{j+1} - \dots - \alpha_q \sum_{t=-k}^k t^{j+q} = 0$$

für $j = 0, \dots, q$ gelten. Als Schätzung für α_0 erhält man hieraus gerade

$$\hat{\alpha}_0 = c_0 + c_1 y_{-k} + c_2 y_{-k+1} + \dots + c_{2k+1} y_k,$$

wobei die Größen $c_0, c_1, \dots, c_{2k+1}$ von k und q abhängig sind.

Der Wert des Polynoms an der Stelle des $(k+1)$ -ten mittleren Glieds ist nun gerade

$$P(0) = \hat{\alpha}_0,$$

so daß c_1, \dots, c_{2k+1} gerade die Gewichte der Glieder des gewogenen gleitenden Durchschnitts an dieser Stelle angeben.

Beispiel: Wir wollen einmal die Gewichte der Glieder eines gewogenen gleitenden Durchschnitts der Ordnung $2k+1=5$, d. h. $k=2$, bei Anpassung eines Polynoms 2. Grades, d. h. $q=2$, bestimmen.

Die partiellen Ableitungen der Funktion S^2 haben hier für $j=0, 1, 2$ die Gestalt

$$\sum_{t=-2}^2 t^j y_t - \alpha_0 \sum_{t=-2}^2 t^j - \alpha_1 \sum_{t=-2}^2 t^{j+1} - \alpha_2 \sum_{t=-2}^2 t^{j+2},$$

so daß sich für $j=0$

$$\sum_{t=-2}^2 y_t - 5\hat{\alpha}_0 - \hat{\alpha}_1 \sum_{t=-2}^2 t - \hat{\alpha}_2 \sum_{t=-2}^2 t^2 = \sum_{t=-2}^2 y_t - 5\hat{\alpha}_0 - 0\hat{\alpha}_1 - 10\hat{\alpha}_2 = 0$$

bzw.

$$\sum_{t=-2}^2 y_t = 5\hat{\alpha}_0 + 10\hat{\alpha}_2,$$

für $j=1$

$$\sum_{t=-2}^2 t y_t - \hat{\alpha}_0 \sum_{t=-2}^2 t - \hat{\alpha}_1 \sum_{t=-2}^2 t^2 - \hat{\alpha}_2 \sum_{t=-2}^2 t^3 = \sum_{t=-2}^2 t y_t - 0\hat{\alpha}_0 - 10\hat{\alpha}_1 - 0\hat{\alpha}_2 = 0$$

bzw.

$$\sum_{t=-2}^2 t y_t = 10\hat{\alpha}_1$$

und für $j=2$

$$\begin{aligned} \sum_{t=-2}^2 t^2 y_t - \hat{\alpha}_0 \sum_{t=-2}^2 t^2 - \hat{\alpha}_1 \sum_{t=-2}^2 t^3 - \hat{\alpha}_2 \sum_{t=-2}^2 t^4 &= \sum_{t=-2}^2 t^2 y_t - 10\hat{\alpha}_0 - 0\hat{\alpha}_1 - 34\hat{\alpha}_2 \\ &= 0 \end{aligned}$$

bzw.

$$\sum_{t=-2}^2 t^2 y_t = 10\alpha_0 + 34\alpha_2$$

ergibt. Daraus erhalten wir dann

$$\begin{aligned}\alpha_0 &= \frac{1}{35} \left(17 \sum_{t=-2}^2 y_t - 5 \sum_{t=-2}^2 t^2 y_t \right) \\ &= \frac{1}{35} (-3y_{-2} + 12y_{-1} + 17y_0 + 12y_1 - 3y_2),\end{aligned}$$

d.h. die Gewichte zur Bestimmung des gewogenen gleitenden Durchschnitts fünfter Ordnung bei Anpassung eines Polynoms 2. Grades sind

$$c_0 = 0, c_1 = -\frac{3}{35}, c_2 = \frac{12}{35}, c_3 = \frac{17}{35}, c_4 = \frac{12}{35} \text{ und } c_5 = -\frac{3}{35}.$$

1.3.3. Die Differenzenmethode

Eine Möglichkeit zur Elimination von Trend- und Saisonkomponente (bei konstanter Saisonfigur) bietet die **Differenzenmethode (variate difference method)**, vgl. etwa auch Gosset (1914), Anderson (1914), Tintner (1960).

Weist eine Zeitreihe y_1, \dots, y_n einen linearen Trend auf, so kann dieser durch Übergang zu den **ersten Differenzen**

$$y_t^* = \nabla y_t = y_t - y_{t-1} \quad \text{für } t = 2, \dots, n$$

eliminiert werden. Die Bereinigung einer Zeitreihe von einem polynomialem Trend vom Grad d kann entsprechend durch Bildung der **d -ten Differenzen**

$$y_t^* = \nabla^d y_t = \nabla^{d-1}(\nabla y_t) \quad \text{für } t = d+1, \dots, n$$

erfolgen. Für $d = 2$ ergibt sich etwa

$$\begin{aligned}y_t^* &= \nabla^2 y_t = \nabla(\nabla y_t) = \nabla(y_t - y_{t-1}) = y_t - y_{t-1} - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \quad \text{für } t = 3, \dots, n,\end{aligned}$$

und für $d = 3$ erhalten wir

$$\begin{aligned}y_t^* &= \nabla^3 y_t = \nabla^2(\nabla y_t) = \nabla(\nabla^2 y_t) = \nabla(y_t - 2y_{t-1} + y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} - (y_{t-1} - 2y_{t-2} + y_{t-3}) \\ &= y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3} \quad \text{für } t = 4, \dots, n.\end{aligned}$$

Was die Differenzenmethode bewirkt, veranschaulicht man sich am besten anhand einer idealen Zeitreihe, die nur durch einen Trend, z. B. durch einen polynomialem Trend 3. Grades gebildet wird:

$$y_t = 1 + t + t^2 + t^3 \quad \text{für } t = 1, \dots, 6;$$

die Zeitreihe sowie die zugehörigen ersten, zweiten und dritten Differenzen sind in Tab. 11 angegeben. Man sieht, daß der Trend bei Übergang zu den dritten Differenzen vollkommen verschwunden ist, er bei den zweiten Differenzen linear und bei den ersten Differenzen noch quadratisch ist.

Bei Zeitreihen mit konstanter Saisonfigur der Periodenlänge p kann die Differenzenmethode zur Elimination der Saisonkomponente verwandt werden. Hierzu geht

Tab. 11: 1.–3. Differenzen zur Zeitreihe $y_t = 1 + t + t^2 + t^3$ für $t = 1, \dots, 6$

t	y_t	∇y_t	$\nabla^2 y_t$	$\nabla^3 y_t$
1	4			
2	15	11		
3	40	25	14	
4	85	45	20	6
5	156	71	26	6
6	259	103	32	6

man von der Zeitreihe y_1, \dots, y_n zu den **saisonalen Differenzen**

$$y_t^* = y_t - y_{t-p} \quad \text{für } t = p+1, \dots, n \\ \text{über.}$$

Natürlich können d-te Differenzen und saisonale Differenzen auch kombiniert gebildet werden. Möchte man etwa einen linearen Trend sowie eine Saisonkomponente (bei Periodenlänge p) eliminieren, so geht man zunächst zu den ersten Differenzen über und wendet auf sie die saisonalen Differenzen an oder umgekehrt. In beiden Fällen erhält man die **bereinigte Zeitreihe**

$$y_t^* = y_t - y_{t-1} - y_{t-p} + y_{t-p-1} \quad \text{für } t = p+2, \dots, n.$$

Beispiel: Bei der Zeitreihe der Arbeitslosen im Baugewerbe, vgl. Tab. 9 und Abb. 7, lässt sich in etwa ein linearer Trend und eine konstante Saisonfigur mit Periodenlänge 12 Monate (= 1 Jahr) feststellen.

Um zu überprüfen, ob der vorhandene lineare Trend zum 5% Niveau signifikant ist, wollen wir hier einmal die approximative Version des im Abschnitt 5.4.3.B des Kap. IV dargestellten **Trendtests** nach Mann anwenden. Und zwar testen wir dazu

$$H_{02}: \text{kein Abwärtstrend} \quad \text{gegen} \quad H_{12}: \text{Abwärtstrend}$$

vermittels der Prüfgröße

$$C = \sum_{t=1}^{n-1} \sum_{j=t+1}^n \operatorname{sgn}(y_j - y_t) = \sum_{t=1}^{50} \Sigma_t$$

wobei sgn die Vorzeichenfunktion signum bezeichnet (d.h. $\operatorname{sgn}(y_j - y_t)$ ist gleich 1, falls $y_j - y_t > 0$, gleich -1 , falls $y_j - y_t < 0$ und gleich 0, falls $y_j - y_t = 0$ ist). Die Größen Σ_t für $t = 1, \dots, 50$ sind in Tab. 9 bereits angegeben und wir erhalten dann

$$C = -469.$$

Nun ist die Größe C^* , die hier den Wert

$$C^* = C / \sqrt{51 \cdot 50 \cdot 107/18} = -469 / \sqrt{272850/18} = -3,809$$

annimmt, approximativ standardnormalverteilt, so daß wir unsere Hypothese H_{02} zum 5% Niveau verwerfen müssen, da gilt

$$C^* = -3,809 < -1,646 = u_{0,05}.$$

Der Abwärtstrend ist also zum 5% Niveau (und sogar zum 0,01% Niveau, $u_{0,0001} = -3,719$) signifikant.

Zur Trendbereinigung wird man daher von der Ursprungsreihe y_1, \dots, y_{51} zu

Tab. 12: Zeitreihe der Arbeitslosen im Baugewerbe, vgl. auch Tab. 9, sowie bereinigte Reihen y_t^* , y_t^{**} und \tilde{y}_t

t	y_t	$y_t^* = y_t - y_{t-1}$	$y_t^{**} = y_t - y_{t-12}$	$\tilde{y}_t = y_t - y_{t-1} - y_{t-12} + y_{t-13}$
1	60572			
2	52461	- 8111		
3	47357	- 5104		
4	48320	963		
5	60219	11899		
6	84418	24199		
7	119916	35498		
8	124350	4434		
9	87309	- 37041		
10	57035	- 30274		
11	39903	- 17132		
12	34053	- 5850		
13	29905	- 4148	- 30667	
14	28068	- 1837	- 24393	6274
15	26634	- 1434	- 20723	3670
16	29259	2625	- 19061	1662
17	38942	9683	- 21272	- 2211
18	65036	26094	- 19382	1890
19	110728	45692	- 9188	10194
20	108931	- 1797	- 15419	- 6231
21	71517	- 37414	- 15792	- 373
22	54428	- 17089	- 2607	13185
23	42911	- 11517	3008	5615
24	37123	- 5788	3070	62
25	33044	- 4079	3139	69
26	30755	- 2289	2687	- 452
27	28742	- 2013	2108	- 579
28	31698	2956	2439	331
29	41427	9729	2485	46
30	63685	22258	- 1351	- 3836
31	99189	35504	- 11539	- 10188
32	104240	5051	- 4691	6848
33	75304	- 28936	3787	8478
34	43622	- 31682	- 10806	- 14593
35	33990	- 9632	- 8921	1885
36	26819	- 7171	- 10304	- 1383
37	25291	- 1528	- 7753	2551
38	24538	- 753	- 6217	1536
39	22685	- 1853	- 6057	160
40	23945	1260	- 7753	- 1696
41	28245	4300	- 13182	- 5429
42	47017	18772	- 16668	- 3486
43	90920	43903	- 8269	8399
44	89340	- 1580	- 14900	- 6631
45	47792	- 41548	- 27512	- 12612
46	28448	- 19344	- 15174	12338
47	19139	- 9309	- 14851	323
48	16728	- 2411	- 10091	4760
49	16523	- 205	- 8768	1323
50	16622	99	- 7916	852
51	15499	- 1123	- 7186	730

ersten Differenzen

$$y_t^* = y_t - y_{t-1} \quad \text{für } t = 2, \dots, 51$$

und zur Saisonbereinigung zu saisonalen Differenzen

$$y_t^{**} = y_t - y_{t-12} \quad \text{für } t = 13, \dots, 51$$

übergehen. Möchte man die Daten von Trend und Saison bereinigen, so wird man die Reihe

$$\begin{aligned}\tilde{y}_t &= y_t^* - y_{t-12}^* = y_t^{**} - y_{t-1}^{**} \\ &= y_t - y_{t-1} - y_{t-12} + y_{t-13} \quad \text{für } t = 14, \dots, 51\end{aligned}$$

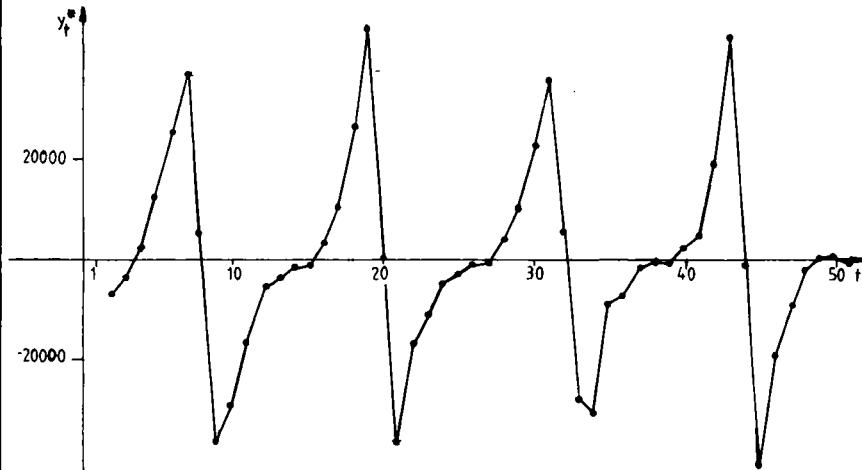


Abb. 8: Erste Differenzen $y_t^* = y_t - y_{t-1}$ zur Zeitreihe der Arbeitslosen im Baugewerbe

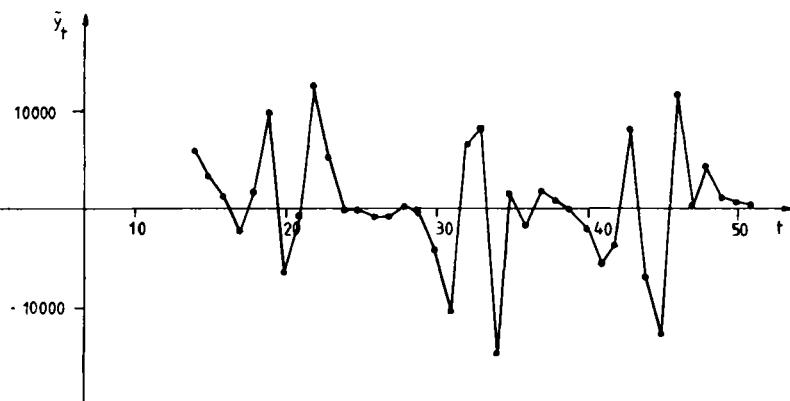


Abb. 9: Reihe der ersten und ersten saisonalen Differenzen $\tilde{y}_t = y_t - y_{t-1} - y_{t-12} + y_{t-13}$ zur Zeitreihe der Arbeitslosen im Baugewerbe

betrachten. In **Tab. 12** sind neben der Ausgangsreihe auch diese drei bereinigten Zeitreihen angegeben und in den **Abb. 8 und 9** sind die Reihen y_2^*, \dots, y_{51}^* bzw. $\tilde{y}_{14}, \dots, \tilde{y}_{51}$ graphisch durch Polygonzüge dargestellt.

1.3.4. Exponentielles Glätten

Unter **exponentiellem Glätten (exponential smoothing)** versteht man den Übergang von einer Zeitreihe mit Werten y_t zu einer Reihe y_t^* , wobei zur Bestimmung von y_t^* alle Vorgänger von y_t verwandt werden. Theoretisch muß hierbei die betrachtete Zeitreihe unendlich lang sein; man bestimmt dann y_t^* als

$$y_t^* = \sum_{j=0}^{\infty} c_j y_{t-j-1},$$

wobei

$$c_j = \alpha(1-\alpha)^j \quad \text{für } j = 0, 1, 2, 3, \dots$$

gesetzt wird. Aus diesen Gewichten c_j erklärt sich auch der Name dieses Verfahrens, denn die Gewichte c_j nehmen mit wachsendem j exponentiell ab, wenn der Parameter α zwischen 0 und 1 gewählt wird; je kleiner α ist, desto ausgeprägter ist die Glättungseigenschaft des Verfahrens. Eine übliche Wahl von α ist $0.1 \leq \alpha \leq 0.3$.

Da eine reale Zeitreihe stets nur aus endlich vielen Werten y_1, \dots, y_n besteht, ist das exponentielle Glätten nur für sehr großes n zu empfehlen. Jedoch hat das Verfahren den Vorteil, daß mit seiner Hilfe der unbekannte Wert y_{n+1} prognostiziert werden kann, falls n hinreichend groß ist:

$$y_{n+1}^* = \sum_{j=0}^{n-1} c_j y_{n-j} = \sum_{j=0}^{n-1} \alpha(1-\alpha)^j y_{n-j}.$$

Jedoch sollte y_1, \dots, y_n hier eine Trend- und Saison-bereinigte Zeitreihe sein.

Setzt man $y_1^* = y_1$, so lassen sich y_2^*, \dots, y_{n+1}^* auch iterativ bestimmen, denn es ist für $t = 1, \dots, n$

$$y_{t+1}^* = \alpha y_t + (1-\alpha) y_t^*.$$

Beispiel: Vermittels des iterativen Verfahrens wollen wir ausgehend von der Trend- und Saison-bereinigten Zeitreihe \tilde{y}_t , $t = 14, \dots, 51$ zur Zeitreihe der Arbeitslosen im Baugewerbe, vgl. Tab. 9 und 12, den Schätzwert \tilde{y}_{52}^* für \tilde{y}_{52} bei exponentieller Glättung bestimmen. Wählen wir einmal $\alpha = 0.2$ und setzen $\tilde{y}_{14}^* = \tilde{y}_{14} = 6274$, so ergeben sich die Werte

$$\tilde{y}_{t+1}^* = \alpha \tilde{y}_t + (1-\alpha) \tilde{y}_t^* \quad \text{für } t = 14, \dots, 51$$

gemäß **Tab. 13**. Die Prognose für den Wert \tilde{y}_{52} ergibt sich also in etwa zu 882. Rechnet man diesen Wert auf die Ausgangszeitreihe um, so erhält man als Prognose für die Anzahl der Arbeitslosen im Baugewerbe am Ende des Monats Oktober 1979 den Wert

$$\hat{y}_{52} = \tilde{y}_{52}^* + y_{51} + y_{40} - y_{39} = 882 + 15499 + 23945 - 22685 = 17641.$$

Tab. 13: Aus exponentiellem Glätten mit $\alpha = 0.2$ resultierende Größen $\tilde{y}_{15}^*, \dots, \tilde{y}_{52}^*$ zur Trend- und Saison-bereinigten Zeitreihe $\tilde{y}_{14}, \dots, \tilde{y}_{51}$ aus Tab. 12

t	\tilde{y}_{t+1}^*	t	\tilde{y}_{t+1}^*	t	\tilde{y}_{t+1}^*
14	6274,000	27	1639,447	40	- 285,640
15	5753,200	28	1377,758	41	- 1314,312
16	4934,960	29	1111,406	42	- 1748,650
17	3505,768	30	121,925	43	280,880
18	3182,614	31	- 1940,060	44	- 1101,496
19	4584,891	32	- 182,448	45	- 3403,597
20	2421,713	33	1549,642	46	- 255,278
21	1862,770	34	- 1678,886	47	- 139,622
22	4127,216	35	- 966,109	48	840,302
23	4424,773	36	- 1049,487	49	936,842
24	3552,218	37	- 329,390	50	919,874
25	2855,574	38	43,688	51	881,899
26	2194,059	39	66,950		

Möchte man eine willkürliche Festlegung des Glättungsparameters α vermeiden, so empfiehlt es sich, denjenigen Wert α zu wählen, der die Summe der Residuenquadrate

$$\sum_{t=1}^n e_t^2(\alpha)$$

bzgl. α ($0 < \alpha < 1$) minimiert. Unter einem Residuum $e_t = e_t(\alpha)$ zur Beobachtung y_t versteht man für $t = 1, \dots, n$ hierbei gerade die Größe

$$\begin{aligned} e_t(\alpha) &= y_t - y_t^* = y_t - \alpha y_{t-1} - (1 - \alpha) y_{t-1}^* \\ &= y_t - \alpha y_{t-1} - (1 - \alpha)(y_{t-1} - e_{t-1}(\alpha)) \\ &= y_t - y_{t-1} + (1 - \alpha)e_{t-1}(\alpha). \end{aligned}$$

Da die Funktion $\sum_{t=1}^n e_t^2(\alpha)$ aber recht flach verläuft, ist die Wahl von α nicht besonders kritisch, vgl. Cox (1961), so daß man auf dieses Verfahren in vielen Fällen verzichten kann.

1.3.5. Lineare Filter

Jede Transformation einer Zeitreihe y_1, \dots, y_n in eine andere Zeitreihe $y_{q+1}^*, \dots, y_{n-s}^*$ der Gestalt

$$y_t^* = \sum_{r=-q}^{+s} a_r y_{t+r} \quad \text{für } t = q + 1, \dots, n - s,$$

wobei a_r für $r = -q, \dots, 0, \dots, +s$ ein Gewicht zum Wert y_{t+r} der Ursprungszreihe bezeichnet, heißt **linearer Filter**. Dementsprechend heißt die geglättete Zeitreihe der y_t^* auch **gefilterte Zeitreihe**, d. h. der Begriff **Filterung** ist synonym zu Glättung.

Wir haben in diesem Abschnitt schon einige spezielle lineare Filter kennengelernt. Z. B. ist ein gleitender Durchschnitt der ungeraden Ordnung $2k + 1$ ein **symmetrischer Filter**, d. h. $q = s = k$ und $a_{-r} = a_r$, mit Gewichten $a_r = 1/(2k + 1)$ für

$r = -k, \dots, k$, und für eine erste Differenz gilt $q = 1, s = 0$ sowie $a_{-1} = -1, a_0 = 1$. In dieser Art und Weise lassen sich alle zuvor behandelten Glättungsverfahren in die Klasse der linearen Filter einordnen.

Wie wir im bisherigen Verlauf des Abschnitts 1.3 gesehen haben, sind die Ziele der Filterung einer Zeitreihe sehr unterschiedlich. Man verwendet sie zur Schätzung der glatten Komponente, zur Schätzung der Saisonkomponente oder etwa auch zur Elimination von Trend und Saison. So gehören gleitende Durchschnitte zur Klasse der Filter, die irreguläre Schwankungen oder saisonale Schwankungen beseitigen. Nichtsaisonale Differenzenfilter beseitigen einen Trend und gehören somit einer ganz anderen Klasse von Filtern an.

Allgemein erfolgt die **Klassifizierung von Filtern** in vier Klassen, nämlich in **Tiefpaß-Filter**, **Hochpaß-Filter**, **Bandsperren** und **Bandpässe**. Tiefpaß-Filter dienen der Bereinigung von lokalen Schwankungen, Hochpaß-Filter der Bereinigung von langfristigen Schwankungen, Bandsperren der Bereinigung von saisonalen Schwankungen und Bandpässe der Sichtbarmachung saisonaler Schwankungen. Demnach ist etwa ein gleitender Durchschnitt ein Tiefpaß-Filter oder eine Bandsperre, ein nicht-saisonaler Differenzenfilter ein Hochpaß-Filter. Wie man allgemein die Klassenzugehörigkeit eines linearen Filters und damit einen für den gewünschten Effekt adäquaten Filter bestimmt, kann erst im Rahmen der Spektralanalyse, vgl. Abschnitt 3, behandelt werden.

Oft werden mehrere Filter nacheinander auf eine Zeitreihe angewandt; man spricht dann von einer **Serienschaltung von Filtern**.

So haben wir etwa im Abschnitt 1.3.3 auf die Zeitreihe der Arbeitslosen im Bauwesen einen ersten Differenzen-Filter sowie einen saisonalen Differenzenfilter mit Periode $p = 12$ angewandt und diese beiden Filter auch kombiniert, was zur Zeitreihe $\tilde{y}_{14}, \dots, \tilde{y}_{51}$ führte. Wenden wir auf eine Zeitreihe y_1, \dots, y_n zunächst erste Differenzen an, so erhalten wir die gefilterte Zeitreihe

$$y_t^* = \nabla y_t = y_t - y_{t-1} = \sum_{r=-1}^0 a_r y_{t+r} \quad \text{mit } a_{-1} = -1, a_0 = 1$$

für $t = 2, \dots, n$. Wenden wir auf y_2^*, \dots, y_n^* dann saisonale Differenzen mit Periode 12 an, so erhalten wir für $t = 14, \dots, n$

$$\tilde{y}_t = y_t^* - y_{t-12}^* = \sum_{r'=-12}^0 b_{r'} y_{t+r'}^* \quad \text{mit } b_{-12} = -1, b_{-11} = \dots = b_{-1} = 0, b_0 = 1$$

bzw.

$$\begin{aligned} \tilde{y}_t &= \sum_{r'=-12}^0 b_{r'} y_{t+r'}^* = \sum_{r'=-12}^0 b_{r'} (\nabla y_{t+r'}) = \sum_{r'=-12}^0 b_{r'} \left(\sum_{r=-1}^0 a_r y_{t+r+r'} \right) \\ &= \sum_{r'=-12}^0 b_{r'} (a_{-1} y_{t+r'-1} + a_0 y_{t+r'}) \\ &= -1(-1y_{t-13} + 1y_{t-12}) + 1(-1y_{t-1} + 1y_t) \\ &= y_t - y_{t-1} - y_{t-12} + y_{t-13} = \sum_{j=-13}^0 c_j y_{t+j} \end{aligned}$$

$$\text{mit } c_{-13} = 1, c_{-12} = -1, c_{-11} = \dots = c_{-2} = 0, c_{-1} = -1, c_0 = 1;$$

das Ergebnis ist also wiederum ein linearer Filter.

Allgemein ergibt sich bei Serienschaltung zweier Filter mit Gewichten a_{-q}, \dots, a_s

und b_{-q}, \dots, b_s ein linearer Filter mit Gewichten $c_{-q-q'}, \dots, c_{s+s'}$ der Gestalt

$$c_j = \sum_{r=-q}^s a_r b_{j-r} \quad \text{für } j = -q - q', \dots, s + s'.$$

Im obigen Beispiel lassen sich die Gewichte also auch direkt ausrechnen:

$$\begin{aligned} c_{-13} &= \sum_{r=-1}^0 a_r b_{-13-r} = a_{-1} b_{-12} + a_0 b_{-13} = (-1)(-1) + 1 \cdot 0 = 1 \\ c_{-12} &= \sum_{r=-1}^0 a_r b_{-12-r} = a_{-1} b_{-11} + a_0 b_{-12} = (-1) \cdot 0 + 1 \cdot (-1) = -1 \\ c_{-11} &= \sum_{r=-1}^0 a_r b_{-11-r} = a_{-1} b_{-10} + a_0 b_{-11} = (-1) \cdot 0 + 1 \cdot 0 = 0 \\ &\vdots \\ c_{-2} &= \sum_{r=-1}^0 a_r b_{-2-r} = a_{-1} b_{-1} + a_0 b_{-2} = (-1) \cdot 0 + 1 \cdot 0 = 0 \\ c_{-1} &= \sum_{r=-1}^0 a_r b_{-1-r} = a_{-1} b_0 + a_0 b_{-1} = (-1) \cdot 1 + 1 \cdot 0 = -1 \\ c_0 &= \sum_{r=-1}^0 a_r b_{-r} = a_{-1} b_1 + a_0 b_0 = (-1) \cdot 0 + 1 \cdot 1 = 1. \end{aligned}$$

1.4. Autokovarianzen, Autokorrelationen und partielle Autokorrelationen

Ein wichtiges Hilfsmittel zur Beurteilung der Eigenschaften einer Zeitreihe y_1, \dots, y_n sind die **empirischen Autokovarianzen** bzw. **Autokorrelationen**. Man versteht darunter Maße für den Zusammenhang zwischen Beobachtungsdaten, die einen bestimmten zeitlichen Abstand zueinander haben. So ist die empirische Autokovarianz bzw. Autokorrelation zum sogenannten **lag** (Zeitverschiebung) 1 ein Maß für den Zusammenhang von y_t und y_{t+1} für $t = 1, \dots, n-1$, die zum lag 2 ein Maß für den Zusammenhang von y_t und y_{t+2} für $t = 1, \dots, n-2$, usw.

Die empirische Autokovarianz zum lag k wird für $k = 0, 1, 2, 3, \dots, n-1$ zu meist definiert als

$$c(k) = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) = c(-k) \quad \text{mit} \quad \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t;$$

mitunter wird allerdings anstelle des Faktors $\frac{1}{n}$ der Faktor $\frac{1}{n-k}$ gewählt. Für $k = 0$ ist $c(k)$ natürlich gerade die empirische Varianz der Zeitreihe y_1, \dots, y_n (allerdings mit dem Faktor $\frac{1}{n}$ anstelle von $\frac{1}{n-1}$).

Entsprechend der sonst üblichen Vorgehensweise in der Korrelationsanalyse wird zur Bestimmung der empirischen Autokorrelation zum lag k die Autokovarianz zu diesem lag mit der empirischen Varianz normiert, d.h. die empirische Autokorrelation zum lag k ist für $k = 0, 1, 2, \dots, n-1$ gerade

$$r(k) = \frac{c(k)}{c(0)} = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} = r(-k);$$

für $k = 0$ ist $r(k) = r(0) = 1$.

Einen Plot von $r(k)$ gegen den lag k für $k = 0, 1, 2, \dots$ nennt man ein **Korrelogramm**. Man sollte es jedoch nur dann verwenden, wenn eine Zeitreihe Trendbereinigt ist, da sonst die Gefahr besteht, daß der Trend alle übrigen Informationen überdeckt.

Beispiel: Für die Zeitreihe der Arbeitslosen im Baugewerbe, vgl. Abschnitt 1.3 und insbesondere Tab. 9 und 12, wollen wir die Autokovarianzen und Autokorrelationen zur Reihe der ersten Differenzen $y_t^* = y_t - y_{t-1}$ berechnen. Diese sind zum lag k in Tab. 14 zusammengestellt, und in Abb. 10 ist das entsprechende Korrelogramm durch einen Polygonzug veranschaulicht.

Am Korrelogramm erkennt man sehr schön die starken Autokorrelationen zu den lags $12, 2 \cdot 12 = 24$ und $3 \cdot 12 = 36$, die auf die starke saisonale 12-monatige Schwankung (Jahresschwankung) hinweisen.

Tab. 14: Empirische Autokovarianzen $c(k)$ und Autokorrelationen $r(k)$ zum lag k , $k = 0, \dots, 49$ für die Reihe der ersten Differenzen $y_t^* = y_t - y_{t-1}$ zur Zeitreihe der Arbeitslosen im Baugewerbe, vgl. Tab. 12

k	$c(k)$	$r(k)$	k	$c(k)$	$r(k)$
0	356550468,2	1,000	25	100440289,9	0,282
1	191998483,8	0,538	26	- 22467516,8	-0,063
2	- 41509056,7	-0,116	27	- 68791094,8	-0,193
3	- 123413595,9	-0,346	28	- 63756471,6	-0,179
4	- 115117049,4	-0,323	29	- 39950410,7	-0,112
5	- 76904209,2	-0,216	30	- 15256261,4	-0,043
6	- 47130877,3	-0,132	31	- 9646812,9	-0,027
7	- 51086042,2	-0,143	32	- 21954742,9	-0,062
8	- 78201345,7	-0,219	33	- 31333606,7	-0,088
9	- 89830980,6	-0,252	34	- 8686041,8	-0,024
10	- 21525433,3	-0,060	35	53555767,0	0,150
11	152221907,4	0,427	36	86921833,5	0,244
12	251002649,3	0,704	37	41476380,1	0,116
13	139540875,4	0,391	38	- 13790789,1	-0,039
14	- 25894867,8	-0,073	39	- 35255520,1	-0,099
15	- 93655695,9	-0,263	40	- 33592979,6	-0,094
16	- 88120662,6	-0,247	41	- 18478748,0	-0,052
17	- 58947129,1	-0,165	42	690196,0	0,002
18	30526355,0	-0,086	43	7765313,8	0,022
19	- 29741098,4	-0,083	44	3828877,1	0,011
20	- 47488001,8	-0,133	45	1510050,2	0,004
21	- 51960828,5	-0,146	46	139713,8	0,000
22	- 20195894,5	-0,057	47	- 192773,6	-0,001
23	90200360,6	0,253	48	- 125636,5	-0,000
24	174928654,3	0,491	49	31944,0	0,000

Faßt man eine Zeitreihe y_1, \dots, y_n als Realisation einer Folge von Zufallsvariablen Y_1, \dots, Y_n auf, so sind $c(k) = c(-k)$ bzw. $r(k) = r(-k)$ Schätzer für die theoretischen Autokovarianzen

$$\gamma(k) = \gamma(-k) = \text{Cov}(Y_t, Y_{t+k}) = \text{Cov}(Y_t, Y_{t-k})$$

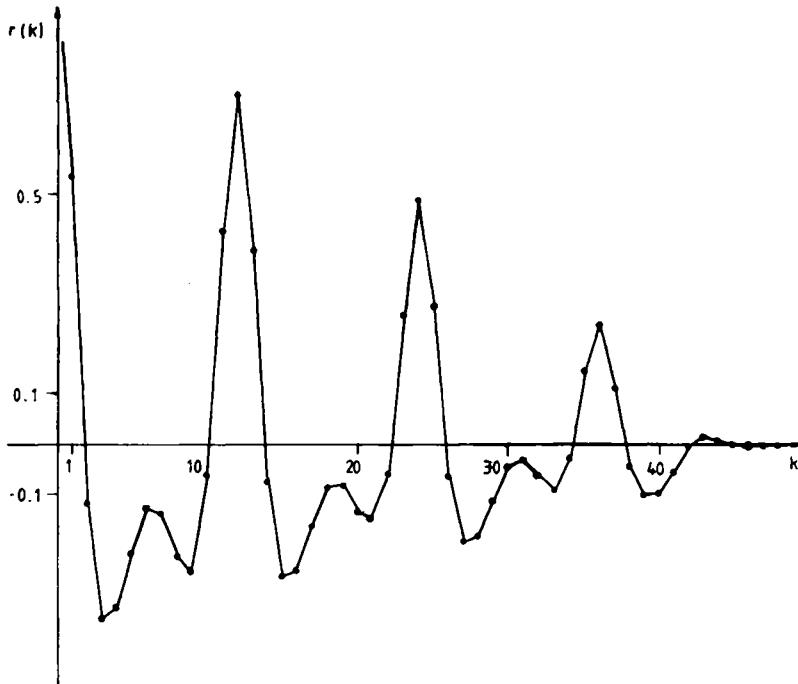


Abb. 10: Autokorrelationen $r(k)$ zum lag k , $k = 1, \dots, 49$, zur Reihe der ersten Differenzen zur Zeitreihe der Arbeitslosen im Baugewerbe

bzw. die theoretischen Autokorrelationen

$$\varrho(k) = \varrho(-k) = \text{Corr}(Y_t, Y_{t+k}) = \text{Corr}(Y_t, Y_{t-k})$$

zum lag k für $k = 0, 1, 2, \dots$. Diese theoretischen Größen sowie die **theoretischen partiellen Autokorrelationen** werden uns im Abschnitt 2 noch begegnen. Die theoretische partielle Autokorrelation zum lag k ist dabei gerade die Korrelation zwischen Y_t und Y_{t+k} bei Ausschaltung des Einflusses der dazwischenliegenden Zufallsvariablen $Y_{t+1}, \dots, Y_{t+k-1}$. Mit den Bezeichnungen

$$P_{k-1} = (\varrho(1), \dots, \varrho(k-1)),$$

$$\tilde{P}_{k-1} = (\varrho(k-1), \dots, \varrho(1)),$$

$$P_{k-1,k-1} = \begin{pmatrix} 1 & \varrho(1) & \varrho(2) & \dots & \varrho(k-2) \\ \varrho(1) & 1 & \varrho(1) & \dots & \varrho(k-3) \\ \varrho(2) & \varrho(1) & 1 & \dots & \varrho(k-4) \\ \vdots & \vdots & \vdots & & \vdots \\ \varrho(k-2) & \varrho(k-3) & \varrho(k-4) & \dots & 1 \end{pmatrix}$$

ist die **partielle Autokorrelation zum lag k** gerade

$$\varrho_{k-1}(k) = \frac{\varrho(k) - P_{k-1} P_{k-1,k-1}^{-1} \tilde{P}_{k-1}^T}{\sqrt{(1 - P_{k-1} P_{k-1,k-1}^{-1} \tilde{P}_{k-1}^T)(1 - \tilde{P}_{k-1} P_{k-1,k-1}^{-1} \tilde{P}_{k-1}^T)}},$$

wobei P_{k-1}^T bzw. \tilde{P}_{k-1}^T gerade den zu P_{k-1} bzw. \tilde{P}_{k-1} transponierten Vektor und $P_{k-1,k-1}^{-1}$ die zu $P_{k-1,k-1}$ inverse Matrix bezeichnet; man vgl. hierzu auch Abschnitt 1.5 von Kap. III in Hartung/Elpelt (1984). Natürlich ist die partielle Autokorrelation zum lag 1 gerade $\varrho(1)$.

Die partielle Autokorrelation wird geschätzt, indem in allen Größen $\varrho(j)$ für $j = 1, \dots, k$ durch die entsprechende empirische Autokorrelation $r(j)$ ersetzt wird; diesen **Schätzer** wollen wir dann mit $r_{k-1}(k)$ bezeichnen.

2. Selbsterklärende Zeitreihenmodelle und die Methode von Box und Jenkins

Jede beobachtete oder geglättete Zeitreihe lässt sich auffassen als endliche Realisation einer Folge korrelierter Zufallsvariablen $\dots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \dots$, die auch ein **stochastischer Prozeß** (Y_t) genannt wird. Ein solcher Prozeß heißt **schwach stationär**, wenn alle Zufallsvariablen Y_t den gleichen Erwartungswert μ_Y und die gleiche Varianz σ_Y^2 (**Homoskedastizität**) besitzen, d. h. falls gilt

$$EY_t = \mu_Y \quad \text{und} \quad \text{Var } Y_t = \sigma_Y^2 \quad \text{für alle } t,$$

und falls zusätzlich gilt

$$EY_{t_1} Y_{t_2} = EY_{t_1+k} Y_{t_2+k} \quad \text{für alle } t_1, t_2 \text{ und } k.$$

Wir werden in den folgenden Abschnitten 2.1–2.4 einige Zeitreihenmodelle betrachten, bei denen ein schwach stationärer Prozeß (Y_t) durch sich selbst und/oder durch einen Prozeß (ε_t) erklärt wird, der ein sogenanntes **weißes Rauschen** (**white noise**) darstellt. Ein Prozeß (ε_t) heißt weißes Rauschen, falls gilt

$$E\varepsilon_t = 0, \quad \text{Var } \varepsilon_t = \sigma_\varepsilon^2 \quad \text{für alle } t \quad \text{und}$$

$$E\varepsilon_{t_1} \varepsilon_{t_2} = 0 \quad \text{für alle } t_1 \neq t_2,$$

d. h. der Erwartungswert von ε_t ist 0, die Varianz von ε_t ist konstant gleich σ_ε^2 und die Zufallsvariablen ε_t sind alle unkorreliert; insbesondere ist (ε_t) also ein schwach stationärer Prozeß. Der Einfachheit halber betrachten wir hier nur **Mittelwert-bereinigte Prozesse** (Y_t), d. h. wir beschränken uns auf den Fall, daß gilt $EY_t = 0$, denn ist $EY_t = \mu_Y$, so ist der Prozeß $(Y_t - \mu_Y)$ ein Mittelwert-bereinigter Prozeß. Ein solcher Prozeß (Y_t) heißt **invertierbar**, falls sich der Prozeß (ε_t) mittels des Prozesses (Y_t) darstellen lässt. Speziell werden wir uns mit **autoregressiven (AR-)Prozessen**, mit **moving average (MA-)Prozessen** und Mischformen dieser beiden Typen (**ARMA-Prozessen**) beschäftigen.

Wird die endliche Realisation y_1, \dots, y_n eines schwach stationären Prozesses (Y_t) beobachtet, so läßt sich mittels des **Box-Jenkins-Verfahrens** zunächst der zugehörige Prozeß **identifizieren**, d. h. man überprüft, welches der erwähnten Zeitreihenmodelle adäquat ist, sodann lassen sich die **Prozeß-Parameter schätzen** und **Prognosen für zukünftige Beobachtungen** y_{n+1}, y_{n+2}, \dots bestimmen. Die Vorgehensweise wird in Abschnitt 2.5 beschrieben und in Abschnitt 2.6 anhand eines Beispiels demonstriert.

2.1. Autoregressive Prozesse (AR-Prozesse)

Ein Prozeß (Y_t) mit

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad \text{für alle } t,$$

wobei ϕ_1, \dots, ϕ_p die Parameter des Prozesses, $\phi_j \neq 0$ für $j = 1, \dots, p$, und der Prozeß der ε_t ein weißes Rauschen mit $E\varepsilon_t = 0$ und $\text{Var } \varepsilon_t = \sigma_\varepsilon^2$ sind, heißt **autoregressiver Prozeß der Ordnung p** (kurz: AR(p)-Prozeß). Y_t wird beim AR(p)-Prozeß also durch ein gewogenes Mittel seiner p Vorgänger Y_{t-1}, \dots, Y_{t-p} und einen zufälligen Rest erklärt. Jeder autoregressive Prozeß ist natürlich invertierbar; damit der AR(p)-Prozeß (Y_t) auch schwach stationär ist, muß zusätzlich gefordert werden, daß die Lösungen z , die auch komplex (vgl. Abschnitt 3.1) sein können, der Gleichung

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0$$

dem Betrage nach größer als 1 sind.

Die **Autokovarianzen** $\gamma(k)$ zum lag k eines AR(p)-Prozesses lassen sich für $k = 0, 1, 2, \dots$ im Falle $EY_t = 0$ für alle t rekursiv aus den sogenannten **Yule-Walker-Gleichungen** bestimmen

$$\begin{aligned}\gamma(0) &= \sum_{j=1}^p \phi_j \gamma(j) + \sigma_\varepsilon^2, \\ \gamma(k) = \gamma(-k) &= \sum_{j=1}^p \phi_j \gamma(k-j) \quad \text{für } k = 1, 2, 3, \dots,\end{aligned}$$

so daß für die **Autokorrelationen** $\varrho(k)$ zum lag k natürlich gilt

$$\begin{aligned}\varrho(0) &= 1, \\ \varrho(k) = \varrho(-k) &= \sum_{j=1}^p \phi_j \gamma(k-j)/\gamma(0) \quad \text{für } k = 1, 2, 3, \dots.\end{aligned}$$

Vermittels dieser Autokorrelationen kann die Varianz $\sigma_Y^2 = \gamma(0)$ eines AR(p)-Prozesses auch geschrieben werden als

$$\text{Var } Y_t = \sigma_Y^2 = \gamma(0) = \sigma_\varepsilon^2 / (1 - \phi_1 \varrho(1) - \dots - \phi_p \varrho(p)).$$

Für die **partiellen Autokorrelationen** $\varrho_{k-1}(k)$ zum lag k eines AR(p)-Prozesses gilt

$$\varrho_{k-1}(k) = \varrho_{k-1}(-k) \begin{cases} \neq 0 & \text{für } k = 0, 1, 2, \dots, p \\ = 0 & \text{für } k > p. \end{cases}$$

Beispiel: Wir wollen nun einmal speziell die Prozesse AR(1) und AR(2) betrachten und, ausgehend von den Yule-Walker-Gleichungen, deren Autokovarianzen bzw. Autokorrelationen sowie deren partielle Autokorrelationen zum lag k bestimmen.

Ein **AR(1)-Prozeß** wird durch

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t \quad \text{mit } \phi_1 \neq 0$$

beschrieben und ist für $-1 < \phi_1 < 1$ schwach stationär. Die zugehörigen Yule-Walker-Gleichungen sind

$$\begin{aligned}\gamma(0) &= \phi_1 \gamma(1) + \sigma_\varepsilon^2, \\ \gamma(k) = \gamma(-k) &= \phi_1 \gamma(k-1) \quad \text{für } k = 1, 2, \dots.\end{aligned}$$

Setzt man

$$\gamma(1) = \phi_1 \gamma(0)$$

in die erste Gleichung ein, so ergibt sich

$$\gamma(0) = \phi_1^2 \gamma(0) + \sigma_\varepsilon^2, \quad \text{d.h. } \gamma(0)(1 - \phi_1^2) = \sigma_\varepsilon^2$$

und somit

$$\begin{aligned}\gamma(0) &= \sigma_\epsilon^2 / (1 - \phi_1^2), \quad \gamma(1) = \phi_1 \gamma(0), \quad \gamma(2) = \phi_1 \gamma(1) = \phi_1^2 \gamma(0), \\ \gamma(3) &= \phi_1 \gamma(2) = \phi_1^3 \gamma(0), \dots.\end{aligned}$$

Für die Autokovarianzen eines AR(1)-Prozesses gilt also

$$\begin{aligned}\gamma(0) &= \sigma_\epsilon^2 / (1 - \phi_1^2), \\ \gamma(k) &= \gamma(-k) = \phi_1^k \gamma(0) \quad \text{für } k = 1, 2, \dots,\end{aligned}$$

so daß sich die Autokorrelationen zu

$$\begin{aligned}\varrho(0) &= 1, \\ \varrho(k) &= \varrho(-k) = \gamma(k) / \gamma(0) = \phi_1^k \gamma(0) / \gamma(0) = \phi_1^k\end{aligned}$$

ergeben. Die partiellen Autokorrelationen eines AR(1)-Prozesses sind für $k > 1$ gerade 0, für $k = 0$ gerade 1 und für $k = 1$ gerade gleich $\varrho(1) = \phi_1$, vgl. die Ausführungen in Abschnitt 1.4.

Für einen AR(2)-Prozeß

$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$ mit $\phi_1 \neq 0, \phi_2 \neq 0$,
der schwach stationär ist, falls

$$\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1 \quad \text{und} \quad -1 < \phi_2 < 1 \quad \text{gilt,}$$

erhalten wir die Yule-Walker-Gleichungen

$$\begin{aligned}\gamma(0) &= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma_\epsilon^2, \\ \gamma(k) &= \gamma(-k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) \quad \text{für } k = 1, 2, \dots;\end{aligned}$$

insbesondere ist also

$$\gamma(1) = \phi_1 \gamma(0) + \phi_2 \gamma(-1) = \phi_1 \gamma(0) + \phi_2 \gamma(1),$$

bzw.

$$\gamma(1) = \phi_1 \gamma(0) / (1 - \phi_2),$$

und somit

$$\begin{aligned}\gamma(2) &= \phi_1 \gamma(1) + \phi_2 \gamma(0) = \phi_1^2 \gamma(0) / (1 - \phi_2) + \phi_2 \gamma(0) \\ &= \gamma(0)(\phi_2 + \phi_1^2 / (1 - \phi_2)).\end{aligned}$$

Setzen wir diese Ausdrücke für $\gamma(1)$ und $\gamma(2)$ in die Yule-Walker-Gleichung für $\gamma(0)$ ein, so erhalten wir

$$\begin{aligned}\gamma(0) &= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma_\epsilon^2 \\ &= \phi_1^2 \gamma(0) / (1 - \phi_2) + \phi_2 \gamma(0)(\phi_2 + \phi_1^2 / (1 - \phi_2)) + \sigma_\epsilon^2 \\ &= \gamma(0)(\phi_1^2 / (1 - \phi_2) + \phi_2^2 + \phi_1^2 \phi_2 / (1 - \phi_2)) + \sigma_\epsilon^2 \\ &= \gamma(0)(\phi_1^2 + \phi_2^2(1 - \phi_2) + \phi_1^2 \phi_2) / (1 - \phi_2) + \sigma_\epsilon^2\end{aligned}$$

und somit

$$\gamma(0) = \frac{(1 - \phi_2) \sigma_\epsilon^2}{1 - \phi_2 - \phi_1^2 - \phi_2^2(1 - \phi_2) - \phi_1^2 \phi_2} = \frac{(1 - \phi_2) \sigma_\epsilon^2}{(1 + \phi_2)((1 - \phi_2)^2 - \phi_1^2)}.$$

Natürlich ist wieder

$$\begin{aligned}\varrho(0) &= 1 \quad \text{und} \quad \varrho(k) = \varrho(-k) = \gamma(k) / \gamma(0) = \phi_1 \varrho(k-1) + \phi_2 \varrho(k-2) \\ &\quad \text{für } k = 1, 2, \dots\end{aligned}$$

und für die partiellen Autokorrelationen zum lag k erhalten wir den Wert 1 für $k = 0$, den Wert $\varrho(1)$ für $k = 1$ und den Wert 0 für $k > 2$. Für $k = 2$ ergibt sich, vgl. Abschnitt 1.4, mit

$$P_1 = P_1^T = \tilde{P}_1 = \tilde{P}_1^T = \varrho(1) \quad \text{und} \quad P_{1,1} = 1 = P_{1,1}^{-1}$$

gerade

$$\varrho_1(2) = \frac{\varrho(2) - \varrho(1) \cdot 1 \cdot \varrho(1)}{1 - \varrho(1) \cdot 1 \cdot \varrho(1)} = \frac{\varrho(2) - \varrho^2(1)}{1 - \varrho^2(1)}.$$

2.2. Moving average Prozesse (MA-Prozesse)

Ein einseitiger, gewogener gleitender Durchschnitt der Ordnung q eines white-noise Prozesses (ε_t) mit $E\varepsilon_t = 0$, $\text{Var } \varepsilon_t = \sigma_\varepsilon^2$ und $E(\varepsilon_t \varepsilon_{t'}) = 0$ für $t \neq t'$, also der stets schwachstationäre Prozeß (Y_t) mit

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad EY_t = 0,$$

wobei die Parameter $\theta_1, \dots, \theta_q$ des Prozesses alle von Null verschieden sind, heißt **moving average Prozeß der Ordnung q** (kurz: **MA(q)-Prozeß**).

Die **Autokovarianzen** $\gamma(k)$ zum lag k eines MA(q)-Prozesses ergeben sich mit $\theta_0 = 1$ wegen der Unkorriertheit der Störterme ε_t zu

$$\gamma(k) = \gamma(-k) = \begin{cases} 0, & \text{für } k > q \\ \sigma_\varepsilon^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k}, & \text{für } k = 0, 1, \dots, q \end{cases},$$

so daß insbesondere

$$\text{Var } Y_t = \sigma_Y^2 = \gamma(0) = \sigma_\varepsilon^2 \sum_{j=0}^q \theta_j^2$$

gilt. Damit lassen sich die **Autokorrelationen** $\varrho(k)$ zum lag k eines MA(q)-Prozesses schreiben als

$$\varrho(k) = \varrho(-k) = \begin{cases} 0, & \text{für } k > q \\ 1, & \text{für } k = 0 \\ \sum_{j=0}^{q-k} \theta_j \theta_{j+k} / \sum_{j=0}^q \theta_j^2, & \text{für } k = 1, \dots, q \end{cases}$$

Beispiel: Wir wollen die Autokorrelationen $\varrho(k)$ zum lag k eines MA(1)-Prozesses bestimmen. Es ergibt sich

$$\begin{aligned} \varrho(0) &= 1, \\ \varrho(1) &= \theta_0 \theta_1 / (\theta_0^2 + \theta_1^2) = \theta_1 / (1 + \theta_1^2) = \varrho(-1), \\ \varrho(k) &= \varrho(-k) = 0 \quad \text{für } k = 2, 3, 4, \dots. \end{aligned}$$

Da gilt

$$\varrho(1) = \theta_1 / (1 + \theta_1^2) = \frac{\theta_1 / \theta_1^2}{(1 + \theta_1^2) / \theta_1^2} = \frac{1 / \theta_1}{1 / \theta_1^2 + 1} = \frac{1 / \theta_1}{1 + (1 / \theta_1)^2},$$

sind die Autokorrelationen eines MA(1)-Prozesses mit Parameter θ_1 identisch mit

denen eines MA(1)-Prozesses mit Parameter $1/\theta_1$. Ein MA(1)-Prozeß läßt sich also nicht eindeutig aufgrund seiner Autokorrelationen identifizieren, es sei denn, man fordert etwa

$$-1 < \theta_1 < 1.$$

Wie im Beispiel für einen MA(1)-Prozeß gezeigt, sind moving average Prozesse q-ter Ordnung auch allgemein nur dann eindeutig durch ihre Autokorrelationen $\rho(k)$ zum lag k bestimmt, wenn man die Menge der zulässigen Parameterwerte $\theta_1, \dots, \theta_q$ einschränkt. Und zwar muß man fordern, daß für alle Lösungen z , die auch komplex (vgl. Abschnitt 3.1) sein können, der Gleichung

$$1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q = 0$$

gilt: $|z| > 1$. Diese Bedingung garantiert auch die Invertierbarkeit eines MA(q)-Prozesses (Y_t); für $q = 1$ entspricht sie gerade der Forderung $-1 < \theta_1 < 1$.

An dieser Stelle sei erwähnt, daß sich jeder schwach stationäre AR-Prozeß (Y_t) auch als MA(∞)-Prozeß, d.h. in der Form

$$Y_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j}$$

darstellen läßt, wobei für die Parameter $\beta_0, \beta_1, \beta_2, \dots$ gilt:

$$\sum_{j=0}^{\infty} |\beta_j| < \infty.$$

Umgekehrt läßt sich jeder invertierbare MA-Prozeß (Y_t) auch als AR(∞)-Prozeß in der Form

$$Y_t = \sum_{j=1}^{\infty} \alpha_j Y_{t-j} + \varepsilon_t$$

mit Parametern $\alpha_1, \alpha_2, \dots$ darstellen.

2.3. Gemischte Prozesse (ARMA-Prozesse)

Ein Prozeß (Y_t), der sich aus einem AR(p)- und einem MA(q)-Prozeß in der Form

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

zusammensetzt, wobei ϕ_1, \dots, ϕ_p die p von Null verschiedenen autoregressiven Parameter, $\theta_1, \dots, \theta_q$ die q von Null verschiedenen moving average Parameter bezeichnen und die ε_t mit $E\varepsilon_t = 0$, $\text{Var } \varepsilon_t = \sigma_\varepsilon^2$ ein weißes Rauschen bilden, heißt ein **gemischter Prozeß mit Ordnungen p und q** (kurz: ARMA(p, q)-Prozeß). Damit die Y_t einen schwach stationären Prozeß bilden, müssen wir weiterhin verlangen, daß die Lösungen z der Gleichung

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0$$

betragmäßig größer als 1 sind. Die Invertierbarkeit des Prozesses wird durch die zusätzliche Forderung, daß die Lösungen \tilde{z} der Gleichung

$$1 + \theta_1 \tilde{z} + \theta_2 \tilde{z}^2 + \dots + \theta_q \tilde{z}^q = 0$$

ebenfalls $|\tilde{z}| > 1$ erfüllen. Die Lösungen z bzw. \tilde{z} können auch komplex, vgl. Abschnitt 3.1, sein.

Gilt $EY_t = 0$ für alle t und setzt man $\theta_0 = 1$, so lassen sich die **Autokovarianzen**

$\gamma(k)$ zum lag k eines ARMA(p, q)-Prozesses in der Form

$$\gamma(k) = \gamma(-k) = \sum_{j=1}^p \phi_j \gamma(k-j) + \sum_{j=0}^q \phi_j \gamma_{Y_t}(k-j)$$

darstellen, wobei $\gamma_{Y_t}(k)$ die Kovarianz zwischen Y_{t-k} und ε_t angibt. Für diese Kovarianzen zum lag k gilt

$$\gamma_{Y_t}(k) = E(Y_{t-k} \cdot \varepsilon_t) \begin{cases} = 0 & \text{für } k > 0 \\ \neq 0 & \text{für } k \leq 0 \end{cases}$$

Für einen ARMA(p, q)-Prozeß mit speziell gewählten p und q lassen sich anhand dieser Angaben die Autokovarianzen $\gamma(k)$ rekursiv bestimmen, wobei insbesondere ausgenutzt wird, daß für $k > q$ gilt:

$$\gamma(k) = \gamma(-k) = \sum_{j=1}^p \phi_j \gamma(k-j).$$

Die benötigten Kovarianzen $\gamma_{Y_t}(k)$ werden hier unter Ausnutzung der Linearität des Erwartungswertes durch Multiplikation der Prozeßgleichung

$$Y_{t-k} = \phi_1 Y_{t-k-1} + \dots + \phi_p Y_{t-k-p} + \varepsilon_{t-k} + \theta_1 \varepsilon_{t-k-1} + \dots + \theta_q \varepsilon_{t-k-q}$$

mit ε_t bestimmt.

Die Autokorrelationen $\rho(k)$ zum lag k ergeben sich bei einem ARMA(p, q)-Prozeß natürlich gerade zu

$$\begin{aligned} \rho(0) &= 1, \\ \rho(k) &= \rho(-k) = \gamma(k)/\gamma(0) \quad \text{für } k = 1, 2, 3, \dots \end{aligned}$$

Es sei an dieser Stelle noch erwähnt, daß ein reiner AR(p)-Prozeß oft auch als ARMA(p, 0)-Prozeß und ein reiner MA(q)-Prozeß oft als ARMA(0, q)-Prozeß bezeichnet wird.

Beispiel: Wir betrachten hier einmal den einfachsten echt gemischten Prozeß, also den ARMA(1, 1)-Prozeß ($p = 1, q = 1$) mit

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1},$$

und berechnen die Autokorrelationen $\gamma(k)$ eines solchen Prozesses. Wir erhalten zunächst

$$\begin{aligned} \gamma(0) &= \phi_1 \gamma(-1) + \gamma_{Y_t}(0) + \theta_1 \gamma_{Y_t}(-1) \\ &= \phi_1 \gamma(1) + \gamma_{Y_t}(0) + \theta_1 \gamma_{Y_t}(-1) \\ \gamma(1) &= \phi_1 \gamma(0) + \gamma_{Y_t}(1) + \theta_1 \gamma_{Y_t}(0) \\ &= \phi_1 \gamma(0) + \theta_1 \gamma_{Y_t}(0), \\ \gamma(k) &= \phi_1 \gamma(k-1) \quad \text{für } k = 2, 3, \dots \end{aligned}$$

und berechnen dann

$$\begin{aligned} \gamma_{Y_t}(0) &= E(Y_t \cdot \varepsilon_t) = E(\phi_1 Y_{t-1} \varepsilon_t + \varepsilon_t^2 + \theta_1 \varepsilon_{t-1} \varepsilon_t) \\ &= \phi_1 E(Y_{t-1} \varepsilon_t) + E(\varepsilon_t^2) + \theta_1 E(\varepsilon_{t-1} \varepsilon_t). \end{aligned}$$

Da (ε_t) ein weißes Rauschen ist, gilt

$$E(\varepsilon_t^2) = \text{Var } \varepsilon_t = \sigma_\varepsilon^2 \quad \text{und} \quad E(\varepsilon_{t-1} \cdot \varepsilon_t) = 0 \quad \text{für alle } t,$$

so daß wir

$$\gamma_{Y_t}(0) = \phi_1 \gamma_{Y_t}(1) + \sigma_e^2 + \theta_1 \cdot 0 = \sigma_e^2$$

erhalten. Weiter bestimmen wir nun

$$\begin{aligned}\gamma_{Y_t}(-1) &= E(Y_{t+1} \cdot \varepsilon_t) = E(\phi_1 Y_t \varepsilon_t + \varepsilon_{t+1} \varepsilon_t + \theta_1 \varepsilon_t^2) \\ &= \phi_1 E(Y_t \varepsilon_t) + E(\varepsilon_{t+1} \varepsilon_t) + \theta_1 E(\varepsilon_t^2) \\ &= \phi_1 \gamma_{Y_t}(0) + 0 + \theta_1 \sigma_e^2 \\ &= \phi_1 \sigma_e^2 + \theta_1 \sigma_e^2,\end{aligned}$$

und setzen $\gamma_{Y_t}(0)$ sowie $\gamma_{Y_t}(-1)$ in obige Gleichungen für $\gamma(0)$ und $\gamma(1)$ ein:

$$\begin{aligned}\gamma(0) &= \phi_1 \gamma(1) + \sigma_e^2 + \theta_1 (\phi_1 \sigma_e^2 + \theta_1 \sigma_e^2), \\ \gamma(1) &= \phi_1 \gamma(0) + \theta_1 \sigma_e^2.\end{aligned}$$

Die zweite Gleichung in die erste eingesetzt liefert dann

$$\begin{aligned}\gamma(0) &= \phi_1 (\phi_1 \gamma(0) + \theta_1 \sigma_e^2) + \sigma_e^2 + \theta_1 (\phi_1 \sigma_e^2 + \theta_1 \sigma_e^2) \\ &= \phi_1^2 \gamma(0) + (2\phi_1 \theta_1 + 1 + \theta_1^2) \sigma_e^2\end{aligned}$$

und somit

$$\gamma(0) = \frac{2\phi_1 \theta_1 + 1 + \theta_1^2}{1 - \phi_1^2} \sigma_e^2,$$

was dann

$$\gamma(1) = \frac{\phi_1 (2\phi_1 \theta_1 + 1 + \theta_1^2)}{1 - \phi_1^2} \sigma_e^2 + \theta_1 \sigma_e^2 = \frac{(1 + \phi_1 \theta_1)(\phi_1 + \theta_1)}{1 - \phi_1^2} \sigma_e^2$$

ergibt.

So wie Beziehungen zwischen MA- und AR-Prozessen bestehen, vgl. Abschnitt 2.2, lassen sich auch Beziehungen zwischen ARMA-Prozessen und AR- sowie MA-Prozessen herstellen. Jeder stationäre ARMA(p, q)-Prozeß läßt sich als MA(∞)-Prozeß mit Parametern $\beta_0, \beta_1, \beta_2, \dots$ derart, daß $\sum_{j=0}^{\infty} |\beta_j| < \infty$ gilt, darstellen und jeder (stationäre) invertierbare ARMA(p, q)-Prozeß hat eine (stationäre) AR(∞)-Darstellung.

2.4. Instationäre stochastische Prozesse, ARIMA- und SARIMA-Prozesse, Box-Cox-Transformation

Ein stochastischer Prozeß (Y_t), der nicht schwach stationär ist, läßt sich vermittels **linearer Filter** vielfach in einen schwach stationären Prozeß (Y_t^*) überführen. Die linearen Filter für konkrete Zeitreihen, die wir im Abschnitt 1 kennengelernt haben, lassen sich natürlich in gleicher Art und Weise auch auf einen stochastischen Prozeß (Y_t) anwenden. Von besonderem Interesse sind hier die nichtsaisonalen und saisonalen Differenzenfilter aus Abschnitt 1.3.3, vgl. auch Abschnitt 1.3.5.

Ist (Y_t) ein nicht schwach stationärer Prozeß derart, daß die d-ten Differenzen

$$Y_t^* = \nabla^d Y_t$$

einen schwach stationären ARMA(p, q)-Prozeß bilden, $p, q \geq 0$, so heißt der Prozeß (Y_t) **integrierter ARMA(p, q)-Prozeß**; man schreibt auch: (Y_t) ist ein **ARIMA(p, d, q)-Prozeß**, wobei p die Ordnung des autoregressiven Teils, q die Ordnung des moving average Teils und d die Differenzenordnung angibt. Ein ARMA(p, q)-Prozeß wird oft auch ARIMA(p, 0, q)-Prozeß genannt.

Beispiel: Für einen ARIMA(1,1,1)-Prozeß (Y_t) ergibt sich die Darstellung

$$\begin{aligned}\nabla Y_t &= Y_t^* = Y_t - Y_{t-1} = \phi_1 \nabla Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \\ &= \phi_1 Y_{t-1}^* + \varepsilon_t + \theta_1 \varepsilon_{t-1} \\ &= \phi_1 Y_{t-1} - \phi_1 Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1},\end{aligned}$$

für einen ARIMA(0,1,1)-Prozeß ergibt sich

$$\nabla Y_t = Y_t^* = Y_t - Y_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1},$$

für einen ARIMA(1,1,0)-Prozeß erhalten wir

$$\begin{aligned}\nabla Y_t &= Y_t^* = Y_t - Y_{t-1} = \phi_1 \nabla Y_{t-1} + \varepsilon_t = \phi_1 Y_{t-1}^* + \varepsilon_t \\ &= \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + \varepsilon_t\end{aligned}$$

und schließlich erhalten wir zum Beispiel für einen ARIMA(0,2,2)-Prozeß

$$\begin{aligned}\nabla^2 Y_t &= \nabla(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2} = Y_t^* \\ &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2};\end{aligned}$$

hier bezeichnete (ε_t) natürlich stets ein weißes Rauschen.

Auch ein Prozeß (Y_t) mit saisonaler Komponente ist natürlich nicht schwach stationär. Besitzt er jedoch eine konstante Saisonfigur mit Periode s , so ist der Prozeß (Y_t^*) mit

$$Y_t^* = Y_t - Y_{t-s}$$

schwach stationär. Ist (Y_t^*) ein ARMA(p, q)-Prozeß, so heißt (Y_t) ein **saisonaler ARMA(p, q)-Prozeß** mit Periode s oder kurz **SARMA_s(p, q)-Prozeß**.

Natürlich treten Trend und Saison mitunter auch gemeinsam auf. Einen solchen Prozeß (Y_t), für den die $Y_t^* = \nabla^d(Y_t - Y_{t-s})$ einen ARMA(p, q)-Prozeß bilden, nennt man dann auch einen **SARIMA_s(p, d, q)-Prozeß**.

Komplexere saisonale ARIMA-Modelle wie z. B. das **multiplikativ-saisonale ARIMA-Modell** wollen wir hier nicht behandeln; der interessierte Leser sei verwiesen auf Box/Jenkins (1976).

Wir haben gesehen, daß man durch Differenzenbildung beim Prozeß (Y_t) Trend- und Saisonkomponente eliminieren kann, um einen Prozeß (Y_t^*) mit $EY_t^* = 0$ bzw. $EY_t = \mu_Y$ für alle t zu erhalten. Bei der schwachen Stationarität werden aber auch gleiche Varianzen σ_Y^2 der Y_t verlangt, d. h. (Y_t) muß homoskedastisch sein. Häufig wachsen jedoch bei einem Prozeß (Y_t) die Schwankungen im Zeitverlauf, d. h. die Varianzen der Y_t werden mit wachsendem t immer größer. In diesem Fall kann eine sogenannte **Box-Cox-Transformation** der Y_t (bzw. der Y_t^*) bewirken, daß der transformierte Prozeß (\tilde{Y}_t) mit

$$\tilde{Y}_t = \begin{cases} \frac{(Y_t + m)^\lambda - 1}{\lambda} & \text{für } \lambda \neq 0 \\ \ln(Y_t + m) & \text{für } \lambda = 0 \end{cases}$$

konstante Varianz $\sigma_{\tilde{Y}}^2$ besitzt, vgl. Box/Cox (1964) und Abschnitt 4.2 und 4.5 in Kap. V. Hier wird m so gewählt, daß gilt: $Y_t + m > 0$ für alle t , und die Wahl des Parameters λ der Box-Cox-Transformation erfolgt derart, daß der Prozeß (\tilde{Y}_t) möglichst homogen ist; häufig wählt man $\lambda = 0$ oder $\lambda = 0,5$. (\tilde{Y}_t) stellt dann einen ARIMA- oder SARIMA-Prozeß dar (bei ‚richtiger‘ Transformation).

2.5. Die Methode von Box und Jenkins

Hat man eine Zeitreihe y_1, \dots, y_n , also eine endliche Realisation eines stochastischen Prozesses (Y_t) beobachtet, so kann an die Zeitreihe ein Prozeß aus der Klasse der bisher behandelten, selbsterklärenden Zeitreihenmodelle angepaßt werden. Wir wollen hier einmal davon ausgehen, daß die Zeitreihe y_1, \dots, y_n mittels der Methoden aus Abschnitt 2.4 derart transformiert wurde, daß sie **mittelwertbereinigt, trend- und saisonbereinigt** sowie **homoskedastisch** ist. Wir beschäftigen uns also mit der Anpassung eines ARMA(p, q)-Prozesses mit $p, q \geq 0$; der Fall $p = q = 0$ würde natürlich dem Fall entsprechen, daß die Zeitreihe y_1, \dots, y_n die Realisation eines weißen Rauschens ist.

Wir haben im Abschnitt 2.2 bzw. 2.3 gesehen, daß sich jeder stationäre, invertierbare ARMA(p, q)-Prozeß auch als MA(∞)- oder AR(∞)-Prozeß darstellen läßt. Wir könnten also auf die Anpassung vom ARMA-Prozeß an die Zeitreihe y_1, \dots, y_n verzichten und statt dessen nur reine AR- oder MA-Prozesse betrachten. Box und Jenkins, vgl. etwa Box/Jenkins (1976), plädieren jedoch nach dem **Parsinomitätsprinzip (Sparsamkeitsprinzip, principle of parsimony)** für ARMA-Prozesse, da dann weniger Parameter notwendig sind, um die Zeitreihe y_1, \dots, y_n zu erklären.

Wird durch die Prozesse AR(p'), MA(q') und ARMA(p, q) eine in etwa gleiche Anpassungsgüte (Erklärungsgüte) für eine Zeitreihe y_1, \dots, y_n erreicht, so gilt stets

$$p + q \leq p' \quad \text{und} \quad p + q \leq q'.$$

Das Ziel, möglichst wenige Parameter zur Erklärung von y_1, \dots, y_n zu benötigen, führt daher, daß man aufgrund der Beobachtungen y_1, \dots, y_n auch zukünftige Werte y_{n+1}, y_{n+2}, \dots der Zeitreihe prognostizieren möchte. Solche Prognosen sind aber umso besser, je weniger Parameterschätzungen in sie einfließen.

Die Methode von **Box und Jenkins** zur Analyse einer Zeitreihe y_1, \dots, y_n vermittels eines ARMA(p, q)-Prozesses gliedert sich nun in vier Teile. Zunächst wird ein adäquates **Modell identifiziert**, d. h. die Ordnungen p und q des ARMA(p, q)-Prozesses werden festgelegt. Sodann werden die **Parameter $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$** des Prozesses geschätzt und die **Modelladäquatheit** wird basierend auf den Schätzungen überprüft. Der vierte und letzte Teil des Box-Jenkins-Verfahrens dient dann der **Prognose zukünftiger Werte** sowie der **Güteprüfung der Prognoseeigenschaften** des gewählten Modells.

2.5.1. Modellidentifikation

Die Problemstellung der Modellidentifikation ist die Bestimmung der Ordnungen p und q des autoregressiven und des moving average Teils des ARMA-Prozesses. Die Festlegung von p und q kann basierend auf den empirischen Autokorrelationen $r(k)$ und partiellen Autokorrelationen $r_{k-1}(k)$ zum lag k, $k = 0, 1, 2, \dots, n - 1$, vgl. Abschnitt 1.4 dieses Kapitels, erfolgen.

Die **Ordnung eines reinen AR-Prozesses** kann mit Hilfe der empirischen partiellen Autokorrelationen erfolgen. Wie wir im Abschnitt 2.1 gesehen haben, zeichnet sich ein AR(p)-Prozeß dadurch aus, daß die (theoretischen) partiellen Autokorrelationen $\rho_{k-1}(k)$ für $k > p$ verschwinden. Für die empirischen Größen $r_{k-1}(k)$ kann man dies natürlich nicht exakt erwarten. Jedoch läßt sich zeigen, daß die empirischen partiellen Autokorrelationen einen Standardfehler von ungefähr $n^{-1/2}$ für eine Zeitreihe y_1, \dots, y_n haben, d. h. sie können als Null angesehen werden, falls sie im Intervall $[-n^{-1/2}, n^{-1/2}]$ liegen. Die Ordnung p eines AR-Prozesses ist also

gleich dem größten Wert k , für den $r_{k-1}(k)$ „wesentlich“ außerhalb dieses Intervalls liegt.

Die **Ordnung eines reinen MA-Prozesses** lässt sich mittels der empirischen Autokorrelationen $r(k)$ zum lag k für $k = 0, 1, 2, \dots, n - 1$ bzw. des Korrelogramms, vgl. Abschnitt 1.4, festlegen. Wie wir im Abschnitt 2.2 gesehen haben, verschwinden die (theoretischen) Autokorrelationen $\rho(k)$ bei einem MA(q)-Prozeß für $k > q$. Bei den empirischen Autokorrelationen $r(k)$ zur Zeitreihe y_1, \dots, y_n werden die Ergebnisse natürlich nicht so ideal sein. Jedoch ist beim MA(q)-Prozeß der Standardfehler der $r(k)$ beim Stichprobenumfang n approximativ gegeben durch

$$SF_q = \left(\frac{1}{n} (1 + 2 \sum_{j=1}^q \rho^2(j)) \right)^{1/2} \quad \text{für } k > q.$$

Die Festlegung der **Ordnungen p und q eines ARMA(p, q)-Prozesses** anhand von empirischen Autokorrelationen und empirischen partiellen Autokorrelationen ist nicht so einfach, da die $r(k)$ und $r_{k-1}(k)$ nicht getrennt betrachtet werden können. Vielmehr beeinflusst der MA-Teil ja auch die partiellen Autokorrelationen und der AR-Teil die Autokorrelationen. Eine Identifikation erfordert daher Erfahrung mit der wechselseitigen Beeinflussung des Verhaltens von Autokorrelationen und partiellen Autokorrelationen. Bei einem ARMA(1,1)-Prozeß etwa nehmen die Autokorrelationen $\rho(k)$ vom ersten lag an exponentiell ab und auch die partiellen Autokorrelationen $\rho_{k-1}(k)$ werden vom lag 1 an von exponentieller Abnahme dominiert. Beim ARMA(1,0)-Prozeß nimmt $\rho(k)$ exponentiell ab und beim ARMA(2,0)-Prozeß haben wir eine exponentielle Abnahme der $\rho(k)$, oder $\rho(k)$ verläuft betrachtet als Funktion von k in gedämpften Sinus-Schwingungen. Ähnlich wie $\rho(k)$ bei ARMA(1,0)- bzw. ARMA(2,0)-Prozessen verhält sich $\rho_{k-1}(k)$ beim ARMA(0,1)- bzw. ARMA(0,2)-Prozeß.

Andere Möglichkeiten zur Festlegung der Ordnungen p und q eines ARMA-Prozesses basieren auf iterativen Schätzprozeduren. Nehmen wir an, daß aufgrund einer endlichen Realisation, einer Zeitreihe y_1, \dots, y_n die Varianz σ_e^2 des weißen Rauschens des zugrundegelegten ARMA(p, q)-Prozesses (Y_t) durch den Maximum-Likelihood-Schätzer $\hat{\sigma}_{p,q}^2$ geschätzt wird; eine solche Schätzung wird für alle „sinnvoll erscheinenden“ Paare (p, q) durchgeführt. Dann schlägt z. B. Akaike (1977) vor, diejenigen Ordnungen p_0 und q_0 des AR- bzw. MA-Teils zu wählen, für die das Kriterium

$$BIC(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{1}{n} (p + q) \ln(n)$$

minimal wird. Hannan/Quinn (1979) etwa schlagen vor, p_0 und q_0 so festzulegen, daß

$$\phi(p, q) = \ln \hat{\sigma}_{p,q}^2 + \frac{1}{n} (p + q) \cdot c \cdot \ln(\ln(n))$$

minimal wird, wobei $c > 2$ eine wählbare Konstante ist. Diese Kriterien führen zu konsistenten Schätzungen für die Ordnungen p und q , wohingegen die Schätzungen aufgrund des Kriteriums

$$AIC(p, q) = \log \hat{\sigma}_{p,q}^2 + \frac{2}{n} (p + q),$$

das von Akaike (1974) vorgeschlagen wurde, dazu tendieren, die Ordnungen zu überschätzen.

2.5.2. Schätzen der Modellparameter

Hat man die Ordnungen p und q eines für die Zeitreihe y_1, \dots, y_n adäquaten ARMA-Prozesses festgelegt, so müssen die Modellparameter $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ dieses Prozesses

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

durch Schätzungen $\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q$ derart ersetzt werden, daß eine möglichst gute Anpassung im Sinne der Methode der kleinsten Quadrate, vgl. Kap. X, an die Zeitreihendaten erreicht wird.

D.h. die Residuen-Quadratsumme

$$S^2(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \sum_{t=-\infty}^n \tilde{\varepsilon}_t^2$$

mit den Residuen

$$\tilde{\varepsilon}_t = y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} - \theta_1 \tilde{\varepsilon}_{t-1} - \dots - \theta_q \tilde{\varepsilon}_{t-q}$$

muß bzgl. $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ minimiert werden. Problematisch hierbei ist, daß zur Bestimmung der $\tilde{\varepsilon}_t$ nichtbeobachtbare Startwerte $\tilde{\varepsilon}_t, y_t$ für $t \leq 0$ erforderlich sind.

Beim **bedingten Least-Squares-Ansatz (conditional least squares, CLS)** setzt man nun

$$\tilde{\varepsilon}_t = y_t = 0 \quad \text{für } t \leq 0$$

und erhält

$$S^2(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \sum_{t=1}^n \tilde{\varepsilon}_t^2,$$

wobei die $\tilde{\varepsilon}_t$ aus der Rekursion

$$\begin{aligned} \tilde{\varepsilon}_1 &= y_1 \\ \tilde{\varepsilon}_2 &= y_2 - \phi_1 y_1 - \theta_1 \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_3 &= y_3 - \phi_1 y_2 - \phi_2 y_1 - \theta_1 \tilde{\varepsilon}_2 - \theta_2 \tilde{\varepsilon}_1 \\ &\vdots \\ \tilde{\varepsilon}_j &= y_j - \phi_1 y_{j-1} - \dots - \phi_p y_{j-p} - \theta_1 \tilde{\varepsilon}_{j-1} - \dots - \theta_q \tilde{\varepsilon}_{j-q} \end{aligned} \quad \text{für } j = \max(p, q), \dots, n$$

berechnet werden können. Die Minimierung von $S^2(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ erfolgt dann mittels numerischer Methoden (z. B. Newton-Raphson-, Marquardt-Algorithmus), vgl. z. B. Box/Jenkins (1976). Den Minimalwert $S^2(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ wollen wir im folgenden mit s^2 bezeichnen, die zugehörigen Residuen mit $\hat{\tilde{\varepsilon}}_t$ für alle t .

Der **unbedingte Least-Squares-Ansatz (unconditional least squares, ULS)** hingegen geht davon aus, daß für genügend großes r approximativ gilt

$$S^2(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) = \sum_{t=-\infty}^n \tilde{\varepsilon}_t^2 = \sum_{t=1-r}^n \tilde{\varepsilon}_t^2.$$

Zunächst werden dann ausgehend vom Startwert $\tilde{\varepsilon}_{t,0} = 0$ für die Residuen $\tilde{\varepsilon}_t$ ($t \leq 0$) aus den Daten y_1, \dots, y_n rückwärts Schätzer $\hat{y}_{t,0}$ für y_t ($t = 0, -1, \dots, 1-r-p$) bestimmt. Mit diesen werden dann die Residuen im ersten Schritt

$\tilde{\varepsilon}_{1-r,1}, \dots, \tilde{\varepsilon}_{n,1}$ berechnet. Durch Minimierung von $S^2(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ bzgl. der Parameter erhält man Schätzungen $\hat{\phi}_{1,1}, \dots, \hat{\phi}_{p,1}, \hat{\theta}_{1,1}, \dots, \hat{\theta}_{q,1}$ und geschätzte Residuen $\hat{\varepsilon}_{1-r,1}, \dots, \hat{\varepsilon}_{n,1}$, bestimmt rückwärts neue Schätzer $\hat{y}_{t,1}$ für y_t ($t = 0, -1, \dots, 1-r-p$), usw. bis das Verfahren konvergiert, d.h. bis S^2 sich nicht mehr „wesentlich“ verringert. Ist dies im k -ten Schritt der Fall, so setzt man

$$\hat{\phi}_j = \hat{\phi}_{j,k} \quad \text{für } j = 1, \dots, p, \quad \hat{\theta}_j = \hat{\theta}_{j,k} \quad \text{für } j = 1, \dots, q$$

und

$$s^2 = S^2(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q);$$

man vgl. hierzu auch Box/Jenkins (1976) und z.B. Schlittgen/Streitberg (1984).

Ein Schätzer für die Varianz σ_ϵ^2 des weißen Rauschens (ϵ_t) ist natürlich bei beiden Ansätzen gegeben durch

$$\hat{\sigma}_\epsilon^2 = s^2/(n - p - q).$$

Wir wollen nun noch eine einfache Methode zur Bestimmung **approximativer Kleinst-Quadrat-Schätzer** $\hat{\phi}_1, \dots, \hat{\phi}_p$ für die Parameter eines stationären ARMA($p, 0$)- bzw. AR(p)-Prozesses angeben. Erinnern wir uns einmal an die Yule-Walker-Gleichungen, vgl. Abschnitt 2.1, mit deren Hilfe sich rekursiv die Autokovarianzen $\gamma(k)$ eines AR(p)-Prozesses bestimmen lassen. Die Yule-Walker-Gleichungen kann man natürlich ebensogut direkt für die Autokorrelationen $\varrho(k)$ zum lag k angeben:

$$\varrho(0) = 1, \quad \varrho(k) = \varrho(-k) = \sum_{j=1}^p \phi_j \varrho(k-j) \quad \text{für } k = 1, 2, \dots.$$

Betrachtet man nun die ersten p Yule-Walker-Gleichungen für $k = 1, \dots, p$ und ersetzt die Autokorrelationen $\varrho(k)$ bzw. $\varrho(-k)$ durch die empirischen Autokorrelationen $r(k)$ einer Zeitreihe y_1, \dots, y_n , vgl. Abschnitt 1.4, so ergeben sich p empirische Yule-Walker-Gleichungen

$$\begin{aligned} r(1) &= \phi_1 + \phi_2 r(1) + \phi_3 r(2) + \dots + \phi_p r(p-1) \\ r(2) &= \phi_1 r(1) + \phi_2 + \phi_3 r(1) + \dots + \phi_p r(p-2) \\ r(3) &= \phi_1 r(2) + \phi_2 r(1) + \phi_3 + \dots + \phi_p r(p-3) \\ &\vdots \\ r(p) &= \phi_1 r(p-1) + \phi_2 r(p-2) + \phi_3 r(p-3) + \dots + \phi_p, \end{aligned}$$

die sich in matrzieller Form auch schreiben lassen als

$$\begin{aligned} \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{pmatrix} &= \begin{pmatrix} 1 & r(1) & r(2) & \dots & r(p-1) \\ r(1) & 1 & r(1) & \dots & r(p-2) \\ \vdots & \vdots & \vdots & & \vdots \\ r(p-1) & r(p-2) & r(p-3) & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} \\ &= R(p) \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}. \end{aligned}$$

Löst man dieses Gleichungssystem nach ϕ_1, \dots, ϕ_p auf, so erhält man Schätzer

$\hat{\phi}_1, \dots, \hat{\phi}_p$ für ϕ_1, \dots, ϕ_p . Ist die Matrix $R(p)$ invertierbar, so ist natürlich

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = R^{-1}(p) \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{pmatrix}.$$

Beispiel: Für einen AR(1)-Prozeß erhält man die empirische Yule-Walker-Gleichung

$$r(1) = \hat{\phi}_1,$$

d. h. der Parameter ϕ_1 wird durch die empirische Autokorrelation $r(1)$ der Zeitreihe y_1, \dots, y_n zum lag 1 geschätzt.

Für einen AR(2)-Prozeß ergibt sich

$$r(1) = \hat{\phi}_1 + \hat{\phi}_2 r(1),$$

$$r(2) = \hat{\phi}_1 r(1) + \hat{\phi}_2$$

bzw.

$$\begin{pmatrix} r(1) \\ r(2) \end{pmatrix} = \begin{pmatrix} 1 & r(1) \\ r(1) & 1 \end{pmatrix} \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix},$$

so daß für $r(1) \neq \pm 1$ gilt

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} = \frac{1}{1 - r^2(1)} \begin{pmatrix} 1 & -r(1) \\ -r(1) & 1 \end{pmatrix} \begin{pmatrix} r(1) \\ r(2) \end{pmatrix}.$$

Es ist also

$$\hat{\phi}_1 = \frac{r(1) - r(1) \cdot r(2)}{1 - r^2(1)}, \quad \hat{\phi}_2 = \frac{r(2) - r^2(1)}{1 - r^2(1)}.$$

2.5.3. Modellüberprüfung

Nachdem die Schätzer $\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q$ für die Parameter eines ARMA(p, q)-Prozesses bestimmt worden sind, wird die Güte der Modellanpassung überprüft. Hierzu schlagen Box/Pierce (1970) einen sogenannten **Portmanteau lack-of-fit-test** vor, der allerdings nicht sehr trennscharf ist. Ausgehend von einer Zeitreihe y_1, \dots, y_n wird mittels dieses Tests überprüft, ob die Residuen

$$\hat{\epsilon}_t = y_t - \hat{y}_t = y_t - \hat{\phi}_1 y_{t-1} - \dots - \hat{\phi}_p y_{t-p} - \hat{\theta}_1 \hat{\epsilon}_{t-1} - \dots - \hat{\theta}_q \hat{\epsilon}_{t-q}$$

für $t = 1, \dots, n$

die endliche Realisation eines weißen Rauschens sind. Konkret treten an die Stelle der \hat{y}_t in obiger Formel Ein-Schritt ex-post Prognosen $\hat{Y}_{t-1,1}$ bzw. an die Stelle der Residuen $\hat{\epsilon}_t$ die entsprechenden Prognosefehler, vgl. Abschnitt 2.5.4. Man testet also die Hypothese H_0 , daß die Autokorrelationen

$$\rho_\epsilon(k) = E(\epsilon_t \epsilon_{t+k})$$

zum lag k für $k = 1, 2, \dots$ alle Null sind. Dabei beschränkt man sich stets auf eine Anzahl K von Autokorrelationen, d. h. der Test bezieht nur die empirische Autokorrelation $r_\epsilon(1), \dots, r_\epsilon(K)$ ein. In der Regel wird man ihn dann für verschiedene

maximale Zeitverzögerungen K durchführen. Die Prüfgröße des Tests mit maximalem lag K

$$Q(K) = n \sum_{k=1}^K r_e^2(k)$$

ist unter der Hypothese H_0 approximativ χ^2 -verteilt mit $K - p - q$ Freiheitsgraden, so daß H_0 zum Niveau α verworfen werden muß, falls

$$Q(K) > \chi^2_{K-p-q; 1-\alpha}.$$

Eine trennschärfere, verbesserte Version dieses Tests wurde von Box/Ljung (1978) vorgeschlagen. Anstelle der empirischen Autokorrelationen betrachten sie die normierten empirischen Autokorrelationen

$$\sqrt{\frac{n+2}{n-k}} r_e(k) \quad \text{für } k = 1, \dots, K,$$

d. h. die Hypothese H_0 wird zum Niveau α verworfen, falls

$$Q^*(k) = n(n+2) \sum_{k=1}^K \frac{1}{n-k} r_e^2(k) > \chi^2_{K-p-q; 1-\alpha}.$$

Sollte die Güteprüfung mittels dieses Tests negativ verlaufen, d. h. muß H_0 auch für „kleine“ Niveaus α verworfen werden, so wird man versuchen, ein anderes ARMA(p,q)-Modell mit besserer Anpassungsgüte für die Zeitreihe y_1, \dots, y_n zu finden.

2.5.4. Prognose und Prognosegüte

Ist eine Zeitreihe y_1, \dots, y_n die endliche Realisation eines stationären und invertierbaren ARMA(p,q)-Prozesses (Y_t) mit

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

so dienen Prognoseverfahren zum einen der Vorhersage zukünftiger Beobachtungswerte y_{n+1}, y_{n+2}, \dots , zum anderen auch der Überprüfung der Prognosefähigkeit des Modells.

Wir wollen mit $\hat{Y}_{t,h}$ die sogenannte optimale h-Schrittprognose ($h \geq 1$) von y_{t+h} bezeichnen, wobei $t = 0, 1, \dots, n$ gewählt werden kann. Die Größe

$$\hat{e}_{t,h} = y_{t+h} - \hat{Y}_{t,h},$$

also die Differenz von beobachtetem Wert und optimaler h-Schrittprognose, heißt dann Prognosefehler der optimalen h-Schrittprognose; natürlich kann dieser Prognosefehler nur für $t+h \leq n$ berechnet werden. Die Größe h heißt auch Prognosehorizont.

Wie bestimmt man nun die optimale h-Schrittprognose $\hat{Y}_{t,h}$ eines stationären, invertierbaren ARMA(p,q)-Prozesses? Man betrachtet dazu die Definitionsgleichung des Prozesses für Y_{t+h} , also

$$Y_{t+h} = \phi_1 Y_{t+h-1} + \dots + \phi_p Y_{t+h-p} + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \dots + \theta_q \varepsilon_{t+h-q}$$

und erhält $\hat{Y}_{t,h}$ rekursiv, indem man

- Y_t, Y_{t-1}, \dots durch die Beobachtungsdaten y_t, y_{t-1}, \dots ersetzt,
- Y_{t+h}, \dots, Y_{t+1} durch $\hat{Y}_{t+h}, \dots, \hat{Y}_{t+1}$ ersetzt,

- $\varepsilon_t, \varepsilon_{t-1}, \dots$ durch die Prognosefehler $\hat{\varepsilon}_{t-1,1}, \hat{\varepsilon}_{t-2,1}, \dots$ der optimalen 1-Schrittprognose ersetzt,
- $\varepsilon_{t+h}, \dots, \varepsilon_{t+1}$ durch 0 ersetzt.

In der Praxis werden zunächst für $s = 0, \dots, t-1$ die 1-Schrittprognosen $\hat{Y}_{s,1}$ bestimmt, indem unbekannte Werte y_t, ε_t für $t \leq 0$ durch 0 ersetzt werden. Sodann werden $\hat{Y}_{t,1}, \hat{Y}_{t,2}, \dots, \hat{Y}_{t,h}$ berechnet.

Es sei hier erwähnt, daß das Prognoseverfahren nur für relativ kleine Werte von h geeignet ist, da $\hat{Y}_{t,h}$ mit wachsendem Prognosehorizont h gegen Null strebt.

Beispiel: Wir betrachten hier einmal einen ARMA(1,1)-Prozeß mit den Parametern $\phi_1 = 0,5$ und $\theta_1 = 0,4$ und wollen die optimalen h -Schrittprognosen $\hat{Y}_{4,h}$ bestimmen. Dazu berechnen wir zunächst für $s = 0, 1, 2, 3$ die optimalen Einschrittprognosen und die zugehörigen Prognosefehler. Es ergibt sich

$$\begin{aligned}\hat{Y}_{0,1} &= 0,5Y_0 + \varepsilon_1 + 0,4\varepsilon_0 = 0,5 \cdot 0 + 0 + 0,4 \cdot 0 = 0, \\ \hat{\varepsilon}_{0,1} &= y_1 - \hat{Y}_{0,1} = y_1 - 0 = y_1, \\ \hat{Y}_{1,1} &= 0,5Y_1 + \varepsilon_2 + 0,4\varepsilon_1 = 0,5y_1 + 0 + 0,4\hat{\varepsilon}_{0,1} = 0,5y_1 + 0,4y_1 = 0,9y_1, \\ \hat{\varepsilon}_{1,1} &= y_2 - \hat{Y}_{1,1} = y_2 - 0,9y_1, \\ \hat{Y}_{2,1} &= 0,5Y_2 + \varepsilon_3 + 0,4\varepsilon_2 = 0,5y_2 + 0 + 0,4\hat{\varepsilon}_{1,1} = 0,5y_2 + 0,4(y_2 - 0,9y_1) \\ &= 0,9y_2 - 0,36y_1, \\ \hat{\varepsilon}_{2,1} &= y_3 - \hat{Y}_{2,1} = y_3 - 0,9y_2 + 0,36y_1, \\ \hat{Y}_{3,1} &= 0,5Y_3 + \varepsilon_4 + 0,4\varepsilon_3 = 0,5y_3 + 0 + 0,4\hat{\varepsilon}_{2,1} \\ &= 0,5y_3 + 0,4(y_3 - 0,9y_2 + 0,36y_1) = 0,9y_3 - 0,36y_2 + 0,144y_1, \\ \hat{\varepsilon}_{3,1} &= y_4 - \hat{Y}_{3,1} = y_4 - 0,9y_3 + 0,36y_2 - 0,144y_1.\end{aligned}$$

Nun berechnen wir beginnend mit $h = 1$ die optimale h -Schrittprognose $\hat{Y}_{4,h}$:

$$\begin{aligned}\hat{Y}_{4,1} &= 0,5Y_4 + \varepsilon_5 + 0,4\varepsilon_4 = 0,5y_4 + 0 + 0,4\hat{\varepsilon}_{3,1} \\ &= 0,5y_4 + 0,4(y_4 - 0,9y_3 + 0,36y_2 - 0,144y_1) \\ &= 0,9y_4 - 0,36y_3 + 0,144y_2 - 0,0576y_1, \\ \hat{Y}_{4,2} &= 0,5Y_5 + \varepsilon_6 + 0,4\varepsilon_5 = 0,5\hat{Y}_{4,1} + 0 + 0,4 \cdot 0 = 0,5\hat{Y}_{4,1}, \\ \hat{Y}_{4,3} &= 0,5Y_6 + \varepsilon_7 + 0,4\varepsilon_6 = 0,5\hat{Y}_{5,1} + 0 + 0,4 \cdot 0 = 0,5(0,5\hat{Y}_{4,1}) \\ &= (0,5)^2\hat{Y}_{4,1}\end{aligned}$$

usw. und erhalten somit als h -Schrittprognose für y_{4+h}

$$\hat{Y}_{4,h} = (0,5)^{h-1}\hat{Y}_{4,1} \quad \text{für } h = 2, 3, \dots$$

Allgemein ergibt sich für einen ARMA(1,1)-Prozeß mit Parametern ϕ_1 und θ_1 :

$$\begin{aligned}\hat{Y}_{0,1} &= 0, \quad \hat{Y}_{s,1} = (\phi_1 + \theta_1) \sum_{j=1}^s (-\theta_1)^{s-j} y_j, \quad \text{für } s = 1, \dots, t \leq n, \\ \hat{Y}_{t,h} &= (\phi_1)^{h-1}(\phi_1 + \theta_1) \sum_{j=1}^t (-\theta_1)^{t-j} y_j = (\phi_1)^{h-1}\hat{Y}_{t,1} \quad \text{für } h = 2, 3, \dots\end{aligned}$$

Wurden die Parameter eines ARMA(p,q)-Prozesses zunächst mit einem Least-Squares Ansatz, vgl. Abschnitt 2.5.2, geschätzt, so läßt sich obiges Prognoseverfahren natürlich ebenfalls anwenden, jedoch sind die Prognosen dann nicht mehr optimal in dem Sinne, daß Prognosen und Prognosefehler unkorreliert sind.

Ist das weiße Rauschen (ε_t) eines ARMA(p,q)-Prozesses normalverteilt

$(\varepsilon_t \sim N(0, \sigma_\varepsilon^2))$, so lassen sich auch **Konfidenzintervalle für die h-Schrittprognosen** $\hat{Y}_{t,h}$ angeben. Bezeichnet u_γ das γ -Quantil der Standardnormalverteilung, vgl. Kap. IV, so ist

$$[\hat{Y}_{t,h} - u_{1-\alpha/2} \sqrt{V(h)}, \hat{Y}_{t,h} + u_{1-\alpha/2} \sqrt{V(h)}]$$

$$\text{mit } V(h) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} \psi_j^2,$$

$$\psi_j = \begin{cases} 1 & \text{für } j = 0 \\ \sum_{i=1}^j \phi_i \psi_{j-i} + \theta_j & \text{für } j > 0 \end{cases},$$

$$\phi_i = 0 \quad \text{für } i > p, \quad \theta_j = 0 \quad \text{für } j > q,$$

ein Konfidenzintervall für $\hat{Y}_{t,h}$ zum Niveau $1 - \alpha$. Diese Konfidenzintervalle sind gegenüber Verletzungen der Normalverteilungsannahme für die ε_t nicht sehr empfindlich und können approximativ auch zumindest noch dann verwandt werden, wenn $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ und σ_ε^2 vermittels des ULS-Ansatzes aus den Daten geschätzt werden; man vgl. hierzu auch z. B. Schlittgen/Streitberg (1984).

Beispiel: Für einen ARMA(1,1)-Prozeß mit Parametern ϕ_1, θ_1 und $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ ergibt sich

$$\begin{aligned} \psi_0 &= 1, \\ \psi_1 &= \phi_1 \psi_0 + \theta_1 = \phi_1 + \theta_1, \\ \psi_2 &= \phi_1 \psi_1 + \theta_2 = \phi_1 \psi_1 = \phi_1(\phi_1 + \theta_1), \\ \psi_3 &= \phi_1 \psi_2 + \phi_2 \psi_1 + \theta_3 = \phi_1 \psi_2 = \phi_1^2(\phi_1 + \theta_1) \end{aligned}$$

usw., also

$$\psi_0 = 1, \quad \psi_j = (\phi_1)^{j-1}(\phi_1 + \theta_1) \quad \text{für } j = 1, 2, \dots, h-1 \quad \text{und für alle } h,$$

so daß wir also für alle h

$$V(h) = \sigma_\varepsilon^2 \left(1 + \sum_{j=1}^{h-1} (\phi_1)^{2j-2} (\phi_1 + \theta_1)^2 \right)$$

erhalten.

Um **zukünftige Beobachtungen** y_{n+h} ($h = 1, 2, \dots$) zu **prognostizieren**, wird man also die h-Schrittprognosen

$$\hat{Y}_{n,h} \quad \text{für } h = 1, 2, \dots$$

berechnen. Um die **Prognosefähigkeit, Prognosegüte** eines Modells zu überprüfen, wird man sogenannte ex-ante und ex-post Prognosen durchführen und die Prognosegüte aufgrund der jeweiligen Prognosefehler beurteilen. Bei **ex-ante Prognosen** werden für verschiedene j basierend auf den ersten $n-j$ Beobachtungen y_1, \dots, y_{n-j} (d. h. die Prozeßparameter werden nur aufgrund der ersten $n-j$ Beobachtungen geschätzt) Prognosen für y_{n-j+1}, \dots, y_n bestimmt, d. h. y_{n-j+1}, \dots, y_n werden mit $\hat{Y}_{n-j,1}, \dots, \hat{Y}_{n-j,j}$ verglichen. Bei **ex-post Prognosen** werden basierend auf allen Beobachtungen y_1, \dots, y_n alle Beobachtungen prognostiziert; man vergleicht also y_1, y_2, \dots, y_n mit den 1-Schritt-Prognosen $\hat{Y}_{0,1}, \hat{Y}_{1,1}, \dots, \hat{Y}_{n-1,1}$.

2.6. Ein Beispiel zur Methode von Box und Jenkins

Im Abschnitt 1 haben wir uns mit der Reihe y_1, \dots, y_{51} der Arbeitslosen im Baugewerbe am Monatsende von Mai 1975 bis September 1979 beschäftigt, vgl. z. B. Tab. 9 und Tab. 12. Insbesondere ist in Tab. 12 auch die differenzierte und saisonal differenzierte Reihe

$$\tilde{y}_t = y_t - y_{t-1} - y_{t-12} + y_{t-13} \quad \text{für } t = 14, \dots, 51$$

angegeben, die in Abb. 9 auch graphisch veranschaulicht wird. In Abb. 9 erkennt man, daß trotz saisonaler Differenzenbildung noch eine Jahresschwingung vorhanden ist, die z. B. durch weiteres Differenzieren beseitigt werden müßte, um eine Realisation eines schwach stationären Prozesses zu erhalten, auf die dann die Methode von Box und Jenkins angewandt werden könnte. Da aber die Reihe bei weiterer Differenzenbildung zu kurz würde, wollen wir das Verfahren von Box und Jenkins anhand der Reihe $\tilde{y}_{14}, \dots, \tilde{y}_{51}$ demonstrieren.

Wir fassen also die Zeitreihe $\tilde{y}_{14}, \dots, \tilde{y}_{51}$ der Länge $\tilde{n} = 38$ aus Tab. 12 als Realisation eines schwach stationären Prozesses (\tilde{Y}_t) mit Erwartungswert μ auf und berechnen zunächst die empirischen Autokorrelationen $r(k)$ sowie partielle Autokorrelationen $r_{k-1}(k)$ dieser Reihe, vgl. Abschnitt 1.4. In Abb. 11 sind $r(k)$ und $r_{k-1}(k)$ als Funktion vom lag k für $k = 0, \dots, 10$ vermittels je eines Polygonzuges dargestellt. Die Identifikation der Ordnungen p und q eines „adäquaten“ ARMA(p, q)-Prozesses allein aufgrund dieser Korrelationen erscheint zumindest schwierig. Darauf wurden vermittels der Prozedur ARIMA der Version SAS/ETS des SAS (Statistical Analysis System) Parameterschätzungen, Box-Ljung-Tests und Prognosen für alle Kombinationen von p und q für $p, q \leq 2$ vorgenommen. Als am geeignetesten, wenn auch sicherlich nicht als „passend“ zu bezeichnen, erwies sich hier die Anpassung eines ARMA(2,1)-Prozesses. Die für diesen Prozeß resultierenden Ergebnisse wollen wir im folgenden schildern.

Wir betrachten das (nicht mittelwertbereinigte) Modell

$$(\tilde{Y}_t - \mu) = \phi_1(\tilde{Y}_{t-1} - \mu) + \phi_2(\tilde{Y}_{t-2} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

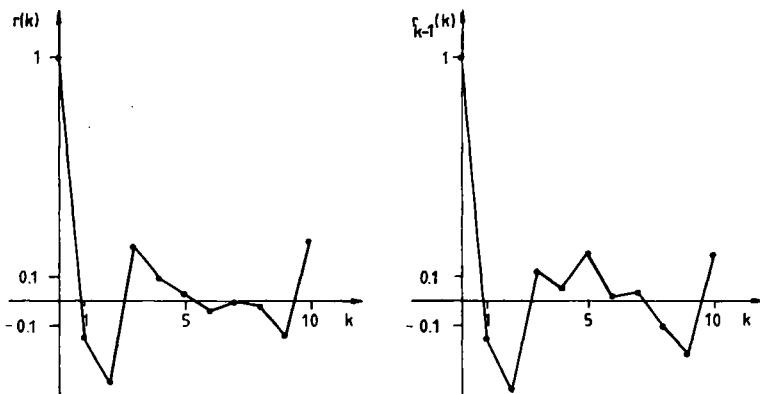


Abb. 11: Autokorrelationen $r(k)$ und partielle Autokorrelationen $r_{k-1}(k)$ bis zum lag $k = 10$ zur Reihe der ersten und ersten saisonalen Differenzen zur Zeitreihe der Arbeitslosen im Baugewerbe

bzw. das Modell

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_0$$

mit

$$\theta_0 = \mu(1 - \phi_1 - \phi_2)$$

und erhalten vermittels der Prozedur ARIMA die folgenden ULS-Schätzer, vgl. Abschnitt 2.5.2, für ϕ_1 , ϕ_2 , θ_1 und μ bzw. θ_0 :

$$\begin{aligned}\hat{\phi}_1 &= -0,522834, \quad \hat{\phi}_2 = -0,402613, \quad \hat{\theta}_1 = 0,373103, \\ \hat{\mu} &= 573,562 \quad \text{bzw.} \quad \hat{\theta}_0 = \hat{\mu}(1 - \hat{\phi}_1 - \hat{\phi}_2) = 1104,36.\end{aligned}$$

Für diese Schätzwerte ergab sich als Summe der quadratischen Abweichungen

$$s^2 = S^2(\hat{\phi}_1, \hat{\phi}_2, \hat{\theta}_1, \hat{\theta}_0) = 1132915627,$$

so daß wir als Schätzung für die Varianz σ_ε^2 des Prozesses (ε_t) erhalten:

$$\hat{\sigma}_\varepsilon^2 = s^2/(\tilde{n} - p - q - 1) = s^2/34 = 33321048;$$

die -1 im Nenner, die in Abschnitt 2.5.2 fehlt, ist auf die Berücksichtigung des Mittelwertes μ im Modell zurückzuführen.

Bevor wir das Modell, wie in Abschnitt 2.5.3 beschrieben, überprüfen, wollen wir Ein-Schritt ex-post Prognosen für $\tilde{Y}_{14}, \dots, \tilde{Y}_{51}$ bzw. für die ursprünglichen Arbeitslosenzahlen Y_{14}, \dots, Y_{51} bestimmen. Dazu verwenden wir das in Abschnitt 2.5.4 geschilderte Verfahren, wobei jetzt allerdings zusätzlich der konstante Term berücksichtigt werden muß. Wir betrachten also das Modell

$$\tilde{Y}_t = \hat{\phi}_1 \tilde{Y}_{t-1} + \hat{\phi}_2 \tilde{Y}_{t-2} + \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1} + \hat{\theta}_0$$

und berechnen die Prognosen, die Schätzungen für $\tilde{y}_{14}, \dots, \tilde{y}_{51}$:

$$\begin{aligned}\hat{y}_{14} &= \hat{\tilde{Y}}_{14-1,1} = \hat{\theta}_0, \quad \hat{y}_{15} = \hat{\tilde{Y}}_{15-1,1} = \hat{\phi}_1 \tilde{y}_{14} + \hat{\theta}_1 (\tilde{y}_{14} - \hat{\tilde{y}}_{14}) + \hat{\theta}_0, \\ \hat{y}_t &= \hat{\tilde{Y}}_{t-1,1} = \hat{\phi}_1 \tilde{y}_{t-1} + \hat{\phi}_2 \tilde{y}_{t-2} + \hat{\theta}_1 (\tilde{y}_{t-1} - \hat{\tilde{y}}_{t-1}) + \hat{\theta}_0 \quad \text{für } t = 16, \dots, 51.\end{aligned}$$

Übertragen auf die ursprünglichen Daten y_{14}, \dots, y_{51} erhalten wir dann die Schätzungen

$$\hat{y}_t = \hat{\tilde{y}}_t + y_{t-1} + y_{t-12} - y_{t-13} \quad \text{für } t = 14, \dots, 51,$$

die in Tab. 15 zusammengestellt sind. Außerdem findet man für $t = 14, \dots, 51$ in Tab. 15 nochmals die Daten y_t , vgl. z. B. Tab. 9, sowie die Prognosefehler $y_t - \hat{y}_t = \tilde{y}_t - \hat{\tilde{y}}_t$, die ja gerade ex-post Schätzungen $\hat{\varepsilon}_t = \hat{\varepsilon}_{t-1,1}$ für ε_t sind. In Abb. 12 sind sowohl die ursprüngliche Zeitreihe y_{14}, \dots, y_{51} als auch die Reihe $\hat{y}_{14}, \dots, \hat{y}_{51}$, letztere als Polygonzug, graphisch dargestellt.

Zur Prüfung der Autokorrelation des Prozesses (ε_t) des ARMA(2,1)-Prozesses wollen wir nun Box-Ljung-Tests, vgl. Abschnitt 2.5.3, durchführen, und zwar unter Berücksichtigung der lags k bis $K = 5$ und bis $K = 10$. Dazu berechnen wir zunächst ausgehend von den Residuen ε_t aus Tab. 15 die empirischen Autokorrelationen $r_\varepsilon(k)$ für $k = 1, \dots, 10$:

$$\begin{aligned}r_\varepsilon(1) &= -0,006, \quad r_\varepsilon(2) = 0,004, \quad r_\varepsilon(3) = 0,074, \quad r_\varepsilon(4) = 0,108, \\ r_\varepsilon(5) &= 0,157, \quad r_\varepsilon(6) = -0,036, \quad r_\varepsilon(7) = -0,073, \quad r_\varepsilon(8) = -0,041, \\ r_\varepsilon(9) &= 0,026 \quad \text{und} \quad r_\varepsilon(10) = 0,135.\end{aligned}$$

Tab. 15: Arbeitslose in Bauberufen y_t , Schätzungen \hat{y}_t mittels Box-Jenkins-Methode und Prognosefehler $\hat{\epsilon}_t = y_t - \hat{y}_t$ für $t = 14, \dots, 51$

t	y_t	\hat{y}_t	$\hat{\epsilon}_t$	t	y_t	\hat{y}_t	$\hat{\epsilon}_t$
14	28068	22898	5170	33	75304	69646	5658
15	26634	22717	3917	34	43622	55691	-10619
16	29259	25718	3541	35	33990	33464	526
17	38942	41079	-2290	36	26819	32854	-7573
18	65036	63878	1153	37	25291	20983	4308
19	110728	101971	8757	38	24538	24937	-399
20	108931	113443	-4512	39	22685	21650	1035
21	71517	70446	1053	40	23945	26429	-2484
22	54428	45444	8984	41	28245	34674	-6429
23	42911	35009	7902	42	47017	52730	-5713
24	37123	32869	4254	43	90920	85502	5418
25	33044	33373	-329	44	89340	96100	-6769
26	30755	32128	-1373	45	47792	59068	-11276
27	28742	30122	-1380	46	28448	22271	6177
28	31698	32441	-743	47	19139	20852	-1713
29	41427	42268	-841	48	16728	7292	9431
30	63685	68184	-4499	49	16523	17204	-681
31	99189	110790	-11601	50	16622	14012	2610
32	104240	101039	3201	51	15499	15869	-370

Für $K = 5$ ergibt sich damit dann

$$Q^*(5) = \tilde{n}(\tilde{n} + 2) \sum_{k=1}^5 \frac{1}{\tilde{n} - k} r_\epsilon^2(k) = 38 \cdot 40 \sum_{k=1}^5 \frac{1}{38 - k} r_\epsilon^2(k) = 1,90,$$

und für $K = 10$ erhalten wir

$$Q^*(10) = 3,33.$$

$Q^*(5)$ wird aufgrund der Berücksichtigung eines konstanten Terms mit dem $1 - \alpha$ Quantil der χ^2 -Verteilung mit $5 - 2 - 1 - 1 = 1$ Freiheitsgrad und $Q^*(10)$ mit dem einer χ^2 -Verteilung mit $10 - 2 - 1 - 1 = 6$ Freiheitsgraden verglichen. Im ersten Fall wird die Hypothese H_0 , daß (ϵ_t) ein weißes Rauschen ist, für Niveaus $\alpha > 0,16$, im zweiten Fall für Niveaus $\alpha > 0,77$ verworfen, da gilt

$$\chi^2_{1;1-0,16} \approx 1,90 \quad \text{und} \quad \chi^2_{6;1-0,77} \approx 3,33.$$

Insbesondere der erste Test spricht also gegen das tatsächliche Vorliegen eines weißen Rauschens (ϵ_t).

Zur Überprüfung der Anpassungsgüte der Daten y_t durch das ARMA(2,1)-Modell für die differenzierten Daten \hat{y}_t wollen wir hier einmal das Bestimmtheitsmaß berechnen:

$$B_y = 1 - \frac{\sum_{t=14}^{51} \hat{\epsilon}_t^2}{\sum_{t=14}^{51} (y_t - \bar{y})^2} = 1 - \frac{1132915627}{29861051280} = 0,96.$$

Dieses Bestimmtheitsmaß mag sehr hoch erscheinen; berücksichtigt man jedoch, daß man bei der Annahme eines weißen Rauschens mit Mittelwert $\mu = \hat{y} = 617,921$

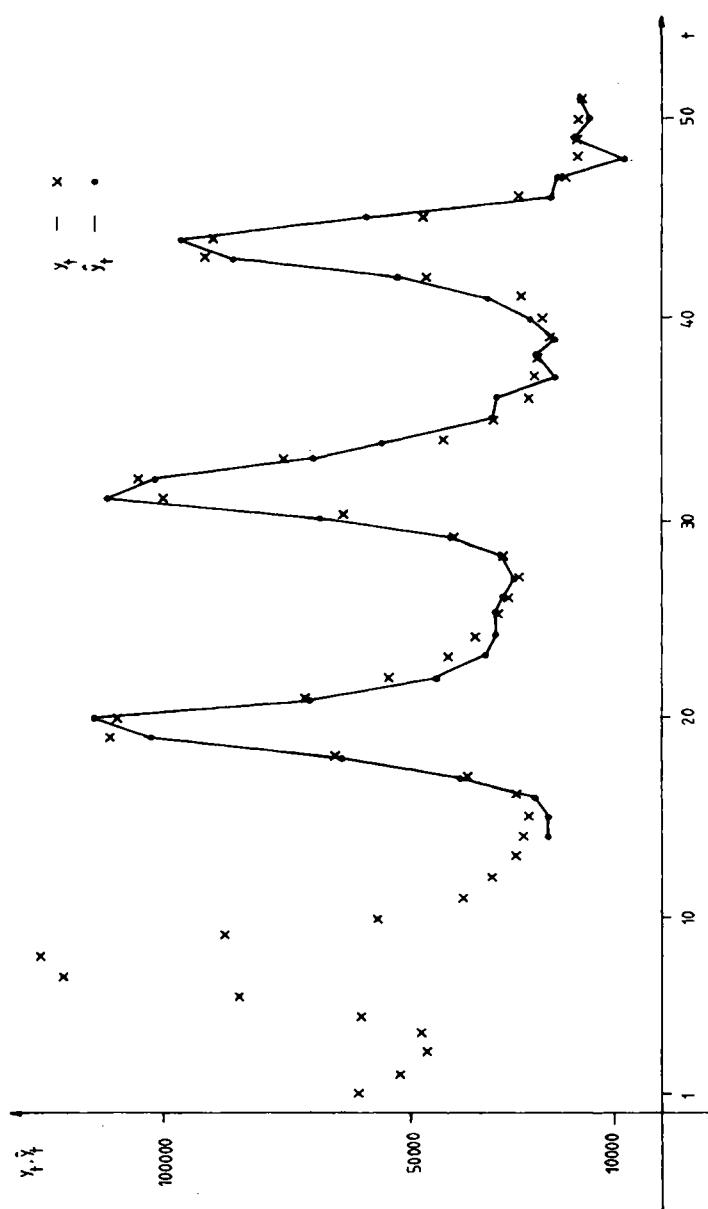


Abb. 12: Arbeitslose im Baugewerbe y_t von Mai 1975 ($t = 1$) bis September 1979 ($t = 51$) und Prognosen \hat{y}_t nach der Methode von Box und Jenkins

für den Prozeß (\tilde{Y}_t), bei der dann $\hat{y}_t = 617,921$ für $t = 14, \dots, 51$ und damit

$$\hat{y}_t = 617,921 + y_{t-1} + y_{t-12} - y_{t-13} \quad \text{für } t = 14, \dots, 51$$

prognostiziert würde, bereits ein Bestimmtheitsmaß von

$$B = 1 - \frac{\sum_{t=14}^{51} (\tilde{y}_t - 617,921)^2}{\sum_{t=14}^{51} (y_t - \bar{y})^2} = 0,95$$

erzielte, so erkennt man, wie gering obiges Bestimmtheitsmaß und damit der Gewinn durch die Prozeßanpassung zu bewerten ist.

Abschließend wollen wir ausgehend vom Zeitpunkt $t = 51$ noch h-Schritt-Prognosen für die Zahl der Arbeitslosen von Oktober 1979 bis September 1980 ($h = 1, \dots, 12$) berechnen. Die h-Schritt-Prognose für \tilde{Y}_{51+h} ist hier gerade

$$\hat{y}_{51+h} = \tilde{Y}_{51,h} = \hat{\theta}_1 \tilde{y}_{51+h-1} + \hat{\theta}_2 \tilde{y}_{51+h-2} + \hat{\theta}_1 (\tilde{y}_{51+h-1} - \hat{y}_{51+h-1}) + \hat{\theta}_0$$

für $h = 1, \dots, 12$, wobei die unbekannten Größen $\tilde{y}_{52}, \dots, \tilde{y}_{62}$ auf der rechten Seite durch die h-Schritt-Prognosen $\hat{y}_{52}, \dots, \hat{y}_{62}$ ersetzt werden. Die Zahl der Arbeitslosen selbst wird dann durch

$$\hat{y}_t = \hat{y}_{t-1} + y_{t-1} + y_{t-12} - y_{t-13} \quad \text{für } t = 52, \dots, 63$$

prognostiziert, wobei unbekannte Größen wiederum durch die entsprechenden Prognosen ersetzt werden. Um zu sehen, wie gut diese Prognosen sind, haben wir in Tab. 16 neben ihnen auch die tatsächlichen Arbeitslosenzahlen in Bauberufen y_{52}, \dots, y_{63} in den Monaten Oktober 1979 bis September 1980 (aus: „Wirtschaft und Statistik“; Hrsg. Statistisches Bundesamt Wiesbaden (1980–1981), Kohlhammer-Verlag, Stuttgart–Mainz) sowie die Prognosefehler $y_t - \hat{y}_t$ für $t = 52, \dots, 63$ angegeben.

Wir sehen, daß die Prognosen für die ersten 6 Monate noch recht gut sind, dann aber bedingt durch einen Strukturbruch (Trendwechsel von fallend auf steigend) in den Arbeitslosenzahlen zunehmend schlechter werden.

Tab. 16: Prognostizierte Arbeitslosenzahlen \hat{y}_t , tatsächliche Arbeitslosenzahlen y_t und Prognosefehler $y_t - \hat{y}_t$ für Bauberufe in den Monaten Oktober 1979 bis September 1980, $t = 52, \dots, 63$

Monat	t	Prognose \hat{y}_t	y_t	$y_t - \hat{y}_t$
Okt. 1979	52	17001	17464	463
Nov. 1979	53	22067	24922	2855
Dez. 1979	54	41315	37683	- 3632
Jan. 1980	55	85524	82681	- 2843
Febr. 1980	56	84487	75552	- 8935
März 1980	57	43394	46234	2840
April 1980	58	24446	30830	6384
Mai 1980	59	15597	25312	9715
Juni 1980	60	13636	23513	9877
Juli 1980	61	13859	24002	10143
Aug. 1980	62	14403	24491	10088
Sept. 1980	63	13724	24315	10591

3. Die Spektralanalyse

Bislang haben wir uns mit der Analyse von stochastischen Prozessen (Y_t) bzw. Zeitreihen y_1, \dots, y_n im Zeitbereich auseinandersetzt. Die Spektralanalyse dient dagegen der Analyse im **Frequenzbereich**. Um die Bedeutung der Spektralanalyse zu verdeutlichen, betrachten wir hier zunächst einen **schwach stationären harmonischen Prozeß** (Y_t) mit $EY_t = 0$ für alle t und endlicher Varianz σ_Y^2 . Ein solcher Prozeß, der insbesondere bei Zeitreihen aus dem Bereich der Naturwissenschaften auftritt, hat die Gestalt

$$Y_t = \sum_{j=1}^m (A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)),$$

wobei $\lambda_1, \dots, \lambda_m$ feste Frequenzen mit $|\lambda_j| < \pi$ ($= 3,14\dots$) und $A_1, \dots, A_m, B_1, \dots, B_m$ voneinander unabhängige Zufallsvariablen mit Erwartungswert 0 und Varianz $\text{Var } A_j = \text{Var } B_j = \sigma_j^2$ sind. Ein schwach stationärer harmonischer Prozeß setzt sich also aus m **Cosinoiden**

$$A_j \cos \lambda_j t + B_j \sin \lambda_j t$$

verschiedener Frequenzen λ_j zusammen.

Jeder Cosinoid ist dabei gerade eine um den **Phasenwinkel** $\phi_j = \arctan(B_j/A_j)$ verschobene Cosinusschwingung mit der **Amplitude** $C_j = \sqrt{A_j^2 + B_j^2}$, d.h.

$$A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t) = C_j \cos(\lambda_j t - \phi_j).$$

Für die Varianz des Prozesses (Y_t) gilt nun

$$\begin{aligned} \text{Var } Y_t &= \sigma_Y^2 = \sum_{j=1}^m [\text{Var}(A_j \cos(\lambda_j t)) + \text{Var}(B_j \sin(\lambda_j t))] \\ &= \sum_{j=1}^m [\cos^2(\lambda_j t) \text{Var } A_j + \sin^2(\lambda_j t) \text{Var } B_j] \\ &= \sum_{j=1}^m \sigma_j^2 (\cos^2(\lambda_j t) + \sin^2(\lambda_j t)) \\ &= \sum_{j=1}^m \sigma_j^2, \end{aligned}$$

denn

$$\cos^2 \omega + \sin^2 \omega = 1 \quad \text{für alle } \omega.$$

Die Varianz σ_j^2 von A_j gibt an, welchen Beitrag der Cosinoid der Frequenz λ_j zur Varianz σ_Y^2 des Prozesses liefert. Aufgrund der Amplitudenvarianzen zu den verschiedenen Frequenzen kann also die Wirkung der Schwingungskomponente auf den Prozeß (Y_t) untersucht werden.

Wie der schwach stationäre harmonische Prozeß läßt sich jeder schwach stationäre stochastische Prozeß (Y_t) als Überlagerung von Cosinoiden darstellen, wobei dann allerdings nicht nur einige Frequenzen λ_j , sondern u.U. alle Frequenzen λ im Intervall $[0, \pi]$ berücksichtigt werden müssen.

Bevor wir uns mit solchen Darstellungen und mit der Analyse im Frequenzbereich allgemein beschäftigen können, muß zunächst der Begriff der **komplexen Zahl** eingeführt werden.

3.1. Komplexe Zahlen

Das Bedürfnis, komplexe Zahlen zu betrachten, röhrt ursprünglich daher, daß man die Gleichung

$$x^2 + 1 = 0$$

lösen möchte. Zu diesem Zweck wird die **imaginäre Einheit i** , für die gilt $i^2 = -1$, eingeführt. Damit ist weiter $i^3 = -i$, $i^4 = 1$, $i^5 = i$, ... und $i^{-1} = \frac{1}{i} = \frac{i}{i^2} = -i$, $i^{-2} = \frac{1}{i} \cdot \frac{1}{i} = (-i)(-i) = i^2 = -1$, $i^{-3} = i$, $i^{-4} = 1$,

Eine **komplexe Zahl z** hat die Gestalt

$$z = a + ib,$$

wobei a und b gewöhnliche, reelle Zahlen sind; a heißt **Realteil** und b **Imaginärteil** von z . Weiterhin bezeichnet man

$$\bar{z} = a - ib$$

als **komplex konjugierte Zahl** zu z und

$$|z| = \sqrt{a^2 + b^2}$$

heißt **Betrag** von z .

Summe und **Produkt** zweier komplexer Zahlen $z_1 = a_1 + ib_1$, $z_2 = a_2 + ib_2$ sind gegeben durch

$$\begin{aligned} z_1 + z_2 &= (a_1 + ib_1) + (a_2 + ib_2) = (a_1 + a_2) + i(b_1 + b_2) \\ \text{bzw. } z_1 \cdot z_2 &= (a_1 + ib_1)(a_2 + ib_2) = a_1 a_2 + ia_1 b_2 + ia_2 b_1 + i^2 b_1 b_2 \\ &= (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + a_2 b_1). \end{aligned}$$

An letzterer Formel erkennt man, daß für $z = a + ib$ gilt:

$$|z| = \sqrt{z \cdot \bar{z}} = \sqrt{(a^2 + b^2) + i(-ab + ab)} = \sqrt{a^2 + b^2}.$$

Weiterhin werden wir noch die folgenden Beziehungen benötigen. Für eine (reelle) Zahl x gilt:

$$e^{ix} = \cos x + i \sin x,$$

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix}),$$

$$\sin x = \frac{1}{2i}(e^{ix} - e^{-ix}) = \frac{i}{2}(e^{-ix} - e^{ix}).$$

Für eine komplexe Zufallsvariable $Z = A + iB$ (A, B reelle Zufallsvariablen) gelten die obigen Regeln gleichermaßen und man definiert

$$EZ = E(A + iB) = EA + iEB,$$

$\text{Cov}(Z_1, Z_2) = E[(Z_1 - EZ_1)(\bar{Z}_2 - EZ_2)]$ für komplexe Zufallsvariablen Z_1 und Z_2 sowie

$$\text{Var } Z = \text{Cov}(Z, Z) = E(|Z - EZ|^2).$$

3.2. Spektrum und Spektraldichte

Kommen wir hier zunächst noch einmal auf den schwach stationären harmonischen Prozeß zurück. Zu diesem Prozeß läßt sich ein komplexer Prozeß (Y_t) definieren mit

$$Y_t = \sum_{j=1}^m Z_j e^{-i\lambda_j t} = \sum_{j=1}^m Z_j (\cos(\lambda_j t) - i \sin(\lambda_j t)),$$

wobei λ_j Zahlen mit $|\lambda_j| < \pi$ und $Z_j = A_j + iB_j$ komplexe Zufallsvariablen mit

$$EZ_j = 0 \quad \text{für } j = 1, \dots, m,$$

$$\text{Var } Z_j = E(Z_j \bar{Z}_j) = \sigma_j^2 \quad \text{für } j = 1, \dots, m,$$

$$E(Z_j \bar{Z}_{j'}) = 0 \quad \text{für } j, j' = 1, \dots, m, j \neq j'$$

sind. Der Realteil des komplexen Prozesses (Y_t) ist nun gerade

$$\sum_{j=1}^m (A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)) = \sum_{j=1}^m \sqrt{Z_j \bar{Z}_j} \cos(\lambda_j t - \arctan(B_j/A_j)),$$

also identisch mit dem eingangs vorgestellten aus m Cosinoiden zusammengesetzten reellen, harmonischen Prozeß; der Imaginärteil ist bis auf eine Phasenverschiebung von $\pi/2$ identisch mit dem Realteil. Für die Autokovarianzen $\gamma(k)$ zum lag k des Prozesses (Y_t) ergibt sich dann

$$\begin{aligned} \gamma(k) &= E(Y_t \bar{Y}_{t+k}) = E\left(\sum_{j=1}^m \sum_{j'=1}^m Z_j \bar{Z}_{j'} e^{-i\lambda_j t} e^{i\lambda_{j'}(t+k)}\right) \\ &= \sum_{j=1}^m \sum_{j'=1}^m e^{i\lambda_j t} e^{i\lambda_{j'}(t+k)} E(Z_j \bar{Z}_{j'}) \\ &= \sum_{j=1}^m e^{i\lambda_j t} e^{i\lambda_j(t+k)} \sigma_j^2 \\ &= \sum_{j=1}^m e^{-i\lambda_j t + i\lambda_j t + i\lambda_j k} \sigma_j^2 \\ &= \sum_{j=1}^m \sigma_j^2 \cdot e^{i\lambda_j k} \\ &= \sum_{j=1}^m f(\lambda_j) \cdot e^{i\lambda_j k}; \end{aligned}$$

die Funktion

$$f(\lambda) = \begin{cases} \sigma_j^2, & \text{für } \lambda = \lambda_j, j = 1, \dots, m \\ 0, & \text{sonst} \end{cases}$$

nennt man die **(diskrete) Spektraldichte** des harmonischen Prozesses (Y_t).

Der eigentlich interessierende Teil von $\gamma(k)$ ist der Realteil, also

$$\sum_{j=1}^m \sigma_j^2 \cos(\lambda_j k);$$

definiert man die symmetrische Funktion

$$f(\lambda) = \begin{cases} \frac{1}{2} \sigma_j^2, & \text{für } \lambda = \lambda_j \text{ und } \lambda = -\lambda_j, j = 1, \dots, m \\ 0, & \text{sonst} \end{cases}$$

so erhält man mit der Festsetzung $\lambda_{-j} = -\lambda_j$ für $j = 1, \dots, m$ als symmetrische Darstellung dieses Realteils von $\gamma(k)$:

$$\sum_{j=-m}^m \tilde{f}(\lambda_j) e^{i\lambda_j k} \quad (\text{mit } \tilde{f}(\lambda_0) = 0).$$

Dieser Realteil von $\gamma(k)$ ist zugleich die Autokorrelation zum lag k des zugrundeliegenden reellen harmonischen Prozesses, und die Funktion $\tilde{f}(\lambda)$ ist die Spektraldichte desselben.

Wir betrachten nun nur noch reelle Prozesse (Y_t) mit stetiger Spektraldichte, für deren Autokovarianzen $\gamma(k)$ gilt $\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$. Diese sind analog in der folgenden Form darstellbar:

$$\gamma(k) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda k} d\lambda \quad \text{mit } f(\lambda) = f(-\lambda),$$

wobei $f(\lambda)$ den Beitrag der Frequenz λ zur Varianz des Prozesses (Y_t) angibt.

Speziell für die Varianz $\gamma(0) = \sigma_Y^2$ des Prozesses (Y_t) gilt somit natürlich

$$\gamma(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda,$$

d.h. $\gamma(0)$ ist die Fläche zwischen dem Graphen von $f(\lambda)$ und der λ -Achse für $-\pi \leq \lambda \leq \pi$.

Die nichtnegative Funktion $f(\lambda)$ heißt **Spektraldichte (Varianzspektrum, Powerspektrum)** des Prozesses (Y_t). Durch sie sind die Autokovarianzen $\gamma(k)$ eindeutig bestimmt und umgekehrt lässt sich somit $f(\lambda)$ aus den $\gamma(k)$ festlegen. Es gilt nämlich

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\lambda} \quad \text{für } -\pi \leq \lambda \leq \pi$$

bzw.

$$f(\lambda) = f(-\lambda) = \frac{\gamma(0)}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \gamma(k) \cos(\lambda k) \quad \text{für } 0 \leq \lambda \leq \pi.$$

Die Spektraldichte enthält also die gleichen Informationen wie die Autokovarianzen $\gamma(k) = \gamma(-k)$, $k = 0, 1, 2, \dots$, jedoch geordnet nach Frequenzen und nicht nach lags.

Neben der Spektraldichte $f(\lambda)$ wird auch oft das **Spektrum $F(\lambda)$** eines Prozesses (Y_t) betrachtet. Die Beziehung zwischen $f(\lambda)$ und $F(\lambda)$ ist dieselbe wie die Beziehung zwischen einer Verteilungsdichte und einer Verteilungsfunktion, d.h. es ist

$$\begin{aligned} F(\lambda) &= \int_{-\infty}^{\lambda} f(\omega) d\omega = \int_{-\pi}^{\lambda} f(\omega) d\omega \\ &= \frac{\gamma(0)}{2\pi} (\lambda + \pi) + \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{\gamma(k)}{k} \sin(\lambda k) \quad \text{für } -\pi \leq \lambda \leq \pi. \end{aligned}$$

Beispiel: Für ein weißes Rauschen (ε_t) ergibt sich wegen $\gamma(0) = \sigma_{\varepsilon}^2 = \text{Var } \varepsilon_t$, $\gamma(k) = \gamma(-k) = 0$ für $k = 1, 2, \dots$

$$f(\lambda) = f(-\lambda) = \frac{\gamma(0)}{2\pi} = \frac{\sigma_{\varepsilon}^2}{2\pi} \quad \text{für } 0 \leq \lambda \leq \pi,$$

d. h. die Spektraldichte ist konstant, und

$$F(\lambda) = \frac{\gamma(0)}{2\pi} (\lambda + \pi) = \frac{\sigma_e^2}{2\pi} (\lambda + \pi) = \frac{\sigma_e^2}{2} + \frac{\sigma_e^2}{2\pi} \lambda \quad \text{für } -\pi \leq \lambda \leq \pi,$$

d. h. das Spektrum ist eine Gerade, die durch die Werte 0 für $\lambda = -\pi$ und σ_e^2 für $\lambda = \pi$ gekennzeichnet ist.

3.3. Spektraldichten gefilterter Prozesse

Im Abschnitt 1.3 haben wir lineare Filter für Zeitreihen y_1, \dots, y_n kennengelernt und in Abschnitt 2 haben wir gesehen, daß sich lineare Filter natürlich ebensogut auf stochastische Prozesse (Y_t) anwenden lassen. Der Eingangsprozeß (Y_t) wird dabei mittels der sogenannten **Impulsresponsefunktion**

$$Y_t^* = \sum_{r=-q}^{+s} a_r Y_{t+r}$$

in den Ausgangsprozeß (Y_t^*) transformiert. Die zur Impulsresponsefunktion korrespondierende **Frequenzresponsefunktion** ist

$$H(\lambda) = \sum_{r=-q}^{+s} a_r e^{-i\lambda r}$$

und

$$G(\lambda) = |H(\lambda)|$$

heißt **Amplitudendiagramm** des linearen Filters.

Ist nun $f_Y(\lambda)$ die Spektraldichte des Eingangsprozesses (Y_t), so ergibt sich die Spektraldichte des (gefilterten) Ausgangsprozesses $f_{Y^*}(\lambda)$ durch Multiplikation mit $G(\lambda)^2$, d. h.

$$f_{Y^*}(\lambda) = G(\lambda)^2 \cdot f_Y(\lambda).$$

Mit Hilfe dieser Beziehung lassen sich die Spektraldichten vieler Prozesse bestimmen.

Betrachten wir zunächst einen **AR(p)-Prozeß** (Y_t), d. h.

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t,$$

wobei (ε_t) ein weißes Rauschen mit Spektraldichte $f_\varepsilon(\lambda) = \sigma_e^2/(2\pi)$, vgl. Abschnitt 3.2, darstellt. Wir können hier auch schreiben

$$\varepsilon_t = Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = \sum_{r=0}^p a_r Y_{t-r} \quad \text{mit } a_0 = 1, a_r = -\phi_r \\ \text{für } r = 1, \dots, p,$$

d. h. (ε_t) entsteht durch Filterung von (Y_t). Die Frequenzresponsefunktion ist hier gegeben als

$$H_\varepsilon(\lambda) = \sum_{r=0}^p a_r e^{-i\lambda r},$$

so daß sich

$$G_\varepsilon(\lambda) = |H_\varepsilon(\lambda)| = \left(\sum_{r=1}^p a_r^2 + 2 \sum_{r=1}^p \sum_{r'=0}^{p-r} a_r a_{r'+r} \cos(\lambda r) \right)^{1/2}$$

und damit

$$f_\varepsilon(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} = G_\varepsilon^2(\lambda) f_Y(\lambda)$$

bzw.

$$\begin{aligned} f_Y(\lambda) &= \sigma_\varepsilon^2 / (2\pi \cdot G_\varepsilon^2(\lambda)) \\ &= \sigma_\varepsilon^2 / [2\pi \left(\sum_{r=0}^p a_r^2 + 2 \sum_{r=1}^p \sum_{r'=0}^{p-r} a_r a_{r'+r} \cos(\lambda r) \right)] \end{aligned}$$

mit $a_0 = 1$ und $a_r = -\phi_r$ für $r = 1, \dots, p$ ergibt.

Ein **MA(q)-Prozeß** (Y_t) entsteht durch Filterung eines weißen Rauschens (ε_t) mit Varianz σ_ε^2 :

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} = \sum_{r=0}^q \theta_r \varepsilon_{t-r}$$

für $\theta_0 = 1$. Mit

$$f_\varepsilon(\lambda) = \sigma_\varepsilon^2 / (2\pi)$$

und

$$H_Y(\lambda) = \sum_{r=0}^q \theta_r e^{-ir\lambda}$$

ergibt sich wegen

$$G_Y^2(\lambda) = |H_Y(\lambda)|^2 = \left| \sum_{r=0}^q \theta_r e^{-ir\lambda} \right|^2 = \sum_{r=0}^q \theta_r^2 + 2 \sum_{r=1}^q \sum_{r'=0}^{q-r} \theta_r \theta_{r'+r} \cos(\lambda r)$$

die Spektraldichte des MA(q)-Prozesses (Y_t) zu

$$\begin{aligned} f_Y(\lambda) &= G_Y^2(\lambda) f_\varepsilon(\lambda) \\ &= \frac{1}{2\pi} \left(\sigma_\varepsilon^2 \sum_{r=0}^q \theta_r^2 + 2\sigma_\varepsilon^2 \sum_{r=1}^q \sum_{r'=0}^{q-r} \theta_r \theta_{r'+r} \cos(\lambda r) \right). \end{aligned}$$

Zur Bestimmung der Spektraldichte $f_Y(\lambda)$ eines **ARMA(p, q)-Prozesses** (Y_t) mit

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

gehen wir von folgender Überlegung aus. Der Prozeß (Y_t^*) mit

$$Y_t^* = Y_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} = \sum_{j=0}^p a_j Y_{t-j}$$

mit $a_0 = 1$, $a_j = -\phi_j$ für $j = 1, \dots, p$ entsteht seinerseits durch Filterung eines weißen Rauschens (ε_t) mit $\text{Var } \varepsilon_t = \sigma_\varepsilon^2$:

$$Y_t^* = \sum_{r=0}^q \theta_r \varepsilon_{t-r} \quad \text{mit } \theta_0 = 1.$$

Daher ist die Spektraldichte des Prozesses (Y_t^*) gerade die eines MA(q)-Prozesses:

$$f_{Y^*}(\lambda) = G_{Y^*}^2(\lambda) f_\varepsilon(\lambda) = \frac{1}{2\pi} \left(\sigma_\varepsilon^2 \sum_{r=0}^q \theta_r^2 + 2\sigma_\varepsilon^2 \sum_{r=1}^q \sum_{r'=0}^{q-r} \theta_r \theta_{r'+r} \cos(\lambda r) \right).$$

Andererseits entsteht der Prozeß (ε_t^*) mit

$$\varepsilon_t^* = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} = \sum_{r=0}^q \theta_r \varepsilon_{t-r} \quad \text{mit} \quad \theta_0 = 1$$

durch Filterung des Prozesses (Y_t):

$$\varepsilon_t^* = \sum_{j=0}^p a_j Y_{t-j} \quad \text{mit} \quad a_0 = 1, a_j = -\phi_j \quad \text{für} \quad j = 1, \dots, p.$$

Daher ist die Spektraldichte des Prozesses (ε_t^*) gerade

$$\begin{aligned} f_{\varepsilon^*}(\lambda) &= G_{\varepsilon^*}^2(\lambda) f_Y(\lambda) = \left| \sum_{j=0}^p a_j e^{-ij\lambda} \right|^2 f_Y(\lambda) \\ &= f_Y(\lambda) \left(\sum_{j=0}^p a_j^2 + 2 \sum_{j=1}^p \sum_{j'=0}^{p-j} a_j a_{j'+j} \cos(\lambda j) \right). \end{aligned}$$

Aus der Prozeß-Gleichung eines ARMA(p, q)-Prozesses folgt weiterhin, daß gilt

$$Y_t^* = \varepsilon_t^* \quad \text{und somit} \quad f_{Y^*}(\lambda) = f_{\varepsilon^*}(\lambda).$$

Wir haben also

$$G_{Y^*}^2(\lambda) f_{\varepsilon^*}(\lambda) = G_{\varepsilon^*}^2(\lambda) f_Y(\lambda),$$

und somit gilt für die Spektraldichte des ARMA(p, q)-Prozesses (Y_t):

$$\begin{aligned} f_Y(\lambda) &= G_{Y^*}^2(\lambda) f_{\varepsilon^*}(\lambda) / G_{\varepsilon^*}^2(\lambda) \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \cdot \frac{\sum_{r=0}^q \theta_r^2 + 2 \sum_{r=1}^q \sum_{r'=0}^{q-r} \theta_r \theta_{r'+r} \cos(\lambda r)}{\sum_{j=0}^p a_j^2 + 2 \sum_{j=1}^p \sum_{j'=0}^{p-j} a_j a_{j'+j} \cos(\lambda j)}, \end{aligned}$$

wobei $\theta_0 = 1$, $a_0 = 1$, $a_j = -\phi_j$ für $j = 1, \dots, p$ ist.

Wir haben hier schon die Nützlichkeit des Amplitudendiagramms $G(\lambda)$ eines linearen Filters bei der Bestimmung von Spektraldichten schwach stationärer stochastischer Prozesse (Y_t) gesehen. Daneben lässt sich am Amplitudendiagramm die **Wirkungsweise eines linearen Filters auf den Ausgangsprozeß** ablesen.

Wir haben im Abschnitt 1.3.5 bereits erwähnt, daß Filter in Klassen eingeteilt werden, je nachdem, was sie bewirken, wenn man sie auf eine **reale Zeitreihe** y_1, \dots, y_n anwendet. Ein **Tiefpaß-Filter** dient etwa der Bereinigung von **lokalen Schwankungen**, die in der Spektraldichte, betrachtet im Intervall $[0, \pi]$, eines Prozesses durch die Amplituden der **hohen Frequenzen** in der Nähe von π zum Ausdruck kommen. Dementsprechend kommen **langfristige Schwankungen** durch die Amplituden **niedriger Frequenzen** λ zum Ausdruck und werden durch **Hochpaß-Filter** eliminiert. Saisonale Schwankungen entsprechen Amplituden bestimmter Frequenzen λ im Frequenzband $[0, \pi]$ und werden durch **Bandsperrern** eliminiert, durch **Bandpässe** sichtbar gemacht.

Das **Amplitudendiagramm** $G(\lambda)$ eines linearen Filters **charakterisiert** nun direkt den **Filter-Typ**. Zum Beispiel hat das Amplitudendiagramm eines idealen Tiefpaß-Filters für „kleine“ Werte von λ einen hohen Wert und für „große“ Werte von λ den Wert 0. Für die Spektraldichte des gefilterten Prozesses bedeutet dies, daß nur die Amplituden „kleiner“ Frequenzen der Eingangsspektraldichte „durchgelassen“ werden. In der Abb. 13 sind die Amplitudendiagramme eines idealen Tiefpasses,

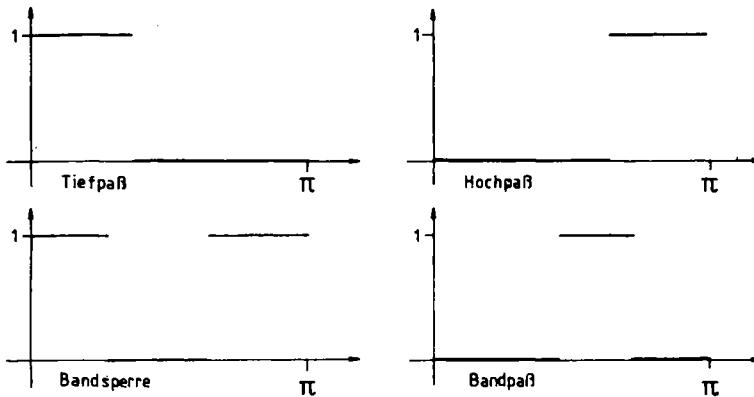


Abb. 13: Amplitudendiagramme von idealem Tiefpaß-, Hochpaß-Filter, idealer Bandsperre und idealem Bandpaß

eines idealen Hochpasses, einer idealen Bandsperre und eines idealen Bandpasses dargestellt.

Solche idealen Filter gibt es real natürlich nicht; man kann lediglich versuchen, sie zu approximieren. Um einen Eindruck davon zu geben, wie das Amplitudendiagramm üblicher Filter aussieht, betrachten wir folgende Beispiele.

Beispiel: (a) Das Amplitudendiagramm des linearen Filters

$$Y_t^* = \frac{1}{4} Y_{t-1} + \frac{1}{2} Y_t + \frac{1}{4} Y_{t+1} = \sum_{r=-1}^1 a_r Y_{t+r}, \quad a_{-1} = a_1 = \frac{1}{4}, \quad a_0 = \frac{1}{2}$$

ergibt sich zu

$$\begin{aligned} G(\lambda) = |H(\lambda)| &= \left| \sum_{r=-1}^1 a_r e^{-i\lambda r} \right| = \left| \frac{1}{4} e^{i\lambda} + \frac{1}{2} e^{-0 \cdot i\lambda} + \frac{1}{4} e^{-i\lambda} \right| \\ &= \left| \frac{1}{2} + \frac{1}{4} (e^{i\lambda} + e^{-i\lambda}) \right| = \left| \frac{1}{2} + \frac{1}{2} \cos \lambda \right| = \frac{1}{2} + \frac{1}{2} \cos \lambda. \end{aligned}$$

Es ist in Abb. 14 für $0 \leq \lambda \leq \pi$ graphisch dargestellt, und wir erkennen, daß es sich um einen Tiefpaß-Filter handelt, der für niedrige Frequenzen stark und für hohe Frequenzen schwach durchlässig ist.

(b) Die ersten Differenzen, mithin der Filter

$$Y_t^* = \nabla Y = Y_t - Y_{t-1} = \sum_{r=-1}^0 a_r Y_{t+r} \quad \text{mit} \quad a_{-1} = -1, \quad a_0 = 1,$$

bilden einen Hochpaß-Filter. Das Amplitudendiagramm

$$\begin{aligned} G(\lambda) = |H(\lambda)| &= \left| \sum_{r=-1}^0 a_r e^{-i\lambda r} \right| = |-e^{i\lambda} + e^{0i\lambda}| = |1 - e^{i\lambda}| \\ &= |1 - \cos \lambda - i \sin \lambda| = ((1 - \cos \lambda)^2 + (-\sin \lambda)^2)^{1/2} \\ &= (1 - 2 \cos \lambda + \cos^2 \lambda + \sin^2 \lambda)^{1/2} = \sqrt{2 - 2 \cos \lambda} \end{aligned}$$

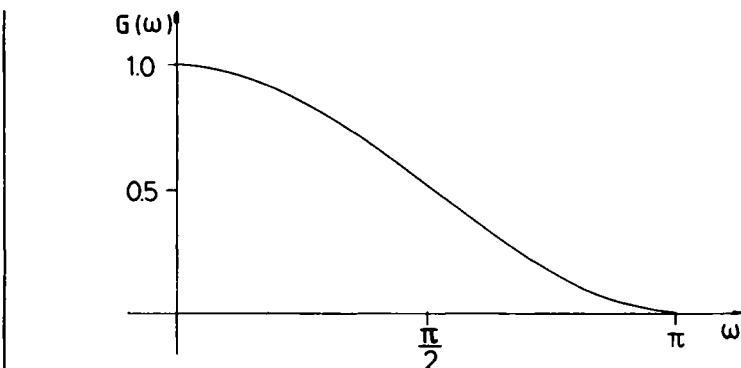


Abb. 14: Amplitudendiagramm $G(\omega)$ des Tiefpaß-Filters $y_t^* = \frac{1}{4}y_{t-1} + \frac{1}{2}y_t + \frac{1}{4}y_{t+1}$

ist in **Abb. 15** dargestellt und lässt niedrige Frequenzen schwach und hohe Frequenzen stark durch.

(c) Ein einfacher gleitender Durchschnitt der Ordnung 3

$$Y_t^* = \frac{1}{3}(Y_{t-1} + Y_t + Y_{t+1}) = \sum_{r=-1}^1 \frac{1}{3} Y_{t+r}$$

hat das Amplitudendiagramm

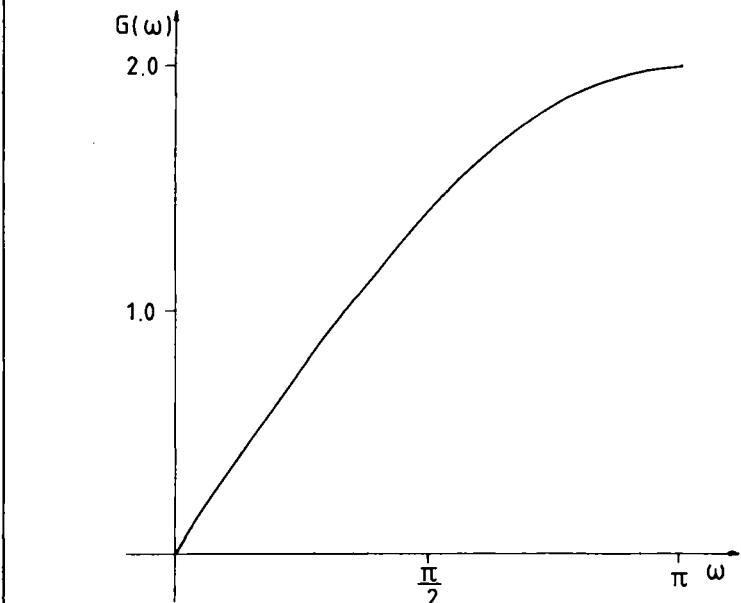


Abb. 15: Amplitudendiagramm $G(\omega)$ des Hochpaß-Filters $y_t^* = y_t - y_{t-1}$

$$\begin{aligned} G(\lambda) = |H(\lambda)| &= \left| \sum_{r=-1}^1 \frac{1}{3} e^{-i\lambda r} \right| = \left| \frac{1}{3} e^{i\lambda} + \frac{1}{3} + \frac{1}{3} e^{-i\lambda} \right| \\ &= \left| \frac{1}{3} + \frac{1}{3} (e^{i\lambda} + e^{-i\lambda}) \right| = \left| \frac{1}{3} + \frac{2}{3} \cos \lambda \right|. \end{aligned}$$

Es ist in Abb. 16 veranschaulicht und zeigt, daß dieser Filter eine Bandsperre ist, bei der Frequenzen λ in der Nähe von $\frac{2}{3}\pi$ nicht durchgelassen werden.

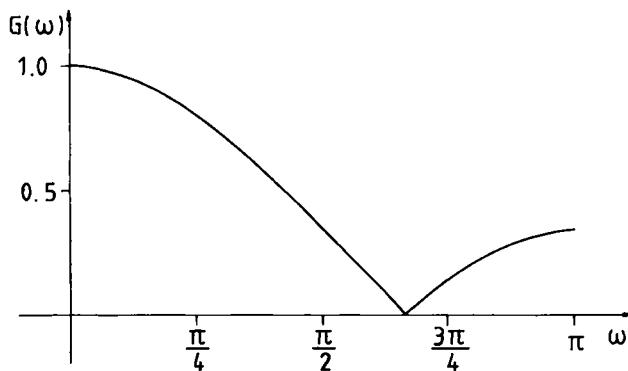


Abb. 16: Amplitudendiagramm $G(\omega)$ der Bandsperre $y_t^* = (y_{t-1} + y_t + y_{t+1})/3$ (gleitender Durchschnitt der Ordnung 3)

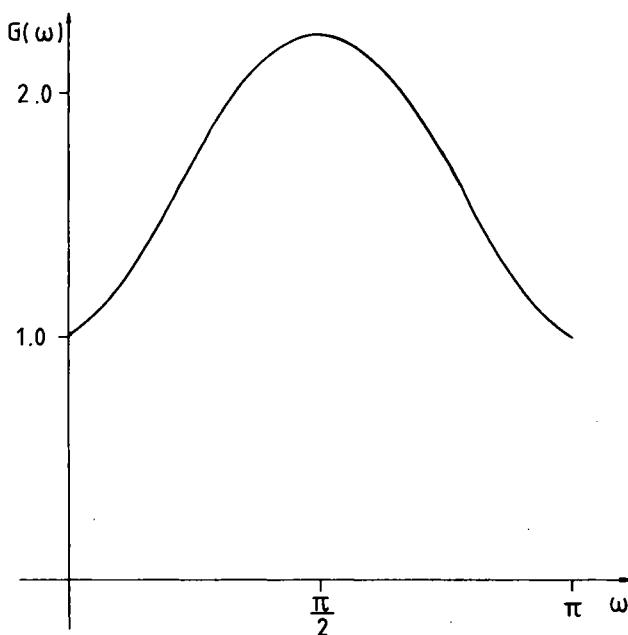


Abb. 17: Amplitudendiagramm $G(\omega)$ des Bandpasses $y_t^* = y_t - y_{t-1} - y_{t-2}$

(d) In der Abb. 17 ist das Amplitudendiagramm des Filters

$$Y_t^* = Y_t - Y_{t-1} - Y_{t-2} = \sum_{r=-2}^0 a_r Y_{t+r} \quad \text{mit} \quad a_{-2} = a_{-1} = -1, a_0 = 1$$

dargestellt. Dieser Bandpaß, der mittlere Frequenzen (λ in der Nähe von $\frac{\pi}{2}$) besonders hervorhebt, ergibt sich mit

$$\begin{aligned} G(\lambda) &= |H(\lambda)| = \left| \sum_{r=-2}^0 a_r e^{-i\lambda r} \right| = |-e^{2i\lambda} - e^{i\lambda} + 1| \\ &= |1 - \cos(2\lambda) - i \sin(2\lambda) - \cos\lambda - i \sin\lambda| \\ &= [(1 - \cos(2\lambda) - \cos\lambda)^2 + (-\sin(2\lambda) - \sin\lambda)^2]^{1/2} \\ &= [1 - 2\cos(2\lambda) - 2\cos\lambda + \cos^2(2\lambda) + 2\cos(2\lambda)\cos\lambda + \cos^2\lambda \\ &\quad + \sin^2(2\lambda) + \sin^2\lambda + 2\sin(2\lambda)\sin\lambda]^{1/2} \\ &= [3 - 2\cos(2\lambda) - 2\cos\lambda + 2\cos(2\lambda)\cos\lambda + 2\sin(2\lambda)\sin\lambda]^{1/2}. \end{aligned}$$

Mit

$$\cos x \cdot \cos y + \sin x \cdot \sin y = \cos(x - y) \quad \text{für alle } x, y$$

erhält man dann weiter

$$\begin{aligned} G(\lambda) &= [3 - 2\cos(2\lambda) - 2\cos\lambda + 2\cos(2\lambda - \lambda)]^{1/2} \\ &= [3 - 2\cos(2\lambda)]^{1/2} = \sqrt{3 - 2\cos(2\lambda)}. \end{aligned}$$

3.4. Schätzen der Spektraldichte

Bis jetzt haben wir uns nur mit der Spektralanalyse schwach stationärer Prozesse (Y_t) beschäftigt. Insbesondere haben wir die (theoretische) Spektraldichte $f_Y(\lambda) = f_Y(\lambda)$ betrachtet. Liegt nun eine endliche Realisation, also eine Zeitreihe y_1, \dots, y_n , eines stochastischen Prozesses (Y_t) vor, so müssen wir die Spektraldichte des Prozesses anhand der Realisationen y_1, \dots, y_n schätzen.

3.4.1. Das Periodogramm und das Stichprobenspektrum

Wir haben die Spektraldichte $f_Y(\lambda)$ eines schwach stationären stochastischen Prozesses (Y_t) mit $EY_t = 0$ für alle t im Abschnitt 3.2 u.a. in der Form

$$f_Y(\lambda) = \frac{\gamma(0)}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \gamma(k) \cos(\lambda k) = f_Y(-\lambda) \quad \text{für } 0 \leq \lambda \leq \pi$$

dargestellt. Eine Spektraldichte, die nur auf einem Ausschnitt Y_1, \dots, Y_n des Prozesses basiert, hat die Gestalt

$$f_{Y,n}(\lambda) = \frac{\gamma(0)}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \gamma(k) \cos(\lambda k).$$

Gesucht ist nun ein erwartungstreuer Schätzer $\tilde{f}_{Y,n}(\lambda)$ für diese Spektraldichte, ausgehend von der Zeitreihe y_1, \dots, y_n .

Dazu betrachten wir das sogenannte **Periodogramm** zur Zeitreihe y_1, \dots, y_n mit $\bar{y} = 0$. Es ist definiert als

$$R^2(\lambda) = A^2(\lambda) + B^2(\lambda)$$

mit

$$A(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^n y_j \cos(\lambda j), \quad B(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^n y_j \sin(\lambda j).$$

Damit ergibt sich

$$\begin{aligned} R^2(\lambda) &= \frac{1}{2\pi n} \left(\left(\sum_{j=1}^n y_j \cos(\lambda j) \right)^2 + \left(\sum_{j=1}^n y_j \sin(\lambda j) \right)^2 \right) \\ &= \frac{1}{2\pi n} \left(\sum_{j=1}^n \sum_{j'=1}^n y_j y_{j'} \cos(\lambda j) \cos(\lambda j') \right. \\ &\quad \left. + \sum_{j=1}^n \sum_{j'=1}^n y_j y_{j'} \sin(\lambda j) \sin(\lambda j') \right) \\ &= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{j'=1}^n y_j y_{j'} \cos(\lambda j - \lambda j'). \end{aligned}$$

Das Periodogramm $R^2(\lambda) = \tilde{f}_{Y,n}(\lambda)$ ist nun ein erwartungstreuer Schätzer für die Spektraldichte $f_{Y,n}(\lambda)$, denn

$$\begin{aligned} E R^2(\lambda) &= E \left[\frac{1}{2\pi n} \sum_{j=1}^n \sum_{j'=1}^n y_j y_{j'} \cos(\lambda j - \lambda j') \right] \\ &= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{j'=1}^n E(y_j y_{j'}) \cos(\lambda j - \lambda j') \\ &= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{j'=1}^n \gamma(j - j') \cos(\lambda j - \lambda j') \\ &= \frac{1}{2\pi n} \sum_{j=-n+1}^{n-1} (n - |j|) \gamma(j) \cos(\lambda j) \\ &= \frac{1}{2\pi} \sum_{j=-n+1}^{n-1} \left(1 - \frac{|j|}{n} \right) \gamma(j) \cos(\lambda j) \\ &= \frac{\gamma(0)}{2\pi} + \frac{1}{\pi} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n} \right) \gamma(j) \cos(\lambda j) \\ &= f_{Y,n}(\lambda). \end{aligned}$$

Anstelle des Periodogramms $R^2(\lambda)$ wird wegen der leichteren Berechenbarkeit zumeist ein verzerrter Schätzer für die Spektraldichte verwandt. Beim sogenannten **Stichprobenspektrum**

$$\hat{f}_{Y,n}(\lambda) = \frac{c(0)}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{n-1} c(k) \cos(\lambda k)$$

werden die theoretischen Autokovarianzen $\gamma(k)$ ersetzt durch $\frac{n}{n-k} c(k)$, wobei $c(k)$ die empirische lag k -Autokovarianz der Zeitreihe y_1, \dots, y_n bezeichnet, vgl. Abschnitt 1.4. Mit wachsendem Stichprobenumfang n wird die Verzerrung des Schätzers $\hat{f}_{Y,n}(\lambda)$ für $f_{Y,n}(\lambda)$ immer kleiner.

Das Stichprobenspektrum, das wir in den Abschnitten 3.4.2 und 3.4.3 als Schätzer für die Spektraldichte zugrundelegen werden, hat gegenüber dem Periodogramm den Vorteil, daß es direkt auf Zeitreihen y_1, \dots, y_n mit $\bar{y} \neq 0$ angewandt

werden kann. Beim Periodogramm muß in einem solchen Fall y_j durch $y_j - \bar{y}$ ersetzt werden.

Einem Vorschlag von Blackman/Tukey (1959) folgend, wird das Stichproben-spektrum im Bereich $0 \leq \lambda \leq \pi$ nur an $M + 1$ Stellen $\lambda_j = \pi_j/M$, $j = 0, 1, 2, \dots, M$, geschätzt. Die Schätzwerte $\hat{f}_{Y,n}(\lambda_j)$ werden für $j = 0, \dots, M$ dann durch einen Polygongzug verbunden. Der Wert des Stichprobenspektrums an der Stelle λ_j , $j \neq 0$, entspricht dann gerade der Stärke der Schwingungskomponente mit Dauer $2M/j$. Hat man etwa Monatswerte beobachtet und berechnet λ_j für $j = 0, \dots, 24 (= M)$, so entspricht $\hat{f}_{Y,n}(\pi/6)$ gerade der Stärke der Jahresschwingung, denn an der Stelle $\lambda_4 = 4\pi/24 = \pi/6$ ergibt sich die Schwingungsdauer $2M/j = 2 \cdot 24/4 = 12$ (Monate).

3.4.2. Geglättete Spektraldichteschätzungen, Spektral- und Lag-Fenster

Das Stichprobenspektrum $\hat{f}_{Y,n}(\lambda)$ ist zwar ein asymptotisch erwartungstreuer Schätzer (die Verzerrung geht für $n \rightarrow \infty$ gegen 0) für die Spektraldichte; jedoch ist es nicht konsistent, denn seine Varianz ist unabhängig vom Stichprobenumfang n .

Um konsistente Schätzer für die Spektraldichte zu erhalten, glättet man das Stichprobenspektrum, d. h. man mittelt benachbarte Werte des Stichprobenspektrums. Ein solches **geglättetes Stichprobenspektrum** berechnet man durch unterschiedliche Gewichtung der empirischen Autokovarianzen $c(k)$ der Zeitreihe y_1, \dots, y_n . D.h. man berechnet anstelle von $\hat{f}_{Y,n}(\lambda)$ nun

$$\hat{f}_{Y,n}^w(\lambda) = \frac{c(0)}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{n-1} w(k)c(k)\cos(\lambda k),$$

wobei $w(k)$ das Gewicht von $c(k)$ bezeichnet.

Allgemein betrachtet man beliebige Gewichtsfunktionen $w_m(k)$ für die gilt

$$w_m(0) = 1, \quad w_m(k) = w_m(-k), \quad w_m(k) = 0 \quad \text{für } k > m.$$

Solche Gewichtsfunktionen $w_m(k)$ nennt man auch **lag-Fenster** der Breite m .

Berechnet werden solche geglätteten Stichprobenspektren dann an $M + 1$ Stellen $\lambda_j = \pi_j/M$, $j = 0, 1, 2, \dots, M$, wobei $M = m, 2m, 3m, \dots$ gewählt wird.

Das wohl einfachste lag-Fenster ist das **Rechteck- oder Daniell-Fenster**

$$w_m(k) = \begin{cases} 1 & \text{für } k = 0, \dots, m \\ 0 & \text{für } k > m \end{cases},$$

das bewirkt, daß im Stichprobenspektrum nur die empirischen Autokovarianzen $c(0), \dots, c(m)$, diese aber alle gleich stark, berücksichtigt werden.

Bevor wir weitere spezielle Fenster betrachten, wollen wir den Begriff des **Spektralfensters** einführen. Das zu einem lag-Fenster korrespondierende Spektralfenster ist gegeben durch

$$W_m(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} w_m(k)e^{-ik\lambda} \quad \text{für } -\pi \leq \lambda \leq \pi.$$

Umgekehrt gilt

$$w_m(k) = \int_{-\pi}^{\pi} W_m(\lambda)e^{ik\lambda} d\lambda$$

und das geglättete Stichprobenspektrum lässt sich auch darstellen als

$$\hat{F}_{Y,n}^w(\lambda) = \int_{-\pi}^{\pi} \hat{f}_{Y,n}(\omega) W_m(\lambda - \omega) d\omega.$$

Jedes Spektralfenster $W_m(\lambda)$ genügt folgenden Bedingungen:

$$\int_{-\pi}^{\pi} W_m(\lambda) d\lambda = 1, \quad W_m(\lambda) = W_m(-\lambda), \quad \lim_{m \rightarrow \infty} W_m(\lambda) = \delta(\lambda) \quad [\text{Dirac-Funktion}].$$

Letztere Bedingung ist erfüllt, wenn W_m für $\lambda = 0$ ein absolutes Maximum besitzt und dann relativ schnell abfällt; mit wachsendem m wird $W_m(0)$ immer größer und der Abfall wird immer schneller ($\delta(\lambda) = \infty$ für $\lambda = 0$ und 0 sonst).

Das Spektralfenster zum (lag-) Rechteck-Fenster hat gerade die Gestalt

$$W_m(\lambda) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^m \cos(\lambda k).$$

Basierend auf diesem Fenster wird auch die **Bandbreite** β eines Fensters definiert. Die Bandbreite β eines Fensters ist die Breite des Spektral-Rechteck-Fensters mit gleicher Varianz wie das betrachtete Fenster. Die Bandbreite sollte möglichst nicht größer gewählt werden als der Abstand zwischen zwei Spitzen (Peaks) im Stichprobenspektrum, die noch zum Ausdruck kommen sollen.

Wir wollen nun noch einige häufig benutzte Fenster betrachten.

Das **Bartlett-Fenster** oder **Dreieck-Fenster** hat in seiner lag-Form die Gestalt

$$w_m(k) = \begin{cases} 1 - \frac{k}{m} & \text{für } k = 0, \dots, m \\ 0 & \text{für } k > m \end{cases};$$

das zugehörige Spektralfenster ist

$$W_m(\lambda) = \frac{1}{2\pi m} \left(\frac{\sin(m\lambda/2)}{\sin(\lambda/2)} \right)^2$$

und seine Bandbreite ist $\beta = \pi/m$.

Das **Tukey-Fenster** mit der Bandbreite $\beta = 2,67\pi/m$ hat die lag-Form

$$w_m(k) = \begin{cases} \frac{1}{2} + \frac{1}{2} \cos \frac{\pi k}{m} & \text{für } k = 0, \dots, m \\ 0 & \text{für } k > m \end{cases}$$

und die Spektralform

$$W_m(\lambda) = \frac{1}{2\pi} \left(\frac{\sin((m+0,5)\lambda)}{2\sin(\lambda/2)} + \frac{\sin((m+0,5)(\lambda+\pi/m))}{4\sin(0,5(\lambda+\pi/m))} \right. \\ \left. + \frac{\sin((m+0,5)(\lambda-\pi/m))}{4\sin(0,5(\lambda-\pi/m))} \right).$$

An der Stelle $\lambda = 0$ hat $W_m(\lambda)$ sein absolutes Maximum

$$W_m(0) = m/(2\pi)$$

und die Nullstellen von $W_m(\lambda)$ liegen bei $\lambda = \pm (j+1)\pi/m$ für $j = 1, 2, \dots, m-1$. Schließlich ist

$$w_m(k) = \begin{cases} 1 - 6(k/m)^2 + 6(k/m)^3 & \text{für } 0 \leq k < m/2 \\ 2(1 - k/m)^3 & \text{für } m/2 \leq k \leq m \\ 0 & \text{für } k > m \end{cases}$$

die lag-Form und

$$W_m(\lambda) = \frac{3}{8\pi m^3} \left(\frac{\sin(m\lambda/4)}{\sin(\lambda/4)} \right)^4$$

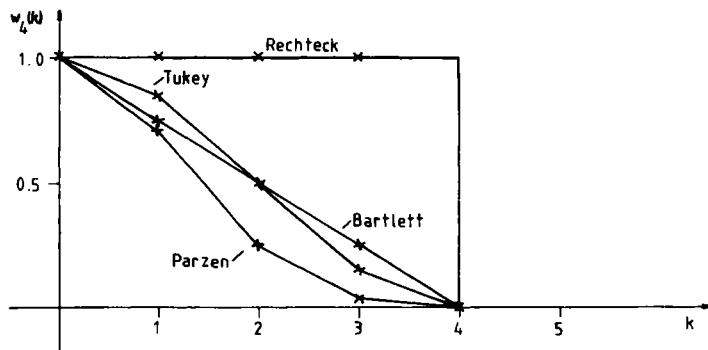


Abb. 18: Bartlett-, Parzen-, Tukey- und Rechteck-lag-Fenster $w_4(k)$ für $m = 4$

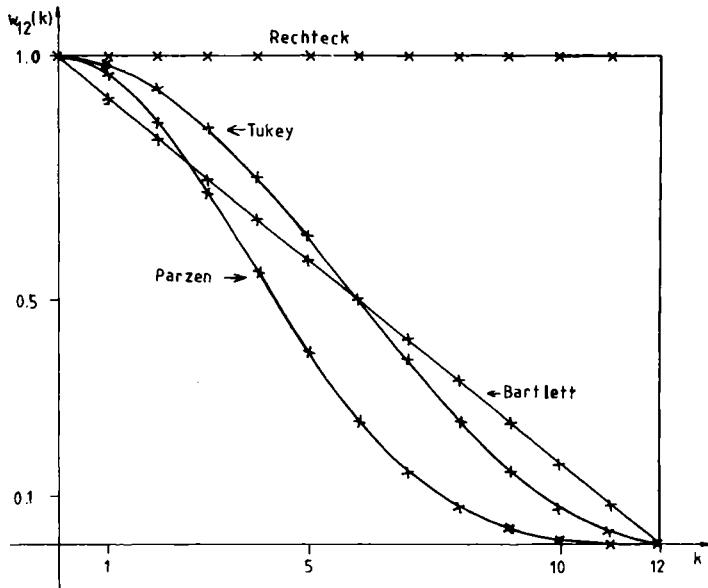


Abb. 19: Bartlett-, Parzen-, Tukey- und Rechteck-lag-Fenster $w_{12}(k)$ für $m = 12$

die Spektralform des **Parzen-Fensters**, das die Bandbreite $\beta = 3,71\pi/m$ hat. Das absolute Maximum des Parzen-Fensters an der Stelle $\lambda = 0$ ist

$$W_m(0) = 3m/(8\pi)$$

und die Nullstellen liegen bei $\lambda = \pm 4\pi j/m$, wobei j die Zahlen $1, 2, \dots$ bis zur größten ganzen Zahl $\leq m/4$ durchläuft.

In Abb. 18 und Abb. 19 sind die lag-Fenster von Bartlett-, Tukey- und Parzen sowie das Rechteck-Fenster für $m = 4$ und $m = 12$ dargestellt und in Abb. 20 bzw. Abb. 21 findet man für $m = 12$ die Spektralfenster von Tukey und Parzen. Man sieht, daß das Spektralfenster von Parzen für alle λ Werte größer oder gleich Null liefert, wohingegen das Tukey-Fenster **negative Seitenfenster** besitzt, die natürlich nicht wünschenswert sind, d. h. für einige λ auch negativ ist.

Eine andere Klasse von Fenstern, die wir hier nicht näher behandeln wollen, sind sogenannte **Splinefenster**, bei denen die Glättung des Stichprobenspektrums mittels Splinefunktionen erfolgt. Stetige und gemischt stetig-diskrete Splinefenster werden in Cogburn/Davis (1974) eingeführt; man vgl. aber auch Hebbel/Heiler (1978), wo zudem ein rein diskretes Splinefenster betrachtet wird.

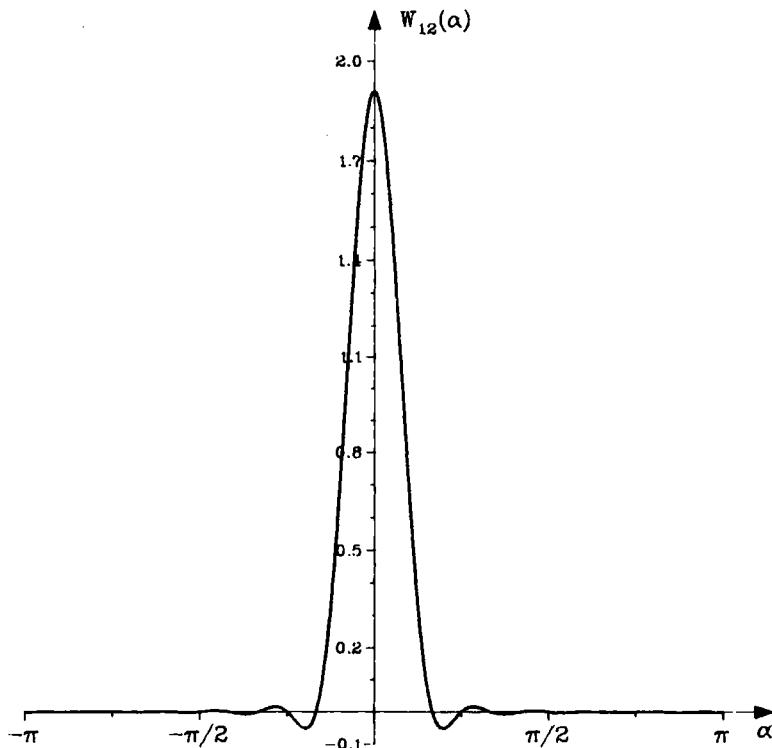
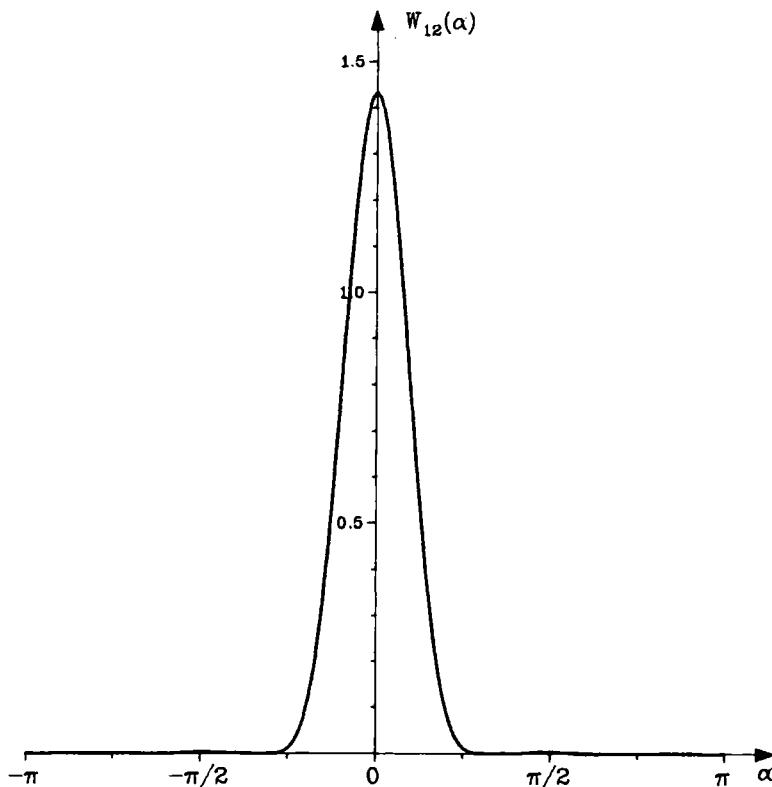


Abb. 20: Tukey-Spektralfenster $W_{12}(\alpha)$ für $m = 12$

Abb. 21: Parzen-Spektralfenster $W_{12}(\alpha)$ für $m = 12$

3.4.3. Ein Beispiel zum Stichprobenspektrum

Im Abschnitt 1 haben wir die Zeitreihe y_1, \dots, y_{51} der Arbeitslosen im Baugewerbe am Monatsende in den Monaten Juli 1975 bis September 1979 betrachtet; die Ausgangsdaten sind in der Tab. 9 angegeben. Insbesondere haben wir im Abschnitt 1.3.3, vgl. Tab. 12 und Abb. 8, die ersten Differenzen $\nabla y_t = y_t^* = y_t - y_{t-1}$ berechnet und im Abschnitt 1.4 die empirischen Autokovarianzen $c(k)$ bzw. die empirischen Autokorrelationen $r(k)$ für $k = 0, \dots, 49$ der Reihe der ersten Differenzen, vgl. Tab. 14, d. h. das Korrelogramm dieser Reihe, vgl. Abb. 9, bestimmt. Am Korrelogramm wurde durch die Peaks an den lags 12, 24 und 36 die stark ausgeprägte Jahresschwankung (12 Monats-Schwankung) deutlich. Um die Auswirkung einer solchen Schwankung im Frequenzbereich deutlich zu machen, wollen wir nun das Stichprobenspektrum der ersten Differenzen der Arbeitslosen berechnen. Dazu verwenden wir natürlich die Autokovarianzen $c(k)$ für $k = 0, \dots, 49$ aus Tab. 14.

Um auch den Glättungseffekt von Fenstern deutlich zu machen, berechnen wir neben dem ungeglätteten Stichprobenspektrum $\hat{f}_{Y^*, 50}(\lambda)$ auch die mittels Tukey- und Parzen-Fenster geglätteten Stichprobenspektren $f_{Y^*, 50}^*(\lambda)$, wobei jeweils $m = 12$ gewählt wird. Die Berechnungen der Stichprobenspektren werden hier jeweils an den Stellen

$$\lambda_j = \pi j / 12 \quad \text{für } j = 0, \dots, 12 = M = m$$

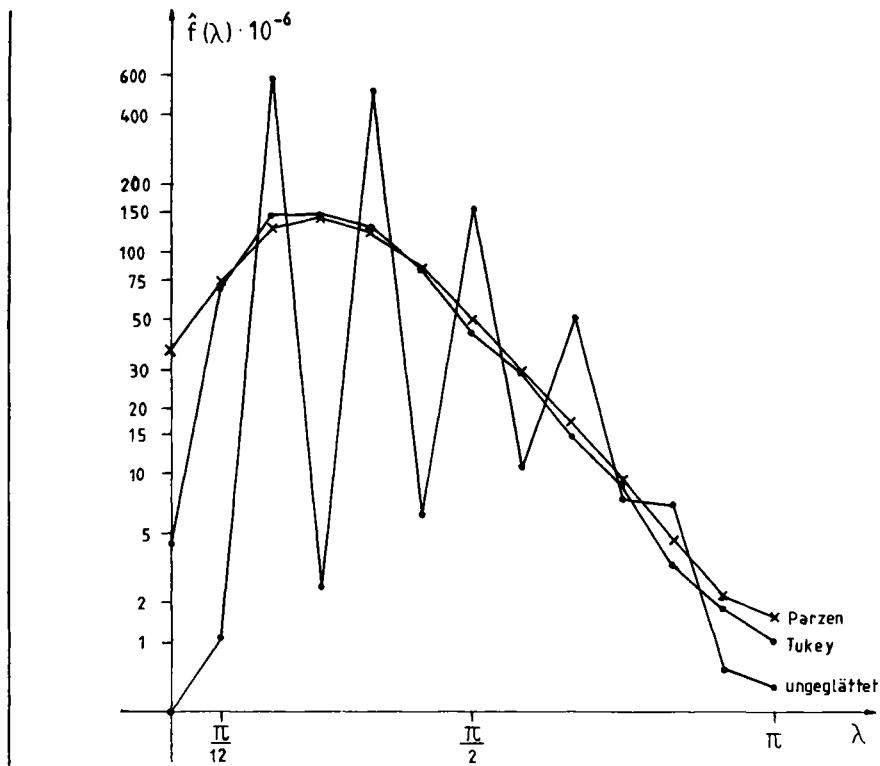


Abb. 22: Ungeglättetes Stichprobenspektrum und geglättete Stichprobenspektren (mittels Tukey- und Parzen-Fenster mit $m = 12$) berechnet an den Stellen $\pi j/12$ für $j = 0, \dots, 12$

d. h. $\lambda_0 = 0, \lambda_1 = \pi/12, \lambda_2 = 2\pi/12, \dots, \lambda_{11} = 11\pi/12, \lambda_{12} = \pi$ vorgenommen und sind in Tab. 17 zusammengestellt und in Abb. 22 graphisch veranschaulicht.

Da wir $M = 12$ gewählt haben, entspricht λ_j gerade der Schwingungskomponente mit Dauer $2M/j = 24/j$ Monate. λ_1 entspricht also der Schwingungskomponente mit Dauer 24 Monate = 2 Jahre, λ_2 der mit Dauer 12 Monate = 1 Jahr, λ_3 der mit Dauer 8 Monate, \dots, λ_{12} der mit Dauer 2 Monate. Im ungeglätteten Stichprobenspektrum sieht man nun sehr schön die Peaks bei $\lambda_2 = \pi/6, \lambda_4 = 2\pi/6, \lambda_6 = 3\pi/6, \lambda_8 = 4\pi/6$ und leicht noch bei $\lambda_{10} = 5\pi/6$. $\lambda_2 = \pi/6$ entspricht gerade der Schwingungskomponente mit Dauer 1 Jahr, drückt also die Jahresschwingung aus; die übrigen Peaks haben keine eigenständige Bedeutung, da sie gerade die Harmonischen der Schwingungskomponente mit Dauer 1 Jahr sind. Allgemein sind die **Harmonischen** oder **Oberschwingungen** einer Frequenz λ gerade die Frequenzen $2\lambda, 3\lambda, 4\lambda, \dots$. Betrachtet man die geglätteten Stichprobenspektren, so verschwinden alle Peaks; lediglich die Peaks bei λ_2 und λ_4 werden zu einem Peak bei λ_3 zusammengezogen, da sie ursprünglich ungefähr gleich stark waren. Eine Trennung zwischen Jahres- und Halbjahresschwingung ist bei der Fensterbreite $m = 12$ hier nicht möglich, da die Bandbreite des Tukey-Fensters etwa $\beta = 2,67\pi/12 = 0,2225\pi$ und die des Parzen-Fensters etwa $\beta = 3,71\pi/12 = 0,30917\pi$ beträgt. Diese Bandbreiten sind größer als $\pi/6$, so daß Peaks mit diesem Abstand nicht mehr einzeln zum Ausdruck kommen können.

Tab. 17: Ungeglättetes und mittels Tukey- bzw. Parzen-Fenster der Breite 12 geglättete Stichprobenspektren für die Reihe der ersten Differenzen der Arbeitslosen im Baugewerbe, vgl. Tab. 9 und Tab. 12 (Werte hier multipliziert mit $10^{-6} = 0,000001$)

j	λ_j	$\hat{F}_{Y^*, 50}(\lambda_j) \cdot 10^{-6}$	$\hat{F}_{Y^*, 50}(\lambda_j) \cdot 10^{-6}$
		Tukey-Fenster	Parzen-Fenster
0	0	0,0000	4,5010
1	$\pi/12$	1,1090	69,8912
2	$\pi/6$	583,9561	146,0097
3	$\pi/4$	2,5440	147,2832
4	$\pi/3$	521,0156	129,9427
5	$5\pi/12$	6,3825	83,4176
6	$\pi/2$	156,8483	43,3293
7	$7\pi/12$	10,8471	29,3740
8	$2\pi/3$	50,6343	15,0405
9	$3\pi/4$	7,5252	8,7286
10	$5\pi/6$	7,1214	3,4046
11	$11\pi/12$	0,5338	1,7861
12	π	0,2847	1,0066

3.5. Die harmonische Analyse einer Zeitreihe

Wie wir im Abschnitt 3.4.3 gesehen haben, wird die Reihe der ersten Differenzen der Arbeitslosen im Baugewerbe, vgl. Tab. 9 und Tab. 12, im wesentlichen durch die Frequenz $\pi/6$ und deren Oberschwingungen bestimmt. Dies bedeutet, daß sich diese Zeitreihe y_t^* , $t = 2, \dots, 51$, durch eine Reihe

$$\hat{y}_t^* = \alpha + \sum_{j=1}^6 \left(\hat{\beta}_j \cos\left(\frac{j\pi}{6} t\right) + \hat{\gamma}_j \sin\left(\frac{j\pi}{6} t\right) \right)$$

mit geschätzten Parametern $\alpha, \beta_1, \dots, \beta_6, \gamma_1, \dots, \gamma_6$ approximieren läßt.

Wie in diesem Beispiel weiß oder erkennt man – vor allem bei Zeitreihen aus dem Bereich der Naturwissenschaften – mitunter, daß sich eine Zeitreihe y_1, \dots, y_n bis auf eine additive Konstante α aus h Sinus- und Cosinusfunktionen der Perioden $\omega_1, \dots, \omega_h$ bzw. der Frequenzen $\lambda_1 = 2\pi/\omega_1, \dots, \lambda_h = 2\pi/\omega_h$ zusammensetzt. Eine solche Zeitreihe läßt sich dann durch

$$\hat{y}_t = \alpha + \sum_{j=1}^h (\hat{\beta}_j \cos(\lambda_j t) + \hat{\gamma}_j \sin(\lambda_j t)), \quad \omega_j = 2\pi/\lambda_j$$

approximieren, wobei die Parameterschätzungen nach der Methode der kleinsten Quadrate, vgl. Kap. X, ausgehend vom Modellansatz

$$y_t = \alpha + \sum_{j=1}^h (\beta_j \cos(\lambda_j t) + \gamma_j \sin(\lambda_j t)) + \varepsilon_t \quad \text{für } t = 1, \dots, n$$

bestimmt werden; hierbei bezeichnet ε_t ein weißes Rauschen. Man nennt dieses Vorgehen auch **harmonische Analyse**.

Besonders einfach ist die Parameterschätzung, falls die Perioden $\omega_1, \dots, \omega_h$ ganzzählige Teiler der Anzahl n der Beobachtungen sind, d. h. n/ω_j ganzzahlig. In diesem Falle erhalten wir als Schätzung für den Parameter α

$$\hat{\alpha} = \frac{1}{n} \sum_{t=1}^n y_t,$$

als Schätzung für die Parameter β_1, \dots, β_h

$$\hat{\beta}_j = \begin{cases} \frac{2}{n} \sum_{t=1}^n y_t \cos(\lambda_j t), & \text{falls } \omega_j > 2 \\ \frac{1}{n} \sum_{t=1}^n (-1)^t y_t, & \text{falls } \omega_j = 2 \end{cases}$$

und als Schätzung für die Parameter $\gamma_1, \dots, \gamma_h$

$$\hat{\gamma}_j = \begin{cases} \frac{2}{n} \sum_{t=1}^n y_t \sin(\lambda_j t), & \text{falls } \omega_j > 2 \\ 0, & \text{falls } \omega_j = 2 \end{cases}$$

Natürlich können nun auch Werte y_{n+1}, y_{n+2}, \dots prognostiziert werden. Dabei nützt man aus, daß \hat{y}_t betrachtet als Funktion der Zeit t streng periodisch mit Periode

$$P = \max \{\omega_1, \dots, \omega_h\}$$

ist, d.h. es gilt

$$\hat{y}_t = \hat{y}_{t+k \cdot P} \quad \text{für alle } t, k = 1, 2, 3, \dots$$

Beispiel: Wir wollen hier noch einmal auf die Zeitreihe y_2^*, \dots, y_{51}^* der ersten Differenzen der Arbeitslosen im Baugewerbe zurückkommen, vgl. Tab. 12, und basierend auf der im Stichprobenspektrum, vgl. Abschnitt 3.4.2, enthaltenen Information eine harmonische Analyse durchführen. Den Frequenzen $\lambda_j = j\pi/6$ für $j = 1, \dots, 6$ entsprechen die Perioden $\omega_j = 12/j$, für $j = 1, \dots, 6$, also die Perioden 12; 6; 4; 3; 2,4 und 2. Betrachten wir anstelle der 50 Beobachtungen y_2^*, \dots, y_{51}^* nur die letzten 48 Beobachtungen y_4^*, \dots, y_{51}^* , so sind alle Perioden ganzzahlige Teiler der Anzahl $n' = 48$ Beobachtungen. Wir können also die relativ einfachen, angegebenen Schätzformeln für die Parameter des Modells

$$y_{t'+3}^* = \alpha + \sum_{j=1}^6 (\beta_j \cos(\lambda_j t' + \gamma_j \sin(\lambda_j t')) + \varepsilon_t \quad \text{für } t' = 1, \dots, 48$$

verwenden und erhalten

$$\hat{\alpha} = \frac{1}{48} \sum_{t'=1}^{48} y_{t'+3}^* = -663,7083333,$$

$$\hat{\beta}_1 = \frac{1}{24} \sum_{t'=1}^{48} y_{t'+3}^* \cos\left(\frac{\pi}{6} t'\right) = 7390,6069420,$$

$$\hat{\beta}_2 = -13975,93750, \quad \hat{\beta}_3 = 9411,333333, \quad \hat{\beta}_4 = -4414,979167,$$

$$\hat{\beta}_5 = 479,4791667,$$

$$\hat{\beta}_6 = \frac{1}{48} \sum_{t'=1}^{48} (-1)^{t'} y_{t'+3}^*$$

$$= \frac{1}{48} (-y_4^* + y_5^* - y_6^* + y_7^* - \dots + y_{49}^* - y_{50}^* + y_{51}^*)$$

$$= 134,3750000,$$

$$\begin{aligned}\hat{\gamma}_1 &= \frac{1}{24} \sum_{t'=1}^{48} y_{t'+3}^* \sin\left(\frac{\pi}{6} t'\right) = 16285,18228, \\ \hat{\gamma}_2 &= -9184,524167, \quad \hat{\gamma}_3 = 189,0833333, \quad \hat{\gamma}_4 = 2551,274756, \\ \hat{\gamma}_5 &= -2028,223938 \quad \text{und} \\ \hat{\gamma}_6 &= 0 \quad \text{wegen} \quad \omega_6 = 2.\end{aligned}$$

In der Tab. 18 sind neben y_t^* für $t = 4, \dots, 51$, vgl. auch Tab. 12, die geschätzten Größen

$$\hat{y}_t^* = \alpha + \sum_{j=1}^6 (\beta_j \cos[\lambda_j(t-3)] + \hat{\gamma}_j \sin[\lambda_j(t-3)]) \quad \text{für } t = 4, \dots, 51,$$

die geschätzten Ursprungsgrößen, also Schätzungen für die Zahl der Arbeitslosen y_t aus Tab. 9,

$$\hat{y}_t = \hat{y}_t^* + y_{t-1} \quad \text{für } t = 4, \dots, 51$$

sowie die Differenzen $y_t - \hat{y}_t$ für $t = 4, \dots, 51$ angegeben. Wir sehen, daß die Schätzungen \hat{y}_t^* natürlich streng periodisch mit Periode 12 sind.

In Abb. 23 sind nun sowohl die Ursprungsdaten y_t als auch zum Vergleich die mittels harmonischer Analyse geschätzten Größen \hat{y}_t (diese zum Polygonzug verbunden) für $t = 4, \dots, 51$ graphisch dargestellt.

Die sehr gute Anpassung, die aus Abb. 23 ablesbar ist, wird bestätigt durch das Bestimmtheitsmaß

$$B_y = 1 - \frac{\sum_{t=4}^{51} (y_t - \hat{y}_t)^2}{\sum_{t=4}^{51} (y_t - \bar{y})^2} = 0,99, \quad \text{wobei} \quad \bar{y} = \frac{1}{48} \sum_{t=4}^{51} y_t = 50502,25.$$

Im Gegensatz zum Box-Jenkins-Ansatz in Abschnitt 2.6, wo sich bezogen auf die Originaldaten ebenso ein recht hohes Bestimmtheitsmaß von 0,96 ergab, ist hier jedoch auch die Güte der Anpassung im eigentlich betrachteten Problem (also die Güte der Anpassung für die ersten Differenzen) sehr hoch:

$$B_{y^*} = 1 - \frac{\sum_{t=4}^{51} (y_t^* - \hat{y}_t^*)^2}{\sum_{t=4}^{51} (y_t^* - \bar{y}_t^*)^2} = 0,97 \quad \text{mit} \quad \bar{y}_t^* = \alpha.$$

Mittels harmonischer Analyse lassen sich natürlich auch zukünftige Werte prognostizieren und zwar gilt wegen

$$P = \max \{\omega_1, \dots, \omega_6\} = \omega_1 = 12$$

für die ersten Differenzen

$$\hat{y}_t^* = \hat{y}_{t+k \cdot 12}^* \quad \text{für } t = 4, 5, \dots, k = 1, 2, \dots$$

Für die Prognose der Arbeitslosen im Baugewerbe selbst besagt dies, daß gilt

$$\begin{aligned}\hat{y}_{n+1} &= \hat{y}_{52} = \hat{y}_{52}^* + y_{51} = \hat{y}_4^* + y_{51} = 1980 + 15499 = 17479, \\ \hat{y}_{n+j} &= \hat{y}_{n+j}^* + y_{n+j-1} \quad \text{für } j = 2, 3, \dots\end{aligned}$$

Da y_{n+j-1} für $j = 2, 3, \dots$ nicht bekannt ist, wird es durch die Schätzung \hat{y}_{n+j-1}

Tab. 18: Erste Differenzen y_t^* der Arbeitslosen im Baugewerbe, Schätzungen \hat{y}_t^* , \hat{y}_t für die ersten Differenzen und die Ursprungsdaten y_t aus Tab. 9 sowie Differenzen $y_t - \hat{y}_t$ für $t = 4, \dots, 51$

t	y_t^*	\hat{y}_t^*	\hat{y}_t	$y_t - \hat{y}_t$
4	963	1980	49337	-1017
5	11899	8886	57206	3013
6	24199	22831	83050	1368
7	35498	40166	124584	-4668
8	4434	1498	121414	2936
9	-37041	-36202	88148	-839
10	-30274	-24626	62683	-5648
11	-17132	-11881	45154	-5251
12	-5850	-5305	34598	-545
13	-4148	-2507	31546	-1641
14	-1837	-1166	28739	-671
15	-1434	-1639	26429	205
16	2625	1980	28614	645
17	9683	8886	38145	797
18	26094	22831	61773	3263
19	45692	40166	105202	5526
20	-1797	1498	112226	-3295
21	-37414	-36202	72729	-1212
22	-17089	-24626	46891	7537
23	-11517	-11881	42547	364
24	-5788	-5305	37606	-483
25	-4079	-2507	34616	-1572
26	-2289	-1166	31878	-1123
27	-2013	-1639	29116	-374
28	2956	1980	30722	976
29	9729	8886	40584	843
30	22258	22831	64258	-573
31	35504	40166	103851	-4662
32	5051	1498	100687	3553
33	-28936	-36202	68038	7266
34	-31682	-24626	50678	-7056
35	-9632	-11881	31741	2249
36	-7171	-5305	28685	-1866
37	-1528	-2507	24312	979
38	-753	-1166	24125	413
39	-1853	-1639	22899	-214
40	1260	1980	24665	-720
41	4300	8886	32831	-4586
42	18772	22831	51076	-4059
43	43903	40166	87183	3737
44	-1580	1498	92418	-3078
45	-41548	-36202	53138	-5346
46	-19344	-24626	23166	5282
47	-9309	-11881	16567	2572
48	-2411	-5305	13834	2894
49	-205	-2507	14221	2302
50	99	-1166	15357	1265
51	-1123	-1639	14983	516

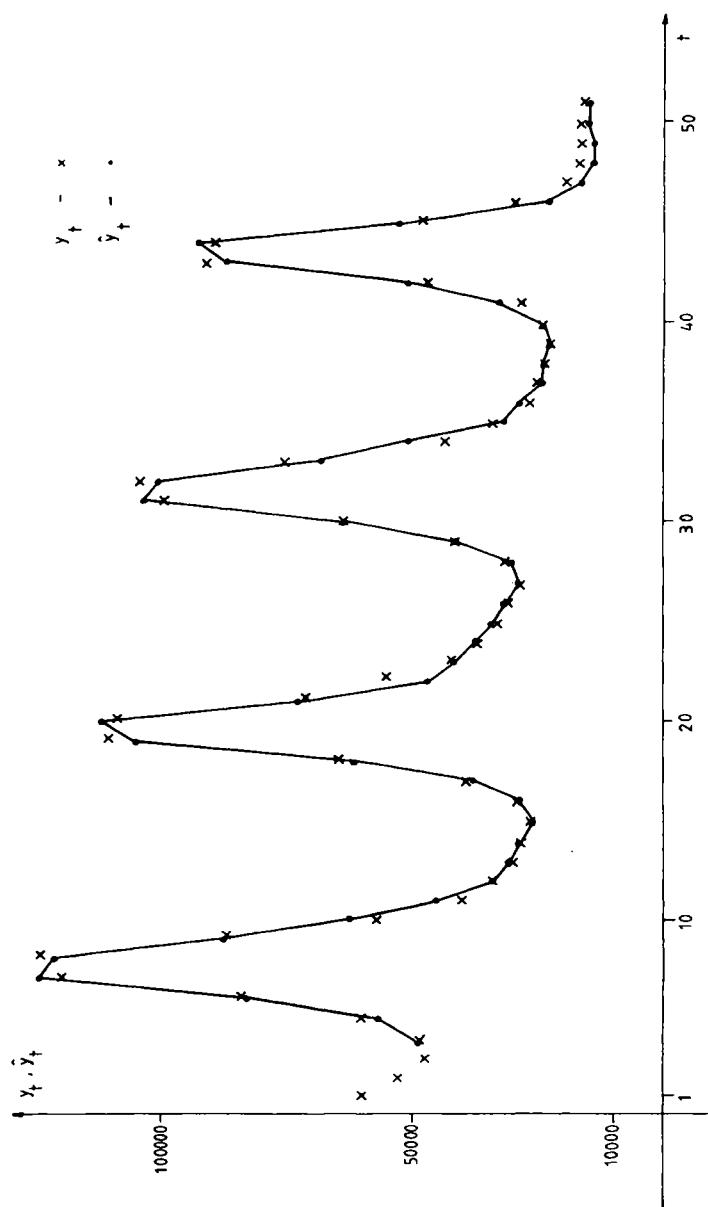


Abb. 23: Arbeitslose im Baugewerbe y_t von Mai 1975 ($t = 1$) bis September 1979 ($t = 51$) und mittels harmonischer Analyse prognostizierte Arbeitslosenzahlen \hat{y}_t

ersetzt. Für $j = 2$ erhalten wir somit z. B.

$$\hat{y}_{53} = \hat{y}_{53}^* + \hat{y}_{52} = \hat{y}_5^* + \hat{y}_{52} = 8886 + 17479 = 26365;$$

in der **Tab. 19** sind die Prognosen für y_{52}, \dots, y_{63} (Oktober 1979–September 1980) zusammengestellt. Zum Vergleich sind auch die echten Arbeitslosenzahlen für diese Monate aufgeführt und die Prognosefehler $y_t - \hat{y}_t$ angegeben. Für die ersten 6 Monate sind die Prognosen zwar noch recht gut; danach macht sich ein Strukturbruch bemerkbar, d.h. der Trend der Arbeitslosigkeit wird steigend, was mit unserem Modell natürlich so nicht erfassbar ist.

Tab. 19: Mittels harmonischer Analyse prognostizierte Arbeitslosenzahlen $\hat{y}_{52}, \dots, \hat{y}_{63}$ und echte Arbeitslosenzahlen y_{52}, \dots, y_{63} im Baugewerbe in den Monaten Oktober 1979 bis September 1980 (aus: „Wirtschaft und Statistik“, Hrsg. Statistisches Bundesamt Wiesbaden (1980–1981), Kohlhammer-Verlag, Stuttgart–Mainz) sowie Prognosefehler

Monat	t	y_t	\hat{y}_t	$y_t - \hat{y}_t$
Okt. 1979	52	17464	17479	— 15
Nov. 1979	53	24922	26365	— 1443
Dez. 1979	54	37683	49196	— 11513
Jan. 1980	55	82681	89362	— 6681
Febr. 1980	56	75552	90860	— 15308
März 1980	57	46234	54658	— 8424
April 1980	58	30830	30032	798
Mai 1980	59	25312	18151	7161
Juni 1980	60	23513	12846	10667
Juli 1980	61	24002	10339	13663
Aug. 1980	62	24491	9173	15318
Sept. 1980	63	24315	7534	16781

Wie im Abschnitt 5 geschildert wird, lassen sich Regressionsansätze wie z. B. die harmonische Analyse mit dem Box-Jenkins-Ansatz kombinieren, indem man die Annahme unabhängig identisch verteilter Fehler $\varepsilon_1, \dots, \varepsilon_n$ aufgibt und stattdessen an die Residuen $\hat{\varepsilon}_t = y_t - \hat{y}_t$ einen adäquaten ARMA(p, q)-Prozeß anpaßt. In unserem Beispiel der Arbeitslosen im Baugewerbe lohnt sich dies nicht, da die Reste $y_t - \hat{y}_t$ für $t = 4, \dots, 51$ aus Tab. 18 nahezu Realisation eines weißen Rauschens sind, das lediglich mit einer geringen Jahresschwankung überlagert ist; vgl. hierzu **Abb. 24**, in der diese Reste graphisch als Polygonzug dargestellt sind.

3.6. Das Berliner Verfahren zur Saisonbereinigung

Das auf Heiler zurückgehende **Berliner Verfahren**, vgl. Heiler (1966, 1970), Nullau et al. (1969), dient der Bereinigung einer Zeitreihe von saisonalen Schwankungen, vgl. auch Abschnitt 1.3. Es liegt derzeit in den verschiedensten Varianten vor; man vgl. etwa Nourney/Söll (1976), Hebbel/Kuhlmeyer (1983), Hebbel/Heiler (1985) sowie die Darstellung in Stier (1980).

Wir wollen hier nur die Grundprinzipien des Verfahrens kurz darstellen. Dazu betrachten wir einmal den Modellansatz der harmonischen Analyse

$$y_t = \alpha + \sum_{j=1}^h (\beta_j \cos(\lambda_j t) + \gamma_j \sin(\lambda_j t)) + \varepsilon_t \quad \text{für } t = 1, \dots, n,$$

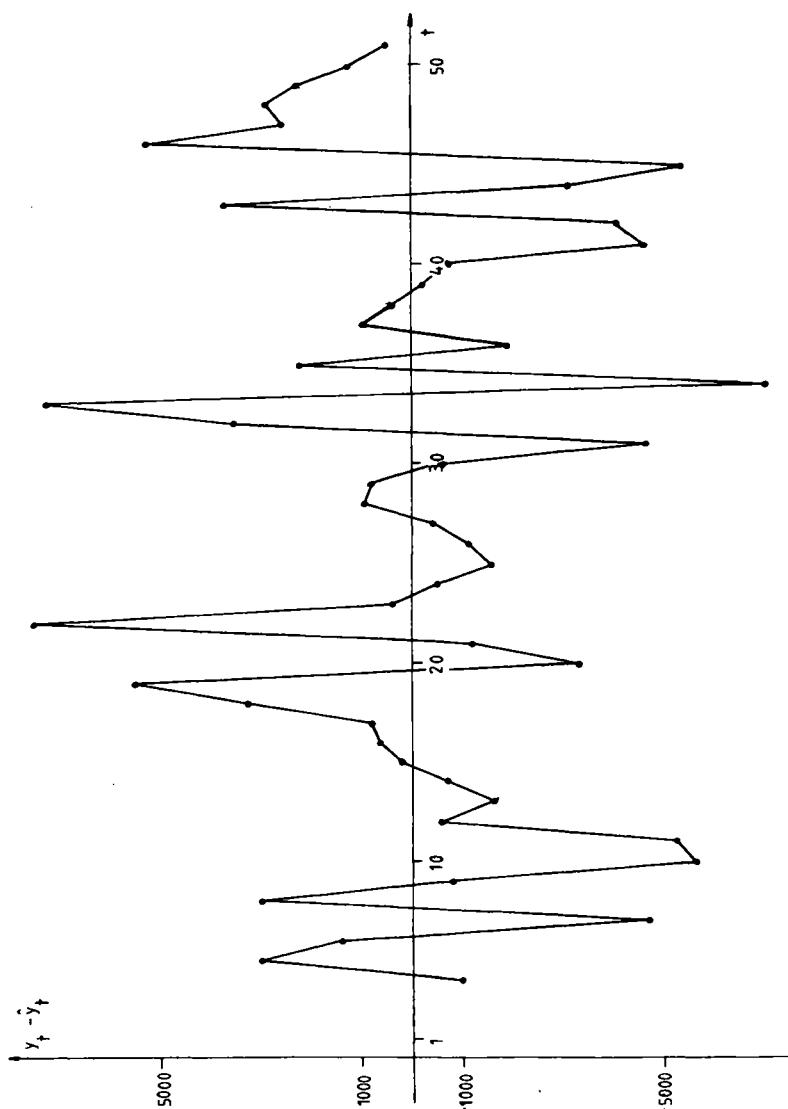


Abb. 24: Residuen $y_i - \hat{y}_i$ bei der Anpassung der Arbeitslosen im Baugewerbe mittels harmonischer Analyse

wobei ε_i ein weißes Rauschen bezeichnet und wie bisher die Zeitabstände zwischen je zwei Beobachtungen äquidistant sind. Beim Berliner Verfahren wird dieser Ansatz insofern verallgemeinert, als daß an die Stelle des Absolutglieds α ein Polynom k -ten Grades

$$\sum_{j=0}^k \alpha_j t^j = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_k t^k$$

tritt. Dieses Polynom steht dann für die glatte Komponente G_t der Zeitreihe, die Summe der h Cosinoide für die saisonale Komponente S_t und der Fehler ε_t für die irreguläre Komponente R_t . Bei der Wahl der Frequenzen λ_j geht man nun von einer konstanten Saisonfigur mit Periode P aus. Man betrachtet die zu dieser Periode gehörige Frequenz $\lambda_1 = 2\pi/P$ sowie zugehörige Oberschwingungen $\lambda_j = 2\pi j/P$, $j = 2, 3, \dots$, wobei j maximal den Wert $P/2$ annehmen kann. Zu beachten ist, daß für $\lambda_j = \pi$ der Parameter γ_j entfällt, da $\sin(\pi t) = 0$ für alle t ist.

Nun betrachtet man den Modellansatz

$$y_t = \sum_{j=0}^k \alpha_j t^j + \sum_{j=1}^h (\beta_j \cos(\lambda_j t) + \gamma_j \sin(\lambda_j t)) + \varepsilon_t$$

nicht für die gesamte Zeitreihe y_1, \dots, y_n sondern nur für einen Ausschnitt y_{t_1}, \dots, y_{t_m} der Länge m , wobei m prinzipiell eine frei wählbare, positive ganze, ungerade Zahl ist. Für y_{t_1}, \dots, y_{t_m} hat das Modell in Matrixschreibweise dann die Gestalt

$$\begin{bmatrix} y_{t_1} \\ \vdots \\ y_{t_m} \end{bmatrix} = X\beta + \begin{bmatrix} \varepsilon_{t_1} \\ \vdots \\ \varepsilon_{t_m} \end{bmatrix}$$

mit dem Parametervektor

$$\beta = (\alpha_0, \alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_h, \gamma_1, \dots, \gamma_h)^T$$

und der Regressorenwertematrix

$$\begin{aligned} X &= \left[\begin{array}{c|ccccc} 1 t_1 \dots t_1^k & \cos(\lambda_1 t_1) & \dots & \cos(\lambda_h t_1) & \sin(\lambda_1 t_1) & \dots & \sin(\lambda_h t_1) \\ 1 t_2 \dots t_2^k & \cos(\lambda_1 t_2) & \dots & \cos(\lambda_h t_2) & \sin(\lambda_1 t_2) & \dots & \sin(\lambda_h t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 t_m \dots t_m^k & \cos(\lambda_1 t_m) & \dots & \cos(\lambda_h t_m) & \sin(\lambda_1 t_m) & \dots & \sin(\lambda_h t_m) \end{array} \right] \\ &= (X_g^{-1} X_s). \end{aligned}$$

Natürlich sollte m stets größer oder gleich $k + 1 + 2h$ bzw., wenn $\lambda_h = \pi$ ist, $k + 2h$ gewählt werden. Existiert dann die Inverse $(X^T X)^{-1}$ von $X^T X$, so ist

$$\hat{\beta} = (X^T X)^{-1} X^T (y_{t_1}, \dots, y_{t_m})^T$$

der Kleinstes-Quadrat-Schätzer für β , wobei „T“ wie üblich für „transponiert“ steht, und es ist

$$\begin{bmatrix} \hat{y}_{t_1} \\ \vdots \\ \hat{y}_{t_m} \end{bmatrix} = X \hat{\beta} = X (X^T X)^{-1} X^T \begin{bmatrix} y_{t_1} \\ \vdots \\ y_{t_m} \end{bmatrix}$$

eine Schätzung für $(y_{t_1}, \dots, y_{t_m})^T$. Die Matrix

$$W = \begin{bmatrix} w_{11} & \dots & w_{1m} \\ \vdots & & \vdots \\ w_{m1} & \dots & w_{mm} \end{bmatrix} = X (X^T X)^{-1} X^T$$

wird auch **Gewichtsmatrix** genannt. Sie ist unabhängig von der Auswahl des Ausschnitts t_1, \dots, t_m der m berücksichtigten Indizes.

Ausgehend von dieser Gewichtsmatrix W wird nun zu jedem Wert $y_t, t = 1, \dots, n$, der Zeitreihe die systematische (glatte + saisonale) Komponente geschätzt. Und zwar ist für

$$\frac{m+1}{2} \leq t \leq n - \frac{m+1}{2} + 1$$

der Schätzwert für die systematische Komponente

$$\hat{y}_t = (w_{(m+1)/2,1}, \dots, w_{(m+1)/2,m})(y_{t-(m-1)/2}, \dots, y_t, \dots, y_{t+(m-1)/2})^T.$$

Für $t = 1, 2, \dots, \frac{m+1}{2} - 1$ berechnet man

$$\hat{y}_t = (w_{1,1}, \dots, w_{tm})(y_1, \dots, y_m)^T,$$

für $t = n - \frac{m+1}{2} + 2$ berechnet man

$$\hat{y}_t = (w_{(m+1)/2+1,1}, \dots, w_{(m+1)/2+1,m})(y_{n-m+1}, \dots, y_n)^T,$$

für $t = n - \frac{m+1}{2} + 3$ berechnet man

$$\hat{y}_t = (w_{(m+1)/2+2,1}, \dots, w_{(m+1)/2+2,m})(y_{n-m+1}, \dots, y_n)^T$$

usw. bis man für $t = n$ schließlich

$$\hat{y}_n = \hat{y}_t = (w_{m,1}, \dots, w_{mm})(y_{n-m+1}, \dots, y_n)^T$$

als Schätzung für die systematische Komponente berechnet. Die Schätzung \hat{y}_t ergibt sich für $t = 1, \dots, n$ also stets als **gewichtetes Mittel** von m Werten der Zeitreihe, wobei die Gewichte jeweils Elemente der Gewichtsmatrix W sind.

Um nun Schätzungen für die glatte Komponente bzw. die saisonale Komponente zu erhalten, verwendet man Gewichtsmatrizen W_g bzw. W_s , die sich wie folgt ergeben. Die **Gewichtsmatrix für die glatte Komponente** ist

$$W_g = (X_g | 0)(X^T X)^{-1} X^T$$

und die für die saisonale Komponente ist

$$W_s = (0 | X_s)(X^T X)^{-1} X^T.$$

Zur konkreten Schätzung \hat{G}_t, \hat{S}_t von glatter Komponente G_t bzw. saisonaler Komponente S_t zum Wert $y_t, t = 1, \dots, n$, der Zeitreihe verwendet man dann – wie oben beschrieben – Zeilen von W_g bzw. W_s . Natürlich gilt

$$\hat{y}_t = \hat{G}_t + \hat{S}_t \quad \text{für } t = 1, \dots, n.$$

Die Reihe $\hat{G}_1, \dots, \hat{G}_n$ heißt dann **Schätzung der glatten Komponente** der Zeitreihe y_1, \dots, y_n und die Reihe $y_1 - \hat{S}_1, \dots, y_n - \hat{S}_n$ heißt **saisonbereinigte Zeitreihe**.

Im Spezialfall, daß die Matrix X quadratisch und regulär ist, d. h. X^{-1} existiert, gilt stets

$$W = X(X^T X)^{-1} X^T = X X^{-1} X^{T-1} X^T = I_m,$$

d.h. W ist die m-dimensionale Einheitsmatrix und

$$\hat{y}_t = y_t \quad \text{für } t = 1, \dots, n.$$

Die Schätzung der systematischen Komponente ist also gerade mit der Zeitreihe selbst identisch. Weiterhin ist dann natürlich

$$W_g = (X_g | 0) X^{-1}, \quad W_s = (0 | X_s) X^{-1}, \quad W_g + W_s = W = I_m,$$

und wegen

$$\hat{y}_t = \hat{G}_t + \hat{S}_t = y_t \quad \text{für } t = 1, \dots, n$$

ist die saisonbereinigte Zeitreihe mit der Schätzung für die glatte Komponente identisch:

$$y_t - \hat{S}_t = \hat{G}_t \quad \text{für } t = 1, \dots, n.$$

Wie wir anhand eines Beispiels im Anschluß noch sehen werden, ergeben sich die Werte \hat{G}_t für $t = \frac{m+1}{2}, \dots, n - \frac{m+1}{2} + 1$ hier – im Falle X^{-1} existiert – gerade als gleitende Durchschnitte der geraden Ordnung $m-1$, wie wir sie in Abschnitt 1.3.1 beschrieben haben.

Beispiel: Bei wirtschaftlichen Zeitreihen ist es üblich, eine Periode von einem Jahr zugrunde zu legen. Werden Monatswerte beobachtet, so ist also $P = 12$, vgl. auch das Beispiel „Arbeitslose im Baugewerbe“ und insbesondere hier die Ausführungen in Abschnitt 3.5. Dieser Periode entspricht die Frequenz $\lambda_1 = 2\pi/12 = \pi/6$ und die entsprechenden Oberschwingungen haben die Frequenzen $\lambda_j = \pi j/6$ für $j = 2, \dots, 6$, also $\lambda_2 = \pi/3$, $\lambda_3 = \pi/2$, $\lambda_4 = 2\pi/3$, $\lambda_5 = 5\pi/6$ und $\lambda_6 = \pi$, so daß der Parameter γ_6 entfällt. Nimmt man weiter für die glatte Komponente einen linearen Trend an, also $k = 2$, und wählt $m = 13$, so ist die Regressorenwertematrix X eine reguläre 13×13 -Matrix, d.h. es gilt

$$W = I_{13} \quad \text{und} \quad \hat{y}_t = y_t \quad \text{für } t = 1, \dots, n.$$

Die Matrizen W_g und W_s ergeben sich zu

$$W_g = (X_g | 0) X^{-1}$$

$$= \frac{1}{24} \begin{bmatrix} 13 & 11 & 9 & 7 & 5 & 3 & 1 & -1 & -3 & -5 & -7 & -9 & -11 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & .2 \\ \vdots & \vdots \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ -11 & -9 & -7 & -5 & -3 & -1 & 1 & 3 & 5 & 7 & 9 & 11 & 13 \end{bmatrix}^T,$$

$$W_s = (0 | X_s) X^{-1} = I_{13} - W_g$$

$$= \frac{1}{24} \begin{bmatrix} 11 & -11 & -9 & -7 & -5 & -3 & -1 & 1 & 3 & 5 & 7 & 9 & 11 \\ -2 & 22 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 \\ -2 & -2 & 22 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 \\ \vdots & \vdots \\ -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & 22 & -2 \\ 11 & 9 & 7 & 5 & 3 & 1 & -1 & -3 & -5 & -7 & -9 & -11 & 11 \end{bmatrix}^T.$$

Wir sehen, daß die mittlere, 7. Zeile von W_g bzw. 7. Spalte von W_g^T gerade

$$\frac{1}{24}(1 \ 2 \ 2 \ \dots \ 2 \ 2 \ 1)$$

ist, so daß also

$$\hat{G}_t = y_t - \hat{S}_t \quad \text{für } t = 7, \dots, n - 6$$

gerade als gleitender Durchschnitt der Ordnung 12, vgl. Abschnitt 1.3.1, geschätzt wird.

Für die Zeitreihe y_1, \dots, y_{51} der Arbeitslosen im Baugewerbe am Monatsende von Mai 1975 bis September 1979, vgl. Tab. 9, sind diese gleitenden Durchschnitte $\hat{G}_1, \dots, \hat{G}_{45}$ bereits in Tab. 9 angegeben; sie wurden dort mit y_7^*, \dots, y_{45}^* bezeichnet. Wir berechnen hier nun noch $\hat{G}_t = y_t^*$ für $t = 1, \dots, 6$ und $t = 46, \dots, 51$ gemäß der Gewichtsmatrix W_g . Diese Größen sind in Tab. 20 zusammengestellt. Für $t = 1$ ergibt sich etwa

$$\begin{aligned} y_1^* &= \hat{G}_1 = \frac{1}{24}(13y_1 + 2y_2 + 2y_3 + \dots + 2y_{12} - 11y_{13}) \\ &= 82048,46 \approx 82048, \end{aligned}$$

und für $t = 49 = n - 2 = n - \frac{m+1}{2} + 5$ erhalten wir

$$\begin{aligned} y_{49}^* &= \hat{G}_{49} = \frac{1}{24}(-7y_{39} + 2y_{40} + 2y_{41} + \dots + 2y_{50} + 9y_{51}) \\ &= 34588,92 \approx 34589. \end{aligned}$$

Tab. 20: Glatte Komponente bzw. saisonbereinigte Zeitreihe $y_t^* = \hat{G}_t = y_t - \hat{S}_t$ für $t = 1, \dots, 6, 46, \dots, 51$ zur Zeitreihe der Arbeitslosen im Baugewerbe, vgl. Tab. 9

t	y_t^*	t	y_t^*
1	82048	46	36385
2	79493	47	35787
3	76937	48	35188
4	74382	49	34589
5	71826	50	33990
6	69271	51	33391

4. Analyse des Zusammenhangs zweier Zeitreihen

Bisher haben wir stets nur eine Zeitreihe bzw. eine Realisation y_1, \dots, y_n eines stochastischen Prozesses (Y_t) betrachtet. Oft ist man jedoch daran interessiert, den **Zusammenhang zweier Zeitreihen** x_1, \dots, x_n und y_1, \dots, y_n bzw. **zweier ihnen zugrundeliegender Prozesse** (X_t) und (Y_t) zu untersuchen. Mit dieser Problemstellung werden wir uns im Folgenden auseinandersetzen.

Modellmäßig betrachten wir hier zwei mittelwertbereinigte, schwach stationäre stochastische Prozesse (X_t) , (Y_t) , deren Autokovarianzen zum lag k wir mit $\gamma_X(k)$ bzw. $\gamma_Y(k)$ und deren Spektraldichten wir mit $f_X(\lambda)$ bzw. $f_Y(\lambda)$ bezeichnen. Weiterhin wollen wir verlangen, daß die Prozesse (X_t) und (Y_t) **schwach stationär korreliert** sind, d. h. die Kovarianz bzw. Korrelation von X_t und Y_{t+k} ist nur vom lag (von der Zeitverschiebung) k und nicht vom Zeitpunkt t abhängig.

In Abschnitt 4.1 beschäftigen wir uns zunächst mit der **Zusammenhangsanalyse im Zeitbereich**, und in Abschnitt 4.2 gehen wir auf die **Kreuzspektralanalyse**, d. h. die **Analyse im Frequenzbereich** ein.

4.1. Analyse im Zeitbereich

Sind die schwach stationären Prozesse (X_t) und (Y_t) mit $E X_t = E Y_t = 0$ schwach stationär korreliert, so lässt sich die **Kovarianz der Prozesse zum lag k** für $k = \dots, -2, -1, 0, 1, 2, \dots$ schreiben als

$$\text{Cov}(X_t, Y_{t+k}) = E(X_t Y_{t+k}) = \gamma_{XY}(k),$$

und ihre **Korrelation zum lag k** ist natürlich gerade

$$\rho_{XY}(k) = \gamma_{XY}(k) / \sqrt{\gamma_X(0) \cdot \gamma_Y(0)};$$

sie gibt die **Stärke des linearen Zusammenhangs zwischen den Prozessen** bei einer Zeitverschiebung k an. Natürlich ist

$$\gamma_{YX}(k) = \text{Cov}(Y_t, X_{t+k}) = \text{Cov}(X_t, Y_{t-k}) = \gamma_{XY}(-k)$$

und entsprechend

$$\rho_{YX}(k) = \rho_{XY}(-k).$$

Sind x_1, \dots, x_n und y_1, \dots, y_n endliche Realisationen der Prozesse (X_t) und (Y_t) , so können die lag k Kovarianzen bzw. Korrelationen für $|k| < n$ aus den Daten auch geschätzt werden. Und zwar ist die **empirische Kovarianz zum lag k**

$$c_{XY}(k) = c_{YX}(-k) = \frac{1}{n} \sum_{t=1}^{n-|k|} x_t y_{t+k}$$

für $k = -(n-1), \dots, 0, \dots, n-1$ ein Schätzer für $\gamma_{XY}(k)$ bzw. $\gamma_{YX}(-k)$, und die **empirische Korrelation zum lag k**

$$r_{XY}(k) = r_{YX}(-k) = \frac{\sum_{t=1}^{n-|k|} x_t y_{t+k}}{\sqrt{\left(\sum_{t=1}^n x_t^2\right) \cdot \left(\sum_{t=1}^n y_t^2\right)}}$$

ist für $k = -(n-1), \dots, 0, \dots, n-1$ ein Schätzer für $\rho_{XY}(k)$ bzw. $\rho_{YX}(-k)$.

Man nennt $r_{XY}(k)$, betrachtet als Funktion des lags k , auch **Kreuzkorrelogramm** und stellt es als Polygonzug graphisch dar. Am Kreuzkorrelogramm kann man dann die empirischen Korrelationen (bzw. deren Veränderungen mit dem lag) der Zeitreihen x_1, \dots, x_n und y_1, \dots, y_n zu den verschiedenen Zeitverschiebungen k ablesen.

4.2. Analyse im Frequenzbereich – Kreuzspektralanalyse

So wie man die in den Autokovarianzen bzw. Autokorrelationen zum lag k eines Prozesses bzw. einer Zeitreihe enthaltene Information vermittels der Spektraldichte bzw. des Stichprobenspektrums im Frequenzbereich darstellen kann, vgl. Abschnitt 3, lässt sich auch die in den Kovarianzen bzw. Korrelationen zum lag k zweier Prozesse bzw. Zeitreihen enthaltene Information über den Zusammenhang in den Frequenzbereich transferieren. Man kommt dabei zur Kreuzspektraldichte, mit deren Hilfe sich einige Kenngrößen für den Zusammenhang zweier Prozesse berechnen lassen. Wir beschäftigen uns in Abschnitt 4.2.1 zunächst mit diesen Grö-

ßen und gehen im Abschnitt 4.2.2 auf ihre Schätzung anhand von endlichen Zeitreihen ein.

4.2.1. Das Kreuzspektrum

Sind (X_t) , (Y_t) schwach stationäre und schwach stationär korrelierte stochastische Prozesse mit $E X_t = E Y_t = 0$ und lag k Kovarianzen $\gamma_{XY}(k) = \gamma_{YX}(-k)$ derart, daß gilt

$$\sum_{k=-\infty}^{\infty} |\gamma_{XY}(k)| < \infty,$$

so läßt sich die in den $\gamma_{XY}(k)$ enthaltene Information in den Frequenzbereich transferieren:

$$\gamma_{XY}(k) = \int_{-\pi}^{\pi} f_{XY}(\lambda) e^{i\lambda k} d\lambda$$

bzw.

$$\begin{aligned} f_{XY}(\lambda) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{XY}(k) e^{-i\lambda k} \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \left(\gamma_{XY}(k) \cos(\lambda k) - i \gamma_{XY}(k) \sin(\lambda k) \right) \end{aligned}$$

für $-\pi \leq \lambda \leq \pi$. Die Funktion $f_{XY}(\lambda)$ heißt **Kreuzspektraldichte**; wegen $\gamma_{XY}(-k) = \gamma_{YX}(k)$, $\cos 0 = 1$, $\sin 0 = 0$, $\cos x = \cos(-x)$ und $\sin x = -\sin(-x)$ für alle x läßt sie sich auch in der Form

$$\begin{aligned} f_{XY}(\lambda) &= \frac{1}{2\pi} \left(\gamma_{XY}(0) + \sum_{k=1}^{\infty} (\gamma_{XY}(k) + \gamma_{YX}(k)) \cos(\lambda k) \right. \\ &\quad \left. - i \frac{1}{2\pi} \sum_{k=1}^{\infty} (\gamma_{XY}(k) - \gamma_{YX}(k)) \sin(\lambda k) \right) \\ &= C_{XY}(\lambda) - i Q_{XY}(\lambda) \end{aligned}$$

darstellen. Der Realteil von $f_{XY}(\lambda)$

$$C_{XY}(\lambda) = \frac{1}{2\pi} \left(\gamma_{XY}(0) + \sum_{k=1}^{\infty} (\gamma_{XY}(k) + \gamma_{YX}(k)) \cos(\lambda k) \right),$$

für den gilt

$$C_{XY}(\lambda) = C_{XY}(-\lambda), \quad C_{XY}(\lambda) = C_{YX}(\lambda),$$

heißt auch **Kospektrum** der Prozesse (X_t) und (Y_t) , und der Imaginärteil der Kreuzspektraldichte

$$Q_{XY}(\lambda) = \frac{1}{2\pi} \sum_{k=1}^{\infty} (\gamma_{XY}(k) - \gamma_{YX}(k)) \sin(\lambda k)$$

mit

$$Q_{XY}(\lambda) = -Q_{XY}(-\lambda), \quad Q_{XY}(\lambda) = -Q_{YX}(\lambda)$$

wird auch **Quadraturspektrum** genannt.

Aus den Eigenschaften von Ko- und Quadraturspektrum ergibt sich nun direkt

und $f_{XY}(\lambda) = C_{XY}(\lambda) - iQ_{XY}(\lambda) = C_{XY}(-\lambda) + iQ_{XY}(-\lambda) = \overline{f_{XY}(-\lambda)}$

$$f_{XY}(\lambda) = C_{XY}(\lambda) - iQ_{XY}(\lambda) = C_{YX}(\lambda) + iQ_{YX}(\lambda) = \overline{f_{YX}(\lambda)},$$

d. h. sowohl $f_{XY}(-\lambda)$ als auch $f_{YX}(\lambda)$ sind zu $f_{XY}(\lambda)$ komplex konjugiert, so daß man wie bei der Spektraldichte in Abschnitt 3 in der Regel nur Frequenzen $0 \leq \lambda \leq \pi$ betrachtet.

Zur Interpretation der Kreuzspektraldichte müssen wir uns Folgendes klar machen. Wie in Abschnitt 3 erwähnt, lassen sich die Prozesse (X_t) und (Y_t) in unendlich viele Schwingungskomponenten unterschiedlicher Dauer zerlegen. Jeder Frequenz λ mit $0 \leq \lambda \leq \pi$ entspricht eine dieser Schwingungskomponenten. Durch die Kreuzspektraldichte wird nun der (lineare) Zusammenhang zwischen den Schwingungskomponenten von (X_t) und (Y_t) gleicher Frequenz λ dargestellt.

Aufgrund der sogenannten **Eulerschen Relation** ($z = |z| \cdot e^{i\phi}$, $\phi = \arctan(b/a)$ für $z = a + ib$) läßt sich die Kreuzspektraldichte $f_{XY}(\lambda)$ auch schreiben als

$$f_{XY}(\lambda) = |f_{XY}(\lambda)| \cdot e^{i\phi_{XY}(\lambda)},$$

wobei

$$|f_{XY}(\lambda)| = \sqrt{C_{XY}^2(\lambda) + Q_{XY}^2(\lambda)} = |f_{YX}(\lambda)|$$

das **Kreuzamplitudenspektrum** und

$$\phi_{XY}(\lambda) = \arctan \frac{-Q_{XY}(\lambda)}{C_{XY}(\lambda)} = -\phi_{YX}(\lambda)$$

das **Phasenspektrum** genannt wird.

Grob gesprochen, gibt das Kreuzamplitudenspektrum den Zusammenhang zwischen den Amplituden der Prozesse (X_t) und (Y_t) zur Frequenz λ an.

Das Phasenspektrum $\phi_{XY}(\lambda)$ gibt die Zeitverschiebung k zwischen den der Frequenz λ entsprechenden Schwingungskomponenten der Prozesse (X_t) und (Y_t) an. Allerdings ist die Zeitverschiebung gegeben als

$$k = (\phi_{XY}(\lambda) \pm 2\pi j)/\lambda, \quad j = 0, 1, 2, \dots,$$

und somit nicht eindeutig. Beschreiben (X_t) und (Y_t) etwa Monatswerte, so entspricht die Frequenz $\lambda = \pi/6$ den Schwingungskomponenten mit Dauer 12 Monate = 1 Jahr, vgl. Abschnitt 3. Wird nun zu dieser Frequenz etwa

$$\phi_{XY}(\lambda) = \phi_{XY}(\pi/6) = \pi/3$$

berechnet, so entspricht die zugehörige Zeitverschiebung

$$k = (\pi/3 \pm 2\pi j)/(\pi/6) = 2 \pm 12j, \quad j = 0, 1, 2, \dots.$$

Für $j = 0$ eilt dann der erste Prozeß (X_t) dem zweiten Prozeß (Y_t) um $k = 2$ Monate voraus, für $j = 1$ der erste dem zweiten Prozeß um 14 Monate voraus oder um 10 Monate nach usw. Die tatsächlich vorhandene Zeitverschiebung der Jahreschwingung läßt sich nur aufgrund zusätzlicher Informationen bestimmen.

Trotz der Mehrdeutigkeit ist das Phasenspektrum, das kurz auch **Phase** genannt wird, eine der drei wesentlichen Kenngrößen zur Beurteilung der Zusammenhänge zweier Prozesse im Frequenzbereich.

Die zweite betrachtete Kenngröße ist der sogenannte **Gain**

$$G_{XY}(\lambda) = \frac{|f_{XY}(\lambda)|}{f_X(\lambda)} = G_{XY}(-\lambda),$$

der das Verhältnis vom Kreuzamplitudenspektrum zum Spektrum $f_X(\lambda)$ des Prozesses (X_t) widerspiegelt. $G_{XY}(\lambda)$ ist gerade der Steigungsparameter der linearen Regression von der Schwingungskomponente der Frequenz λ des Prozesses (Y_t) auf die Schwingungskomponente gleicher Frequenz des Prozesses (X_t), vgl. Abschnitt 1 im Kap. X. Der Steigungsparameter der umgekehrten linearen Regression vom Prozeß (X_t) auf den Eingangsprozeß (Y_t) zur Frequenz λ ist natürlich gerade der Gain

$$G_{YX}(\lambda) = \frac{|f_{YX}(\lambda)|}{f_Y(\lambda)} = G_{YX}(-\lambda).$$

Die dritte Kenngröße, die **Kohärenz** $K_{XY}^2(\lambda) = K_{YX}^2(\lambda)$ entspricht dem Bestimmtheitsmaß der linearen Regression vom Prozeß (X_t) auf (Y_t) oder von (Y_t) auf (X_t):

$$\begin{aligned} K_{XY}^2(\lambda) &= K_{XY}^2(-\lambda) = K_{YX}^2(\lambda) = K_{YX}^2(-\lambda) = G_{XY}(\lambda) \cdot G_{YX}(\lambda) \\ &= \frac{|f_{XY}(\lambda)|^2}{f_X(\lambda) \cdot f_Y(\lambda)}, \quad 0 \leq K_{XY}^2(\lambda) \leq 1; \end{aligned}$$

sie mißt also den linearen Zusammenhang zwischen den Schwingungskomponenten gleicher Frequenz λ der Prozesse (X_t) und (Y_t).

Beispiel: Wir wollen hier einmal die Kenngrößen der Prozesse (X_t) und (Y_t) mit

$$\begin{aligned} X_t &= U_t + \alpha_1 U_{t-1}, \\ Y_t &= V_t + \beta_1 V_{t-1} + \beta_2 V_{t-2} + \alpha_2 U_t \end{aligned}$$

bestimmen, wobei die Prozesse (U_t) und (V_t) als voneinander unabhängige weiße Rauschen mit Varianzen σ_U^2 bzw. σ_V^2 angenommen seien.

Zunächst berechnen wir die Autokovarianzen zum lag k sowie die Spektraldichten der Prozesse (X_t) und (Y_t). Hier ergibt sich zunächst für den MA(1)-Prozeß (X_t)

$$\gamma_X(k) = \begin{cases} \sigma_U^2 + \alpha_1^2 \sigma_U^2, & \text{für } k = 0 \\ \alpha_1 \sigma_U^2, & \text{für } k = 1, \\ 0, & \text{für } k = 2, 3, \dots \end{cases} \quad \gamma_X(-k) = \gamma_X(k)$$

und somit für $0 \leq \lambda \leq \pi$

$$\begin{aligned} f_X(\lambda) &= f_X(-\lambda) = \frac{\gamma_X(0)}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \gamma_X(k) \cos(\lambda k) \\ &= \frac{1}{2\pi} (\sigma_U^2 + \alpha_1^2 \sigma_U^2) + \frac{1}{\pi} \alpha_1 \sigma_U^2 \cos \lambda \\ &= \frac{1}{\pi} ((1 + \alpha_1^2)/2 \cdot \sigma_U^2 + \alpha_1 \sigma_U^2 \cos \lambda); \end{aligned}$$

für den Prozeß (Y_t) erhalten wir aufgrund der Unabhängigkeit von (U_t) und (V_t)

$$\gamma_Y(k) = \begin{cases} \sigma_V^2 + \beta_1^2 \sigma_V^2 + \beta_2^2 \sigma_V^2 + \alpha_2^2 \sigma_U^2, & \text{für } k = 0 \\ \beta_1 \sigma_V^2 + \beta_1 \beta_2 \sigma_V^2, & \text{für } k = 1 \\ \beta_2 \sigma_V^2, & \text{für } k = 2 \\ 0, & \text{für } k = 3, 4, \dots \end{cases}, \quad \gamma_Y(-k) = \gamma_Y(k),$$

sowie für $0 \leq \lambda \leq \pi$

$$\begin{aligned} f_Y(\lambda) = f_Y(-\lambda) &= \frac{\gamma_Y(0)}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \gamma_Y(k) \cos(\lambda k) \\ &= \frac{1}{2\pi} (\sigma_v^2 + \beta_1^2 \sigma_v^2 + \beta_2^2 \sigma_v^2 + \alpha_2^2 \sigma_u^2) \\ &\quad + \frac{1}{\pi} ((\beta_1 \sigma_v^2 + \beta_1 \beta_2 \sigma_v^2) \cos \lambda + \beta_2 \sigma_v^2 \cos(2\lambda)) \\ &= \frac{1}{\pi} ((1 + \beta_1^2 + \beta_2^2) \sigma_v^2 + \alpha_2^2 \sigma_u^2)/2 + (1 + \beta_2) \beta_1 \sigma_v^2 \cos \lambda + \beta_2 \sigma_v^2 \cos(2\lambda). \end{aligned}$$

Die Kovarianzen von (X_i) und (Y_i) zum lag k sind gerade

$$\gamma_{XY}(k) = \gamma_{YX}(-k) = \begin{cases} \alpha_2 \sigma_u^2, & \text{für } k = 0 \\ \alpha_1 \alpha_2 \sigma_u^2, & \text{für } k = -1, \\ 0, & \text{sonst} \end{cases}$$

so daß wir als Kreuzspektraldichte für $-\pi \leq \lambda \leq \pi$

$$\begin{aligned} f_{XY}(\lambda) &= \frac{1}{2\pi} \left(\gamma_{XY}(0) + \sum_{k=1}^{\infty} (\gamma_{XY}(k) + \gamma_{YX}(k)) \cos(\lambda k) \right) \\ &\quad - i \frac{1}{2\pi} \sum_{k=1}^{\infty} (\gamma_{XY}(k) - \gamma_{YX}(k)) \sin(\lambda k) \\ &= \frac{1}{2\pi} (\alpha_2 \sigma_u^2 + \alpha_1 \alpha_2 \sigma_u^2 \cos \lambda) - i \frac{1}{2\pi} (-\alpha_1 \alpha_2 \sigma_u^2 \sin \lambda) \\ &= \frac{1}{2\pi} (\alpha_2 \sigma_u^2 + \alpha_1 \alpha_2 \sigma_u^2 \cos \lambda) + i \frac{1}{2\pi} \alpha_1 \alpha_2 \sigma_u^2 \sin \lambda \end{aligned}$$

erhalten. Das Kospektrum der Prozesse (X_i) und (Y_i) ist somit

$$C_{XY}(\lambda) = \frac{1}{2\pi} (\alpha_2 \sigma_u^2 + \alpha_1 \alpha_2 \sigma_u^2 \cos \lambda),$$

und ihr Quadraturspektrum hat die Gestalt

$$Q_{XY}(\lambda) = -\frac{1}{2\pi} \alpha_1 \alpha_2 \sigma_u^2 \sin \lambda.$$

Wir wollen nun die drei wesentlichen Kenngrößen Phase, Gain und Kohärenz der Prozesse (X_i) und (Y_i) bestimmen, wobei wir noch das Kreuzamplitudenspektrum

$$\begin{aligned} |f_{XY}(\lambda)| &= \sqrt{C_{XY}^2(\lambda) + Q_{XY}^2(\lambda)} \\ &= \frac{1}{2\pi} \sqrt{(\alpha_2 \sigma_u^2 + \alpha_1 \alpha_2 \sigma_u^2 \cos \lambda)^2 + (-\alpha_1 \alpha_2 \sigma_u^2 \sin \lambda)^2} \\ &= \frac{\sigma_u^2}{2\pi} \sqrt{\alpha_2^2 + \alpha_1^2 \alpha_2^2 \cos^2 \lambda + 2\alpha_1 \alpha_2^2 \cos \lambda + \alpha_1^2 \alpha_2^2 \sin^2 \lambda} \\ &= \frac{\sigma_u^2}{2\pi} \sqrt{\alpha_2^2 + \alpha_1^2 \alpha_2^2 + 2\alpha_1 \alpha_2^2 \cos \lambda} = \frac{\alpha_2 \sigma_u^2}{2\pi} \sqrt{1 + \alpha_1^2 + 2\alpha_1 \cos \lambda} \end{aligned}$$

benötigen. Wir erhalten für die Phase

$$\begin{aligned}\phi_{XY}(\lambda) &= \arctan \frac{-Q_{XY}(\lambda)}{C_{XY}(\lambda)} = \arctan \frac{\alpha_1 \alpha_2 \sin \lambda}{\alpha_2 + \alpha_1 \alpha_2 \cos \lambda} \\ &= \arctan \frac{\alpha_1 \sin \lambda}{1 + \alpha_1 \cos \lambda},\end{aligned}$$

für den Gain für $0 \leq \lambda \leq \pi$

$$\begin{aligned}G_{XY}(\lambda) &= \frac{|f_{XY}(\lambda)|}{f_X(\lambda)} \\ &= \frac{\alpha_2 \sigma_U^2}{2\pi} \sqrt{1 + \alpha_1^2 + 2\alpha_1 \cos \lambda} \left(\frac{\sigma_V^2}{\pi} ((1 + \alpha_1^2)/2 + \alpha_1 \cos \lambda) \right) \\ &= \alpha_2 \sqrt{1 + \alpha_1^2 + 2\alpha_1 \cos \lambda} / (1 + \alpha_1^2 + 2\alpha_1 \cos \lambda) \\ &= \alpha_2 / \sqrt{1 + \alpha_1^2 + 2\alpha_1 \cos \lambda}\end{aligned}$$

bzw.

$$\begin{aligned}G_{YX}(\lambda) &= \frac{|f_{YX}(\lambda)|}{f_Y(\lambda)} \\ &= \frac{\alpha_2 \sigma_U^2}{2\pi} \sqrt{1 + \alpha_1^2 + 2\alpha_1 \cos \lambda} \left[\left(\frac{1}{\pi} [((1 + \beta_1^2 + \beta_2^2)\sigma_V^2 + \alpha_2^2 \sigma_U^2)/2 \right. \right. \\ &\quad \left. \left. + (1 + \beta_2)\beta_1 \sigma_V^2 \cos \lambda + \beta_2 \sigma_V^2 \cos(2\lambda)] \right) \right] \\ &= \alpha_2 \sigma_U^2 \sqrt{1 + \alpha_1^2 + 2\alpha_1 \cos \lambda} / ((1 + \beta_1^2 + \beta_2^2)\sigma_V^2 + \alpha_2^2 \sigma_U^2 \\ &\quad + 2(1 + \beta_2)\beta_1 \sigma_V^2 \cos \lambda + 2\beta_2 \sigma_V^2 \cos(2\lambda))\end{aligned}$$

und somit für die Kohärenz für $0 \leq \lambda \leq \pi$

$$\begin{aligned}K_{XY}(\lambda) &= K_{YX}(\lambda) = G_{XY}(\lambda) G_{YX}(\lambda) \\ &= \alpha_2^2 \sigma_U^2 / [\alpha_2^2 \sigma_U^2 + (1 + \beta_1^2 + \beta_2^2 + 2(1 + \beta_2)\beta_1 \cos \lambda \\ &\quad + 2\beta_2 \cos(2\lambda)) \sigma_V^2].\end{aligned}$$

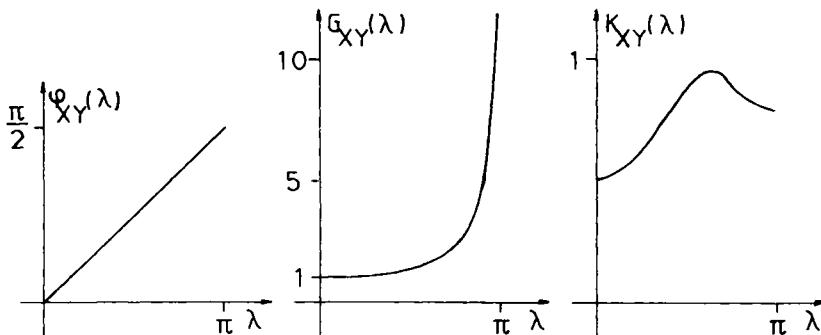


Abb. 25: Phase $\phi_{XY}(\lambda)$, Gain $G_{XY}(\lambda)$ und Kohärenz $K_{XY}(\lambda)$ für die beiden Prozesse des Beispiels

In der Abb. 25 sind Phase, Gain $G_{XY}(\lambda)$ und Kohärenz der Prozesse (X_t) und (Y_t) für $\sigma_U^2 = 2$, $\sigma_V^2 = 1$, $\alpha_1 = \beta_1 = 1$ und $\alpha_2 = \beta_2 = 2$ dargestellt. Man erkennt, daß zwischen den Prozessen eine konstante Zeitverschiebung $k = (\phi_{XY}(\lambda) \pm 2\pi j)/\lambda = 2 \pm 2\pi j/\lambda$, $j = 0, 1, 2, \dots$ vorhanden ist, da $\phi_{XY}(\lambda)/\lambda = 2$ für alle λ ist. Am Gain erkennt man, daß der Steigungsparameter der linearen Regression von der Schwingungskomponente des Prozesses (X_t) auf die des Prozesses (Y_t) mit wachsendem λ größer wird und gegen unendlich strebt, wenn λ gegen π strebt. Die Kohärenz ist das Bestimmtheitsmaß der linearen Regression von den Komponenten von (X_t) auf die von (Y_t) . Man erkennt, daß dieses ausgehend von $K_{XY}(0) = 0,5$ zunächst mit λ ansteigt, etwa für $\lambda = 5\pi/8$ (exakt für $\lambda = \arccos(-3/8)$) seinen Maximalwert von rund 0,95 annimmt, um dann wieder abzufallen auf den Wert $K_{XY}(\pi) = 0,8$.

4.2.2. Schätzung kreuzspektraler Größen

So wie das (ungeglättete) Stichprobenspektrum ein asymptotisch unverzerrter, aber inkonsistenter Schätzer für die Spektraldichte eines schwach stationären Prozesses ist, haben wir mit dem Stichprobenkreuzspektrum einen asymptotisch unverzerrten, inkonsistenten Schätzer für die Kreuzspektraldichte $f_{XY}(\lambda)$ zweier schwach stationärer, schwach stationär korrelierter und mittelwertbereinigter Prozesse (X_t) und (Y_t) . Sind x_1, \dots, x_n und y_1, \dots, y_n endliche Realisationen von (X_t) und (Y_t) , so haben wir die lag k Kovarianz $\gamma_{XY}(k)$ der Prozesse in Abschnitt 4.1 vermittels

$$c_{XY}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} x_t y_{t+k} \quad \text{für } |k| < n$$

geschätzt. Unter Verwendung dieser Schätzer ist das (ungeglättete) **Stichprobenspektrum** gerade

$$\hat{f}_{XY,n}(\lambda) = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} c_{XY}(k) e^{-ik\lambda} = \hat{C}_{XY,n}(\lambda) - i\hat{Q}_{XY,n}(\lambda),$$

wobei

$$\hat{C}_{XY,n}(\lambda) = \frac{1}{2\pi} \left(c_{XY}(0) + \sum_{k=1}^{n-1} (c_{XY}(k) + c_{YX}(k)) \cos(\lambda k) \right)$$

und

$$\hat{Q}_{XY,n}(\lambda) = \frac{1}{2\pi} \sum_{k=1}^{n-1} (c_{XY}(k) - c_{YX}(k)) \sin(\lambda k).$$

Um konsistente Schätzungen für $f_{XY}(\lambda)$ zu erhalten, wird das Stichprobenkreuzspektrum mittels der in Abschnitt 3.4.2 eingeführten lag- bzw. Spektralfenster geblättet. Bezeichnet w_m das verwandte lag-Fenster der Breite m , so ergibt sich das **geglättete Stichprobenspektrum** zu

$$\hat{f}_{XY,n}^w(\lambda) = \frac{1}{2\pi} \sum_{k=-m}^m w_m(k) c_{XY}(k) e^{-ik\lambda} = \hat{C}_{XY,n}^w(\lambda) - i\hat{Q}_{XY,n}^w(\lambda)$$

mit dem Realteil

$$\hat{C}_{XY,n}^w(\lambda) = \frac{1}{2\pi} \left(c_{XY}(0) + \sum_{k=1}^m w_m(k) (c_{XY}(k) + c_{YX}(k)) \cos(\lambda k) \right)$$

und dem Imaginärteil

$$\hat{Q}_{XY,n}^w(\lambda) = \frac{1}{2\pi} \sum_{k=1}^m w_m(k) (c_{XY}(k) - c_{YX}(k)) \sin(\lambda k).$$

Phase, Gain und Kohärenz werden dann ausgehend von diesen Größen und geglätteten Stichprobenspektren $\hat{f}_{X,n}^w(\lambda)$ bzw. $\hat{f}_{Y,n}^w(\lambda)$ der Zeitreihen x_1, \dots, x_n und y_1, \dots, y_n geschätzt:

$$\begin{aligned}\hat{\phi}_{XY,n}^w(\lambda) &= \arctan \frac{-\hat{Q}_{XY,n}^w(\lambda)}{\hat{C}_{XY,n}^w(\lambda)}, \\ \hat{G}_{XY,n}^w(\lambda) &= \sqrt{(\hat{C}_{XY,n}^w(\lambda))^2 + (\hat{Q}_{XY,n}^w(\lambda))^2} / \hat{f}_{X,n}^w(\lambda), \\ \hat{G}_{YX,n}^w(\lambda) &= \sqrt{(\hat{C}_{YX,n}^w(\lambda))^2 + (\hat{Q}_{XY,n}^w(\lambda))^2} / \hat{f}_{Y,n}^w(\lambda)\end{aligned}$$

und

$$(\hat{K}_{XY,n}^w(\lambda))^2 = ((\hat{C}_{XY,n}^w(\lambda))^2 + (\hat{Q}_{XY,n}^w(\lambda))^2) / (\hat{f}_{X,n}^w(\lambda) \cdot \hat{f}_{Y,n}^w(\lambda)).$$

Man berechnet diese Schätzungen dann konkret an den $M+1$ Stellen $\lambda_j = j\pi/M$ für $j = 0, \dots, M$, wobei M ein ganzzahliges Vielfaches der Fensterbreite m ist, und verbindet bei der graphischen Darstellung die Punkte jeweils durch Polygonzüge.

Natürlich kann man bei den Schätzungen von Phase, Gain und Kohärenz anstelle von geglätteten Stichprobenspektren und -kreuzspektren auch die entsprechenden ungeglätteten Größen verwenden.

5. Gemischte Regressions-Zeitreihen-Modelle

Im Abschnitt 2 haben wir schwach stationäre stochastische Prozesse (Y_t) betrachtet, die im wesentlichen vom Typ ARMA(p, q) mit $p \geq 0, q \geq 0, p+q \geq 1$ waren:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

wobei (ε_t) ein weißes Rauschen war. Hingegen haben wir in Kap. X Regressionsmodelle mit (nichtstochastischen) Regressoren X_1, \dots, X_k

$$Y_t = x_t^\top \beta + \varepsilon_t \quad \text{für } t = 1, \dots, n$$

mit festen Regressorenwerten $x_t = (x_{1t}, \dots, x_{kt})^\top$, Parametervektor $\beta = (\beta_1, \dots, \beta_k)^\top$ und unabhängig identisch verteilten Fehlern ε_t mit Erwartungswert 0 und Varianz σ_ε^2 betrachtet. Mit

$$Y = (Y_1, \dots, Y_n)^\top, \quad X = (x_1, \dots, x_n)^\top \quad \text{und} \quad \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$$

lässt sich der Zufallsvektor Y hier auch in der Form

$$Y = X\beta + \varepsilon$$

darstellen, wobei

$$EY = X\beta, \quad E((Y - X\beta)(Y - X\beta)^\top) = E(\varepsilon\varepsilon^\top) = \text{Cov}(\varepsilon)$$

$$= \begin{pmatrix} \text{Var}(\varepsilon_1) & \text{Cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{Cov}(\varepsilon_1, \varepsilon_n) \\ \text{Cov}(\varepsilon_2, \varepsilon_1) & \text{Var}(\varepsilon_2) & \dots & \text{Cov}(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(\varepsilon_n, \varepsilon_1) & \text{Cov}(\varepsilon_n, \varepsilon_2) & \dots & \text{Var}(\varepsilon_n) \end{pmatrix} = \sigma_\varepsilon^2 I_n$$

und I_n die $n \times n$ -Einheitsmatrix bezeichnet; a^\top sei der zum Vektor a transponierte

Vektor, d. h. der zum Zeilen-(Spalten-)Vektor a gehörige Spalten-(Zeilen-)Vektor, und X^T die zu X transponierte Matrix. Wie in Kap. X setzen wir stets voraus, daß die Matrix X vollen Spaltenrang k hat, d. h. $(X^T X)^{-1}$ existiert. Der Erwartungswert eines Vektors oder einer Matrix wird elementweise gebildet.

Hier wollen wir uns mit der Kombination der beiden Modelltypen beschäftigen. Zunächst betrachten wir im Abschnitt 5.1 **Regressionsmodelle mit korrelierten Fehlern**

$$Y_t = x_t^T \beta + \varepsilon_t$$

insbesondere derart, daß der Prozeß der ε_t ein mittelwertbereinigter, stationärer und invertierbarer ARMA(p, q)-Prozeß ist. Ausführlich werden wir auf den Fall

$$\varepsilon_t = \varrho \varepsilon_{t-1} + u_t, \quad |\varrho| < 1$$

eingehen, d. h. die ε_t bilden einen stationären AR(1)-Prozeß mit Parameter ϱ ; der Prozeß der u_t ist hier natürlich ein weißes Rauschen.

Im Abschnitt 5.2 betrachten wir dann noch kurz sogenannte **autoregressive Regressionsmodelle (lagged dependent variable models)** der Gestalt

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + x_t^T \beta + \varepsilon_t,$$

wobei der Prozeß der ε_t entweder ein weißes Rauschen oder ein ARMA(p, q)-Prozeß ist.

5.1. Regressionsmodelle mit korrelierten Fehlern

5.1.1. Allgemeine Vorgehensweise

In einem Regressionsmodell mit korrelierten Fehlern

$$Y_t = x_t^T \beta + \varepsilon_t \quad \text{für } t = 1, \dots, n \quad \text{bzw. } Y = X\beta + \varepsilon$$

hat der Vektor $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ zwar den Erwartungswert 0, jedoch ist im Gegensatz zum multiplen Regressionsmodell aus Kap. X seine Kovarianzmatrix nicht $\sigma_\varepsilon^2 I_n$, sondern vielmehr von der Struktur

$$\text{Cov}(\varepsilon) = \sigma_\varepsilon^2 \Sigma,$$

wobei Σ eine symmetrische, positiv definite Matrix ist, d. h.

$$z^T \Sigma z > 0 \quad \text{für alle Vektoren } z = (z_1, \dots, z_n)^T \neq 0 \quad \text{und} \quad \Sigma = \Sigma^T.$$

Ist Σ bekannt, so wird der Parametervektor $\beta = (\beta_1, \dots, \beta_k)^T$ bei Vorliegen einer Realisation y von Y mittels des sogenannten **Aitken-Schätzers**, eines **generalisierten Least-Squares-Schätzers (GLSE)** $\tilde{\beta}$ geschätzt:

$$\tilde{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y.$$

In der Regel ist Σ jedoch unbekannt und muß seinerseits zunächst geschätzt werden. Ist $\hat{\Sigma}$ ein konsistenter Schätzer für Σ , so ist

$$\tilde{\beta} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} y$$

ein zulässiger GLSE für β . Dieses Schätzverfahren kann auch zu einer iterativen Prozedur ausgebaut werden: Unter Verwendung des Schätzers $\tilde{\beta}$ für β wird Σ neu geschätzt, dann wird wiederum β geschätzt usw., bis sich die Schätzer für Σ und β stabilisieren.

Bilden speziell die ε_t einen ARMA(p, q)-Prozeß, also

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \dots + \phi_p \varepsilon_{t-p} + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q},$$

wobei die u_t natürlich ein weißes Rauschen darstellen, so wird zunächst der **gewöhnliche Least-Squares-Schätzer (OLSE)**

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

für β bestimmt, sodann werden die Residuen

$$\hat{\varepsilon}_t = y_t - x_t^T \hat{\beta} \quad \text{für } t = 1, \dots, n$$

berechnet, die dann als Realisation des Fehlerprozesses betrachtet werden. Vermittelst der in Abschnitt 2.5.2 angesprochenen Schätzverfahren werden dann ausgehend von $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ die Parameter $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ geschätzt und die dann resultierenden Schätzungen für die Autokorrelationen $\varrho(1), \dots, \varrho(n-1)$ in die Korrelationsmatrix

$$\hat{\varrho} = \begin{bmatrix} 1 & \varrho(1) & \dots & \varrho(n-1) \\ \varrho(1) & 1 & \dots & \varrho(n-2) \\ \vdots & & \ddots & \vdots \\ \varrho(n-1) & \varrho(n-2) & \dots & 1 \end{bmatrix}$$

von ε eingesetzt. Mit diesem Schätzer $\hat{\varrho}$ für ϱ bestimmt man dann den Schätzer $\tilde{\beta}$ für β . Natürlich können auch hier dann iterativ die Schätzungen für β und ϱ verbessert werden, indem man neue Residuen

$$\hat{\varepsilon}_t = y_t - x_t^T \tilde{\beta} \quad \text{für } t = 1, \dots, n$$

bestimmt, daraus ϱ neu schätzt zu $\hat{\varrho}$, anschließend

$$\tilde{\beta} = (X^T \hat{\varrho}^{-1} X)^{-1} X^T \hat{\varrho}^{-1} y$$

berechnet usw.

Wir haben hier vorausgesetzt, daß man die Ordnungen p und q des ARMA(p, q)-Prozesses kennt. In der Regel ist dies natürlich nicht der Fall, so daß man anhand der Residuen zunächst, wie in Abschnitt 2.5 beschrieben, den Prozeß der ε_t mit dem Box-Jenkins-Vorfahren identifizieren und überprüfen muß.

Zur **Prognose** zukünftiger Realisationen y_{n+h} bzw. zur **Modellüberprüfung** mittels ex-post und ex-ante Prognosen für y_1, \dots, y_n geht man hier so vor, daß man zunächst entsprechende Prognosen für die Residuen vermittels des in Abschnitt 2.5.4 geschilderten Verfahrens bestimmt und damit dann Prognosen für den Regressanden berechnet. Ist $\hat{\varepsilon}_j(h)$ eine h-Schritt-Prognose ($h = 1, 2, \dots$) für ε_{j+h} ($j = 1, \dots, n$) und ist zum Zeitpunkt $j+h$ der Regressorenvektor $x_{j+h} = (x_{1j+h}, \dots, x_{kj+h})^T$ vorgegeben, so ist für $j = 0, \dots, n, h = 1, 2, \dots$

$$\hat{y}_j(h) = x_{j+h}^T b + \hat{\varepsilon}_j(h),$$

wobei b einen Schätzer für den Parametervektor β bezeichnet, die Prognose für y_{j+h} vom Zeitpunkt j auf den Zeitpunkt $j+h$.

5.1.2. Regressionsmodelle mit AR(1)-Fehler-Prozeß

Nachdem wir in Abschnitt 5.1.1 die allgemeine Vorgehensweise bei der Parameterschätzung in Regressionsmodellen mit korrelierten Fehlern kurz beschrieben ha-

ben, wollen wir hier speziell das Modell

$$Y_t = x_t^T \beta + \varepsilon_t \quad \text{mit} \quad \varepsilon_t = \varrho \varepsilon_{t-1} + u_t, \quad |\varrho| < 1 \quad (\text{für } t = 1, \dots, n)$$

betrachten. Der Prozeß der u_t sei hier ein weißes Rauschen mit $\text{Var } u_t = \sigma_u^2$, so daß der Fehler-Prozeß ε_t gerade ein stationärer AR(1)-Prozeß mit Parameter ϱ ist.

Aus Abschnitt 2.1 wissen wir, daß für einen mittelwertbereinigten, stationären AR(1)-Prozeß (ε_t) gilt

$$\begin{aligned} E\varepsilon_t &= 0, \quad \text{Var } \varepsilon_t = \sigma_\varepsilon^2 = \sigma_u^2 / (1 - \varrho^2) \quad \text{und} \\ \text{Cov}(\varepsilon_t, \varepsilon_{t+k}) &= \sigma_\varepsilon^2 \varrho^k = \sigma_\varepsilon^2 \varrho(-k) = \sigma_\varepsilon^2 \varrho^k, \end{aligned}$$

d.h. mit $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$

$$\text{Cov}(\varepsilon) = \sigma_\varepsilon^2 \mathbf{\Sigma} = \sigma_u^2 \frac{1}{1 - \varrho^2} \begin{bmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{n-1} \\ \varrho & 1 & \varrho & \dots & \varrho^{n-2} \\ \varrho^2 & \varrho & 1 & \dots & \varrho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varrho^{n-1} & \varrho^{n-2} & \varrho^{n-3} & \dots & 1 \end{bmatrix}.$$

Ist nun ϱ bekannt und $y = (y_1, \dots, y_n)^T$ eine Realisation von $Y = (Y_1, \dots, Y_n)^T$, so ist

$$\hat{\beta} = (X^T \mathbf{\Sigma}^{-1} X)^{-1} X^T \mathbf{\Sigma}^{-1} y$$

der Aitken-Schätzer für β , wobei gilt

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} 1 & -\varrho & & & & \\ -\varrho & 1 + \varrho^2 & -\varrho & & & 0 \\ & -\varrho & \ddots & 1 + \varrho^2 & -\varrho & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & -\varrho & 1 + \varrho^2 & -\varrho \\ & & & & -\varrho & 1 \end{bmatrix}.$$

Ist ϱ unbekannt, so muß dieser Parameter natürlich zunächst geschätzt werden. Dabei kann man, wie in Abschnitt 5.1.1 für beliebige ARMA-Prozesse geschildert, hier mit $p = 1, q = 0$ vorgehen. Speziell für AR(1)-Fehler-Prozesse gibt es jedoch explizite Vorgehensweisen, von denen wir hier die Cochrane-Orcutt-Methode in Verbindung mit dem Durbin-Watson-Test zur Modellüberprüfung vorstellen wollen.

Natürlich setzen wir hier voraus, daß die ε_t tatsächlich einen AR(1)-Prozeß bilden, was in konkreten Fällen selbstverständlich zunächst überprüft werden muß.

A. Die Cochrane-Orcutt-Methode

Im Fall eines Regressionsmodells mit stationärem AR(1)-Fehler-Prozeß haben Cochrane/Orcutt (1949) vorgeschlagen, vom Modell

$$Y_t = x_t^T \beta + \varepsilon_t, \quad \varepsilon_t = \varrho \varepsilon_{t-1} + u_t, \quad |\varrho| < 1, \quad t = 1, \dots, n,$$

wobei $x_t = (x_{1t}, \dots, x_{kt})^T$ und die u_t ein weißes Rauschen mit Varianz σ_u^2 bilden, überzugehen zum Modell

$$Y_t^* = x_t^{*T} \beta + \varepsilon_t^* \quad \text{für } t = 2, \dots, n,$$

mit

$$Y_t^* = Y_t - \varrho Y_{t-1}, \quad x_t^* = (x_{1t}^*, \dots, x_{kt}^*)^T$$

$$\text{mit } x_{it}^* = x_{it} - \varrho x_{i,t-1} \quad \text{für } i = 1, \dots, k, \quad \varepsilon_t^* = \varepsilon_t - \varrho \varepsilon_{t-1},$$

in dem die Fehler ε_t^* dann unkorreliert sind. Schreibt man das Ausgangsmodell mit $Y = (Y_1, \dots, Y_n)^T$, $X = (x_1, \dots, x_n)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ in Matrixform

$$Y = X\beta + \varepsilon,$$

so ergibt sich das transformierte Modell gerade zu

$$GY = GX\beta + G\varepsilon$$

mit der $(n-1) \times n$ -dimensionalen **Cochrane-Orcutt-Transformationsmatrix**

$$G = \begin{bmatrix} -\varrho & 1 & & & \\ & -\varrho & 1 & & 0 \\ & & -\varrho & \ddots & 1 \\ & 0 & & \ddots & -\varrho & 1 \end{bmatrix}.$$

Im Falle eines bekannten Parameters ϱ wird der Parametervektor β hier natürlich vermittels (Invertierbarkeit vorausgesetzt)

$$\hat{\beta}_G = (X^T G^T G X)^{-1} X^T G^T G y$$

geschätzt, wenn die Realisation $y = (y_1, \dots, y_n)^T$ von Y beobachtet wird. Ist nun ϱ unbekannt, so wird zunächst der OLSE für β bestimmt

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

Dann werden die Residuen

$$\hat{\varepsilon}_t = y_t - x_t^T \hat{\beta} \quad \text{für } t = 1, \dots, n$$

berechnet und als Realisation des AR(1)-Fehler-Prozesses betrachtet. Als konsistenter Schätzer für ϱ ergibt sich nun

$$\hat{\varrho} = \frac{\sum_{t=2}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^n \hat{\varepsilon}_{t-1}^2}$$

und mit

$$\hat{G} = \begin{bmatrix} -\hat{\varrho} & 1 & & & \\ & -\hat{\varrho} & 1 & & 0 \\ & & \ddots & \ddots & \ddots & 0 \\ & 0 & & \ddots & -\hat{\varrho} & 1 \end{bmatrix}$$

erhält man in Form von

$$\hat{\beta}_{\hat{G}} = (X^T \hat{G}^T \hat{G} X)^{-1} X^T \hat{G}^T \hat{G} y$$

einen GLSE für den Parametervektor β .

Auch hier ist natürlich ein iteratives Vorgehen möglich, indem man neue Residuen

$$\hat{\varepsilon}_t = y_t - x_t^T \hat{\beta} \quad \text{für } t = 1, \dots, n$$

berechnet, damit ϱ und somit G neu schätzt, daraus einen neuen Schätzer für β bestimmt usw.

B. Der Durbin-Watson-Test

Der **Durbin-Watson-Test** dient der Überprüfung der Autokorrelation des AR(1)-Prozesses, d. h. man testet hiermit, ob der Parameter ϱ des Fehler-Prozesses überhaupt signifikant von Null verschieden ist. Sonst könnte man ja direkt davon ausgehen, daß die ε_t wie in den Regressionsmodellen aus Kap. X ein weißes Rauschen bilden und den OLSE $\hat{\beta}$ als Schätzer für β verwenden.

Voraussetzung für den Durbin-Watson-Test ist – neben der Annahme, daß die Regressorenwerte nichtstochastisch, also fest vorgegeben sind –, daß die ε_t einer Normalverteilung ($\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$) genügen und daß es sich um ein Regressionsmodell mit Absolutglied, ein sogenanntes **inhomogenes Modell** handelt. Konkret bedeutet letztere Voraussetzung, daß gilt, vgl. auch Kap. X,

$$x_{1t} = 1 \quad \text{für alle } t.$$

Die **Teststatistik** des Durbin-Watson-Tests, der natürlich auch allgemein im Falle eines AR(1)-Prozesses anstelle des Box-Pierce- oder Box-Ljung-Tests aus Abschnitt 2.5.3 angewandt werden kann, ist

$$d = \sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2 / \sum_{t=1}^n \hat{\varepsilon}_t^2,$$

wobei

$$\hat{\varepsilon}_t = y_t - x_t^T \hat{\beta} \quad \text{für } t = 1, \dots, n$$

die gewöhnlichen Least-Squares-Residuen, vgl. Abschnitt A, bezeichnen. Da gilt

$$d \approx 2(1 - \hat{\varrho})$$

ist

$$d \approx 0 \quad \text{für } \hat{\varrho} = 1, \quad d \approx 4 \quad \text{für } \hat{\varrho} = -1 \quad \text{und} \quad d \approx 2 \quad \text{für } \hat{\varrho} = 0.$$

Unter Verwendung dieser Teststatistik können sowohl die **einseitigen Tests** von

$$H'_0: \varrho = 0 \quad \text{gegen} \quad H'_1: \varrho > 0$$

und

$$H''_0: \varrho = 0 \quad \text{gegen} \quad H''_1: \varrho < 0$$

als auch der **zweiseitige Test**

$$H_0: \varrho = 0 \quad \text{gegen} \quad H_1: \varrho \neq 0$$

durchgeführt werden.

Jedoch lassen sich die **kritischen Werte** $d_{n,k;\alpha}$ dieser Tests zum Niveau α , die vom Stichprobenumfang n und von der Zahl k der Regressoren abhängen, nicht exakt bestimmen. Vielmehr können nur **obere und untere Schranken (upper and lower bounds)**

$$d_{L,n,k;\alpha} \leq d_{n,k;\alpha} \leq d_{U,n,k;\alpha}$$

für sie angegeben werden. Diese Schranken sind bei Durbin/Watson (1951) und in

extensiverer Form ($n = 6, \dots, 200, k \leq 20$) bei Savin/White (1977) für verschiedene Niveaus α vertafelt.

Für den Test von H_0' gegen H_1' zum Niveau α lautet die Entscheidungsregel dann:

Nehme H_0' an, falls $d \geq d_{U,n,k;\alpha}$,

nehme H_1' an, falls $d \leq d_{L,n,k;\alpha}$.

Für $d \in (d_{L,n,k;\alpha}, d_{U,n,k;\alpha})$ kann keine Entscheidung getroffen werden.

Im Test von H_0'' gegen H_1'' zum Niveau α ist die Entscheidungsregel:

Nehme H_0'' an, falls $d \leq 4 - d_{U,n,k;\alpha}$,

nehme H_1'' an, falls $d \geq 4 - d_{L,n,k;\alpha}$,

treffe keine Entscheidung, falls $d \in (4 - d_{U,n,k;\alpha}, 4 - d_{L,n,k;\alpha})$,

und schließlich ergibt sich für den zweiseitigen Test zum Niveau α :

Nehme H_0 an, falls $d_{U,n,k;\alpha/2} \leq d \leq 4 - d_{U,n,k;\alpha/2}$,

nehme H_1 an, falls $d \leq d_{L,n,k;\alpha/2}$ oder $d \geq 4 - d_{L,n,k;\alpha/2}$.

Für $d \in (d_{L,n,k;\alpha/2}, d_{U,n,k;\alpha/2})$ und $d \in (4 - d_{U,n,k;\alpha/2}, 4 - d_{L,n,k;\alpha/2})$ kann hier keine Entscheidung getroffen werden.

C. Prognose und Prognosegüte

Wurden z. B. mittels der Cochrane-Orcutt-Prozedur aus Abschnitt A Schätzer, sagen wir b und r , für β bzw. ϱ bestimmt, so ist es möglich, zu vorgegebenen Regressorenwerten Prognosen für den Regressanden zu berechnen. Ausgehend vom Zeitpunkt j ($j = 0, \dots, n$) können Prognosen $\hat{y}_j(h)$ für y_{j+h} (für $j = n$ handelt es sich natürlich um echte Prognosen zukünftiger Beobachtungen) für $h = 1, 2, \dots$ wie folgt bestimmt werden. Mit der Prognose

$$\tilde{\varepsilon}_j(h) = r^h \tilde{\varepsilon}_j, \quad \tilde{\varepsilon}_0 = 0, \quad \tilde{\varepsilon}_j = y_j - x_j^T b \quad \text{für } j = 1, \dots, n,$$

für y_{j+h} wird bei vorgegebenem Regressorenwertevektor x_{j+h} dann y_{j+h} prognostiziert zu

$$\hat{y}_j(h) = x_{j+h}^T b + \tilde{\varepsilon}_j(h).$$

Ex-ante Prognosen $\hat{y}_j(h)$ für $h = 1, \dots, n-j$ und ex-post Prognosen $\hat{y}_j(1)$ für $j = 0, \dots, n-1$ dienen der Überprüfung der Prognoseeigenschaften des Modells. Die Beurteilung erfolgt natürlich aufgrund der Prognosefehler

und $y_{j+h} - \hat{y}_j(h) \quad \text{für } h = 1, \dots, n-j \quad \text{bei ex-ante Prognosen}$

$y_{j+1} - \hat{y}_j(1) \quad \text{für } j = 0, \dots, n-1 \quad \text{bei ex-post Prognosen.}$

D. Ein Beispiel

Gehen wir davon aus, daß die Nachfrage Y_t nach einem Gut A in der Periode t im wesentlichen nur von seinem Preis x_{2t} in der gleichen Periode abhängt, so kommen wir bei Berücksichtigung eines Absolutgliedes ($x_{1t} = 1$ für alle t) zum Regressionsmodell

$$Y_t = \beta_1 + \beta_2 x_{2t} + \varepsilon_t = (1, x_{2t}) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon_t,$$

wobei wir annehmen, daß die ε_t einen stationären AR(1)-Fehler-Prozeß bilden:

mit $\varepsilon_t = \rho \varepsilon_{t-1} + u_t, \quad |\rho| < 1,$
 $E u_t = 0, \quad \text{Var } u_t = \sigma_u^2, \quad E(u_t \cdot u_{t'}) = 0 \quad \text{für } t \neq t'.$

In Tab. 21 sind für $t = 1, \dots, 15$ realisierte Nachfragen y_t bei Preisen x_{2t} für das Gut A angegeben.

Tab. 21: Nachfrage y_t , Preis x_{2t} , Residuen $\hat{\varepsilon}_t$, $\tilde{\varepsilon}_t$ und ex-post Prognosen $\tilde{\varepsilon}_{t-1}(1)$, $\hat{y}_{t-1}(1)$ für ε_t , y_t sowie Prognosefehler $y_t - \hat{y}_{t-1}(1)$

t	y_t	x_{2t}	$\hat{\varepsilon}_t$	$\tilde{\varepsilon}_t$	$\tilde{\varepsilon}_{t-1}(1)$	$\hat{y}_{t-1}(1)$	$y_t - \hat{y}_{t-1}(1)$
1	11,0	1,9	0,43	0,68	0,000	10,320	0,680
2	11,2	1,8	0,06	0,36	0,173	11,013	0,187
3	10,1	2,0	0,10	0,30	0,092	9,892	0,208
4	8,4	2,3	0,11	0,16	0,077	8,317	0,083
5	9,2	2,1	-0,23	-0,08	0,041	9,321	-0,121
6	9,7	2,0	-0,30	-0,10	-0,020	9,780	-0,080
7	8,5	2,3	0,21	0,26	-0,026	8,214	0,286
8	7,4	2,5	0,25	0,20	0,066	7,266	0,134
9	7,3	2,5	0,15	0,10	0,051	7,251	0,049
10	7,3	2,4	-0,42	-0,42	0,026	7,746	-0,446
11	8,0	2,3	-0,29	-0,24	-0,107	8,133	-0,133
12	8,7	2,2	-0,16	-0,06	-0,061	8,699	-0,699
13	8,8	2,2	-0,06	0,04	-0,015	8,745	-0,045
14	9,8	2,0	-0,20	0,00	0,010	9,810	-0,010
15	9,0	2,2	0,14	0,24	0,000	8,760	0,240

Aus Tab. 21 ergibt sich mit

$$y = (y_1, \dots, y_{15})^T, \quad X = \begin{pmatrix} 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{215} \end{pmatrix}, \quad \beta = (\beta_1, \beta_2)^T$$

zunächst der gewöhnliche Least-Squares-Schätzer für β zu

$$\hat{\beta} = (X^T X)^{-1} X^T y = (21,4, -5,7)^T,$$

so daß sich die gewöhnlichen Least-Squares-Residuen, wie bereits in Tab. 21 angegeben, zu

$$\hat{\varepsilon}_t = y_t - (1, x_{2t}) \begin{bmatrix} 21,4 \\ -5,7 \end{bmatrix} \quad \text{für } t = 1, \dots, 15$$

berechnen. Der Schätzer $\hat{\varrho}$ für ϱ ist damit dann

$$\hat{\varrho} = \sum_{t=2}^{15} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} / \sum_{t=1}^{15} \hat{\varepsilon}_t^2 = 0,255,$$

so daß sich mit der 14×15 -Matrix

$$\hat{G} = \begin{bmatrix} -0,255 & 1 & & & & & \\ & -0,255 & 1 & \ddots & & & 0 \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & 0 & & & \ddots & -0,255 & 1 \end{bmatrix}$$

dann

$$\hat{\beta}_{\hat{G}} = (\mathbf{X}^T \hat{G}^T \hat{G} \mathbf{X})^{-1} \mathbf{X}^T \hat{G}^T \hat{G} \mathbf{y} = \begin{bmatrix} 20,2 \\ -5,2 \end{bmatrix}$$

als GLSE für β ergibt. Wir führen hier das iterative Verfahren zur Verbesserung der Schätzer für ϱ und β nicht mehr durch und verwenden für Prognosezwecke die Schätzer $b = \hat{\beta}_{\hat{G}}$, $r = \hat{\varrho}$ für β bzw. ϱ .

Bevor wir jedoch konkrete Prognosen bestimmen, wollen wir mittels Durbin-Watson-Test zum 5% Niveau prüfen, ob der Parameter ϱ signifikant größer als Null ist, d.h. wir betrachten das Testproblem

$$H_0': \varrho = 0 \quad \text{gegen} \quad H_1': \varrho > 0.$$

Aus der Tab. 21 berechnen wir die Prüfgröße

$$d = \sum_{t=2}^{15} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2 / \sum_{t=1}^{15} \hat{\varepsilon}_t^2 = 1,24$$

und können mit $n = 15$ und $k = 2$, d.h.

$$1,08 = d_{L,15,2;0,05} < d < d_{U,15,2;0,05} = 1,36,$$

hier keine Entscheidung treffen.

Unter Verwendung der Schätzer

$$b = \begin{bmatrix} 20,2 \\ -5,2 \end{bmatrix} \quad \text{und} \quad r = 0,255$$

für β und ϱ wollen wir nun zunächst für $t = 1, \dots, 15$ die Residuen

$$\tilde{\varepsilon}_t = y_t - (1, x_{2t}) b, \quad \tilde{\varepsilon}_0 = 0$$

und damit dann die ex-post Prognosen

$$\tilde{\varepsilon}_{t-1}(1) = r \cdot \tilde{\varepsilon}_{t-1} = 0,255 \tilde{\varepsilon}_{t-1}$$

für ε_t und

$$\hat{y}_{t-1}(1) = (1, x_{2t})^T b + \tilde{\varepsilon}_{t-1}(1)$$

für y_t berechnen. Diese Größen sowie die Prognosefehler

$$y_t - \hat{y}_{t-1}(1) \quad \text{für} \quad t = 1, \dots, 15$$

sind bereits in Tab. 21 mitangegeben.

Wir wollen nun noch ausgehend von y_{15} die Werte y_{16} , y_{17} und y_{18} bei vorgegebenen Preisen $x_{216} = 2,0$, $x_{217} = 1,8$, $x_{218} = 1,9$ prognostizieren. Wegen

$$\tilde{\varepsilon}_{15} = 0,24$$

vgl. Tab. 21, ergibt sich

$$\begin{aligned} \tilde{\varepsilon}_{15}(1) &= 0,255 \cdot 0,24 = 0,061, & \tilde{\varepsilon}_{15}(2) &= 0,255^2 \cdot 0,24 = 0,016 \\ \tilde{\varepsilon}_{15}(3) &= 0,255^3 \cdot 0,24 = 0,004 \end{aligned}$$

und somit

$$\hat{y}_{15}(1) = (1, 2,0) b + 0,061 = 9,861 \quad \text{als Prognose für } y_{16},$$

$$\hat{y}_{15}(2) = (1, 1,8) b + 0,016 = 10,856 \quad \text{als Prognose für } y_{17},$$

$$\hat{y}_{15}(3) = (1, 1,9) b + 0,004 = 10,324 \quad \text{als Prognose für } y_{18}.$$

5.2. Autoregressive Regressionsmodelle

Bisher sind wir davon ausgegangen, daß die Regressorenwerte x_{it} für $i = 1, \dots, k$ und $t = 1, \dots, n$, $x_t = (x_{1t}, \dots, x_{kt})^T$, feste, von außen vorgegebene Größen sind und der Regressand sich durch diese Werte plus einen Fehler-Prozeß erklären läßt. Bei Zeitreihen läßt sich, vgl. z. B. Abschnitt 2, aber häufig Y_t nur unter zusätzlicher Zuhilfenahme des Regressanden in den p Vorperioden $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ erklären. Dadurch kommt man zu sogenannten **autoregressiven Regressionsmodellen** der Form

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + x_t^T \beta + \varepsilon_t \quad \text{für } t = 1, \dots, n,$$

wobei ϕ_1, \dots, ϕ_p die autoregressiven Parameter, $\beta = (\beta_1, \dots, \beta_k)^T$ den Parametervektor zu den Regressoren und ε_t den Fehlerterm bezeichnet.

Man nennt die Vektoren x_t , da sie von außen vorgegeben werden, auch **exogene Variablen** und die Y_t **endogene Variablen**. Die Werte Y_{t-j} , $j = 1, \dots, p$, die als erklärende Variablen für Y_t mit der **Zeitverzögerung**, dem **lag** j im Modell auftreten, heißen auch **verzögerte Variablen**. Aus dem Auftreten der verzögerten Variablen erklärt sich auch die englische Bezeichnung „**lagged dependent variable model**“ für ein autoregressives Regressionsmodell.

Der **Prozeß der Fehlerterme** ε_t kann beim autoregressiven Regressionsmodell ein weißes Rauschen oder z. B. auch ein ARMA(p, q)-Prozeß mit $E \varepsilon_t = 0$ und $\text{Var } \varepsilon_t = \sigma_\varepsilon^2$ sein.

Wir wollen hier nur eine einfache Vorgehensweise zur Schätzung der Parametervektoren $\phi = (\phi_1, \dots, \phi_p)^T$, $\beta = (\beta_1, \dots, \beta_k)^T$ sowie der Parameter des Fehlerprozesses darstellen. Der interessierte Leser sei z. B. verwiesen auf die Ausführungen in Eicker (1963), Schönfeld (1971) oder Fomby et al. (1984).

Beim einfachsten Schätzverfahren beobachtet man nicht nur Realisationen y_1, \dots, y_n von Y_1, \dots, Y_n , sondern zusätzlich p vorhergehende Realisationen $y_{-(p-1)}, \dots, y_{-1}, y_0$ und behandelt dann die (beobachteten) verzögerten Variablen des Modells so, als ob sie Regressorenwerte wären, d. h. man behandelt das Modell

$$Y = \tilde{X}\tilde{\beta} + \varepsilon$$

mit den Methoden aus Abschnitt 5.1, wobei jetzt gesetzt wird:

$$Y = (Y_1, \dots, Y_n)^T, \quad \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T, \quad \tilde{\beta} = (\phi^T, \beta^T)^T = (\phi_1, \dots, \phi_p, \beta_1, \dots, \beta_k)^T$$

$$\text{und } \tilde{X} = \begin{bmatrix} y_0 & y_{-1} & y_{-2} & \dots & y_{-(p-1)} & x_{11} & x_{21} & \dots & x_{k1} \\ y_1 & y_0 & y_{-1} & \dots & y_{-(p-2)} & x_{12} & x_{22} & \dots & x_{k2} \\ y_2 & y_1 & y_0 & \dots & y_{-(p-3)} & x_{13} & x_{23} & \dots & x_{k3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ y_{n-1} & y_{n-2} & y_{n-3} & \dots & y_{n-p} & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}.$$

6. Tests auf Stationarität einer Zeitreihe [Dickey-Fuller Unit Root Test] und eines korrelierten Zeitreihenpanels

Aus methodischen und technischen Gründen werden Tests auf Stationarität erst im Abschnitt 11 von Kapitel XV behandelt.

Kapitel XIII: Analyse von Lebensdauern und Zuverlässigkeit von Systemen

In diesem Kapitel wird zunächst die **Zuverlässigkeit von allgemeinen Systemen**, die z. B. technischer, ökonomischer oder biologisch-medizinischer (System = Mensch) Natur sein können, untersucht.

Dabei werden in Abschnitt 1 zunächst ausgehend von der Zuverlässigkeit der Komponenten eines Systems Methoden zur Bestimmung der Systemzuverlässigkeit vorgestellt. Insbesondere sind hier die **Fehlerbaumanalyse (fault-tree analysis)** und die **Analyse mehrphasiger Missionen (phased missions)** als Instrumente zur Charakterisierung und Untersuchung von komplexen Systemen und Abläufen zu nennen (**Strukturanalyse**).

In Abschnitt 2 wird dann die Zuverlässigkeit in Abhängigkeit von der **Lebensdauer eines Systems** betrachtet. Unter der Zuverlässigkeit eines Systems versteht man dann die Wahrscheinlichkeit dafür, daß das System einen bestimmten Zeitpunkt intakt überlebt, vgl. auch Kap. IV, Abschnitt 4. Zunächst werden nun verschiedene **Klassen von Lebensdauerverteilungen** vorgestellt, die im Prinzip nach steigenden und fallenden Ausfallraten ordnen, und **Tests** angegeben, die eine Zuordnung von spezifischen Einheiten zu diesen Klassen ermöglichen; diese Tests dienen insbesondere auch zur **Beurteilung von Behandlungserfolgen bei langwierigen Krankheiten** (z. B. Krebs), vgl. auch die Beispiele in diesem Abschnitt.

Im dritten Abschnitt beschäftigen wir uns dann mit dem **Schätzen der Parameter einiger spezieller Lebensdauerverteilungen**. Solche Schätzungen erlauben es etwa, mittlere Lebensdauern von Systemen zu bestimmen, vgl. auch Abschnitt 4 in Kap. IV.

Danach wird in Abschnitt 4 auf die Analyse von Lebensdauern extrem langlebiger Einheiten wie z. B. Kondensatoren eingegangen. Durch Veränderung der Umweltbedingungen wird man schneller in die Lage versetzt, Lebensdauern zu beurteilen. Die Analyse solcher **beschleunigter Lebensdauerprüfungen (accelerated life tests)**, wie z. B. **Burn-Ins**, erfolgt dann mittels des **Eyring-Modells**, des **Arrhenius-Modells** und des **veralgemeinerten Eyring-Modells**.

In Abschnitt 5 wird auf Begriffe wie **Wartung, Wartbarkeit und Erneuerung** von Systemen eingegangen, und als ein umfassendes Beispiel zur Zuverlässigkeitsanalyse von Systemen werden verschiedene **Straßenverkehrssignalanlagen** untersucht und miteinander verglichen.

Schließlich gehen wir in Abschnitt 6 noch auf die **Verfügbarkeit** von Systemen ein und stellen einige spezielle **Instandhaltungsstrategien** vor.

In Kap. IV werden für verschiedene Verteilungen Tests über die Parameter sowie Anpassungstests vorgestellt. Speziell werden im dortigen Abschnitt 6 **sequentielle Tests** angegeben, durch die auch eine Verringerung des erwarteten Prüfaufwandes (z. B. bei Lebensdauerprüfungen) erreicht wird.

1. Die Zuverlässigkeit von Komponenten und Systemen

Unter **Zuverlässigkeit** versteht man im allgemeinen – in Anlehnung an DIN 40041 – die Fähigkeit einer **Betrachtungseinheit**, die sich aus ihrem Verwendungszweck ergebenden Anforderungen, zumeist während einer gegebenen Zeitdauer, zu erfüllen.

Als Betrachtungseinheit ist hierbei ein kleinstes elektromechanisches Bauteil ebenso denkbar wie beispielsweise ein kompliziertes Sicherheitsüberwachungssystem bei Kernreaktoren.

Sofern wir das Zusammenwirken von baugleichen oder verschiedenen Einheiten (Elementen, Komponenten) betrachten, wird der Begriff **System** verwendet.

Wir wollen uns im Folgenden, wenn nicht ausdrücklich anders vermerkt, darauf beschränken, daß eine Komponente bzw. ein System nur die beiden komplementären Zustände intakt (funktionsfähig) oder defekt (funktionsunfähig) annehmen kann. Diese Bestimmung des Zuverlässigkeitssbegriffes schließt die Möglichkeit des Auftretens von Zwischenstufen der Funktionsfähigkeit jedoch nicht aus. Ein vermindertes Erfüllsein der geforderten Eigenschaften ist in diesem Fall nicht mit einem **Totalausfall** gleichzusetzen. Dies kann man sich am einfachen Beispiel einer Fernsehübertragung verdeutlichen, bei der Tonausfall vorliegt, die Bildübertragung dagegen einwandfrei ist, so daß die betrachtete Übertragungsleitung weder als intakt noch als defekt anzusehen ist; hier spricht man vom **Änderungsausfall**.

Auf die Untersuchung von Systemen mit mehr als zwei Zuständen (**multistate systems**) werden wir hier nur kurz eingehen; vgl. hierzu etwa Barlow/Wu (1978), El Newehi et al. (1978), Gaede (1977).

In diesem Abschnitt wollen wir stets davon ausgehen, daß alle Komponenten eines Systems in ihrer Funktionsweise voneinander unabhängig sind und daß sich – mit Hilfe einer **Strukturfunktion** – der Zustand des Systems eindeutig aus den Zuständen seiner Komponenten bestimmen läßt.

Als **quantitatives Maß** für den qualitativen Begriff der **Zuverlässigkeit** einer Komponente bzw. eines Systems wird im folgenden die **Wahrscheinlichkeit** verwendet, daß eine **Komponente** bzw. ein **System** intakt ist.

1.1. Die Zuverlässigkeit elementarer Systeme

Zunächst wollen wir hier zwei grundlegende Systeme – nämlich Parallel- und Seriensysteme – betrachten, aus denen man komplexere Systeme aufbauen kann.

Von einem (unabhängigen) **Parallelsystem** aus n unabhängigen Komponenten spricht man, wenn das betrachtete System dann und nur dann intakt ist, wenn mindestens eine der n Systemkomponenten intakt ist. Ein solches Parallelsystem wird häufig wie in **Abb. 1** graphisch veranschaulicht; man muß sich aber davor hüten, diese Darstellung mit einer physikalischen Schaltung zu verwechseln.

Im Gegensatz dazu spricht man von einem (unabhängigen) **Seriensystem** aus n unabhängigen Komponenten, wenn das betrachtete System dann und nur dann intakt ist, wenn alle n Komponenten intakt sind. Bei der graphischen Darstellung eines Seriensystems, vgl. **Abb. 2**, muß man sich wiederum von der Vorstellung frei machen, daß es sich dabei notwendigerweise um ein physikalisches Schaltbild handelt.

| **Beispiel:** Die parallele Zusammenschaltung eines Kondensators mit Kapazität C

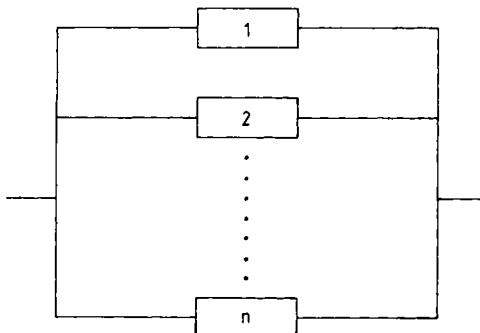


Abb. 1: Graphische Darstellung eines Parallelsystems aus n Komponenten

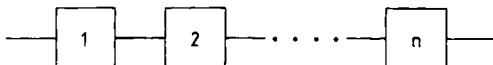


Abb. 2: Graphische Darstellung eines Seriensystems aus n Komponenten

und einer Spule mit Induktivität L zur Realisierung eines LC-Schwingkreises liefert aus Zuverlässigkeitssicht ein Seriensystem, da zur Funktionsfähigkeit des Schwingkreises beide Komponenten (Kondensator und Spule) intakt sein müssen.

Bezeichnen wir mit p_j für $j = 1, \dots, n$ die Wahrscheinlichkeit dafür, daß die j-te Komponente intakt ist – diese Wahrscheinlichkeit wollen wir im Folgenden **Komponentenzuverlässigkeit** nennen –, so ergibt sich aufgrund der Unabhängigkeitsvoraussetzungen über die Komponenten, daß

$$\begin{aligned} h_{Par} &= P(\text{mindestens eine der Komponenten intakt}) \\ &= 1 - P((1. \text{ Komponente defekt}) \text{ und } \dots \\ &\quad \text{und } (n\text{-te Komponente defekt})) \\ &= 1 - \prod_{j=1}^n (1 - p_j) \end{aligned}$$

die **Zuverlässigkeit eines Parallelsystems** aus n Komponenten angibt. Die **Zuverlässigkeit eines Seriensystems** aus n Komponenten dagegen ist gerade gegeben durch

$$\begin{aligned} h_{Ser} &= P((1. \text{ Komponente intakt}) \text{ und } \dots \text{ und } (n\text{-te Komponente intakt})) \\ &= \prod_{j=1}^n p_j. \end{aligned}$$

Bei der Bestimmung von h_{Par} und h_{Ser} haben wir natürlich berücksichtigt, daß sich Intakt- und Defektwahrscheinlichkeit der Komponenten und des Systems zu Eins addieren.

Zur Verdeutlichung dieser Ergebnisse sind in **Abb. 3** Systemzuverlässigkeiten eines Parallelsystems aus n Komponenten und in **Abb. 4** Systemzuverlässigkeiten eines Seriensystems aus n Komponenten dargestellt. Bei diesen Darstellungen für verschiedene n sind wir davon ausgegangen, daß die Zuverlässigkeit (Intaktwahrscheinlichkeit) aller n Komponenten eines Systems gleich sind, d.h.

$$p_1 = p_2 = \dots = p_n.$$

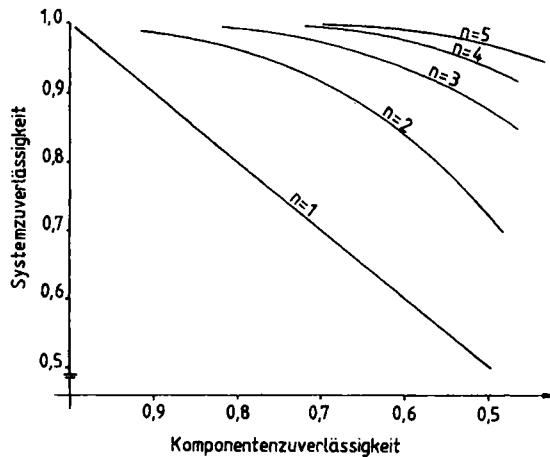


Abb. 3: Systemzuverlässigkeit eines Parallelsystems als Funktion der Komponentenzuverlässigkeit für verschiedene Komponentenzahlen n

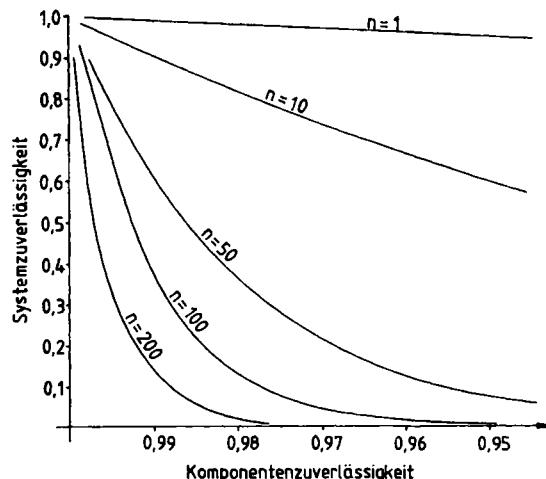


Abb. 4: Systemzuverlässigkeit eines Seriensystems als Funktion der Komponentenzuverlässigkeit für verschiedene Komponentenzahlen n

1.2. Zuverlässigkeitsschaltbilder

Zur graphischen Veranschaulichung des Zuverlässigkeitsverhaltens eines Systems werden gewöhnlich **bi-terminale, orientierte Graphen** verwendet. Bi-terminal bedeutet, daß der Graph einen Anfangspunkt I (Input) und einen Endpunkt 0 (Output) besitzt; orientiert heißt, daß die Verbindung zwischen zwei Komponenten nur in einer vorgegebenen Richtung (von links nach rechts) besteht. Bei der Darstellung geht man von folgendem Analogon aus:

Die Systemkomponenten werden als Elemente betrachtet, die entweder Strom durchlassen können, was einer intakten Komponente entspricht, oder aber Strom

nicht durchlassen können, was einer defekten Komponente gleichkommt. Die Komponenten werden durch Rechtecke oder Kreise symbolisiert, in die eine die Komponenten identifizierende Bezeichnung eingetragen wird. Für das aus diesen Elementen aufgebaute Schaltbild gilt, daß es nur dann als intakt angesehen wird, falls Stromfluß von Punkt I nach 0 möglich ist.

Im Gegensatz zu in der Technik gebräuchlichen Schaltbildern ist bei diesen sogenannten Zuverlässigkeitsschaltbildern die mehrfache Darstellung einzelner Komponenten erlaubt. Es sei jedoch die Vereinbarung getroffen, daß identisch bezeichnete Komponenten im Zuverlässigkeitsschaltbild entweder alle intakt oder alle defekt sind.

In Abb. 5 ist ein Beispiel für die Darstellung eines Systems in einem Zuverlässigkeitsschaltbild gegeben.

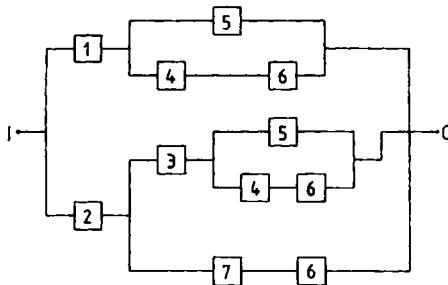


Abb. 5: Systemdarstellung durch ein Zuverlässigkeitsschaltbild

Es sei noch erwähnt, daß die Bezeichnungen I und 0 des Anfangs- und Endpunktes des Graphen häufig nicht eingetragen werden. Man hat dann davon auszugehen, daß links I und rechts 0 ist.

Eine Darstellung des Systems aus Abb. 5, bei der jede Systemkomponente nur einmal vorkommt, ist in Abb. 6 gegeben. Hierbei sind die Verzweigungsstellen als Knoten „◦“ eingezeichnet und das Symbol aus Abb. 6a bedeutet, daß der Ausgang C genau dann intakt ist, wenn mindestens einer der Eingänge A und B intakt ist.

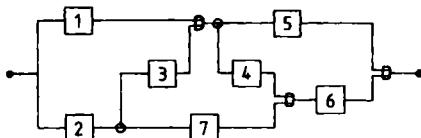


Abb. 6: Mögliche Darstellung des Systems aus Abb. 5, bei der jede Systemkomponente nur einmal vorkommt

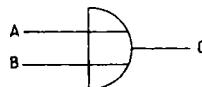


Abb. 6a: Schaltsymbol: Der Ausgang C ist intakt, wenn mindestens einer der Eingänge A und B intakt ist

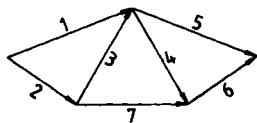


Abb. 7: Darstellung des Systems aus Abb. 5 mit Hilfe eines gerichteten azyklischen Graphen

Eine weitere Möglichkeit der Systemrepräsentation kann mit Hilfe von (zumeist auch gerichteten, azyklischen) Graphen realisiert werden, in denen die Komponenten des Systems hierbei den Kanten des Graphen entsprechen, vgl. auch Abb. 7.

1.3. Darstellung monotoner Systeme mittels minimaler Pfade und minimaler Schnitte

Häufig empfiehlt es sich, Systembetrachtungen unter besonderer Beachtung spezieller Mengen von Komponenten innerhalb des monotonen Systems durchzuführen. Von einem **monotonen System** spricht man dann, wenn durch die Verschlechterung des Zustands einer Systemkomponente keine Verbesserung des Systemzustandes erreicht wird.

Zum einen sind solche kleinste Mengen von Komponenten interessant, deren gemeinsame Funktionsfähigkeit die des ganzen Systems sichert; eine solche Menge heißt **minimale Pfadmenge**. Eine minimale Pfadmenge ist gerade dann ein intaktes Teilsystem, falls alle ihre Komponenten intakt sind, d. h. sie hat eine Serienstruktur; man bezeichnet sie daher auch als **minimale Pfadserienstruktur**. Das ganze System funktioniert immer dann, wenn mindestens eines der Teilsysteme, das eine minimale Pfadmenge darstellt, intakt ist. Daher können wir jedes monotone System mit bekannten minimalen Pfadmengen als Parallelsystem dieser Pfadmengen darstellen.

Zum anderen ist eine Komponentenmenge in einem System besonders hervorzuheben, wenn es sich um eine minimale Menge von Komponenten handelt, deren gemeinsamer Ausfall den Ausfall des Gesamtsystems zur Folge hat. Solche Mengen nennt man auch **minimale Schnittmengen**. Jede minimale Schnittmenge lässt sich also als Teilsystem mit Parallelstruktur darstellen; daher nennt man sie auch **minimale Parallelschnittstruktur**. Hat ein System m Teilsysteme, die eine minimale Parallelschnittstruktur tragen, so fällt es immer dann aus, wenn eines dieser m Teilsysteme ausfällt. Jedes monotone System lässt sich somit auch als Serienstruktur seiner minimalen Parallelschnittstrukturen darstellen.

Beispiel: Abb. 8 zeigt das Zuverlässigkeitsschaltbild einer speziellen Brückenschaltung.

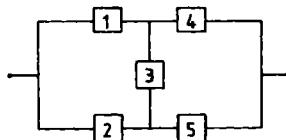


Abb. 8: Zuverlässigkeitsschaltbild einer Brückenschaltung

Dieses System nennt man ein **System mit Brückenschaltung**, da die Komponente 3 in beiden Richtungen durchlaufen werden kann. Wir wollen diese Brückenschaltung nun zunächst mittels der minimalen Pfadmengen beschreiben.

Diese sind

$P_1 = \{1, 4\}$, $P_2 = \{2, 5\}$, $P_3 = \{1, 3, 5\}$ und $P_4 = \{2, 3, 4\}$, so daß sich die Darstellung aus Abb. 9 ergibt.

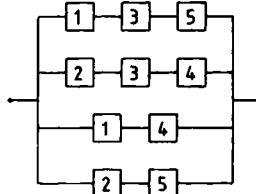


Abb. 9: Darstellung der Brückenschaltung aus Abb. 8 als Parallel-Pfadserienstruktur

Stellt man nun die Brückenschaltung aus Abb. 8 mittels ihrer minimalen Schnittmengen

$S_1 = \{1, 2\}$, $S_2 = \{4, 5\}$, $S_3 = \{1, 3, 5\}$ und $S_4 = \{2, 3, 4\}$ dar, so erhält man Abb. 10.

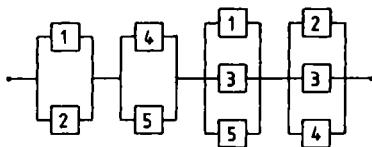


Abb. 10: Darstellung der Brückenschaltung aus Abb. 8 als Serien-Schnittparallelstruktur

1.4. Systemstrukturfunktion und Systemzuverlässigkeitsfunktion

1.4.1. Die Strukturfunktion monotoner Systeme

Die Unterscheidung zwischen den beiden Zuständen „intakt“ und „defekt“ – sowohl bei den einzelnen Komponenten als auch beim System – legt die Kennzeichnung des Zustandes der j-ten Komponente eines Systems aus n Komponenten durch die **binäre Indikatorvariable**

$$x_j = \begin{cases} 1 & \text{falls die } j\text{-te Komponente intakt ist} \\ 0 & \text{falls die } j\text{-te Komponente defekt ist} \end{cases}$$

für $j = 1, \dots, n$ nahe. Der Zustand des gesamten Systems wird dann durch die binäre Variable

$$\varphi = \varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{falls das System intakt ist,} \\ 0 & \text{falls das System defekt ist,} \end{cases}$$

gekennzeichnet. $\varphi(x_1, \dots, x_n)$ heißt auch die **Strukturfunktion** des Systems.

Beispiel: (a) Die Strukturfunktion eines Seriensystems aus n Komponenten ist

$$\varphi(x_1, \dots, x_n) = \min_{1 \leq j \leq n} \{x_j\} \quad (= \prod_{j=1}^n x_j),$$

denn es muß nach Definition eines Seriensystems gelten

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1, & \text{falls } x_1 = x_2 = \dots = x_n = 1 \\ 0, & \text{sonst} \end{cases}$$

(b) Ein Parallelsystem aus n Komponenten ist intakt, wenn mindestens eine Komponente intakt ist, d.h. aber die Strukturfunktion eines Parallelsystems ist

$$\varphi(x_1, \dots, x_n) = \max_{1 \leq j \leq n} \{x_j\} \quad (= 1 - \prod_{j=1}^n (1 - x_j)).$$

(c) Wir wollen hier nun ein **k-aus-n-System** betrachten; ein solches System aus n Komponenten ist dann intakt, wenn mindestens k der n Komponenten intakt sind. Ist etwa ein Flugzeug mit 3 Triebwerken ausgerüstet und bleibt es flugfähig solange mindestens 2 Triebwerke intakt sind, so haben wir ein 2-aus-3-System vor uns. Wie sieht nun die Strukturfunktion eines solchen **k-aus-n-Systems** aus? Bezeichnet $(x_{(1)}, \dots, x_{(n)})$ den Vektor der nach der Größe geordneten Komponentenzustände (zuerst die Komponenten mit Zustand 0, dann die mit Zustand 1), so ist

$$\varphi(x_1, \dots, x_n) = x_{(n-k+1)}.$$

Für ein 2-aus-3-System heißt das: das System ist in dem Zustand, in dem sich die $(n - k + 1)$ -schlechteste, also die zweitschlechteste Komponente befindet.

A. Bestimmung einer Strukturfunktion mittels disjunktiver Normalform

Allgemein lässt sich die Systemstrukturfunktion eines Systems mit n Komponenten mittels der **disjunktiven Normalform**

$$\varphi(x_1, \dots, x_n) = \sum_{(y_1, \dots, y_n)} \left(\prod_{j=1}^n x_j^{y_j} (1 - x_j)^{1-y_j} \right) \varphi(y_1, \dots, y_n), \quad y_j = 0 \text{ oder } 1,$$

bestimmen; hier durchläuft der Vektor (y_1, \dots, y_n) alle binären Vektoren (Vektoren aus 0 und 1) der Ordnung n , und es wird $0^0 = 1$ gesetzt.

Beispiel: Mit Hilfe der disjunktiven Normalform wollen wir die Systemstrukturfunktion des in Abb. 11 dargestellten Systems aus $n = 4$ Komponenten bestimmen.

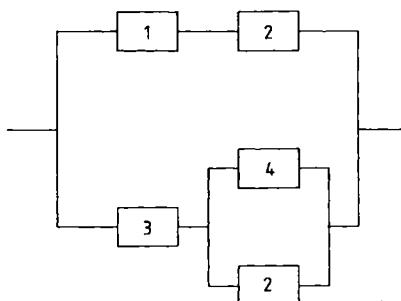


Abb. 11: Darstellung des Systems aus vier Komponenten im Beispiel

Es ergibt sich hier:

$$\begin{aligned}
 \varphi(x_1, x_2, x_3, x_4) &= \sum_{(y_1, \dots, y_4)} \left(\prod_{j=1}^4 x_j^{y_j} (1-x_j)^{1-y_j} \right) \varphi(y_1, \dots, y_4) \\
 &= x_1^0 (1-x_1)^1 x_2^0 (1-x_2)^1 x_3^0 (1-x_3)^1 x_4^0 (1-x_4)^1 \varphi(0, 0, 0, 0) \\
 &\quad + x_1^1 (1-x_1)^0 x_2^0 (1-x_2)^1 x_3^0 (1-x_3)^1 x_4^0 (1-x_4)^1 \varphi(1, 0, 0, 0) \\
 &\quad + x_1^0 (1-x_1)^1 x_2^1 (1-x_2)^0 x_3^0 (1-x_3)^1 x_4^0 (1-x_4)^1 \varphi(0, 1, 0, 0) \\
 &\quad + x_1^0 (1-x_1)^1 x_2^0 (1-x_2)^1 x_3^1 (1-x_3)^0 x_4^0 (1-x_4)^1 \varphi(0, 0, 1, 0) \\
 &\quad + x_1^0 (1-x_1)^1 x_2^0 (1-x_2)^1 x_3^0 (1-x_3)^1 x_4^1 (1-x_4)^0 \varphi(0, 0, 0, 1) \\
 &\quad + x_1^1 (1-x_1)^0 x_2^1 (1-x_2)^0 x_3^0 (1-x_3)^1 x_4^0 (1-x_4)^1 \varphi(1, 1, 0, 0) \\
 &\quad + x_1^1 (1-x_1)^0 x_2^0 (1-x_2)^1 x_3^1 (1-x_3)^0 x_4^0 (1-x_4)^1 \varphi(1, 0, 1, 0) \\
 &\quad + x_1^1 (1-x_1)^0 x_2^0 (1-x_2)^1 x_3^0 (1-x_3)^1 x_4^1 (1-x_4)^0 \varphi(1, 0, 0, 1) \\
 &\quad + x_1^0 (1-x_1)^1 x_2^1 (1-x_2)^0 x_3^1 (1-x_3)^0 x_4^0 (1-x_4)^1 \varphi(0, 1, 1, 0) \\
 &\quad + x_1^0 (1-x_1)^1 x_2^1 (1-x_2)^0 x_3^0 (1-x_3)^1 x_4^1 (1-x_4)^0 \varphi(0, 1, 0, 1) \\
 &\quad + x_1^0 (1-x_1)^1 x_2^0 (1-x_2)^1 x_3^1 (1-x_3)^0 x_4^1 (1-x_4)^0 \varphi(0, 0, 1, 1) \\
 &\quad + x_1^1 (1-x_1)^0 x_2^1 (1-x_2)^0 x_3^1 (1-x_3)^0 x_4^0 (1-x_4)^1 \varphi(1, 1, 1, 0) \\
 &\quad + x_1^1 (1-x_1)^0 x_2^1 (1-x_2)^0 x_3^0 (1-x_3)^1 x_4^1 (1-x_4)^0 \varphi(1, 1, 0, 1) \\
 &\quad + x_1^1 (1-x_1)^0 x_2^0 (1-x_2)^1 x_3^1 (1-x_3)^0 x_4^1 (1-x_4)^0 \varphi(1, 0, 1, 1) \\
 &\quad + x_1^0 (1-x_1)^1 x_2^1 (1-x_2)^0 x_3^1 (1-x_3)^0 x_4^1 (1-x_4)^0 \varphi(0, 1, 1, 1) \\
 &\quad + x_1^1 (1-x_1)^0 x_2^1 (1-x_2)^0 x_3^1 (1-x_3)^0 x_4^1 (1-x_4)^0 \varphi(1, 1, 1, 1) \\
 &= (1-x_1)(1-x_2)(1-x_3)(1-x_4) \cdot 0 + x_1(1-x_2)(1-x_3)(1-x_4) \cdot 0 \\
 &\quad + (1-x_1)x_2(1-x_3)(1-x_4) \cdot 0 + (1-x_1)(1-x_2)x_3(1-x_4) \cdot 0 \\
 &\quad + (1-x_1)(1-x_2)(1-x_3)x_4 \cdot 0 + x_1x_2(1-x_3)(1-x_4) \cdot 1 \\
 &\quad + x_1(1-x_2)x_3(1-x_4) \cdot 0 + x_1(1-x_2)(1-x_3)x_4 \cdot 0 \\
 &\quad + (1-x_1)x_2x_3(1-x_4) \cdot 1 + (1-x_1)x_2(1-x_3)x_4 \cdot 0 \\
 &\quad + (1-x_1)(1-x_2)x_3x_4 \cdot 1 + x_1x_2x_3(1-x_4) \cdot 1 \\
 &\quad + x_1x_2(1-x_3)x_4 \cdot 1 + x_1(1-x_2)x_3x_4 \cdot 1 + (1-x_1)x_2x_3x_4 \cdot 1 \\
 &\quad + x_1x_2x_3x_4 \cdot 1 \\
 &= x_1x_2(1-x_3)(1-x_4) + (1-x_1)x_2x_3(1-x_4) \\
 &\quad + (1-x_1)(1-x_2)x_3x_4 + x_1x_2x_3(1-x_4) + x_1x_2(1-x_3)x_4 \\
 &\quad + x_1(1-x_2)x_3x_4 + (1-x_1)x_2x_3x_4 + x_1x_2x_3x_4 \\
 &= x_1x_2 - x_1x_2x_3 - x_1x_2x_4 + x_1x_2x_3x_4 + x_2x_3 - x_1x_2x_3 \\
 &\quad - x_2x_3x_4 + x_1x_2x_3x_4 + x_3x_4 - x_1x_3x_4 - x_2x_3x_4 \\
 &\quad + x_1x_2x_3x_4 + x_1x_2x_3 - x_1x_2x_3x_4 + x_1x_2x_4 - x_1x_2x_3x_4 \\
 &\quad + x_1x_3x_4 - x_1x_2x_3x_4 + x_2x_3x_4 - x_1x_2x_3x_4 + x_1x_2x_3x_4 \\
 &= x_1x_2 + x_2x_3 + x_3x_4 - x_1x_2x_3 - x_2x_3x_4.
 \end{aligned}$$

Um die Berechnungen zu erleichtern, kann man sich leicht überlegen, daß gilt

$$\begin{aligned}\varphi(x_1, \dots, x_n) &= \sum_{(y_1, \dots, y_n)} \left(\prod_{j=1}^n x_j^{y_j} (1-x_j)^{1-y_j} \right) \varphi(y_1, \dots, y_n) \\ &= \sum_{\substack{(y_1, \dots, y_n): \\ \varphi(y_1, \dots, y_n)=1}} \prod_{j=1}^n x_j^{y_j} (1-x_j)^{1-y_j};\end{aligned}$$

man überlegt sich also zunächst, für welche binären Vektoren (y_1, \dots, y_n) der Ordnung n $\varphi(y_1, \dots, y_n) = 1$ gilt, und summiert dann nur noch über diese.

Beispiel: Für unser Beispiel, vgl. Abb. 11, heißt dies, daß wir zunächst eine Tabelle wie in Tab. 1 erstellen.

Tab. 1: Wert der Systemstrukturfunktion für das System aus Abb. 11 für alle Kombinationen von Komponentenzuständen

y_1	y_2	y_3	y_4	$\varphi(y_1, \dots, y_4)$
0	0	0	0	0
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
1	1	0	0	1
1	0	1	0	0
1	0	0	1	0
0	1	1	0	1
0	1	0	1	0
0	0	1	1	1
1	1	1	0	1
1	1	0	1	1
1	0	1	1	1
0	1	1	1	1
1	1	1	1	1

Dann wird

$$\begin{aligned}\varphi(x_1, x_2, x_3, x_4) &= \sum_{\substack{(y_1, \dots, y_n): \\ \varphi(y_1, \dots, y_n)=1}} \prod_{j=1}^n x_j^{y_j} (1-x_j)^{1-y_j} \\ &= x_1 x_2 (1-x_3) (1-x_4) + (1-x_1) x_2 x_3 (1-x_4) \\ &\quad + (1-x_1) (1-x_2) x_3 x_4 + x_1 x_2 x_3 (1-x_4) + x_1 x_2 (1-x_3) x_4 \\ &\quad + x_1 (1-x_2) x_3 x_4 + (1-x_1) x_2 x_3 x_4 + x_1 x_2 x_3 x_4 \\ &= x_1 x_2 + x_2 x_3 + x_3 x_4 - x_1 x_2 x_3 - x_2 x_3 x_4\end{aligned}$$

berechnet.

B. Bestimmung einer Strukturfunktion mittels minimaler Pfad- oder Schnittmengen

Die Strukturfunktion $\varphi(x_1, \dots, x_n)$ eines Systems aus n Komponenten lässt sich auch mittels minimaler Pfade oder minimaler Schnitte aus dem zugehörigen Zuverlässigkeitsschaltbild gewinnen. Bezeichnet man mit $x_j = 1$ oder 0 den Zustand (in-

takt oder defekt) der j-ten Komponente, mit P_1, \dots, P_m die minimalen Pfadmengen des Systems und mit S_1, \dots, S_l die minimalen Schnittmengen des Systems, so ergibt sich die Systemstrukturfunktion als

$$\varphi(x_1, \dots, x_n) = 1 - \prod_{i=1}^m \left(1 - \prod_{j \in P_i} x_j\right)$$

oder äquivalent dazu als

$$\varphi(x_1, \dots, x_n) = \prod_{i=1}^l \left(1 - \prod_{j \in S_i} (1 - x_j)\right).$$

Beispiel: Die Strukturfunktion des Systems aus Abb. 8 ergibt sich mittels minimaler Pfadmengen zu

$$\begin{aligned}\varphi(x_1, \dots, x_5) &= 1 - \prod_{i=1}^4 \left(1 - \prod_{j \in P_i} x_j\right) \\ &= 1 - (1 - x_1 x_4) (1 - x_2 x_5) (1 - x_1 x_3 x_5) (1 - x_2 x_3 x_4)\end{aligned}$$

bzw. mittels minimaler Schnittmengen zu

$$\begin{aligned}\varphi(x_1, \dots, x_5) &= \prod_{i=1}^4 \left(1 - \prod_{j \in S_i} (1 - x_j)\right) \\ &= (1 - (1 - x_1) (1 - x_2)) \cdot (1 - (1 - x_4) (1 - x_5)) \\ &\quad \cdot (1 - (1 - x_1) (1 - x_3) (1 - x_5)) \\ &\quad \cdot (1 - (1 - x_2) (1 - x_3) (1 - x_4)).\end{aligned}$$

Durch Umrechnung ersieht man, daß beide Ausdrücke identisch sind.

C. Gewinnung des Zuverlässigkeitsschaltbildes zu einer Strukturfunktion

Umgekehrt kann man aus einer gegebenen Strukturfunktion ein Zuverlässigkeitsschaltbild aufbauen. Dies wollen wir an einem Beispiel vorführen.

Beispiel: Ein System aus drei Komponenten besitze die Strukturfunktion

$$\varphi(x_1, x_2, x_3) = x_1 x_3 + x_2 x_3 - x_1 x_2 x_3.$$

Ist die Komponente 3 im Zustand 0, d.h. ist $x_3 = 0$, so ist auch das System im Zustand 0. Die Komponente 3 muß also funktionieren, damit das System überhaupt funktioniert, und es ist dann $x_3 = 1$ und

$$\varphi(x_1, x_2, 1) = x_1 + x_2 - x_1 x_2.$$

$\varphi(x_1, x_2, 1)$ ist immer dann gleich 1, wenn $x_1 = 1$ oder $x_2 = 1$ oder $x_1 = x_2 = 1$ ist; das heißt: Das System ist intakt, wenn neben Komponente 3 mindestens eine der Komponenten 1 und 2 intakt ist. Daraus ergibt sich das Zuverlässigkeitsschaltbild aus Abb. 12.

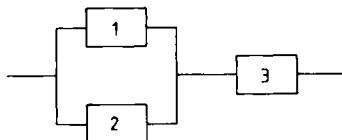


Abb. 12: Zuverlässigkeitsschaltbild zur Strukturfunktion $\varphi(x_1, x_2, x_3)$ im Beispiel

1.4.2. Die Zuverlässigungsfunktion monotoner Systeme

Sind, wie wir verlangt haben, die Komponenten eines Systems unabhängig, so ist die **Zuverlässigkeit der j-ten Komponente** für $j = 1, \dots, n$ des Systems gerade

$$p_j = E x_j = P(x_j = 1)$$

und die **Zuverlässigkeit des Systems** aus n Komponenten ist gerade

$$h = E \varphi(x_1, \dots, x_n) = P(\varphi(x_1, \dots, x_n) = 1).$$

Wegen der Komponentenunabhängigkeit kann man die Systemzuverlässigkeit auch als Funktion der Komponentenzuverlässigkeiten des Systems angeben:

$$h = h(p_1, \dots, p_n).$$

$h(p_1, \dots, p_n)$ nennt man auch die **Zuverlässigungsfunktion** zur Strukturfunktion $\varphi(x_1, \dots, x_n)$.

Unter Verwendung der Strukturfunktion $\varphi(x_1, \dots, x_n)$ eines Systems aus n Komponenten lässt sich die Zuverlässigkeit des Systems dann leicht berechnen: Sind p_1, \dots, p_n die Komponentenzuverlässigkeiten des Systems, so ist

$$\begin{aligned} h &= \sum_{\substack{(x_1, \dots, x_n): \\ x_j = 0, 1}} \prod_{j=1}^n p_j^{x_j} (1 - p_j)^{1-x_j} \cdot \varphi(x_1, \dots, x_n) \\ &= \sum_{\substack{(x_1, \dots, x_n): \\ \varphi(x_1, \dots, x_n) = 1}} \prod_{j=1}^n p_j^{x_j} (1 - p_j)^{1-x_j} \end{aligned}$$

die **disjunktive Normalform** für die Systemzuverlässigungsfunktion $h(p_1, \dots, p_n)$. Es ist hier zweckmäßig, $\varphi(x_1, \dots, x_n)$ unter Zuhilfenahme einer Darstellung des Systems durch minimale Pfade oder Schnitte zu bestimmen.

Bei Verwendung und nur bei Verwendung der disjunktiven Normalform ist also

$$h(p_1, \dots, p_n) = \varphi(p_1, \dots, p_n).$$

Beispiel:

- (a) Wir betrachten noch einmal das System aus dem Beispiel zur disjunktiven Normalform für die Systemstrukturfunktion, vgl. auch Abb. 11, und wollen nun deren Systemzuverlässigungsfunktion unter der Annahme, daß p_1, p_2, p_3, p_4 die Komponentenzuverlässigkeiten des Systems sind, berechnen. Die dazu benötigten Werte der Systemstrukturfunktion $\varphi(x_1, \dots, x_4)$ sind in Tab. 1 bereits angegeben; somit ergibt sich nun

$$\begin{aligned} h = h(p_1, p_2, p_3, p_4) &= \sum_{\substack{(x_1, \dots, x_4): \\ \varphi(x_1, \dots, x_4) = 1}} \prod_{j=1}^4 p_j^{x_j} (1 - p_j)^{1-x_j} \\ &= p_1^1 (1 - p_1)^0 p_2^1 (1 - p_2)^0 p_3^0 (1 - p_3)^1 p_4^0 (1 - p_4)^1 \\ &\quad + p_1^0 (1 - p_1)^1 p_2^1 (1 - p_2)^0 p_3^1 (1 - p_3)^0 p_4^0 (1 - p_4)^1 \\ &\quad + p_1^0 (1 - p_1)^1 p_2^0 (1 - p_2)^1 p_3^1 (1 - p_3)^0 p_4^1 (1 - p_4)^0 \\ &\quad + p_1^1 (1 - p_1)^0 p_2^1 (1 - p_2)^0 p_3^1 (1 - p_3)^0 p_4^0 (1 - p_4)^1 \\ &\quad + p_1^1 (1 - p_1)^0 p_2^1 (1 - p_2)^0 p_3^0 (1 - p_3)^1 p_4^1 (1 - p_4)^0 \end{aligned}$$

$$\begin{aligned}
& + p_1^1 (1 - p_1)^0 p_2^0 (1 - p_2)^1 p_3^1 (1 - p_3)^0 p_4^1 (1 - p_4)^0 \\
& + p_1^0 (1 - p_1)^1 p_2^1 (1 - p_2)^0 p_3^1 (1 - p_3)^0 p_4^1 (1 - p_4)^0 \\
& + p_1^1 (1 - p_1)^0 p_2^1 (1 - p_2)^0 p_3^1 (1 - p_3)^0 p_4^1 (1 - p_4)^0 \\
& = p_1 p_2 (1 - p_3) (1 - p_4) + (1 - p_1) p_2 p_3 (1 - p_4) \\
& + (1 - p_1) (1 - p_2) p_3 p_4 + p_1 p_2 p_3 (1 - p_4) + p_1 p_2 (1 - p_3) p_4 \\
& + p_1 (1 - p_2) p_3 p_4 + (1 - p_1) p_2 p_3 p_4 + p_1 p_2 p_3 p_4 \\
& = p_1 p_2 + p_2 p_3 + p_3 p_4 - p_1 p_2 p_3 - p_2 p_3 p_4.
\end{aligned}$$

Wir wollen an diesem Beispiel noch einmal verdeutlichen, daß die Systemzuverlässigkeitfunktion sich nicht aus einer beliebigen Darstellung $\varphi(p_1, \dots, p_n)$ der Strukturfunktion ergibt. Würde man etwa die Berechnungsformel bei minimalen Pfaden verwenden, so ergäbe sich mit

$$P_1 = \{1, 2\}, \quad P_2 = \{3, 4\} \quad \text{und} \quad P_3 = \{3, 2\}:$$

$$\begin{aligned}
\varphi(p_1, \dots, p_4) &= 1 - \prod_{i=1}^3 \left(1 - \prod_{j \in P_i} p_j\right) \\
&= 1 - (1 - p_1 p_2) (1 - p_3 p_4) (1 - p_2 p_3) \\
&= 1 - 1 + p_1 p_2 + p_3 p_4 + p_2 p_3 - p_1 p_2 p_3 p_4 - p_1 p_2 p_2 p_3 \\
&\quad - p_2 p_3 p_3 p_4 + p_1 p_2 p_2 p_3 p_3 p_4 \\
&= p_1 p_2 + p_3 p_4 + p_2 p_3 - p_1 p_2 p_3 p_4 - p_1 p_2^2 p_3 - p_2 p_3^2 p_4 \\
&\quad + p_1 p_2^2 p_3^2 p_4 \\
&\neq h(p_1, \dots, p_4).
\end{aligned}$$

- (b) Wir wollen die Zuverlässigkeit des Systems aus Abb. 8 mit 5 Komponenten, deren Zuverlässigkeiten p_1, \dots, p_5 sind, berechnen. Es ist hier

$$\varphi(x_1, \dots, x_5) = 1,$$

Tab. 2: Wert der Systemstrukturfunktion für das System aus Abb. 8 für alle möglichen Kombinationen von Komponentenzuständen

x_1	x_2	x_3	x_4	x_5	$\varphi(x_1, \dots, x_5)$	x_1	x_2	x_3	x_4	x_5	$\varphi(x_1, \dots, x_5)$
0	0	0	0	0	0	1	1	1	0	0	0
1	0	0	0	0	0	1	1	0	1	0	1
0	1	0	0	0	0	1	1	0	0	1	1
0	0	1	0	0	0	1	0	1	1	0	1
0	0	0	1	0	0	1	0	1	0	1	1
0	0	0	0	1	0	1	0	0	1	1	1
1	1	0	0	0	0	0	1	1	1	0	1
1	0	1	0	0	0	0	1	1	0	1	1
1	0	0	1	0	1	0	1	0	1	1	1
1	0	0	0	1	0	0	0	1	1	1	0
0	1	1	0	0	0	1	1	1	1	0	1
0	1	0	1	0	0	1	1	1	0	1	1
0	1	0	0	1	1	1	1	0	1	1	1
0	0	1	1	0	0	1	0	1	1	1	1
0	0	1	0	1	0	0	1	1	1	1	1
0	0	0	1	1	0	1	1	1	1	1	1

falls mindestens die Komponenten 1 und 4 oder 2 und 5 oder 1, 3 und 5 oder 2, 3 und 4 intakt sind. In Tab. 2 ist der Wert von $\varphi(x_1, \dots, x_5)$ für alle möglichen Zustandskombinationen von x_1, \dots, x_5 eingetragen.

Mit Hilfe dieser Tabelle können wir nun die Systemzuverlässigkeit berechnen:

$$\begin{aligned}
 h &= \sum_{\substack{(x_1, \dots, x_5): \\ \varphi(x_1, \dots, x_5) = 1}} \prod_{j=1}^5 p_j^{x_j} (1-p_j)^{1-x_j} \\
 &= p_1^1 (1-p_1)^0 p_2^0 (1-p_2)^1 p_3^0 (1-p_3)^1 p_4^1 (1-p_4)^0 p_5^0 (1-p_5)^1 + \dots \\
 &\quad + p_1^1 (1-p_1)^0 p_2^1 (1-p_2)^0 p_3^1 (1-p_3)^0 p_4^1 (1-p_4)^0 p_5^1 (1-p_5)^0 \\
 &= p_1 (1-p_2) (1-p_3) p_4 (1-p_5) + (1-p_1) p_2 (1-p_3) (1-p_4) p_5 \\
 &\quad + p_1 p_2 (1-p_3) p_4 (1-p_5) + p_1 p_2 (1-p_3) (1-p_4) p_5 \\
 &\quad + p_1 (1-p_2) p_3 p_4 (1-p_5) + p_1 (1-p_2) p_3 (1-p_4) p_5 \\
 &\quad + p_1 (1-p_2) (1-p_3) p_4 p_5 + (1-p_1) p_2 p_3 p_4 (1-p_5) \\
 &\quad + (1-p_1) p_2 p_3 (1-p_4) p_5 + (1-p_1) p_2 (1-p_3) p_4 p_5 \\
 &\quad + p_1 p_2 p_3 p_4 (1-p_5) + p_1 p_2 p_3 (1-p_4) p_5 + p_1 p_2 (1-p_3) p_4 p_5 \\
 &\quad + p_1 (1-p_2) p_3 p_4 p_5 + (1-p_1) p_2 p_3 p_4 p_5 + p_1 p_2 p_3 p_4 p_5 \\
 &= (p_1 p_4 - p_1 p_2 p_4 - p_1 p_3 p_4 - p_1 p_4 p_5 + p_1 p_2 p_3 p_4 + p_1 p_2 p_4 p_5 \\
 &\quad + p_1 p_3 p_4 p_5 - p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_2 p_5 - p_1 p_2 p_5 - p_2 p_3 p_5 - p_2 p_4 p_5 + p_1 p_2 p_3 p_5 + p_1 p_2 p_4 p_5 \\
 &\quad + p_2 p_3 p_4 p_5 - p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_1 p_2 p_4 - p_1 p_2 p_3 p_4 - p_1 p_2 p_4 p_5 + p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_1 p_2 p_5 - p_1 p_2 p_3 p_5 - p_1 p_2 p_4 p_5 + p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_1 p_3 p_4 - p_1 p_2 p_3 p_4 - p_1 p_3 p_4 p_5 + p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_1 p_3 p_5 - p_1 p_2 p_3 p_5 - p_1 p_3 p_4 p_5 + p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_1 p_4 p_5 - p_1 p_2 p_4 p_5 - p_1 p_3 p_4 p_5 + p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_2 p_3 p_4 - p_1 p_2 p_3 p_4 - p_2 p_3 p_4 p_5 + p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_2 p_3 p_5 - p_1 p_2 p_3 p_5 - p_2 p_3 p_4 p_5 + p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_2 p_4 p_5 - p_1 p_2 p_4 p_5 - p_2 p_3 p_4 p_5 + p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_1 p_2 p_3 p_4 - p_1 p_2 p_3 p_4 p_5) + (p_1 p_2 p_3 p_5 - p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_1 p_2 p_4 p_5 - p_1 p_2 p_3 p_4 p_5) + (p_1 p_3 p_4 p_5 - p_1 p_2 p_3 p_4 p_5) \\
 &\quad + (p_2 p_3 p_4 p_5 - p_1 p_2 p_3 p_4 p_5) + (p_1 p_2 p_3 p_4 p_5) \\
 &= p_1 p_4 + p_2 p_5 + p_2 p_3 p_4 + p_1 p_3 p_5 - p_1 p_3 p_4 p_5 - p_1 p_2 p_3 p_4 \\
 &\quad - p_1 p_2 p_3 p_5 - p_1 p_2 p_4 p_5 - p_2 p_3 p_4 p_5 + 2 p_1 p_2 p_3 p_4 p_5.
 \end{aligned}$$

Sind etwa $p_1 = 0,75$, $p_2 = 0,85$, $p_3 = 0,90$, $p_4 = 0,95$ und $p_5 = 0,80$, so ergibt sich die Zuverlässigkeit des Systems aus Abb. 8 zu

$$\begin{aligned}
 h &= 0,75 \cdot 0,95 + 0,85 \cdot 0,80 + 0,85 \cdot 0,90 \cdot 0,95 + 0,75 \cdot 0,90 \cdot 0,80 \\
 &\quad - 0,75 \cdot 0,90 \cdot 0,95 \cdot 0,80 - 0,75 \cdot 0,85 \cdot 0,90 \cdot 0,95
 \end{aligned}$$

$$\begin{aligned}
 & - 0,75 \cdot 0,85 \cdot 0,90 \cdot 0,80 - 0,75 \cdot 0,85 \cdot 0,95 \cdot 0,80 \\
 & - 0,85 \cdot 0,90 \cdot 0,95 \cdot 0,80 \\
 & + 2 \cdot 0,75 \cdot 0,85 \cdot 0,90 \cdot 0,95 \cdot 0,80 = 0,9483875.
 \end{aligned}$$

1.5. Stochastische Assoziiertheit von Systemkomponenten

In der Praxis tritt häufig die Situation auf, daß Komponenten stets gemeinsam belastet werden; diese Komponenten sind dann nicht mehr unabhängig. Vielmehr nennt man sie **stochastisch assoziiert**, vgl. Barlow/Proschan (1978).

Beispiel: Zur Veranschaulichung denke man beispielsweise an einen Selengleichrichter in Fernsehempfängern oder elektrolytischen Anlagen, der sich durch eine niedrige Sperrspannung auszeichnet.

Das Durchschlagen eines Gleichrichterelementes bewirkt eine Erhöhung der Sperrspannung an den verbleibenden Elementen.

Wir wollen hier nicht näher auf den Begriff der stochastischen Assoziiertheit eingehen. Es sei lediglich erwähnt, daß bei gemeinsamer Belastung aller Komponenten, die bei Ausfall einer beliebigen Komponente eine erhöhte Beanspruchung der restlichen Komponenten bewirkt, die oben vorgestellten Zuverlässigkeitsberechnungen für ein Seriensystem eine Unterschätzung, für eine Parallelsystem eine Überschätzung der Systemzuverlässigkeit darstellt. Die angegebenen Zuverlässigkeiten lassen sich folglich als untere bzw. obere Schranke der Zuverlässigkeit von Systemen mit assoziierten Komponenten verwenden.

1.6. Klassifizierung der Zuverlässigungsfunktionen monotoner Systeme mit Komponenten gleicher Zuverlässigkeit

Die Systemzuverlässigungsfunktionen monotoner Systeme mit mindestens zwei Komponenten gleicher Zuverlässigkeit p werden nach Moore/Shannon (1956) in drei Klassen eingeteilt. Bezeichnet $h_{sys}(p)$ die Systemzuverlässigkeit, so besteht die Klasse I aus den Systemen mit

$$h_{sys}(p) > p, \quad \text{für } 0 < p < 1,$$

die Klasse III aus Systemen mit s-förmiger Zuverlässigungsfunktion in Abhängigkeit von p , d.h. mit einem bestimmten $p^* \in (0, 1)$ gilt:

$$h_{sys}(p) \begin{cases} < p, & \text{für } 0 < p < p^* \\ = p^*, & \text{für } p = p^* \\ > p, & \text{für } p^* < p < 1 \end{cases}$$

und die Systeme der Klasse II haben Zuverlässigungsfunktionen

$$h_{sys}(p) < p, \quad \text{für } 0 < p < 1.$$

In Abb. 13 sind Repräsentanten der 3 Klassen graphisch dargestellt.

Insgesamt kann man sagen, daß in solchen monotonen Systemen die Funktion der Systemzuverlässigkeit (in Abhängigkeit von p) die Gerade $h_{sys}(p) = p$ höchstens einmal (und dann von unten) durchstößt.

Seriensysteme sind Vertreter der Klasse II, Parallelsysteme hingegen Vertreter der Klasse I. Das Brückensystem aus Abb. 8 schließlich ist ein Vertreter der Klasse III. Die Klassifizierung von Systemen in die drei Klassen läßt sich auch durch

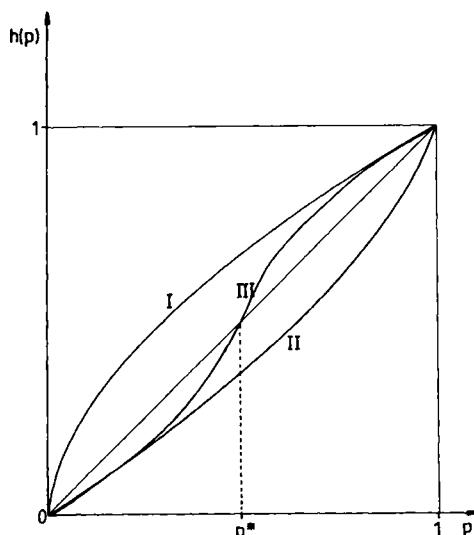


Abb. 13: Graphische Veranschaulichung von Repräsentanten der drei Klassen von Systemen

Bestimmung der Zahl ihrer minimalen Pfadmengen und minimalen Schnittmengen vornehmen, vgl. **Tab. 3**.

Tab. 3: Klassifizierung monotoner Systeme mittels minimaler Pfade und Schnitte

Klasse der Zuverlässigkeitsfunktion	Anzahl der minimalen Pfade mit einer einzigen Komponente	Anzahl der minimalen Schnitte mit einer einzigen Komponente
I	≥ 1	0
II	0	≥ 1
III (S-Kurve)	0	0

1.7. Methoden zur Erhöhung der Zuverlässigkeit – Redundanz bei Komponenten und Systemen, Systeme mit heißer und kalter Reserve

Als intuitive Maßnahme zur Erhöhung der Zuverlässigkeit eines Systems bietet es sich an, unabhängig voneinander arbeitende Komponenten mit gleicher Funktionsaufgabe gleichzeitig in ein System einzubauen und im Zuverlässigkeitssinne parallel arbeiten zu lassen. Ein solches System, das funktionsfähig sein kann, ohne daß sämtliche Komponenten funktionsfähig sind, heißt **redundant**. Ein total redundantes System ist natürlich das Parallelsystem, denn es funktioniert, wenn mindestens eine seiner Komponenten intakt ist.

Beispiel: Durch Verdoppelung des Bremskreises beim Auto oder des Zündstromkreises bei Flugzeugen wird dem Prinzip der Redundanz Rechnung getragen. Der Einsatz dieses Prinzips ist aufgrund wirtschaftlicher und technologischer Überlegungen allerdings nur begrenzt möglich.

Eine andere natürliche Begrenzung der Zuverlässigkeitserhöhung durch Parallelredundanz ist immer dann gegeben, wenn verschiedenartige Ursachen zum Ausfall einer Komponente führen können.

Die Untersuchung der Zuverlässigkeit von Relais birgt diese Schwierigkeit, da ein solches elektromechanisches Bauteil entweder Strom durchzulassen oder den Stromfluß zu unterbrechen hat, woraus sich zwei gegensätzliche Ausfallarten ergeben:

- (i) Der Stromkreis bleibt generell offen, so daß kein Stromfluß möglich ist (Ausfall durch Leerlauf).
- (ii) Der Stromkreis bleibt generell geschlossen, so daß ständiger Stromfluß vorliegt (Ausfall durch Kurzschluß).

Daß die beiden Ausfallarten einander ausschließen, ist leicht einzusehen. Wir wollen nun die Frage erörtern, inwieweit sich hier Parallelredundanz (Parallelschaltung von n Relais) lohnt.

Wir bezeichnen mit $p_{j,L}$ die Wahrscheinlichkeit dafür, daß die j-te Komponente ($j = 1, \dots, n$) durch Leerlauf ausgefallen ist und mit $p_{j,K}$ die Wahrscheinlichkeit dafür, daß sie durch Kurzschluß ausgefallen ist. Weiterhin sollen die Wahrscheinlichkeiten $p_{j,L}$ bzw. $p_{j,K}$ für alle n parallelgeschalteten Komponenten gleich sein, d.h.

$$p_{1,L} = p_{2,L} = \dots = p_{n,L} = p_L; \quad p_{1,K} = p_{2,K} = \dots = p_{n,K} = p_K.$$

Da p_L und p_K Ausfallwahrscheinlichkeiten darstellen, ist

$$0 < p_L < 1 \quad \text{und} \quad 0 < p_K < 1.$$

Ferner muß gefordert werden

$$p_L + p_K \leq 1,$$

da sonst die Intaktwahrscheinlichkeit der Komponenten nicht mehr definiert ist.

Die Zuverlässigkeit des Parallelsystems mit n Komponenten in Abhängigkeit von der Komponentenzahl n ist dann

$$\begin{aligned} R(n) &= P(\text{Parallelsystem mit } n \text{ Komponenten ist intakt}) \\ &= (1 - p_K)^n - p_L^n. \end{aligned}$$

Um die Frage nach der Effektivität der Parallelredundanz zu beantworten, muß man eine ganze Zahl $n = n^*$ bestimmen, die $R(n)$ – also die Systemzuverlässigkeit – maximiert. Dazu bildet man für $n = 1, 2, 3, \dots$ die Differenzen

$$\begin{aligned} D(n) &= R(n+1) - R(n) = (1 - p_K)^{n+1} - p_L^{n+1} - (1 - p_K)^n + p_L^n \\ &= (1 - p_L)p_L^n - p_K(1 - p_K)^n \end{aligned}$$

und sucht die kleinste Zahl $n = n^*$, für die $D(n) \leq 0$ ist. Das ist gleichbedeutend damit, daß man $n = n^*$ als kleinste ganze Zahl bestimmt, für die

$$\left(\frac{1 - p_K}{p_L} \right)^n \geq \frac{1 - p_L}{p_K}$$

ist. Bei einer Parallelschaltung von n^* Komponenten erreicht das System dann maximale Zuverlässigkeit.

Für einige Werte p_K und p_L ist die Zahl n^* in Tab. 4 ablesbar.

Tab. 4: Zuverlässigkeiten $R(n)$ für einige Kombinationen von p_K und p_L ; n^* ist jeweils gleich der Zahl n , deren zugehörige Zuverlässigkeit fett gedruckt ist.

p_L	0,01		0,05		0,1			0,15			
n	p_K	0,001	0,005	0,005	0,01	0,005	0,01	0,05	0,005	0,01	0,05
1		0,9890	0,9850	0,9450	0,9400	0,8950	0,8900	0,8500	0,8450	0,8400	0,8000
2		0,9979	0,9899	0,9875	0,9776	0,9800	0,9701	0,8925	0,9675	0,9576	0,8800
3		0,9970	0,9851	0,9849	0,9702	0,9841	0,9693	0,8564	0,9817	0,9669	0,8540
4		0,9959	0,9801	0,9801	0,9606	0,9800	0,9605	0,8144	0,9796	0,9601	0,8140
5		0,9950	0,9752	0,9752	0,9510	0,9752	0,9510	0,7738	0,9752	0,9509	0,7737

p_L	0,2			0,25			0,3				
n	p_K	0,01	0,05	0,1	0,01	0,05	0,1	0,01	0,05	0,1	0,15
1		0,7900	0,7500	0,7000	0,7400	0,7000	0,6500	0,6900	0,6500	0,6000	0,5500
2		0,9401	0,8625	0,7700	0,9176	0,8400	0,7475	0,8901	0,8125	0,7200	0,6325
3		0,9623	0,8494	0,7210	0,9547	0,8418	0,7134	0,9433	0,8304	0,7020	0,5871
4		0,9590	0,8129	0,6545	0,9567	0,8106	0,6522	0,9525	0,8064	0,6480	0,5139
5		0,9507	0,7735	0,5902	0,9500	0,7728	0,5895	0,9486	0,7714	0,5881	0,4413

In Tab. 4 haben wir uns aus gutem Grunde auf den Fall beschränkt, daß Leerlauf die überwiegende Ausfallursache ist, da andernfalls die Zuverlässigkeit einer einzelnen Komponente durch Parallelschaltung weiterer Komponenten nicht erhöht werden kann. Ist die überwiegende Ausfallursache eines Relais jedoch Kurzschluß, so empfiehlt es sich, weitere Komponenten in Serie zu schalten (**Serienredundanz**); vgl. auch Gaede (1977). Überwiegt keine der Ausfallarten eindeutig, so verwendet man zweckmäßigerweise eine gekoppelte Parallel-Serienschaltung (**Quadredundanz**) zur Erhöhung der Zuverlässigkeit, denn sie bringt sowohl eine Kompensation des Ausfalls durch Kurzschluß als auch durch Leerlauf mit sich.

Die Frage, ob es wirksamer ist, redundante Elemente auf Komponentenebene parallelzuschalten oder Parallelredundanz auf Systemebene dadurch zu betreiben, daß das gesamte System noch einmal parallel hinzugeschaltet wird, kann bei Systemen, die nur einer Ausfallart unterliegen, eindeutig damit beantwortet werden, daß Komponentenparallelredundanz zu einem größeren Zuverlässigkeitsgewinn führt.

Bei den bisher betrachteten redundanten Systemen gingen wir stets davon aus, daß alle redundanten Komponenten gleichzeitig in Betrieb genommen werden. Solche Systeme heißen **Systeme mit heißer Reserve**. Entsprechend versteht man unter einem **System mit kalter Reserve** eine Einrichtung, die einen auftretenden Fehler durch ein Kontrollgerät erkennt und über eine Umschalteinrichtung das entsprechende Reserveelement einschaltet. Nachteile der Systeme mit kalter Reserve sind neben dem steigenden Aufwand durch die hinzukommenden Signalisierungs- und Umschalteinrichtungen zum einen deren Ausfallwahrscheinlichkeiten und zum anderen die Zeitverzögerung bis zur Inbetriebnahme des Ersatzelements. Falls das System keiner noch so kurzen Funktionsunterbrechung unterworfen sein darf, kommt die Anwendung der kalten Reserve nicht in Frage. Der entscheidene Vorteil eines Systems mit kalter Reserve liegt darin, daß hiermit ein höherer Zuverlässig-

keitsgewinn erzielt werden kann, sofern die Zuverlässigkeiten der Kontroll- und Umschalteinrichtungen die des Systems wesentlich überschreiten, da die Reserveelemente bei kalter Reserve nicht seit Inbetriebnahme des Geräts die geforderte Funktion parallel ausüben müssen, sondern erst nach erfolgter Umschaltung beansprucht werden.

Eine vergleichende Zuverlässigkeitsuntersuchung, auf die wir in Abschnitt 5 dieses Kapitels noch ausführlich eingehen werden, für Systeme mit heißer und kalter Reserve am Beispiel von Straßenverkehrssignalanlagen findet sich bei Hartung/Kalin (1980). Dort wird auch die Unterscheidung zwischen kalter Reserve auf Systemebene (**kalte Macro-Reserve**) und kalter Reserve auf Komponentenebene (**kalte Micro-Reserve**) gemacht.

1.8. Systeme mit mehr als zwei Zuständen (Multi-State-Systems)

Die bisher betrachteten monotonen Systeme lassen sich auf der Basis minimaler Pfad- und Schnittbetrachtungen dahingehend verallgemeinern, daß nicht nur die Funktionszustände „intakt“ und „defekt“ möglich sind, sondern auch andere Zustände; vgl. Barlow/Wu (1978), Gaul/Hartung (1985) sowie dortige Literatur.

1.8.1. Die Bestimmung des Systemzustandes mittels minimaler Pfad- oder Schnittmengen

Die m Systemkomponenten können $n + 1$ verschiedene Zustände annehmen, die von Totalausfall (Zustand 0) über abgestufte Zwischenzustände bis zur perfekten Funktionsfähigkeit (Zustand n) reichen. Der Zustand des gesamten Systems ist dann der Zustand der „schlechtesten“ Komponente in der „besten“ minimalen Pfadmenge bzw. gleichbedeutend damit der Zustand der „besten“ Komponente in der „schlechtesten“ minimalen Schnittmenge.

Beispiel: In Abb. 14 ist ein System mit 6 möglichen Zuständen $(0, 1, \dots, 5)$ dargestellt. Die Kanten des Graphen stellen die Systemkomponenten $1, \dots, 7$ dar. In Klammern sind beispielhaft die Zustände dieser Komponenten angegeben.

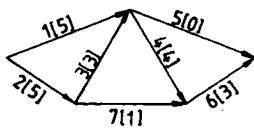


Abb. 14: System mit sechs verschiedenen möglichen Zuständen (Komponentenzustände in Klammern)

Da jede Komponente des Systems 6 verschiedene Zustände annehmen kann, gibt es auch 6 mögliche Systemzustände. Mittels minimaler Pfade und Schnitte wollen wir den Systemzustand bestimmen, wenn sich die Komponenten in den in Abb. 14 angegebenen Zuständen befinden.

Die minimale Pfadmengen des Systems sind

$$\begin{aligned} P_1 &= \{1, 5\}, \quad P_2 = \{1, 4, 6\}, \quad P_3 = \{2, 3, 5\}, \quad P_4 = \{2, 7, 6\} \quad \text{und} \\ P_5 &= \{2, 3, 4, 6\}; \end{aligned}$$

die zugehörigen Komponentenzustände sind

$$Z_1 = \{5, 0\}, \quad Z_2 = \{5, 4, 3\}, \quad Z_3 = \{5, 3, 0\}, \quad Z_4 = \{5, 1, 3\} \quad \text{und} \\ Z_5 = \{5, 3, 4, 3\}.$$

Aufgrund der Serienstruktur der minimalen Pfade liefert jeweils die schlechteste Komponente eines Pfades den Pfadzustand, d.h. hier

$$Z_{P_1} = \min \{5, 0\} = 0, \quad Z_{P_2} = 3, \quad Z_{P_3} = 0, \quad Z_{P_4} = 1, \quad Z_{P_5} = 3.$$

Die besten Pfade sind also P_2 und P_5 , deren schlechteste Komponente jeweils im Zustand 3 ist. Somit ist das Gesamtsystem im Zustand $\varphi(5, 5, 3, 4, 0, 3, 1) = 3$. Arbeitet man mit minimalen Schnittmengen, so kommt man natürlich zum gleichen Ergebnis. Die minimalen Schnitte des Systems aus Abb. 14 sind

$$S_1 = \{1, 2\}, \quad S_2 = \{5, 6\}, \quad S_3 = \{1, 3, 6\}, \quad S_4 = \{1, 3, 7\}, \\ S_5 = \{2, 4, 5\} \quad \text{und} \quad S_6 = \{4, 5, 7\};$$

die zugehörigen Mengen der Komponentenzustände sind gerade

$$Z_1 = \{5, 5\}, \quad Z_2 = \{0, 3\}, \quad Z_3 = \{5, 3, 3\}, \quad Z_4 = \{5, 3, 1\}, \\ Z_5 = \{5, 4, 0\} \quad \text{und} \quad Z_6 = \{4, 0, 1\}.$$

Der Zustand jedes minimalen Schnittes ergibt sich aufgrund der Parallelstruktur minimaler Schnittmengen als derjenige Zustand, in dem sich die „beste“ Komponente befindet:

$$Z_{S_1} = 5, \quad Z_{S_2} = 3, \quad Z_{S_3} = 5, \quad Z_{S_4} = 5, \quad Z_{S_5} = 5 \quad \text{und} \quad Z_{S_6} = 4.$$

Der Zustand φ des Gesamtsystems entspricht dann dem Zustand der „schlechtesten“ minimalen Schnittmenge, d.h. dem Zustand 3 der Schnittmenge S_2 .

1.8.2. Kritische Pfadvektoren bei Multi-State-Systemen

Multi-State-Systeme werden nicht nur durch minimale Pfad- und Schnittmengen charakterisiert; vielmehr gibt man hier auch sogenannte **obere und untere kritische Pfadvektoren** an. Besteht ein System mit $n + 1$ möglichen Zuständen aus m Komponenten $1, 2, \dots, m$, so heißt ein Komponentenzustandsvektor (x_1, \dots, x_m) ein **oberer kritischer Pfadvektor für den Zustand j** ($j = 1, \dots, n$), falls gilt:

$\varphi(x_1, \dots, x_m) = j$, und durch Verschlechterung des Zustandes einer beliebigen der m Komponenten wird der Systemzustand verschlechtert.

(x_1, \dots, x_m) heißt ein **unterer kritischer Pfadvektor für den Zustand j** ($j = 0, \dots, n - 1$), falls gilt:

$\varphi(x_1, \dots, x_m) = j$, und durch Verbesserung des Zustandes einer beliebigen der m Komponenten wird der Systemzustand verbessert.

Ist $n = 1$, so entspricht ein oberer kritischer Pfadvektor für den Zustand 1 gerade einem minimalen Pfad und ein unterer kritischer Pfadvektor für den Zustand 0 gerade einem minimalen Schnitt.

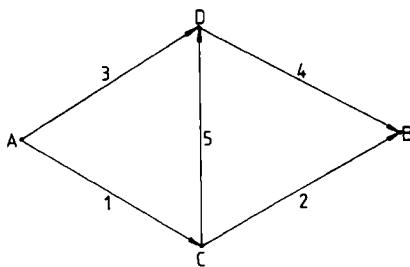
Beispiel: Es soll eine Fernsehübertragung eines Fußballspiels vom Ort A zum Ort B stattfinden, vgl. **Abb. 15**.

Die Übertragungswege können die folgenden $n + 1 = 3$ Zustände annehmen:

$0 \hat{=} \text{die Verbindung ist unterbrochen},$

$1 \hat{=} \text{nur der Ton wird übertragen},$

$2 \hat{=} \text{Bild und Ton werden übertragen}.$

**Abb. 15:** Übertragungswägen von Ort A nach Ort B

Die oberen kritischen Pfadvektoren für den Zustand j ($j = 1, 2$) sind in **Tab. 5** und die unteren kritischen Pfadvektoren für den Zustand j ($j = 0, 1$) sind in **Tab. 6** angegeben.

Tab. 5: Obere kritische Pfadvektoren für den Zustand j ($j = 1, 2$)

x_1	x_2	x_3	x_4	x_5
0	0	j	j	0
j	j	0	0	0
j	0	0	j	j

Tab. 6: Untere kritische Pfadvektoren für den Zustand j ($j = 0, 1$)

x_1	x_2	x_3	x_4	x_5
j	2	j	2	2
j	2	2	j	2
2	j	2	j	2
2	j	j	2	j

1.9. Die Fehlerbaumanalyse

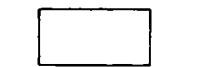
Zur Analyse sehr komplexer Systeme verwenden Zuverlässigkeitsspezialisten sogenannte Fehlerbäume. Das aus der Sicherungstechnik hervorgegangene Konzept der Fehlerbäume findet beispielsweise Anwendung als Hilfsmittel bei der Bestimmung möglicher Ursachen für einen Unfall. Bevor wir auf konkrete Beispiele eingehen, sei noch erwähnt, daß die Methode der **Fehlerbaumanalyse (fault-tree analysis)** ein bequemes und rationelles Mittel für die Berechnung der Wahrscheinlichkeit für einen Systemausfall sein kann. Diese detaillierte deduktive Analyse erfordert jedoch gewöhnlich beträchtliche Informationen über das zugrundeliegende System.

Unter einem **Fehlerbaum** wollen wir ein Modell verstehen, welches graphisch sowie logisch die verschiedenen Kombinationen möglicher Ereignisse in einem System, das in einem **Hauptereignis** gipfelt, darstellt. Der Begriff **Ereignis** beschreibt hierbei eine dynamische Zustandsänderung eines Systemelements, wobei relevante Einflußfaktoren den Systemelementen zuzuschreiben sind. Ein System kann beispielsweise menschlichen, technischen (Hardware- und Software-) sowie Umge-

bungseinflüssen unterliegen. Das Eintreten des Hauptereignisses entspricht dem **Systemausfall**.

Unterscheiden wir lediglich zwischen zwei Zuständen, so ist eine duale Umwandlung des Fehlerbaumes in einen „**Ereignisbaum**“ („**Erfolgsbaum**“), der die Funktionsfähigkeit eines Systems widerspiegelt, möglich.

Zur Konstruktion von Fehlerbäumen, die wir den vom praktischen Gesichtspunkt weniger interessanten Ereignisbäumen vorziehen wollen, benötigen wir die in Abb. 16 dargestellten Symbole.



Hauptereignis an der Baumspitze
und Zwischenereignisse



Basisereignis



UND - Glied



ODER - Glied

Abb. 16: Symbole zur Konstruktion von Fehlerbäumen

Das **Ausgangsereignis** eines **UND-Gliedes** tritt dann und nur dann ein, wenn alle **Eingangsergebnisse** des UND-Gliedes eintreten. Das Ausgangsereignis eines **ODER-Gliedes** hingegen tritt schon dann ein, wenn mindestens eines der Eingangsergebnisse des ODER-Gliedes eintritt.

Beispiel: Mit der Symbolik aus Abb. 16 haben wir ein sogenanntes 2-aus-3-System in Abb. 17 durch einen Fehlerbaum dargestellt. Ein solches System funktioniert genau dann, wenn mindestens 2 der insgesamt 3 Systemkomponenten funktionieren. Dies ist etwa bei einem mit 3 Triebwerken ausgerüsteten Flugzeug der Fall, das auch noch fliegt, wenn nur 2 Triebwerke intakt sind und erst ausfällt, wenn nur 0 oder 1 Triebwerk funktionieren.

Das nachfolgende Beispiel aus dem Bereich der Medizin zeigt zwei wichtige Vorteile der Fehlerbaumanalyse: Zum einen zwingt die Konstruktion eines solchen Baumes den Analytiker, ein profundes Verständnis des Systems zu erlangen, zum anderen wird durch die Anschaulichkeit des Fehlerbaumes die Vermittlung der Systeminformation erleichtert.

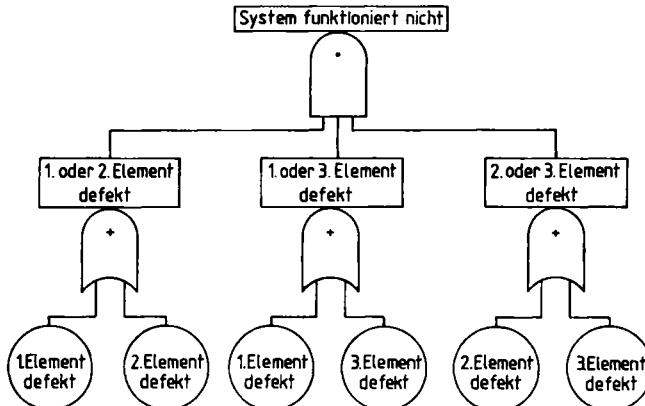


Abb. 17: Fehlerbaumdarstellung eines 2-aus-3-Systems

Beispiel: (vgl. Simonaitis et al. (1972)).

Die Gefahr eines elektrischen Schocks für einen an ein Intra-Aorta-Ballon-Kreislaufentlastungssystem angeschlossenen Patienten soll mit Hilfe der Fehlerbaumanalyse untersucht werden. Dieses Kreislaufentlastungsgerät hat die Aufgabe, bei Störungen der Blutzirkulation aufgrund mangelhafter Herztätigkeit eine zeitweilige Kreislaufhilfe zu liefern. Im Falle der Anwendung wird ein Ballonkatheder in der Thoraxaorta synchron zur Herztätigkeit aufgeblasen bzw. es wird aus ihm die Luft ausgelassen, was eine verminderte Beanspruchung der Herztätigkeit, sich erhöhenden koronaren Blutfluß sowie eine Unterstützung des allgemeinen Kreislaufes zur Folge hat. Durch das Überwachen des Elektrokardiogramms (EKG) des Patienten auf einer Kontrolltafel wird die Synchronisation des Ballons mit dem Herzen erzielt. Die Kontrolltafel dient der Anzeige sämtlicher zu überwachender Funktionen sowie möglicher Störungen des Gerätes. Für den Fall einer Anomalie im EKG oder im Funktionieren des Ballons ist eine zusätzliche Vorsorge für das automatische Luftablassen an dem Ballon getroffen.

Größte Gefahr droht dem Patienten durch einen elektrischen Schock. Dieser kann als sogenannter Makroschock (> 20 Milliampère) durch Ströme, die in den Körper durch Kontakt mit der Haut gelangen, einen Herzstillstand bewirken oder als sogenannter Mikroschock (ca. $10 - 20$ Mikroampère) durch Ströme über einen leitenden Weg direkt in die Nähe des Herzens einen Herzausfall zur Folge haben.

Weitere Gefahren drohen dem Patienten durch Ballonaufblähung während der Kontraktionsperiode des Herzens sowie durch einen überhöhten Druck im Ballon. Letztere Gefahren werden in dem abgebildeten Fehlerbaum, vgl. Abb. 18, nicht weiter analysiert, stellen jedoch keinesfalls Basisereignisse dar. Die Verwendung des Dreiecks im Fehlerbaum symbolisiert, daß der entsprechende Teil an anderer Stelle weiter auszuführen ist.

Hält man sich vor Augen, daß für eine sorgfältige Fehlerbaumanalyse gewöhnlicher chemischer Prozesse 2–3 Mannjahre gerechnet werden, so wird klar, daß man verstärkt versucht, die Fehlerbaumkonstruktion zu automatisieren. Die automatisierte Konstruktion ersetzt aber nicht den Analytiker, sondern kann nur zur Überprüfung und Belebung seiner Arbeit verwendet werden. Dies ist insbesondere dann zu beachten, wenn ein Systemausfall Menschenleben gefährdet.

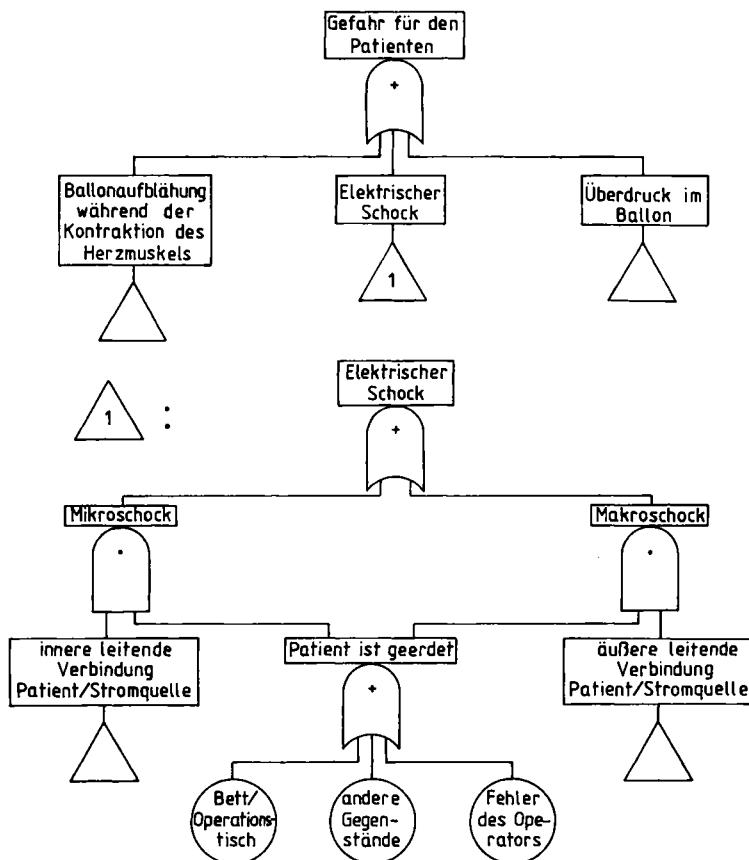


Abb. 18: Teil einer Fehlerbaumdarstellung für das Beispiel „Kreislaufentlastungssystem“

Liefert die Fehlerbaumauswertung keine akzeptable Sicherheitsstufe, ist also die Gefahr für das Eintreten des Systemausfalls zu groß, so müssen weitere Maßnahmen ergriffen werden. Hierzu gehört in erster Linie das Identifizieren „kritischer Ereignisse“ (dazu gehören beispielsweise Komponentenausfälle), die wesentlich zur hohen Systemausfallwahrscheinlichkeit beitragen.

Es muß nun versucht werden, die Zuverlässigkeit dieser Ereignisse zu erhöhen. Wie dies allgemein zu realisieren ist und welche Faktoren eine Erhöhung der Zuverlässigkeit beschränken, ist in Abschnitt 1.7 erläutert worden. Nachdem sämtliche Änderungsmaßnahmen durchgeführt worden sind, wird der Fehlerbaum erneut ausgewertet und eine Entscheidung darüber getroffen, ob der verbesserte Systementwurf einen akzeptablen Sicherheits- und (oder) Zuverlässigkeitssgrad erreicht.

Die Möglichkeit des mehrfachen Auftretens eines Basisereignisses stellt eine besondere Schwierigkeit bei der Berechnung der Systemausfallwahrscheinlichkeit dar. Dies Problem ist nicht fehlerbaumspezifisch, sondern tritt bei allen Systemdarstellungen auf, die das wiederholte Erscheinen eines im realen System genau einmal vorhandenen Systemteiles zulassen.

In der praktischen Zuverlässigkeitssanalyse wird häufig so verfahren, daß das System zunächst in disjunkte Teilsysteme zerlegt wird und danach Berechnungen der Ausfallwahrscheinlichkeiten der einzelnen Teilsysteme angestellt werden, woraus im Endeffekt die Systemausfallwahrscheinlichkeit zu bestimmen ist.

Zu finden ist dazu eine **modulare Repräsentation** des Fehlerbaumes, wenn wir einen **Modul** im Zusammenhang mit monotonen Strukturen als Menge von Komponenten ansehen, die als eine Art „Superkomponente“ in dem Sinne fungiert, daß der Zustand des Moduls ebensoviel Information zur Bestimmung des Gesamtsystemzustandes liefert wie die Zustände aller Modulkomponenten zusammen. Dies hat den Vorteil, daß die einzelnen Komponentenzustände nicht bekannt zu sein brauchen. Die angesprochene Problematik wird ausführlich bei Chatterjee (1975) diskutiert.

Die wichtigsten Argumente bei der Diskussion der Fehlerbaumanalysetechnik sind bereits angeklungen.

Als Hauptschwierigkeiten sind zusammenfassend anzuführen, daß die Fehlerbaumanalyse

- sich als zeitraubend erweist;
- bei sehr komplexen Systemen ihre Übersichtlichkeit verliert;
- oft nicht alle möglichen Gefahren miteinbeziehen kann;
- häufig einem Mangel an sachdienlichen Daten unterliegt;
- ungenaue Ergebnisse liefern kann, die beispielsweise aus einer ungenügenden Quantifizierung menschlicher Einflüsse resultieren.

Demgegenüber ist es bei sorgfältiger Anwendung der Fehlerbaumanalyse möglich,

- in Frage kommende Ursachen für einen Unfall oder für eine Katastrophe durch die Entdeckung zusätzlicher Ausfallkombinationen zu bestimmen;
- Systemschwachstellen aufzufinden;
- die Systeminformation mit Hilfe des visuellen Kommunikationsmittels „Fehlerbaum“ zu vermitteln;
- weitergehende Entscheidungen, z. B. darüber, ob der Systementwurf adäquat ist, zu treffen.

Nicht zuletzt die Tatsache, daß Systemausfälle in der heutigen Zeit katastrophale Auswirkungen haben können, macht deutlich, daß man auf Techniken wie die der Fehlerbaumanalyse nicht verzichten kann, vgl. Cummings (1975).

1.10. Systembetrachtungen bei mehrphasigen Missionen

1.10.1. Mehrphasige Missionen

In den bisherigen Betrachtungen gingen wir davon aus, daß die Darstellung eines Systems sich nicht verändert. Etwa bei einem Raumfahrzeug ist es jedoch so, daß das System (Raumfahrzeug) in verschiedenen Phasen verschiedene Teilaufgaben (**mehrphasige Missionen**) zu erfüllen hat. Diese verschiedenen Phasen unterliegen einer zeitlichen Reihenfolge, und die Aufgabenerfüllung einer speziellen Systemkomponente in einer Phase ist von der Aufgabenerfüllung in den vorangegangenen Phasen abhängig.

Die Analyse solcher Systeme mit mehrphasigen Missionen (**phased missions**) erlangt besondere Bedeutung bei der Planung von Reaktionen auf mögliche Katastrophen.

Beispiel: (vgl. Esary/Ziehms (1975)).

Eine Feuerwehr besitzt drei Fahrzeuge: Eine Mehrzweckspritze (M), einen Tankwagen (T) und ein Löschfahrzeug (L).

Die Feuerbekämpfungsausrüstung einer nahen chemischen Fabrik besteht aus einer Berieselungsanlage (B), einem Hydranten (H) und einem Spezialgerät zum Löschen brennender Chemikalien (S). Es soll nun eine Strategie erarbeitet werden, die im Falle eines Feuerausbruchs in der chemischen Fabrik zur Anwendung kommen kann:

(i) Während der Anfangsphase des Brandes genügt die Mehrzweckspritze, die mit einem geringen Wasservorrat versehen ist, oder das Löschfahrzeug (nur, wenn die Berieselungsanlage intakt ist), um die Evakuierung des Fabrikgeländes zu ermöglichen.

(ii) Zur Eindämmung des Feuers wird das Spezialgerät zusammen mit gewissen Hilfsfähigkeiten der Mehrzweckspritze bzw. des Löschwagens benötigt. Die Wasserzuführung erfolgt über den Hydranten oder bei Funktionsunfähigkeit über den Tankwagen durch die mit Pumpen versehene Mehrzweckspritze.

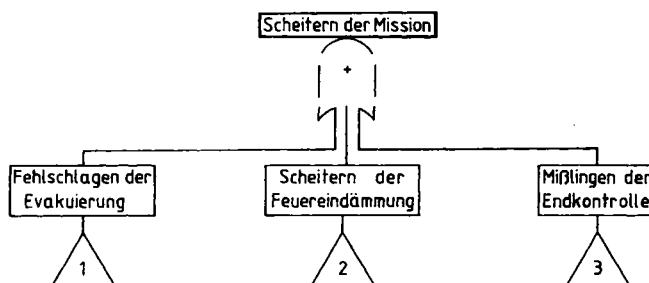
(iii) Nach der Feuereindämmung erfolgt die endgültige Kontrolle durch das Spezialgerät oder die Mehrzweckspritze. Die Wasserzufuhr erfolgt wiederum über den Hydranten oder gegebenenfalls über den Tankwagen durch die Pumpen der Mehrzweckspritze.

Betrachten wir die Mission als mißlungenen, sobald eine Teilmision scheitert, so läßt sich das Beispiel anhand des Fehlerbaumes aus Abb. 19 illustrieren.

Im Fehlerbaum, vgl. Abb. 19, beschreibt M beispielsweise die Funktionsunfähigkeit der Mehrzweckspritze M. Das gestrichelte ODER-Glied unterhalb der Baumspritze soll auf die besonderen Schwierigkeiten bei der Zuverlässigkeitsanalyse mehrphasiger Missionen hinweisen.

1.10.2. Phasenzuverlässigkeit und Missionszuverlässigkeit

Das Resultat der Zuverlässigkeitsanalyse der gesamten Mission ist nicht einfach durch Multiplikation der Phasenzuverlässigkeiten zu erhalten. Die stillschweigende Annahme, daß jede der 6 Komponenten zu Beginn einer Phase funktionsfähig ist, braucht keine Gültigkeit zu besitzen, was zu verschiedenen Zuverlässigkeiten für



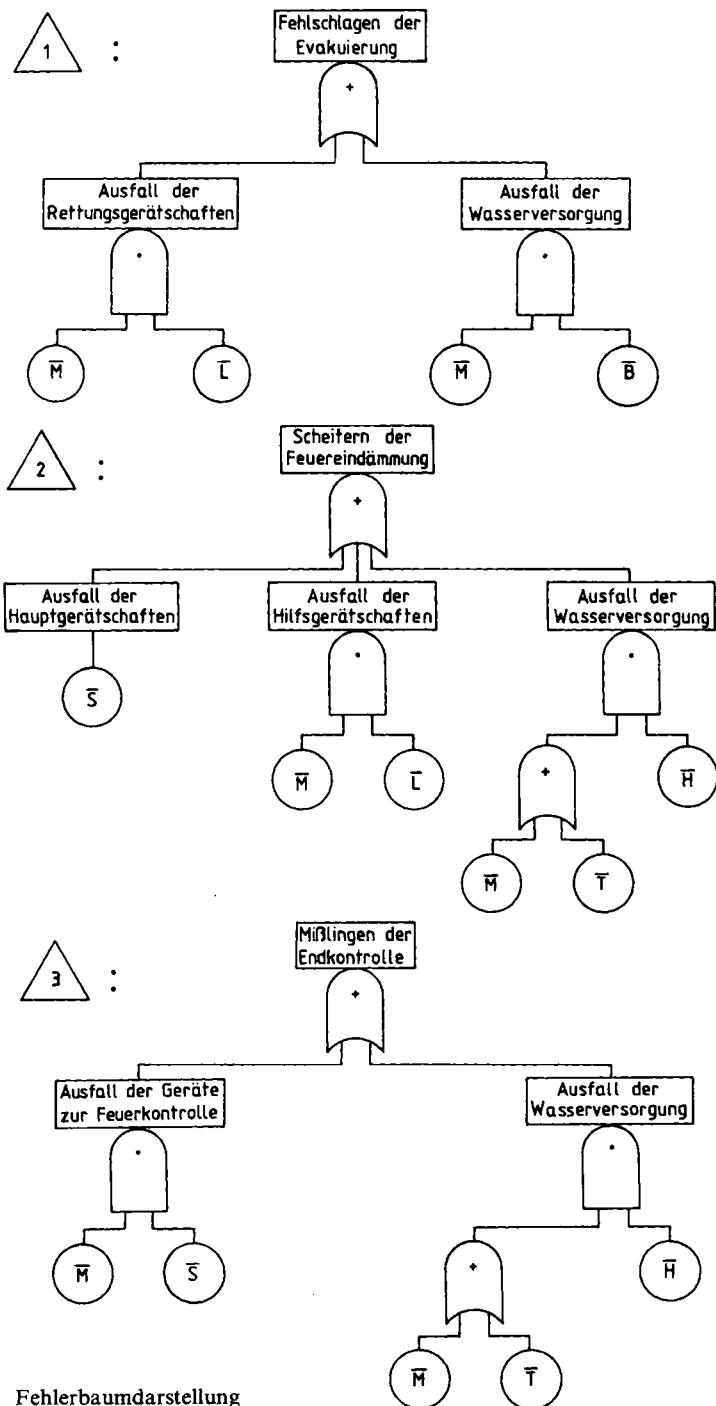


Abb. 19: Fehlerbaumdarstellung für das Feuerwehrbeispiel

die gesamte Mission führen kann. Diese Problematik wird an folgendem kleinen Beispiel deutlich.

Beispiel: Ein System mit 3 unabhängigen Komponenten K_1 , K_2 , K_3 soll eine zweiteilte Aufgabe erfüllen, deren Teilaufgaben durch das Zuverlässigkeitsschaltbild aus Abb. 20 beschrieben sind.

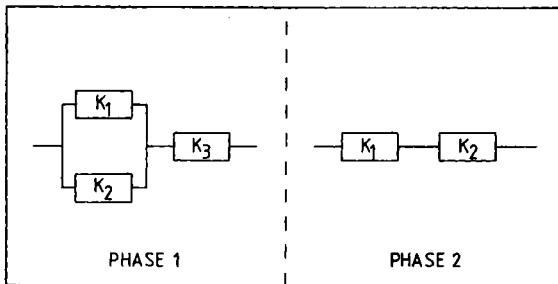


Abb. 20: Zuverlässigkeitsschaltbild eines Systems aus drei Komponenten in zwei Phasen

Wir wollen mit p_{j1} für $j = 1, 2, 3$ die Wahrscheinlichkeit dafür bezeichnen, daß die Komponente K_j in der Phase 1 funktioniert und mit p_{j2} für $j = 1, 2, 3$ die Wahrscheinlichkeit dafür, daß die Komponente K_j in Phase 2 funktioniert, wenn sie in der Phase 1 funktioniert hat. Weiterhin sollen alle diese Wahrscheinlichkeiten kleiner als 1 sein. Die Systemzuverlässigkeit in Phase 1 ergibt sich dann zu

$$\begin{aligned} h_1 &= p_{11}p_{31}(1-p_{21}) + p_{21}p_{31}(1-p_{11}) + p_{11}p_{21}p_{31} \\ &= p_{11}p_{31} + p_{21}p_{31} - p_{11}p_{21}p_{31}, \end{aligned}$$

vgl. Abschnitt 1.1 und 1.3, und die Systemzuverlässigkeit in Phase 2 ist unter der Voraussetzung, daß K_1 , K_2 und K_3 in Phase 1 funktioniert haben, durch

$$h_2 = p_{12}p_{22}$$

gegeben. Das Produkt der Systemzuverlässigkeiten ist somit

$$h_1 \cdot h_2 = (p_{11}p_{31} + p_{21}p_{31} - p_{11}p_{21}p_{31}) \cdot (p_{12}p_{22}).$$

Die korrekte Missionszuverlässigkeit ergibt sich jedoch aufgrund der Abhängigkeit der beiden Phasen zu

$$h_M = p_{31} \cdot p_{11}p_{12}p_{21}p_{22},$$

denn das System funktioniert ja nur dann, wenn K_3 in Phase 1 und K_1 und K_2 in beiden Phasen funktioniert haben. Genauer heißt das: In Phase 1 arbeitet das System zwar noch, wenn eine der beiden Komponenten K_1 oder K_2 ausfällt; damit das System aber in Phase 2 noch arbeitsfähig ist, darf in Phase 1 keine Komponente ausfallen; das gesamte System ist also nur intakt, wenn K_1 und K_2 in jeder Phase und K_3 in Phase 1 arbeiten, was gerade einer Serienstruktur entspricht, wie sie durch die Berechnungsformel für h_M ausgedrückt wird. Die Missionszuverlässigkeit ist hier natürlich stets geringer als das Produkt der Phasenzuverlässigkeiten.

1.10.3. Die Transformation mehrphasiger Missionen

Unsere Aufgabe ist es nun, eine allgemeingültige Transformation zu erarbeiten, die jede mehrphasige Mission in eine einphasige der gleichen Zuverlässigkeit überführt.

Das erwähnte Abhängigkeitsproblem, das zwischen der Erfüllung einer Komponentenaufgabe in einer bestimmten Phase und der Erfüllung in vorangegangenen Phasen besteht, kann gelöst werden, wenn wir uns den speziellen Abhängigkeitscharakter vor Augen führen. Es ist zu beachten, daß eine Komponente in Phase i genau dann funktioniert, wenn sie in Phase 1, dann in Phase 2, ..., dann in Phase $i - 1$ funktioniert hat und schließlich in Phase i funktioniert.

Somit ist es sinnvoll, eine Komponente K_j in Phase i durch ein Seriensystem unabhängig voneinander arbeitender Komponenten K_{j1}, \dots, K_{ji} zu ersetzen.

Im Zuverlässigkeitsschaltbild wird also eine Komponente K_j in Phase i ersetzt durch i gleichartige in Serie geschaltete Komponenten K_{j1}, \dots, K_{ji} .

Im Fehlerbaum wird ein Ereignis \bar{K}_j , das den Ausfall der Komponente K_j beschreibt, in Phase i durch das Symbol aus Abb. 21 ersetzt.



Abb. 21: Fehlerbaumdarstellung des Ereignisses \bar{K}_j in der i -ten Phase

Beispiel: Transformiert man nun den Fehlerbaum des Feuerwehrbeispiels in der beschriebenen Art und Weise und übersetzt ihn in ein Zuverlässigkeitsschaltbild, das getrennt für jede Phase gelesen werden kann, so ergibt sich Abb. 22.

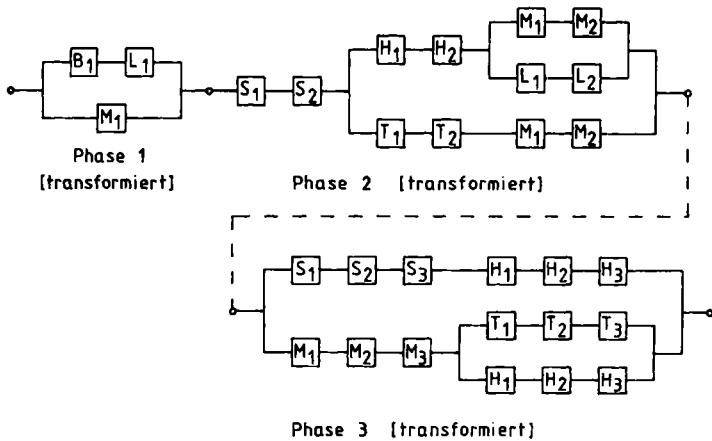


Abb. 22: Zuverlässigkeitsschaltbild für das Feuerwehrbeispiel nach erfolgter Transformation in jeder Phase

Wenden wir zusätzlich unsere Überlegungen zur Darstellung monotoner Systeme durch minimale Schnitte an, so gelangen wir zu einer einfacheren Zuverlässigkeitsdarstellung des Feuerwehrbeispiels, wenn wir uns zusätzlich folgende Regel zunutze machen:

Ein minimaler Schnitt in einer Phase kann gestrichen werden, d. h. aus der Liste minimaler Schnitte für diese Phase entfernt werden, falls er einen minimalen Schnitt einer späteren Phase enthält.

Beispiel: Unser dreiphasiges Feuerwehrbeispiel liefert unter Zuhilfenahme von Abb. 22 die minimalen Schnittmengen $\{M, L\}$, $\{M, B\}$ in Phase 1, $\{S\}$, $\{H, T\}$, $\{H, M\}$, $\{M, L\}$ in Phase 2 und $\{S, M\}$, $\{H, M\}$, $\{H, T\}$ in Phase 3.

Nach Anwendung oben angegebener Regel verbleiben lediglich die Schnittmengen $\{M, B\}$ in Phase 1, $\{S\}$, $\{M, L\}$ in Phase 2 und $\{S, M\}$, $\{H, M\}$, $\{H, T\}$ in Phase 3. Hieraus ergeben sich die vereinfachten Zuverlässigkeitsschaltbilder für die 3 Phasen, die in Abb. 23 dargestellt sind.

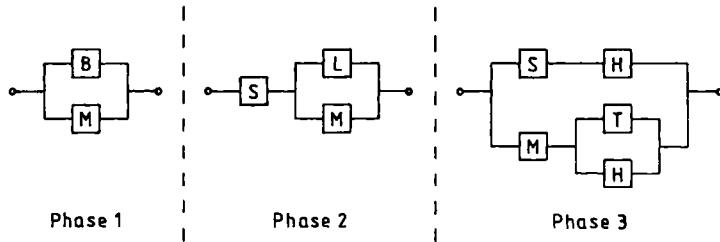


Abb. 23: Zuverlässigkeitsschaltbilder für die einzelnen Phasen des Feuerwehrbeispiels nach Anwendung der Regel zur Streichung minimaler Schnitte

Nach dieser Streichung kann erneut die schon angegebene Transformation der Komponenten in den drei Phasen vorgenommen werden, d. h. die Komponente K_j wird in der i -ten Phase durch in Serie geschaltete Komponenten K_{j1}, \dots, K_{ji} ersetzt.

Beispiel: Die Zuverlässigkeitsschaltbilder aus Abb. 23 sollen nun in den einzelnen Phasen noch einmal transformiert werden. Es ergibt sich dabei die Darstellung aus Abb. 24.

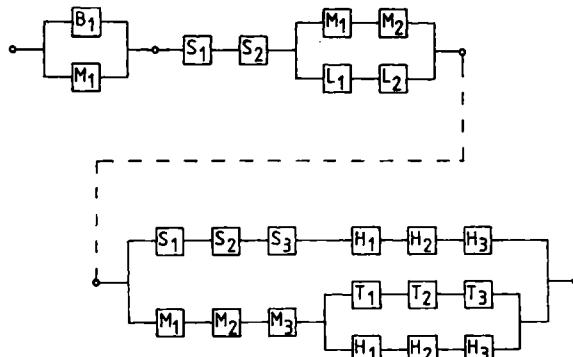


Abb. 24: Endgültiges Missionszuverlässigkeitsschaltbild

2. Klassen von Lebensdauerverteilungen

In Abschnitt 1 sind wir stets davon ausgegangen, daß die Zuverlässigkeit eines Systems (einer Einheit) im interessierenden Zeitraum konstant ist. Im allgemeinen wird es jedoch so sein, daß sich die Zuverlässigkeit im Laufe der Zeit ändert. Bei physikalischen und biologischen Einheiten etwa treten mit fortschreitender Lebens-

zeit einer Einheit Ermüdungserscheinungen (z. B. Korrosion, Verschleiß) auf, die die Zuverlässigkeit verringern.

Diese Abhängigkeit der Zuverlässigkeit von der Lebenszeit wollen wir im folgenden berücksichtigen. Da die Lebensdauer T einer Einheit als zufällig angesehen werden kann, untersuchen wir deren Verteilung mit der Verteilungsfunktion

$$F(t) = P(T \leq t).$$

Geht man davon aus, daß sich die Zuverlässigkeit einer Einheit im Laufe der Zeit ändert, so tritt an die Stelle des Zuverlässigkeitstyps synonym der Begriff der Überlebenswahrscheinlichkeit. Ist $F(t)$ die Verteilungsfunktion der Lebensdauer T einer Einheit, so heißt

$$\bar{F}(t) = P(T > t) = 1 - F(t)$$

die **Überlebenswahrscheinlichkeit** der Einheit, vgl. auch Abschnitt 4 in Kap. IV. Natürlich geht man hier davon aus, daß die Einheit zum Zeitpunkt $t = 0$ funktioniert (lebt) und zu einer Zeit $t > 0$ ausfällt; dies bedeutet gerade, daß gilt

$$\bar{F}(0) = 1 \quad \text{und} \quad \bar{F}(\infty) = 0.$$

Im folgenden wollen wir verlangen, daß die Verteilung einer Lebensdauer T eine Dichte $f(t)$ mit höchstens endlich vielen Sprungstellen besitzt. Damit ist sichergestellt, daß z. B. die folgende Situation erfaßt werden kann: Die Lebensdauer einer Maschine wird untersucht; natürlich arbeitet diese nicht kontinuierlich, sondern wird ein- und ausgeschaltet, was zu Sprungstellen in der Dichte der Lebensdauerverteilung führen kann.

Neben der Überlebenswahrscheinlichkeit ist die (**bedingte**) **Ausfallrate** oder **Hazardrate**, vgl. auch Kap. IV, Abschnitt 4,

$$\begin{aligned} r(x) &= \lim_{\substack{t \rightarrow 0 \\ [t > 0]}} \frac{1}{t} P(T \leq x + t | T > x) \\ &= \lim_{\substack{t \rightarrow 0 \\ [t > 0]}} \frac{1}{t} \frac{P(T \leq x + t \text{ und } T > x)}{P(T > x)} \\ &= \lim_{\substack{t \rightarrow 0 \\ [t > 0]}} \frac{1}{t} \frac{F(x + t) - F(x)}{\bar{F}(x)} \\ &= \frac{f(x)}{\bar{F}(x)} \left[= - \frac{d}{dx} \ln \bar{F}(x) \right] \end{aligned}$$

ein nützliches Instrument zur Beschreibung der Alterung einer Einheit. Das Produkt $t \cdot r(x)$ entspricht in etwa der Wahrscheinlichkeit, daß eine Einheit, die das Alter x bereits erreicht hat, in einem „kleinen“ Zeitabschnitt $(x, x + t]$ ausfällt.

Ein weiterer Begriff zur Beschreibung der Alterung ist die **bedingte Überlebenswahrscheinlichkeit**

$$\bar{F}(t|x) = \frac{F(t+x)}{\bar{F}(x)},$$

d.h. die Wahrscheinlichkeit, daß eine Einheit, die bereits das Alter x erreicht hat, den Zeitpunkt x um mindestens t überlebt. Für einige spezielle Verteilungen sind die Ausfallrate und die Überlebenswahrscheinlichkeit in Abschnitt 4 von Kap. IV bereits angegeben. Dort ergab sich z.B., daß die Ausfallrate bei $\text{Ex}(\lambda)$ -verteilten Lebensdauern konstant gleich dem Parameter λ ist und somit nicht vom bereits erreichten Alter abhängt.

Hier sollen nun ausgehend von Ausfallraten und Überlebenswahrscheinlichkeiten verschiedene Klassen von Lebensdauerverteilungen vorgestellt werden. Auch werden jeweils Tests zur Überprüfung der Klassenzugehörigkeit der Lebensdauerverteilung von Einheiten einer Grundgesamtheit angegeben.

Diese Tests werden stets auf der Basis einer Zufallsstichprobe von n Einheiten aus einer Grundgesamtheit durchgeführt, deren Lebensdauern eine unbekannte Verteilungsfunktion F besitzen mit $F(0) = 0$.

Beobachtet werden die Lebensdauern x_1, \dots, x_n dieser n Einheiten, aufgrund ihrer Entscheidungen getroffen werden.

Zu erwähnen ist, daß die im folgenden angegebenen Tests im wesentlichen nur für die in der jeweiligen Alternativ-Hypothese formulierten Verteilungen als konsistent angesehen werden können; d.h. liegt eine beliebige der betreffenden Verteilungen aus der Alternative dem konkreten Problem zugrunde, so geht mit zunehmendem Stichprobenumfang n der Fehler 2. Art gegen Null. Implizit unterstellt man ohnehin, daß eine zu prüfende Verteilung auch tatsächlich in einer der im jeweiligen Testproblem konkurrierenden Verteilungsklassen liegt.

2.1. IFR- und DFR-Verteilungen

2.1.1. Die Verteilungsklassen IFR und DFR

Die Einteilung von Lebensdauerverteilungen in IFR (**Increasing Failure Rate**)- und DFR (**Decreasing Failure Rate**)-Verteilungen erfolgt, wie die Namen schon besagen, ausgehend von der Ausfallrate oder, äquivalent dazu, von der bedingten Überlebenswahrscheinlichkeit.

Ist die Ausfallrate einer Lebensdauerverteilung steigend (im Sinne von nicht fallend), so spricht man von einer **IFR-Verteilung**. Die bedingte Überlebenswahrscheinlichkeit einer solchen Verteilung fällt mit zunehmender Zeit.

Eine Lebensdauerverteilung heißt **DFR-Verteilung**, falls ihre Ausfallrate fallend (im Sinne von nicht steigend) ist bzw. die bedingte Überlebenswahrscheinlichkeit mit fortschreitender Zeit wächst.

Zu beiden Verteilungsklassen zugleich gehören lediglich die Exponentialverteilungen, die ja die einzigen Verteilungen mit konstanter Ausfallrate sind:

$$r(x) = \frac{f(x)}{\bar{F}(x)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

DFR-Verteilungen sind etwa vorstellbar bei Metallen, deren Bruchfestigkeit bei Betriebsbelastung ansteigt.

Beispiel: Die Weibullverteilung mit Verteilungsfunktion

$$F(t) = 1 - e^{-\alpha t^\beta} \quad \text{für } \alpha, \beta, t > 0$$

und Dichte

$$f(t) = \alpha \beta t^{\beta-1} e^{-\alpha t^\beta}$$

ist, vgl. Kap. IV, Abschnitt 4.2, eine häufig verwandte Lebensdauerverteilung. Ihre Ausfallrate ist

$$r(x) = \frac{f(x)}{F(x)} = \alpha \beta x^{\beta-1};$$

somit ergibt sich für

- $\beta = 1$ eine Exponentialverteilung mit Parameter α , für
- $\beta \geq 1$ eine IFR-Verteilung und für
- $\beta \leq 1$ eine DFR-Verteilung.

2.1.2. IFR- und DFR-Tests

Will man für die Lebensdauerverteilung in einer Grundgesamtheit die Hypothese

$$H_0: \text{DFR} \text{ gegen } H_1: \text{IFR} \text{ (und nicht exponential)}$$

(IFR-Test), oder

$$H'_0: \text{IFR} \text{ gegen } H'_1: \text{DFR} \text{ (und nicht exponential)}$$

(DFR-Test) zum Niveau α testen, so können verschiedene Tests verwandt werden. Sind x_1, \dots, x_n die an einer Stichprobe beobachteten Lebensdauern von n Objekten und bezeichnen

$$0 = x_{(0)} \leqq x_{(1)} \leqq \dots \leqq x_{(n)}$$

die geordneten Lebensdauern, so berechnet man zunächst für $j = 1, \dots, n$ die **normalisierten Abstände**

$$D_j = (n - j + 1) (x_{(j)} - x_{(j-1)})$$

sowie die zugehörigen Rangzahlen

$$R_j \text{ für } j = 1, \dots, n.$$

A. Der Proschan-Pyke-Test

Der von Proschan/Pike (1967) vorgeschlagene (vom Skalenparameter unabhängige) Test verwirft die Nullhypothese H_0 , falls gilt

$$V_n \geqq v_{n; 1-\alpha}$$

und die Nullhypothese H'_0 , falls gilt

$$V_n \leqq v_{n; \alpha}.$$

Dabei ist

$$V_n = \sum_{j < i} V_{ji} \quad \text{mit} \quad V_{ji} = \begin{cases} 1, & \text{falls } D_j \geqq D_i; \\ 0, & \text{sonst} \end{cases}$$

die kritischen Werte $v_{n; \gamma}$ des Proschan-Pyke-Tests müssen aus der Beziehung

$$P_0(V_n \geqq v_{n; \gamma}) = 1 - \gamma$$

gewonnen werden, wobei P_0 die Wahrscheinlichkeit bei Zugrundelegung einer Exponentialverteilung bezeichnet.

B. Der cttot-Test nach Epstein

Epstein (1960) schlägt den **cttot-Test (Cumulative Total Time on Test-Test)** zur Prüfung von H_0 und H'_0 vor, der auf den **kumulativen Gesamtprüfungszeiten**

$$T(x_{(i)}) = \sum_{j=1}^i D_j \quad \text{für } i = 1, \dots, n$$

basiert.

Bei diesem Test, der ebenfalls unabhängig vom Skalenparameter ist, wird die Hypothese H_0 verworfen, falls für die **cttot (Cumulative Total Time on Test)-Statistik**

$$K_n = \frac{\sum_{j=1}^{n-1} \sum_{i=1}^j D_i}{\sum_{t=1}^n D_t} = \frac{\sum_{j=1}^{n-1} (n-j) D_j}{\sum_{t=1}^n D_t} \geq k_{n; 1-\alpha}$$

gilt, und die Hypothese H'_0 wird verworfen, falls gilt

$$K_n \leq k_{n; \alpha}.$$

Einige von Barlow (1968) simulierte kritische Werte $k_{n; \gamma}$ sind in **Tabelle 7** angegeben.

Tab. 7: Kritische Werte $k_{n; \gamma}$ der cttot-Teststatistik K_n

$n \backslash \gamma$	0,05	0,10	0,90	0,95
2	0,05	0,10	0,90	0,95
3	0,32	0,45	1,55	1,68
4	0,68	0,84	2,15	2,33
5	1,04	1,25	2,75	2,95
6	1,43	1,65	3,34	3,57
7	1,83	2,08	3,90	4,15
8	2,24	2,52	4,49	4,75
9	2,65	2,94	5,06	5,35
10	3,06	3,38	5,62	5,92

Stehen keine kritischen Werte zur Verfügung, so kann man die gute Approximation

$$k_{n; \gamma} \approx \frac{n-1}{2} + \sqrt{\frac{n}{12}} u_\gamma,$$

wobei u_γ das γ -Quantil der $N(0, 1)$ -Verteilung bezeichnet, verwenden; die hier nach approximierten Werte weichen von den entsprechenden simulierten der Tab. 7 nicht wesentlich ab.

Asymptotische Untersuchungen von Bickel/Doksum (1969) haben gezeigt, daß die asymptotische relative Effizienz (ARE) des Tests mit K_n größer ist als die des Tests mit der Teststatistik V_n des Proschan-Pyke-Tests.

Bickel/Doksum (1969) geben auch noch einige weitere Prüfgrößen für IFR- bzw. DFR-Tests an, wie z.B. das Ranggegenstück

$$W_n = \sum_{j=1}^n j \ln \left(1 - \frac{R_j}{n+1} \right)$$

zur cttot-Teststatistik.

Beispiel: Bei der Untersuchung von $n = 17$ Kühlaggregaten ergaben sich die (der Größe nach geordneten) Lebensdauern (in Jahren) aus **Tab. 8**. Dort sind auch die normalisierten Abstände

Tab. 8: Lebensdauern von $n = 17$ Kühlaggregaten und zugehörige normalisierte Abstände

(j)	Lebensdauer $x_{(j)}$	normalisierter Abstand D_j
0	0,00	—
1	0,05	0,85
2	0,11	0,96
3	0,24	1,95
4	0,61	5,18
5	1,22	7,93
6	1,42	2,40
7	1,49	0,77
8	1,53	0,40
9	1,92	3,51
10	1,93	0,08
11	2,15	1,54
12	2,53	2,28
13	2,83	1,50
14	3,61	3,12
15	3,80	0,57
16	6,23	4,86
17	9,93	3,70

$$D_j = (n - j + 1) (x_{(j)} - x_{(j-1)})$$

angegeben.

Mittels des cttot-Tests wollen wir zum Niveau $\alpha = 0,05$ die Hypothese H_0 , daß die Lebensdauer der Kühlaggregate in der Grundgesamtheit eine DFR-Verteilung besitzt, testen. Dabei wird die angegebene Approximation der kritischen Werte $k_{n; 1-\alpha}$ des Tests verwendet. Da gilt

$$K_n = \frac{\sum_{j=1}^{16} (17-j) D_j}{\sum_{i=1}^{17} D_i} = \frac{326,14}{41,60} = 7,84 < 9,958 = \sqrt{\frac{17}{12}} u_{0,95} + \frac{16}{2} \approx k_{17; 0,95},$$

kann die Hypothese H_0 zum 5% Niveau nicht verworfen werden.

2.2. NBU- und NWU-Verteilungen

2.2.1. Die Verteilungsklassen NBU und NWU

Die Klassifizierung einer Lebensdauerverteilung in die Klassen NBU und NWU basiert auf dem Vergleich der Überlebenswahrscheinlichkeit $\bar{F}(x)$ bis zum Alter x für eine neue Komponente und der bedingten Überlebenswahrscheinlichkeit

$$\bar{F}(x|y) = \frac{\bar{F}(x+y)}{\bar{F}(y)}.$$

Eine Lebensdauerverteilung heißt **NBU- (New Better than Used)-Verteilung**, falls für ihre Verteilungsfunktion F gilt:

$$\bar{F}(x) \geq \frac{\bar{F}(x+y)}{\bar{F}(y)} \quad \text{für alle } x \text{ und } y \geq 0.$$

Gilt die umgekehrte Relation, so spricht man von einer **NWU-(New Worse than Used)-Verteilung**. Bei NBU-Verteilungen ist die Überlebenswahrscheinlichkeit einer neuen Einheit stets größer oder gleich der einer alten; bei NWU-Verteilungen ist die Überlebenswahrscheinlichkeit einer neuen Einheit stets geringer oder gleich der einer alten Einheit.

Die Klasse der IFR-Verteilungen ist eine Teilmenge der NBU-Verteilungen und die Klasse der DFR-Verteilungen eine Teilmenge der NWU-Verteilungen. Beiden Klassen sind wiederum alle Exponentialverteilungen gemeinsam.

Die Klasse der NBU-Verteilungen hat sich als besonders nützlich bei **Instandhaltungsproblemen** erwiesen, da sie sich als größte Klasse ergibt, bei der sowohl Altersersatzstrategie als auch Gruppenersatzstrategie die Anzahl der in einem beliebigen Zeitintervall $[0; t]$ aufgetretenen Ausfälle verringert. **Altersersatzstrategie** ist hierbei dadurch gekennzeichnet, daß eine Einheit ersetzt wird, sobald sie ausfällt oder ein bestimmtes Alter T erreicht hat, je nachdem, welches Ereignis früher eintritt. **Gruppenersatzstrategie** bedeutet, daß die Einheit zu den Zeitpunkten $T, 2T, 3T, \dots$ sowie nach Ausfall erneuert wird, vgl. Abschnitt 5 dieses Kapitels.

2.2.2. NBU- und NWU-Tests: Der Hollander-Proschan-Test

Will man ausgehend von einer Stichprobe vom Umfang n die Hypothese

$$H_0: \text{NWU} \quad \text{gegen} \quad H_1: \text{NBU (und nicht exponential)}$$

oder

$$H'_0: \text{NBU} \quad \text{gegen} \quad H'_1: \text{NWU (und nicht exponential)}$$

über die Lebensdauerverteilung in der Grundgesamtheit zum Niveau α testen, so kann man den **Hollander-Proschan-Test** verwenden, vgl. Hollander/Proschan (1972).

Sind x_1, \dots, x_n die in einem Experiment beobachteten Lebensdauern von n Einheiten aus einer interessierenden Grundgesamtheit, so ist die Teststatistik dieses Tests gerade

$$J_n = \frac{2}{n(n-1)(n-2)} \cdot T_n,$$

wobei T_n die Anzahl derjenigen Lebensdauern ist, die größer als die Summe zweier anderer sind, d. h.

$$T_n = \sum_{\substack{j_1=1 \\ j_1+j_2 \\ j_2 < j_3}}^n \sum_{j_2=1}^n \sum_{\substack{j_3=1 \\ j_1+j_3}}^n \psi(x_{j_1}, x_{j_2} + x_{j_3})$$

$$\text{mit } \psi(x_{j_1}, x_{j_2} + x_{j_3}) = \begin{cases} 1, & \text{falls } x_{j_1} > x_{j_2} + x_{j_3} \\ 0, & \text{sonst} \end{cases}.$$

Beispiel: In einer Stichprobe von 5 Objekten werden die Lebensdauern $x_1 = 17$, $x_2 = 24$, $x_3 = 39$, $x_4 = 55$ und $x_5 = 62$ beobachtet.

Da gilt $55 \geq 17 + 24$, $62 \geq 17 + 24$, $62 \geq 17 + 39$, ist $T_n = 3$ und somit der Wert der Teststatistik des Hollander-Proschan-Tests gerade

$$J_n = \frac{2}{5 \cdot 4 \cdot 3} \cdot 3 = \frac{1}{30} \cdot 3 = \frac{1}{10} = 0,1.$$

Die Hypothese H_0 muß nun zum Niveau α verworfen werden, falls

$$J_n \leq j_{n; \alpha}$$

ist und H'_0 muß zum Niveau α verworfen werden, falls gilt

$$J_n \geq j_{n; 1-\alpha}.$$

Einige simulierte kritische Werte $j_{n; \gamma}$ sind in **Tab. 9** angegeben.

Tab. 9: Kritische Werte $j_{n; \gamma}$ des Hollander-Proschan-Tests

$n \backslash \gamma$	0,05	0,10	0,90	0,95
8	0,1488	0,1786	0,3095	0,3155
9	0,1587	0,1865	0,3016	0,3095
10	0,1667	0,1917	0,2972	0,3056
11	0,1737	0,1960	0,2949	0,3030
12	0,1772	0,1984	0,2924	0,3000
13	0,1830	0,2028	0,2902	0,2972
14	0,1868	0,2042	0,2885	0,2949
15	0,1897	0,2066	0,2872	0,2938
16	0,1923	0,2083	0,2857	0,2923
17	0,1946	0,2103	0,2843	0,2907
18	0,1961	0,2116	0,2831	0,2896
19	0,1985	0,2129	0,2820	0,2883
20	0,2003	0,2140	0,2810	0,2871
25	0,2068	0,2184	0,2780	0,2835
30	0,2113	0,2220	0,2750	0,2800
35	0,2147	0,2238	0,2729	0,2783
40	0,2171	0,2260	0,2716	0,2763
45	0,2194	0,2275	0,2702	0,2751
50	0,2214	0,2286	0,2691	0,2736

Beispiel: (vgl. Hovsepian et al. (1975), Gail/Gastwirth (1978)).

Bei der Untersuchung von 24 an Prostatakrebs erkrankten Personen ergaben sich die folgenden Überlebenszeiten (in Monaten nach Behandlungsbeginn): 0, 1, 2, 3, 3, 4, 5, 6, 7, 11, 11, 14, 18, 22, 28, 32, 33, 35, 36, 39, 41, 42, 51, 64.

Man könnte nun vermuten, daß erkrankte Patienten, die bereits eine gewisse Zeitspanne seit Behandlungsbeginn überlebt haben, eine größere Überlebenswahrscheinlichkeit besitzen als andere (am Anfang der Behandlung). Zur Überprüfung dieser Vermutung wollen wir die Hypothese

H_0' : NBU gegen H_1' : NWU (und nicht exponential)

über die Verteilung der Überlebenszeiten nach Erkrankung zum 5% Niveau testen.
Mit $T_n = 1549$ ergibt sich

$$J_n = \frac{2}{24 \cdot 23 \cdot 22} \cdot 1549 = 0,255.$$

Da der Wert $j_{n;1-\alpha} = j_{24;0,95}$ nicht vertafelt ist, müssen wir den zum nächstkleineren n gehörigen kritischen Wert – also $j_{20;0,95} = 0,2871$ – verwenden. Da gilt

$$J_n = 0,255 < 0,2871 = j_{20;0,95},$$

können wir die Hypothese H_0' nicht verwerfen, d. h. unsere Vermutung eines längerfristigen Behandlungserfolges wird nicht bestätigt.

2.3. IFRA- und DFRA-Verteilungen

2.3.1. Die Verteilungsklassen IFRA und DFRA

Betrachten wir ein aus IFR-Einheiten zusammengesetztes System, so wird man auch für das System wachsende Ausfallrate vermuten, da die Ausfallrate jeder einzelnen Komponente mit der Zeit wächst und das System durch den Ausfall einer Komponente strukturell nicht stärker wird. Letztere sinnvolle Eigenschaft ist die wichtigste Voraussetzung für den gebräuchlichen Begriff der monotonen Systeme, die wir hier zugrundelegen.

Das folgende Beispiel zeigt, daß ein System aus IFR-Komponenten die IFR-Eigenschaft nicht zu besitzen braucht, die IFR-Eigenschaft also nicht abgeschlossen bzgl. der Bildung monotoner Systeme ist.

Beispiel: Wir untersuchen ein Parallelsystem aus 2 unabhängigen Komponenten, welche die IFR-Lebensdauerverteilungen

$$F_1(t) = 1 - e^{-t} \quad \text{bzw.} \quad F_2(t) = 1 - e^{-2t}$$

besitzen.

Die Überlebenswahrscheinlichkeit des Systems ergibt sich zu

$$\bar{F}(t) = 1 - (1 - e^{-t})(1 - e^{-2t}).$$

Für die Ausfallrate gilt

$$r(t) = \frac{e^{-t} + 2e^{-2t} - (1+2)e^{-(1+2)t}}{e^{-t} + e^{-2t} - e^{-(1+2)t}} = \frac{e^{-t} + 2e^{-2t} - 3e^{-3t}}{e^{-t} + e^{-2t} - e^{-3t}}.$$

Zur Untersuchung der IFR-Eigenschaft betrachtet man die erste Ableitung der Ausfallratefunktion des Systems:

$$r'(t) = \frac{-e^{-3t} + 4e^{-4t} + e^{-5t}}{(e^{-t} + e^{-2t} - e^{-3t})^2}.$$

Da $r'(0) = 4$ und $r'(2) = -0,047762$ ist, besitzt $r'(t)$ im Intervall $(0;2)$ eine Nullstelle und $r(t)$ ein relatives Minimum oder ein relatives Maximum. Somit besitzt die Lebensdauer des betrachteten Systems keine IFR-Verteilung.

Die Ermittlung der kleinsten Klasse von Verteilungen, die sämtliche Systeme mit

IFR-Komponenten enthält, führt auf die Klasse der Verteilungen mit wachsender mittlerer Ausfallrate (**Increasing Failure Rate Average – IFRA**).

Eine Lebensdauerverteilung ist eine **IFRA-Verteilung**, falls für $t \geq 0$ die Funktion (**mittlere kumulierte Ausfallrate**)

$$\frac{1}{t} \int_0^t r(x) dx$$

wächst, bzw. eine **DFRA-Verteilung**, falls diese Funktion für $t \geq 0$ fällt.

Die Bedeutung der IFRA-Eigenschaft als Grundlage einer in der Zuverlässigkeitstheorie enorm wichtigen Verteilungsklasse wird bei der Betrachtung sogenannter **Schockmodelle** erkennbar. Als einfaches Schockmodell sei das **kumulative Beschädigungsmodell** angeführt, das dann vorliegt, wenn in Sprungstellen eines Poisson-Prozesses mit Intensität λ Schocks auftreten, die zufällige Beschädigungen des Geräts hervorrufen.

Dabei spricht man von einem **Poisson-Prozeß mit Intensität λ** , wenn gilt:

In aufeinanderfolgenden zufälligen Zeitpunkten des Intervalls $[0; \infty)$ mögen homogene Ereignisse (Schocks) eintreten. Es seien die Abstände zwischen aufeinanderfolgenden Zeitpunkten des Eintretens dieser Ereignisse als unabhängig und exponentialverteilt mit Verteilungsfunktion $1 - e^{-\lambda t}$ vorausgesetzt. Mit $N(t)$ sei weiterhin die Anzahl der im Intervall $[0; t]$ eingetretenen Ereignisse bezeichnet. Der stochastische Prozeß $\{N(t); t \geq 0\}$ heißt dann Poisson-Prozeß mit Intensität λ .

Die Beschädigungen akkumulieren sich bis zu einer kritischen Schwelle, deren Überschreiten den Geräteausfall zur Folge hat. Die Überlebenswahrscheinlichkeit in diesem kumulativen Beschädigungsmodell, also die Wahrscheinlichkeit für ein fehlerfreies Funktionieren des Gerätes im Intervall $[0; t]$, die gegeben ist durch

$$H_F(t) = \sum_{k=0}^{\infty} [\lambda t]^k / k! e^{-\lambda t} F^{(k)}(y) \quad \text{für } 0 \leq t < \infty,$$

ist die Überlebenswahrscheinlichkeit einer IFRA-Lebensdauerverteilung, vgl. Barlow/Proschan (1978). Hierbei bedeutet $e^{-\lambda t} (\lambda t)^k / k!$ die Poissonsche Wahrscheinlichkeit dafür, daß sich genau k Schocks ereignen. $F^{(k)}(y)$ bezeichnet die k -fache Faltung (Verteilungsfunktion der Summe von k unabhängigen Zufallsvariablen mit gleicher Verteilungsfunktion F), vgl. Abschnitt 7.2.B in Kap. II, die die Wahrscheinlichkeit dafür angibt, daß die Schwelle y nicht durch die von den k Schocks erzielte Beschädigung überschritten wird.

Die IFRA-Eigenschaft bleibt sogar dann noch bestehen, wenn wir eine **Erweiterung dieses kumulativen Beschädigungsmodells** dahingehend vornehmen, daß die wiederum zeitlich gemäß einem Poisson-Prozeß (unabhängig voneinander) eintretenden Schocks immer heftigere Beschädigungen oder Ermüdungen bewirken. Verzichtet man auf die Voraussetzung der Unabhängigkeit aufeinanderfolgender Beschädigungen und bezieht man Beschädigungen mit ein, die nicht identisch verteilt sind, so geht die IFRA-Eigenschaft für die Lebensdauerverteilung nicht verloren. Bei dem hier zugrundeliegenden physikalischen Modell wird vorausgesetzt, daß durch die Anhäufung der Beschädigungen die Widerstandsfähigkeit gegenüber weiteren Beschädigungen vermindert wird und daß die Auswirkung von später eintretenden Schocks bei Akkumulation der Beschädigungen der Wahrscheinlichkeit nach größer ist als bei früheren Schocks.

2.3.2. IFRA- und DFRA-Tests: Der cttot-Test

Zum Testen von Hypothesen über IFRA- und DFRA-Verteilungen kann man den aus Abschnitt 2.2.1 bekannten cttot-Test, vgl. Barlow (1968), verwenden. Im Test von

$$H_0: \text{DFRA} \text{ gegen } H_1: \text{IFRA (und nicht exponential)}$$

wird H_0 zum Niveau α verworfen, falls gilt

$$K_n \geq k_{n; 1-\alpha}$$

und im Test

$$H'_0: \text{IFRA} \text{ gegen } H'_1: \text{DFRA (und nicht exponential)}$$

wird die Hypothese H'_0 verworfen, wenn

$$K_n \leq k_{n; \alpha}$$

2.4. NBUE- und NWUE-Verteilungen

2.4.1. Die Verteilungsklassen NBUE und NWUE

Bei NBUE und NWUE-Verteilungen wird die Forderung „neu besser (schlechter) als alt“ abgeschwächt zu „neu im Mittel besser (schlechter) als alt“.

Bei einer **NBUE (New Better than Used in Expectation)-Verteilung** ist also die erwartete Lebensdauer einer neuen Einheit größer als die erwartete Restlebensdauer einer intakten Einheit des Alters $t > 0$. Eine **NWUE (New Worse than Used in Expectation)-Lebensdauerverteilung** ist im Gegensatz dazu dadurch charakterisiert, daß die erwartete Lebensdauer einer neuen Einheit kleiner ist als die erwartete Restlebensdauer einer intakten Einheit des Alters $t > 0$.

Ist also F eine Verteilungsfunktion mit existierendem endlichen Erwartungswert μ und ist für $t \geq 0$

$$\int_t^{\infty} \bar{F}(x) dx \leq \mu \bar{F}(t),$$

so liegt eine NBUE-Verteilung vor; ist für $t \geq 0$

$$\int_t^{\infty} \bar{F}(x) dx \geq \mu \bar{F}(t),$$

so liegt eine NWUE-Verteilung vor (Hier muß μ nicht unbedingt endlich sein).

2.4.2. NBUE- und NWUE-Tests: Der Hollander-Proshan-Test

Zum Testen der Hypothese

$$H_0: \text{NWUE} \text{ gegen } H_1: \text{NBUE (und nicht exponential)}$$

bzw.

$$H'_0: \text{NBUE} \text{ gegen } H'_1: \text{NWUE (und nicht exponential)}$$

kann man ausgehend von den Lebensdauern x_1, \dots, x_n von n Objekten aus der interessierenden Grundgesamtheit den auf Hollander/Proshan (1975) zurückgehenden **Hollander-Proshan-Test** für NBUE- und NWUE-Alternativen verwenden. Man ordnet die n Lebensdauern der Größe nach:

$$0 = x_{(0)} \leq x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

und berechnet die normalisierten Abstände

$$D_j = (n - j + 1) (x_{(j)} - x_{(j-1)}) \quad \text{für } j = 1, \dots, n.$$

Die Hypothese H_0 wird dann verworfen, wenn gilt

$$K_n^* = \frac{\sum_{j=1}^n \left(\frac{3n+1}{2} - 2j \right) x_{(j)}}{n \sum_{j=1}^n D_j} \geq \sqrt{\frac{1}{12n}} u_{1-\alpha}$$

und H'_0 wird verworfen, falls gilt

$$K_n^* \leq \sqrt{\frac{1}{12n}} u_\alpha.$$

Dabei bezeichnet u_γ das γ -Quantil der Standardnormalverteilung und α gibt das Niveau der Tests an. Zwischen der Teststatistik K_n^* und der cttot-Teststatistik K_n aus Abschnitt 2.1.1 B besteht die Beziehung

$$K_n^* = \frac{1}{n} \left(K_n - \frac{n-1}{2} \right).$$

Beispiel: (vgl. Bryson/Siddiqui (1969)).

Tab. 10: Geordnete Überlebenszeiten $x_{(j)}$ von $n = 43$ Patienten und zugehörige normalisierte Abstände D_j

(j)	$x_{(j)}$	D_j	(j)	$x_{(j)}$	D_j
0	0	—	22	702	1144
1	7	301	23	715	273
2	47	1680	24	779	1280
3	58	451	25	881	1938
4	74	640	26	900	342
5	177	4017	27	930	510
6	232	2090	28	968	608
7	273	1517	29	1077	1635
8	285	432	30	1109	448
9	317	1120	31	1314	2665
10	429	3808	32	1334	240
11	440	363	33	1367	363
12	445	160	34	1534	1670
13	455	310	35	1712	1602
14	468	390	36	1784	576
15	495	783	37	1877	651
16	497	56	38	1886	54
17	532	945	39	2045	795
18	571	1014	40	2056	44
19	579	200	41	2260	612
20	581	48	42	2429	338
21	650	1587	43	2509	80

In Tab. 10 sind die geordneten Überlebenszeiten $x_{(j)}$ (in Tagen) von $n = 43$ an chronischer granuloytischer Leukämie erkrankten Personen angegeben. Als Zeitpunkt „0“ wurde hier der Behandlungsbeginn gewählt. Außerdem sind in Tab. 10 die normalisierten Abstände D_j angegeben. Wir wollen nun die Hypothese

H_0 : NWUE gegen H_1 : NBUE (und nicht exponential)

über die Überlebenszeiten der Patienten zum 10% Niveau testen. Es ist

$$\begin{aligned} K_{43}^* &= \frac{1}{43} \left(K_{43} - \frac{42}{2} \right) = \frac{1}{43} \left(\frac{\sum_{j=1}^{42} (43-j) D_j}{\sum_{j=1}^{43} D_j} - 21 \right) \\ &= \frac{1}{43} \left(\frac{959008}{39780} - 21 \right) \\ &= 0,07227 > 0,0564 = \sqrt{\frac{1}{12 \cdot 43}} u_{0,90} = \sqrt{\frac{1}{12n}} u_{1-\alpha}, \end{aligned}$$

so daß wir die Hypothese der NWUE-verteilten Überlebenszeiten zu Gunsten von NBUE verwerfen müssen; d. h. es ist zum Niveau $\alpha = 0,10$ signifikant, daß neu in die Behandlung eintretende Patienten zu Behandlungsbeginn eine höhere Überlebenserwartung haben als nach einer gewissen Behandlungsdauer; die Behandlung hat also nur bedingt Erfolg.

2.5. Beziehungen zwischen den Verteilungsklassen

Jede IFR-Verteilung ist auch eine IFRA-Verteilung, jede IFRA-Verteilung eine NBU-Verteilung und jede NBU-Verteilung eine NBUE-Verteilung. Die Klasse der NBUE-Verteilungen ist somit in dieser Gruppe am umfassendsten.

Jede DFR-Verteilung ist eine DFRA-Verteilung, jede DFRA-Verteilung eine NWU-Verteilung und jede NWU-Verteilung eine NWUE-Verteilung.

Die Exponentialverteilungen sind die einzigen Verteilungen die allen Klassen gleichzeitig angehören.

Abb. 25 zeigt die Überlebenswahrscheinlichkeit einer IFR-Verteilung, **Abb. 26**

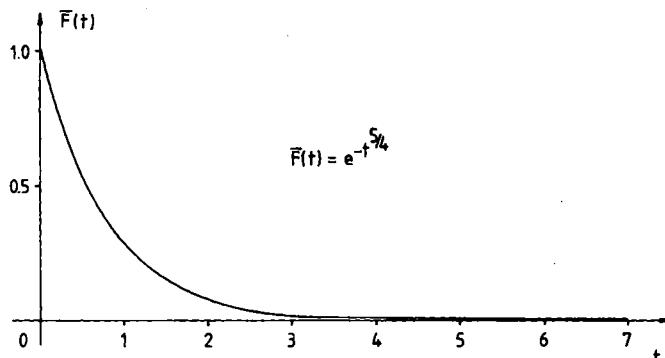


Abb. 25: Überlebenswahrscheinlichkeit einer IFR-Verteilung

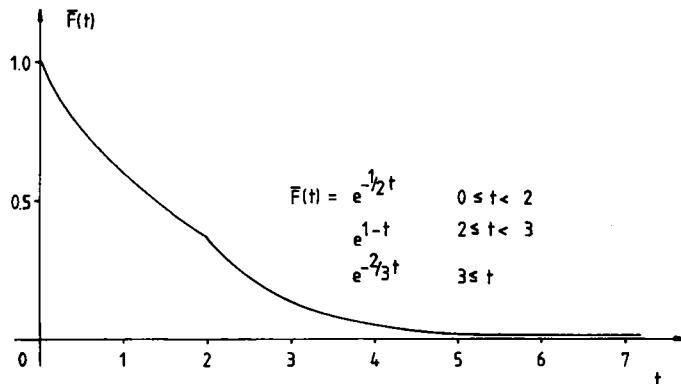


Abb. 26: Überlebenswahrscheinlichkeit einer IFRA-Verteilung, die nicht die IFR-Eigenschaft besitzt

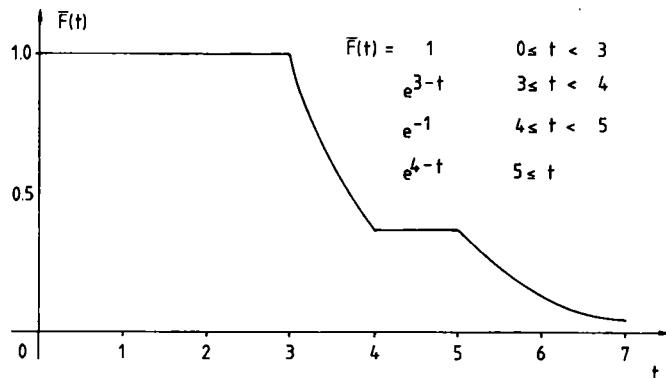


Abb. 27: Überlebenswahrscheinlichkeit einer NBU-Verteilung, die nicht die IFRA-Eigenschaft besitzt

die einer IFRA-Verteilung, die nicht IFR-Verteilung ist, und Abb. 27 zeigt eine NBU-Verteilung, die keine IFRA-Verteilung ist.

2.6. Zensierte Lebensdauerprüfungen

Ein wichtiges Problem bei Lebensdaueruntersuchungen resultiert aus der Tatsache, daß man oft nicht bis zum Ausfall des letzten der Lebensdauerprüfungen unterworfenen Elementes abwarten möchte bzw. kann. Der Grund hierfür liegt insbesondere in der mitunter hohen Lebensdauer der Elemente. Andererseits können finanzielle Gründe gegen eine Zerstörung aller zu untersuchenden Elemente sprechen.

Für exponentialverteilte Lebensdauern etwa nimmt die erwartete Zeit bis zum Ausfall des letzten Elementes einer Stichprobe von n Elementen mit wachsendem n wie $\ln(n)$ zu. Häufig geht man aus diesem Grunde zu zensierten Lebensdauerprüfungen über. Eine **zensierte (censored) Lebensdaueruntersuchung** ist eine Prüfung, die nach dem Ausfall des k -ten von insgesamt n Elementen abgebrochen wird. Von einer **zeitzensierten (truncated, time-censored) Prüfung** spricht man, wenn die Untersuchung spätestens zu einem festgelegten Zeitpunkt abgebrochen wird. Sie wird

schon vorher abgebrochen, wenn bereits zuvor eine festgelegte Zahl k von Elementen ausgefallen ist.

Zensierte Lebensdauerprüfungen sind vor allem dann interessant, wenn Parameter einer Verteilung geschätzt werden sollen, vgl. Abschnitt 3 dieses Kapitels und Abschnitt 4 in Kap. IV.

Für IFR(A)- und DFR(A)-Tests bei zensierten Lebensdauerprüfungen läßt sich als Teststatistik weiterhin die ctot-Teststatistik K_n verwenden; vgl. hierzu Barlow/Proschan (1969), Barlow et al. (1972).

3. Punkt- und Intervallschätzungen für die Parameter einiger spezieller Lebensdauerverteilungen

Bei der Untersuchung der Lebensdauer T in einer Grundgesamtheit gleichartiger Objekte (Einheiten) kennt man in der Regel die Parameter der Verteilung von T nicht und muß sie daher aufgrund einer Stichprobe von n Objekten schätzen.

Wir wollen uns hier speziell mit der Punkt- und Intervallschätzung der Parameter exponential-, Weibull-, lognormal- und Hjorth-verteilter Lebensdauern beschäftigen. Einige der angegebenen Punkt- und Konfidenzschätzer sind uns bereits in Kap. IV begegnet, sollen der Vollständigkeit halber hier aber noch einmal aufgeführt werden.

3.1. Das Modell exponentialverteilter Lebensdauern

Die Bedeutung der Exponentialverteilung bei der Beschreibung von Lebensdauern wurde in den vorausgegangenen Abschnitten schon deutlich. Exponentialverteilungen sind die einzigen Verteilungen mit konstanter Ausfallrate, und der Parameter λ einer Exponentialverteilung ist, vgl. auch Abschnitt 4.1 in Kap. IV, identisch mit dieser Ausfallrate.

Will man den Parameter λ der in einer Grundgesamtheit exponentialverteilten Lebensdauer schätzen, so wird man zunächst n Objekte zufällig aus dieser Grundgesamtheit für ein Experiment auswählen und deren Lebensdauern beobachten.

Aus Zeit- und Kostengründen wird es in der Praxis oft nicht möglich sein, abzuwarten, bis auch das letzte der n Objekte ausfällt. Daher führt man sogenannte **zensierte Lebensdauer-Experimente** durch: Zum einen kann man vor Beginn des Experimentes die Gesamtbeobachtungszeit T_0 festlegen, zum anderen ist es möglich, die Zahl der abzuwartenden Ausfälle festzusetzen.

In beiden Fällen unterscheidet man solche Experimente, bei denen ausgefallene Objekte durch andere (aus der Grundgesamtheit) ohne Zeitverzögerung ersetzt werden, und solche, bei denen das Ersetzen unterbleibt.

3.1.1. Punkt- und Intervallschätzungen bei fest vorgegebener Beobachtungsduer T_0

A. Experimente mit Ersetzung ausgefallener Objekte

Zunächst wollen wir auf den Fall eingehen, daß ausgefallene Objekte sofort durch neue ersetzt werden. Treten im Beobachtungszeitraum $[0; T_0]$ dann insgesamt k Ausfälle ein, so ist, da k hier eine poissonverteilte Zufallsvariable mit Erwartungswert $n \cdot \lambda \cdot T_0$ beschreibt,

$$\hat{\lambda} = \frac{k}{n \cdot T_0}$$

ein erwartungstreuer Schätzer für die Ausfallrate λ in der Grundgesamtheit. Ein $1 - \alpha - \beta$ Konfidenzintervall für λ ist dann

$$\left[\frac{\chi^2_{2k; \beta}}{2n \cdot T_0}; \frac{\chi^2_{2k+2; 1-\alpha}}{2n \cdot T_0} \right],$$

wobei $\chi^2_{v; \gamma}$ das γ -Quantil der χ^2 -Verteilung mit v Freiheitsgraden bezeichnet, vgl. Tabelle im Anhang. Es sei noch erwähnt, daß es bei dieser Art von Experiment natürlich möglich ist, daß $k > n$ gilt, d. h. daß die Anzahl der Ausfälle im Beobachtungszeitraum größer ist als die Anzahl gleichzeitig funktionierender Objekte.

B. Experimente ohne Ersetzung ausgefallener Objekte

Ersetzt man ausgefallene Objekte nicht, so kommt man zur Schätzung von λ , wie sie im Abschnitt 4.1 des Kap. IV ausführlich dargestellt wurde. Fallen bis zum Zeitpunkt T_0 gerade k der insgesamt n beobachteten Objekte aus, so ist

$$\hat{\lambda} = \frac{1}{T_0} \ln \frac{n}{n-k}$$

ein Punktschätzer für den Parameter λ und

$$\left[\frac{1}{T_0} \ln \left(1 + \frac{k \cdot F_{2k, 2(n-k+1); \beta}}{n-k+1} \right); \frac{1}{T_0} \ln \left(1 + \frac{(k+1) \cdot F_{2(k+1), 2(n-k); 1-\alpha}}{n-k} \right) \right]$$

ein Konfidenzintervall zum Niveau $1 - \alpha - \beta$ für λ . Hier bezeichnet $F_{v_1, v_2; \gamma}$ das γ -Quantil der F_{v_1, v_2} -Verteilung; diese Quantile sind im Anhang vertafelt.

3.1.2. Punkt- und Intervallschätzungen bei fest vorgegebener Zahl von Ausfällen

Legt man anstelle der Dauer des Experiments die Anzahl k der abzuwartenden Ausfälle fest, so bestimmt man Punkt- und Intervallschätzungen für den Parameter λ einer Exponentialverteilung aufgrund der Ausfallzeitpunkte t_1, t_2, \dots, t_k .

A. Experimente mit Ersetzung ausgefallener Objekte

Ersetzt man ausgefallene Objekte sofort durch neue Objekte aus der interessierenden Grundgesamtheit und setzt man $t_0 = 0$, so sind die Abstände zwischen zwei Ausfallzeitpunkten t_j, t_{j-1} , für $j = 1, \dots, k$, exponentialverteilt mit Erwartungswert $1/(n\lambda)$. Daher läßt sich $1/(n\lambda)$ erwartungstreu durch

$$\frac{1}{k} \sum_{j=1}^k (t_j - t_{j-1}) = \frac{t_k}{k}$$

schätzen. Somit ist

$$\hat{\lambda} = \frac{k}{n \cdot t_k}$$

ein Punktschätzer für den Parameter λ der exponentialverteilten Lebensdauer in der Grundgesamtheit, und

$$\left[\frac{\chi^2_{2k; \beta}}{2n \cdot t_k}; \frac{\chi^2_{2k+2; 1-\alpha}}{2n \cdot t_k} \right]$$

ist ein $1 - \alpha - \beta$ Konfidenzintervall für die Ausfallrate λ .

B. Experimente ohne Ersetzung ausgefallener Objekte

Werden nun ausgefallene Objekte nicht durch neue ersetzt, so ist

$$\sum_{j=1}^k t_j + (n - k) \cdot t_k = T_{akk}$$

die **akkumulierte Beobachtungszeit** für alle n Objekte in der Stichprobe und

$$\hat{\lambda} = \frac{k}{(n - k)t_k + \sum_{j=1}^k t_j} = \frac{k}{T_{akk}}$$

ist eine Punktschätzung für den Parameter λ . Ein Konfidenzintervall für λ zum Niveau $1 - \alpha - \beta$ ergibt sich hier nun zu

$$\left[\frac{\chi^2_{2k; \beta}}{2T_{akk}}, \frac{\chi^2_{2k+2; 1-\alpha}}{2T_{akk}} \right].$$

Beispiel: $n = 50$ Objekte aus einer Grundgesamtheit mit exponentialverteilter Lebensdauer werden für ein Experiment ausgewählt. Ausgehend von den $k = 10$ ersten Ausfallzeitpunkten in der Stichprobe (ausgefallene Objekte werden nicht ersetzt) sollen Punkt- und Intervallschätzer für den Parameter λ der Exponentialverteilung bestimmt werden. Die 10 Ausfallzeiten (in Stunden) sind in **Tab. 11** angegeben.

Tab. 11: Ausfallzeiten (Lebensdauern) von 10 Objekten

Objekt j	1	2	3	4	5	6	7	8	9	10
Ausfallzeit t_j	58	64	137	194	302	376	457	501	709	849

Mittels dieser Daten ergibt sich als Punktschätzung für den Parameter λ gerade

$$\hat{\lambda} = \frac{k}{T_{akk}} = \frac{10}{(50 - 10)t_{10} + \sum_{j=1}^{10} t_j} = \frac{10}{40 \cdot 849 + 3647} = \frac{10}{37607} = 0,000266.$$

Die Ausfallrate wird also mit $2,66 \cdot 10^{-4}$ Objekte pro Stunde geschätzt.

Als Konfidenzintervall zum Niveau $1 - \alpha - \beta = 1 - 0,05 - 0,05 = 0,90$ für λ erhält man hier

$$\left[\frac{\chi^2_{2 \cdot 10; 0,05}}{2 \cdot 37607}, \frac{\chi^2_{2 \cdot 10 + 2; 0,95}}{2 \cdot 37607} \right] = \left[\frac{10,85}{75214}, \frac{33,92}{75214} \right] = [0,000144; 0,000451] = [1,44 \cdot 10^{-4}; 4,51 \cdot 10^{-4}].$$

3.2. Das Modell Weibull-verteilter Lebensdauern

Ist man vornehmlich daran interessiert, die mittlere Lebensdauer von Objekten aus einer Grundgesamtheit zu schätzen, so empfiehlt es sich, vom Ansatz exponentialverteilter Lebensdauern Abstand zu nehmen und statt dessen eine Weibull-Verteilung mit der Dichte

$$f(t) = \alpha \cdot \beta \cdot t^{\beta-1} \cdot e^{-\alpha t^\beta} \quad \text{für } t > 0, \alpha > 0 \text{ und } \beta > 0$$

der Lebensdauern zugrunde zu legen (als Spezialfall ($\beta = 1$) ist hier die Exponentialverteilung enthalten).

Der Erwartungswert einer solchen Weibull-Verteilung ist, vgl. Abschnitt 4.2 in Kap. IV,

$$\alpha^{-1/\beta} \cdot \Gamma\left(1 + \frac{1}{\beta}\right),$$

wobei $\Gamma(x)$ die Gamma-Funktion an der Stelle x bezeichnet, vgl. Abschnitt 1.2 in Kap. IV und Tabelle im Anhang.

Werden in einem Experiment ausgefallene Objekte, die zufällig aus einer Grundgesamtheit mit Weibull-verteilter Lebensdauer ausgewählt wurden, nicht sofort durch neue ersetzt, so können die Schätzer für die Parameter α und β nicht direkt angegeben werden. Für Experimente, bei denen ausgefallene Objekte sofort durch heiße (genauso alte, jedoch noch funktionsfähige) Objekte ersetzt werden, kann man ein von Störmer (1961) angegebenes Verfahren zur Ermittlung der Schätzwerte für α und β verwenden, auf das wir hier nicht näher eingehen wollen.

Vielmehr soll hier ein schnelles und anschauliches Verfahren, das relativ häufig verwendet wird, zur Schätzung von α und β herangezogen werden.

3.2.1. Schätzen der Parameter α und β bei fest vorgegebener Anzahl von Ausfällen ohne Ersetzung

n aus einer Grundgesamtheit zufällig ausgewählte Objekte werden einer Lebensdauerprüfung unterzogen. Sind t_1, \dots, t_k die Ausfallzeitpunkte der ersten k ausgefallenen Objekte, so ist

$$\widehat{F}(t_j) = 1 - \frac{j-1/2}{n} \quad \text{für } j = 1, \dots, k$$

ein Schätzer für die Überlebenswahrscheinlichkeit in der Grundgesamtheit an der Stelle t_j , wenn die ausgefallenen Objekte nicht durch neue ersetzt werden. Da gilt

$$\ln \ln \widehat{F}(t)^{-1} = \ln \alpha + \beta \ln t,$$

können ausgehend von den k Werten $\widehat{F}(t_j)$ die Parameter α und β der zugrundeliegenden Weibullverteilung mit Methoden der linearen Regressionsrechnung, vgl. Kap. IX, Abschnitt 1, geschätzt werden. Mit

$$x_j = \ln t_j, \quad y_j = \ln \ln \widehat{F}(t_j)^{-1}$$

ergibt sich aus

$$y_j = \ln \alpha + \beta x_j + e_j \quad \text{für } j = 1, \dots, k,$$

mit Fehlertermen e_1, \dots, e_k , nach der Methode der kleinsten Quadrate als Schätzer für β der Wert

$$\hat{\beta} = \frac{\sum_{j=1}^k (y_j - \bar{y})(x_j - \bar{x})}{\sum_{j=1}^k (x_j - \bar{x})^2}$$

und als Schätzer für α gerade

$$\hat{\alpha} = e^{\bar{y} - \hat{\beta}\bar{x}}.$$

Beispiel: Um die Verwendungsfähigkeit einer Raketenkomponente zu überprüfen, soll die mittlere Lebensdauer dieser Komponente geschätzt werden. Hierzu werden aus der Grundgesamtheit aller solcher gleichartiger Komponenten 50 Stück zufällig für ein Laborexperiment, in dem spezielle Einsatzbedingungen simuliert werden, ausgewählt. Die ausfallenden Komponenten werden nicht ersetzt, und der Versuch wird nach dem 10. Ausfall abgebrochen. In Tab. 12 sind die $k = 10$ Ausfallzeitpunkte t_j , die geschätzten Überlebenswahrscheinlichkeiten $\hat{F}(t_j)$ sowie die transformierten Größen $x_j = \ln t_j$ und $y_j = \ln \ln \hat{F}(t_j)^{-1}$ eingetragen.

Tab. 12: Ausfallzeitpunkte von $k = 10$ Raketenkomponenten und Hilfsgrößen für die Schätzung der Parameter α und β

j	$\hat{F}(t_j)$	t_j	x_j	y_j
1	0,99	18	2,890	-4,600
2	0,97	36	3,584	-3,491
3	0,95	40	3,689	-2,970
4	0,93	53	3,970	-2,623
5	0,91	71	4,263	-2,361
6	0,89	90	4,500	-2,150
7	0,87	106	4,663	-1,971
8	0,85	127	4,844	-1,817
9	0,83	149	5,004	-1,680
10	0,81	165	5,106	-1,557

Legt man für die Lebensdauer der Raketenkomponenten in der Grundgesamtheit eine Weibullverteilung mit Parametern α und β zugrunde, so kann man zunächst diese Parameter mittels der in Tab. 12 angegebenen Hilfsgrößen schätzen. Es ergibt sich

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{j=1}^{10} (y_j - \bar{y})(x_j - \bar{x})}{\sum_{j=1}^{10} (x_j - \bar{x})^2} = \frac{\sum_{j=1}^{10} x_j y_j - 10 \bar{x} \bar{y}}{\sum_{j=1}^{10} x_j^2 - 10 \bar{x}^2} \\ &= \frac{-101,26441 - 10 \cdot 4,2513 \cdot (-2,5220)}{185,3091 - 10 \cdot 4,2513^2} \\ &= \frac{-101,26441 + 107,21779}{185,3091 - 180,73552} = 1,302 \end{aligned}$$

und somit

$$\hat{\alpha} = e^{\bar{y} - \hat{\beta}\bar{x}} = e^{-2,5220 - 1,302 \cdot 4,2513} = e^{-8,057} = 0,00032.$$

Die mittlere Lebensdauer der Raketenkomponente kann dann durch

$$\begin{aligned}\hat{\alpha}^{-1/\hat{\beta}} \cdot \Gamma(1 + 1/\hat{\beta}) &= (0,00032)^{-1/1,302} \cdot \Gamma(1 + 1/1,302) \\ &= 483,2998 \cdot \Gamma(1,7680) = 483,2998 \cdot 0,923 \\ &= 446,086\end{aligned}$$

geschätzt werden. Die Ausfallrate

$$r(t) = \hat{\alpha} \hat{\beta} t^{\hat{\beta}-1}$$

kann nun für die Raketenkomponenten geschätzt werden:

$$r(t) = \hat{\alpha} \cdot \hat{\beta} \cdot t^{\hat{\beta}-1} = 0,00032 \cdot 1,302 \cdot t^{0,302} = 0,00041664 \cdot t^{0,302}.$$

Nach einer Minute ($t = 1$) fallen die Komponenten mit der Rate 0,0004163 pro Minute aus, wohingegen nach 1000 Minuten ($t = 1000$) die Ausfallrate auf 0,00336 pro Minute angewachsen ist.

3.2.2. Schätzen der Parameter α und β bei unzensierten Lebensdauerprüfungen

Beobachtet man in einem Experiment n Objekte mit Weibull-verteilter Lebensdauer so lange, bis auch das letzte Objekt ausgefallen ist, so lassen sich aus dem Gleichungssystem

$$n - \hat{\alpha} \cdot \sum_{j=1}^n t_j^{\hat{\beta}} = 0$$

$$\frac{n}{\hat{\beta}} + \sum_{j=1}^n \ln t_j - \hat{\alpha} \sum_{j=1}^n t_j^{\hat{\beta}} \ln t_j = 0$$

Maximum-Likelihood-Schätzer für die Parameter α und β bestimmen; hier bezeichnen t_1, \dots, t_n die Lebensdauern der n beobachteten Objekte. Dieses Gleichungssystem lässt sich nur mittels numerischer Verfahren lösen, vgl. Himmelblau (1972), Blum/Oettli (1975), Stoer (1979). Als Startwert $(\hat{\alpha}_0, \hat{\beta}_0)$ für iterative Verfahren wählt man $\hat{\beta}_0$ als Lösung der Gleichung

$$\frac{1}{n} \sum_{j=1}^n t_j = \left(\frac{1}{\hat{\alpha}_0} \right)^{1/\hat{\beta}_0} \cdot \Gamma \left(1 + \frac{1}{\hat{\beta}_0} \right), \quad \text{wobei } \hat{\alpha}_0 = \frac{n}{\sum_{j=1}^n t_j^{\hat{\beta}_0}} \text{ gesetzt wird.}$$

3.3. Das Modell der Lognormalverteilung

Falls Objekte aus einer Grundgesamtheit lognormalverteilte Lebensdauern besitzen, so bedeutet dies, daß die natürlichen Logarithmen der Lebensdauern normalverteilt sind, vgl. Kap. IV. Ist also die Lebensdauer T lognormalverteilt, gemäß

$$\ln T \sim N(\mu, \sigma^2),$$

so kann man Varianz und Erwartungswert dieser lognormalverteilten Lebensdauern schätzen. Dazu wählt man n Objekte zufällig aus der interessierenden Grundgesamtheit aus und beobachtet in einem Experiment deren Ausfallzeitpunkte t_1, \dots, t_n . Dann schätzt man Erwartungswert und Varianz der logarithmierten Ausfallzeitpunkte mittels der in Kap. IV, Abschnitt 1.3, angegebenen Methoden. Ein Schätzer für μ ist etwa

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n \ln t_j$$

und ein Schätzer für σ^2 ist

$$s^2 = \frac{1}{n-1} \sum_{j=1}^n \left(\ln t_j - \frac{1}{n} \sum_{j=1}^n \ln t_j \right)^2.$$

Es sind dann $e^{\bar{x} + s^2/2}$ bzw. $(e^{s^2} - 1)e^{2\bar{x} + s^2}$ Schätzer für Mittelwert und Varianz der lognormalverteilten Lebensdauern in der Grundgesamtheit.

Die Ausfallrate der Lognormalverteilung ist nicht monoton steigend; sie eignet sich daher nicht zur Beschreibung von Alterserscheinungen. Bei Überlebenszeiten nach extremen Belastungen (etwa Giftgabe an lebende Organismen) zum Beispiel ist die Lognormalverteilung häufig die angemessene Verteilung; vgl. auch S. 151.

3.4. Das Modell der Hjorth-Verteilung (IDB-Verteilung)

Die Hjorth-Verteilung, vgl. Hjorth (1980), wurde in Abschnitt 4.3 des Kap. IV ausführlich beschrieben. Da ihre Ausfallrate

$$r(t) = \alpha t + \frac{\gamma}{1 + \beta t}$$

(für wachsendes t) aus einem wachsenden und einem fallenden Teil zusammengesetzt ist, eignet sich die Hjorth-Verteilung zur Beschreibung wachsender, fallender, konstanter und **badewannenförmiger Ausfallraten**. Methoden zur Bestimmung der Maximum-Likelihood-Schätzer für die Parameter α , β und γ sind in Kap. IV schon angegeben worden.

Gegenüber dem Weibull-Modell zur Beschreibung von Lebensdauern hat die Hjorth-Verteilung den Vorteil, daß sich mit ihr eben auch Lebensdauern mit badewannenförmiger Ausfallrate modellieren lassen:

Ergibt eine auf dem Weibull-Modell basierende Untersuchung, vgl. Abschnitt 3.3, eine fallende Ausfallrate, so wird man folgern, daß ein zwar altes, aber noch intaktes Objekt nicht durch ein neues zu ersetzen ist. Das Hjorth-Modell hingegen vermag auch später einsetzende Verschleißerscheinungen zu erkennen, was zu gelegentlichen, jedoch oft realistischeren Erneuerungsmaßnahmen führen kann. Die praktische Bedeutung des Hjorth-Modells wird erst bei solchen Wartungs- und Erneuerungsüberlegungen voll sichtbar, vgl. Abschnitt 5.

4. Zur Problematik zeitraffender Zuverlässigsprüfungen

Bei den in Abschnitt 2 vorgestellten Tests zur Überprüfung der Zuordnung einer Lebensdauerverteilung zu einer der behandelten Verteilungsklassen treten Schwierigkeiten bei ihrer Anwendung auf, wenn die Objekte (Beobachtungseinheiten) in der interessierenden Grundgesamtheit extrem langlebig sind. Bei Halbleitern etwa muß man von einer mittleren Lebensdauer von mehreren Jahrhunderten ausgehen; natürlich ist es unter diesen Umständen unmöglich, in einem gewöhnlichen Experiment die Ausfallzeitpunkte von einigen zufällig ausgewählten Halbleitern zu beobachten, um Aussagen über die Lebensdauerverteilung solcher Halbleiter zu gewinnen.

Ähnliche Probleme treten bei der experimentellen Bestimmung von Zuverlässigkeitscharakteristika (wie z. B. mittlere Lebensdauer und Ausfallrate) elektronischer Geräte auf.

Es sei hier noch erklärend erwähnt, daß die Ausfallrate in einer Grundgesamtheit aus der sogenannten **Ausfallquote** geschätzt wird, wenn die Lebensdauerverteilung $F(t)$ in der Grundgesamtheit unbekannt ist. In solchen Fällen wollen wir die Ausfallrate kurz mit λ bezeichnen. Der Schätzwert für λ ist dann (im Zeitintervall $[t_j; t_{j+1}]$)

$$q(\Delta t_j, t_j) = \frac{(\text{Bestand zum Zeitpunkt } t_j) - (\text{Bestand zum Zeitpunkt } t_{j+1})}{(\text{Bestand zum Zeitpunkt } t_j) \cdot \Delta t_j}$$

mit $\Delta t_j = t_{j+1} - t_j$. Die (geschätzte) Ausfallquote wird in Ausfälle pro Stunde oder pro 1000 Stunden angegeben. Die Ausfallquote wird auch als **empirische Ausfallrate** bezeichnet, vgl. auch Kap. I, Abschnitt 8.

Beispiel: Der Bestand an Objekten mit unbekannter Lebensdauerverteilung ist 50 zum Zeitpunkt $t_0 = 0$ und 48 zum Zeitpunkt $t_1 = 150$ (Stunden). Die Ausfallrate wird dann wegen

$$\Delta t_0 = 150 - 0 = 150$$

durch die Ausfallquote

$$q(\Delta t_0, t_0) = q(150, 0) = \frac{50 - 48}{50 \cdot 150} = \frac{2}{750} = 0,00027$$

geschätzt; d.h. die Ausfallquote (in den ersten 150 Stunden) ist $2,7 \cdot 10^{-4}$ pro Stunde. Leben zum Zeitpunkt $t_2 = 500$ Stunden noch 40 der ursprünglich 50 Objekte, so wird die Ausfallrate im Zeitintervall [150; 500] durch

$$q(\Delta t_1, t_1) = q(t_2 - t_1, t_1) = q(350, 150) = \frac{48 - 40}{48 \cdot 350} = 0,00048$$

geschätzt, d.h. der Schätzer für die Ausfallrate zwischen 150. und 500. Stunde ist $4,8 \cdot 10^{-4}$ Objekte pro Stunde.

Ein weiteres Beispiel für den oft enormen Aufwand bei Lebensdaueruntersuchungen lieferte Tomášek (1971). Die Ablehnung der Hypothese $\lambda \leq 10^{-7}$ pro Stunde bei Silicium-Transistoren ist zum Niveau 0,2 überhaupt nur möglich, wenn bei einer Experimentierdauer von 16000 Stunden mindestens 1000 Transistoren beobachtet werden. Erhöht man die Dauer des Experiments auf 160000 Stunden, so müssen immerhin noch mehr als 100 Transistoren für das Experiment ausgewählt werden.

Häufig ist es also so, daß sich Zuverlässigkeitsprüfungen entweder unökonomisch oder zeitraubend gestalten. Um die so entstehenden Probleme zu bewältigen, ist es oft unerlässlich, sogenannte zeitraffende Tests (Experimente) zu verwenden, deren Sinn in einer **Beschleunigung des Alterungsprozesses** von Objekten besteht. Aufgabe des Prüftechnikers ist es, solche Kurzzeittests zu entwickeln, die den langfristigen Alterungsprozeß gut simulieren; vgl. auch Kap. IX, Abschnitt 2.1.

Die einfachste Möglichkeit zur Beschleunigung eines Alterungsvorganges ist die Erhöhung der Prüfungstemperatur. Für Halbleiterelemente etwa erreicht man durch eine Prüftemperatur von 70°C , daß

- die Löslichkeit der Materialien abnimmt,
- die mechanische Stabilität geringer wird,
- die Eigenleitfähigkeit zunimmt,
- die Arbeitspunktstabilität abnimmt,
- die chemische Reaktionsgeschwindigkeit zunimmt,

vgl. Schäfer (1980). Für Transistoren und integrierte Schaltungen sind, vgl. Werner (1980), als Auswirkungen erhöhter Temperatur und Stromdichten zu nennen

- Materialermüdung,
- Interdiffusion benachbarter Metallschichten,
- Materialwanderung (Elektro-Migration, Thermo-Migration),
- Korrosion,
- Schubkräfte infolge thermischer Fehlanpassung (unterschiedliche Ausdehnungskoeffizienten).

Das Problem besteht nun darin, eine Extrapolation der zeitraffenden Prüfungen an normale Lebensbedingungen zu gewinnen.

4.1. Die Extrapolation der zeitraffenden Prüfung an Normalbedingungen

4.1.1. Das Eyring-Modell

Da thermische Beeinflussungen Oxydationsprozesse hervorrufen, kann man zur näherungsweisen Extrapolation die aus den Prinzipien der Quantenmechanik ableitete Eyring-Gleichung

$$\frac{dM}{dt} = C \cdot T \cdot e^{-\frac{E_a}{kT}}, \text{ wobei}$$

M die zur Verfügung stehende Reaktionsmenge,

T die absolute Temperatur in °Kelvin ($273,15 + ^\circ\text{C}$),

C eine Konstante,

E_a die Aktivierungsenergie in Joule (J) oder Elektronenvolt (eV) und

k die Boltzmann-Konstante [$1,38 \cdot 10^{-23} \text{ J}/^\circ\text{K}$ oder $8,63 \cdot 10^{-5} \text{ eV}/^\circ\text{K}$]

bezeichnet, verwenden. $\frac{dM}{dt}$ ist die spezifische Geschwindigkeit der chemischen Reaktion; $\frac{1}{k} E_a$ heißt auch Aktivierungstemperatur.

4.1.2. Das Arrhenius-Modell

Einfacher gestaltet sich die Beschreibung der Reaktionsgeschwindigkeit (in Abhängigkeit von der Temperatur) durch die **Arrhenius-Gleichung**, die aufgrund einer Analogie aus der Thermodynamik abgeleitet worden ist. Strenggenommen hat diese nur für wässrige Lösungen Gültigkeit; sie findet jedoch auch in anderen Fällen ihre Anwendung, da sie sich als gute Näherung (ca. 10% Abweichungsfehler) erweist. Die Arrhenius-Gleichung lautet

$$\frac{dM}{dt} = A \cdot e^{-\frac{E_a}{k \cdot T}}, \text{ wobei } A \text{ eine Konstante ist.}$$

Bei Zuverlässigskeitsbetrachtungen empfiehlt es sich, die zur Verfügung stehende Reaktionsmenge M zu quantifizieren:

Ist N die Anzahl der betrachteten Objekte zum Zeitpunkt t, so gilt

$$\frac{dM}{dt} = -\frac{1}{N} \frac{dN}{dt} = A \cdot e^{-\frac{E_a}{k \cdot T}},$$

und die Ausfallrate λ_0 ist (bei angenommener Exponentialverteilung) bei Temperatur T_0 in °Kelvin gerade

$$\lambda_0 = A \cdot e^{-\frac{E_a}{k \cdot T_0}}.$$

Will man nun ausgehend von einer Ausfallrate λ_0 bei Temperatur T_0 die Ausfallrate λ der Elemente in der Grundgesamtheit bei einer festgelegten Temperatur T bestimmen, so kann man die Gleichung

$$\lambda = \lambda_0 \cdot e^{\left(1 - \frac{T_0}{T}\right) \frac{E_a}{k \cdot T_0}} = \lambda_0 \cdot e^{\left(\frac{1}{T_0} - \frac{1}{T}\right) \frac{E_a}{k}}$$

verwenden.

Beobachtet man bei einem Objekt nun die Lebensdauer t_0 (t_0 kann auch die aufgrund des Tests bei Temperatur T_0 geschätzte mittlere Lebensdauer der Elemente aus der Grundgesamtheit sein) bei einer Prüftemperatur T_0 , so kann obige Gleichung in logarithmierter Form auch zur Schätzung dieser Lebensdauer t bei einer Temperatur T verwandt werden. Es ist

$$\ln \hat{t} = \ln t_0 - \left(\frac{1}{T_0} - \frac{1}{T} \right) E_a / k;$$

diese Gleichung beschreibt also eine lineare Regression, vgl. Kap. X, mit Steigungsparameter 1 und Absolutglied

$$c = - \left(\frac{1}{T_0} - \frac{1}{T} \right) E_a / k = -11587,486 \frac{\text{°K}}{\text{eV}} \cdot E_a \left(\frac{1}{T_0} - \frac{1}{T} \right).$$

Hat man $\ln \hat{t}$ berechnet, so ist natürlich

$$\hat{t} = e^{\ln \hat{t}}$$

der gesuchte Schätzwert für die Lebensdauer t bei Temperatur T .

(In einigen Bereichen ist es üblich, statt des hier verwandten Logarithmus zur Basis e den zur Basis 10 zu verwenden:

$$\log \hat{t} = \log t_0 - 5032 \frac{\text{°K}}{\text{eV}} \left(\frac{1}{T_0} - \frac{1}{T} \right) E_a, \quad \text{d.h.} \quad \hat{t} = 10^{\log \hat{t}}.)$$

Beispiel: Richard (1979) ermittelte für Transistoren eine mittlere Lebenszeit von $t_0 = 24000$ Stunden bei einer Testtemperatur von

$T_0 = 280^\circ\text{C} = (280 + 273)^\circ\text{K} = 553^\circ\text{K}$ und einer Aktivierungsenergie von $E_a = 1,27 \text{ eV}$.

Die Schätzung für die entsprechende mittlere Lebensdauer t bei einer Temperatur $T = 165^\circ\text{C} = 438^\circ\text{K}$ ist dann wegen

$$\begin{aligned} \ln \hat{t} &= \ln(24000) - 11587,486 \cdot 1,27 \left(\frac{1}{553} - \frac{1}{438} \right) \\ &= 10,085809 - (-6,987013) = 17,072822 \end{aligned}$$

gerade

$$\hat{t} = e^{\ln \hat{t}} = e^{17,072822} = 25979596.$$

Die mittlere Lebensdauer der Transistoren bei $T = 438^\circ\text{K}$ wird also mit 25,979596 Millionen Stunden geschätzt; das sind ca. 3000 Jahre.

Bei der Schätzung von \hat{t} wurde die Regressionsgleichung

$$\ln \hat{t} = \ln t_0 + c$$

benutzt. Die Größe

$$e^c = e^{-11587,486 \cdot E_a \left(\frac{1}{T_0} - \frac{1}{T} \right)} = B(T_0, T)$$

heißt auch **Beschleunigungsfaktor**. $B(T_0, T)$ ist der Faktor, der angibt, wieviel länger eine Lebenszeit t bei Temperatur T ist als eine Lebenszeit t_0 bei Temperatur T_0 :

$$t = t_0 B(T_0, T).$$

Beispiel: Im obigen Beispiel der Transistoren ist der Beschleunigungsfaktor bei einer Umrechnung von $T_0 = 553^\circ\text{K}$ in $T = 438^\circ\text{K}$ für $E_a = 1,27 \text{ eV}$ gerade

$$\begin{aligned} B(T_0, T) &= B(553, 438) = e^{-11587,486 \cdot 1,27 (1/553 - 1/438)} \\ &= 1082,4832. \end{aligned}$$

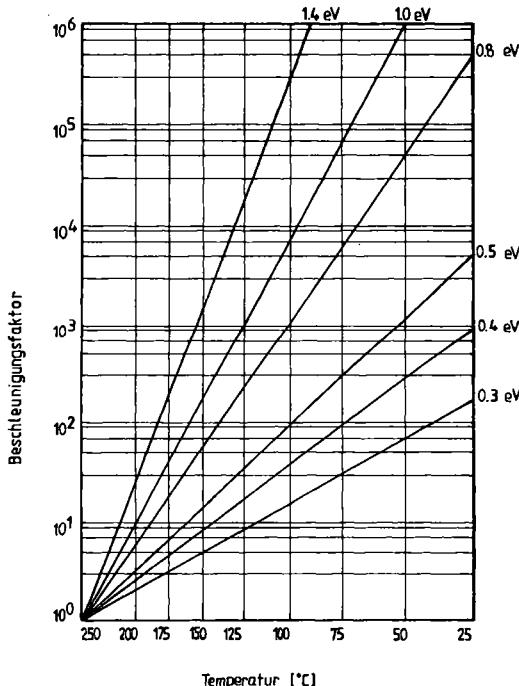


Abb. 28: Beschleunigungsfaktor $B(T_0, T)$ für $T_0 = 250^\circ\text{C}$ bei verschiedenen Aktivierungsenergien; Ablesbeispiel: Bei einer Aktivierungsenergie von 1 eV ist der Beschleunigungsfaktor gegenüber einer Temperatur $T = 125^\circ\text{C}$ ungefähr $10^3 = 1000$

Um einen Eindruck von der Größe des Beschleunigungsfaktors $B(T_0, T)$ zu gewinnen, ist er in Abb. 28 für $T_0 = 250^\circ\text{C} = 523^\circ\text{K}$ und für verschiedene Aktivierungsenergien E_a graphisch dargestellt.

Eine andere Methode (graphischer Art) zur Analyse von Zeitraffungstests auf der Basis des Arrhenius-Modells findet man bei Thomas (1965).

4.1.3. Das verallgemeinerte Eyring-Modell

Die anfangs erwähnte Eyring-Gleichung aus der Quantenmechanik kann dahingehend erweitert werden, daß ein nicht-thermaler Einfluß, z.B. die Spannung, mitberücksichtigt wird. Dieses erweiterte Eyring-Modell mit den 4 unbekannten Parametern A, B, C und D zur Analyse von Daten aus zeitraffenden Zuverlässigkeitsprüfungen ist gegeben durch die **Verallgemeinerte Eyring-Gleichung**

$$\frac{dM}{dt} = A \cdot T \cdot e^{-\frac{B}{kT} + CV + \frac{DV}{kT}} = \lambda,$$

wobei T wiederum den thermalen Einfluß (Temperatur in $^\circ\text{K}$) und V den nichtthermalen Einfluß bezeichnet, vgl. Goldberg (1964), Singpurwalla et al. (1975). Thomas/Gorton (1963) entwickeln ein sequentielles Verfahren zur graphischen Analyse zeitraffender Prüfungen in diesem erweiterten Eyring-Modell.

Die unbekannten Parameter A und B sind mit der im Kurzzeittest notwendigen Temperaturerhöhung verbunden, die Parameter C und D hängen mit der Veränderung des nichtthermalen Einflusses zusammen.

Will man in diesem Modell die Ausfallrate λ_0 unter Testbedingungen, d.h. bei Temperatur T_0 und nichtthermalen Einfluß V_0 , umrechnen in die Ausfallrate λ bei Normalbedingungen (Temperatur T, nichtthermaler Einfluß V), so muß man die Beziehung

$$\lambda = \lambda_0 \cdot B((T, V), (T_0, V_0))$$

ausnutzen, wobei der Beschleunigungsfaktor durch

$$B((T, V), (T_0, V_0)) = \frac{T}{T_0} \cdot e^{-\frac{B}{k} \left(\frac{1}{T} - \frac{1}{T_0} \right) + C(V - V_0) + \frac{D}{k} \left(\frac{V}{T} - \frac{V_0}{T_0} \right)}$$

gegeben ist. Es sei noch erwähnt, daß auch das **erweiterte Eyring-Modell** nur innerhalb gewisser Bereiche angewandt werden kann, da sonst möglicherweise veränderte Ausfallmechanismen zum Tragen kommen.

Legen wir für eine Grundgesamtheit von Objekten eine exponentialverteilte Lebensdauer zugrunde und führen Zeitraffungstests auf p verschiedenen Stufen (T_j, V_j) für $j = 1, \dots, p$ von thermalen und nichtthermalen Einflüssen durch, so können wir mittels Methoden der multiplen Regressionsanalyse, vgl. Abschnitt 5 in Kap. X, die Ausfallrate λ in der Grundgesamtheit unter Normalbedingungen schätzen. Die p verschiedenen Stufen (T_j, V_j) müssen natürlich so gewählt sein, daß keine veränderten Ausfallmechanismen eintreten.

Der j-ten Stufe (T_j, V_j) werden nun zufällig n_j (für $j = 1, \dots, p$) Objekte aus der interessierenden Grundgesamtheit zugeordnet. Zudem wird für die j-te Stufe festgelegt, wieviele der n_j Objekte ausgefallen sein müssen, um den Test beenden zu können. Diese Anzahl bezeichnen wir mit r_j . Nun werden die Prüfungen auf den p Belastungsstufen durchgeführt und die Ausfallzeitpunkte t_{1j}, \dots, t_{rj} für $j = 1, \dots, p$ beobachtet.

Da wir exponentialverteilte Lebensdauern in der Grundgesamtheit unter Normalbedingungen und Gleichbleiben des Ausfallmechanismus auf allen p Stufen gefordert haben, ist für $j = 1, \dots, p$ die Lebensdauer der Objekte auf der j -ten Belastungsstufe exponentialverteilt mit Parameter λ_j . Ausgehend von den p Testergebnissen können wir λ_j für $j = 1, \dots, p$ schätzen durch

$$\hat{\lambda}_j = \frac{r_j}{(n_j - r_j)t_{rjj} + \sum_{i=1}^{r_j} t_{ij}},$$

wenn ausgefallene Objekte nicht durch neue ersetzt werden, vgl. auch Abschnitt 3.

Will man einen Schätzer für die Ausfallrate λ unter Normalbedingungen bestimmen, so betrachtet man nun das Modell

$$\ln \hat{\lambda}_j = \ln A + \ln T_j - \frac{B}{kT_j} + CV_j + \frac{DV_j}{kT_j} + \ln \varepsilon_j \quad \text{für } j = 1, \dots, p.$$

Setzt man voraus, daß die Fehlerterme $\ln \varepsilon_j$ für $j = 1, \dots, p$ unabhängig sind mit Erwartungswert 0 und Varianz σ^2 , so kann man die Parameter A, B, C und D optimal nach der gewöhnlichen Methode der kleinsten Quadrate schätzen. Mit

$$X = \begin{bmatrix} 1 & V_1 & \frac{V_1}{kT_1} & \frac{1}{kT_1} \\ 1 & V_2 & \frac{V_2}{kT_2} & \frac{1}{kT_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & V_p & \frac{V_p}{kT_p} & \frac{1}{kT_p} \end{bmatrix}$$

ergibt sich als Schätzer für $\alpha = (\ln A, C, D, -B)^T$:

$$\hat{\alpha} = (X^T X)^{-1} X^T z = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)^T, \quad \text{wobei } z = (\ln(\hat{\lambda}_1/T_1), \dots, \ln(\hat{\lambda}_p/T_p))^T$$

ist; vgl. auch Abschnitt 5 in Kap. X, wo diese und eine weitere, oft numerisch stabilere Vorgehensweise zum Schätzen der Regressionskoeffizienten ausführlich dargestellt werden. Hier bezeichnet X^T die zu X transponierte (gestürzte) Matrix und y^T den zum Zeilenvektor y gehörigen Spaltenvektor.

Ein Schätzer für die Ausfallrate unter Normalbedingungen (T, V) ist dann

$$\hat{\lambda} = T \cdot e^{\hat{\alpha}_1 + \hat{\alpha}_2 V + (\hat{\alpha}_3 V/(kT)) + (\hat{\alpha}_4/(kT))}$$

Nehmen wir zusätzlich an, daß die Fehlerterme $\ln \varepsilon_j$ für $j = 1, \dots, p$ normalverteilt sind, so ist durch

$$[\hat{\lambda} e^{-G}, \hat{\lambda} e^G]$$

ein Prognoseintervall zum Niveau $1 - \gamma$ für die Ausfallrate λ unter Normalbedingungen gegeben, wenn man

$$G = t_{p-4; 1-\gamma/2} \cdot \sqrt{s^2(1 + H^T(X^T X)^{-1}H)} \quad \text{mit}$$

$$s^2 = \frac{1}{p-4} z^T (I_p - X(X^T X)^{-1} X^T) z = \frac{1}{p-4} \sum_{j=1}^p (\ln \hat{\lambda}_j - \ln \lambda_j^*)^2,$$

$$\ln \lambda_j^* = \ln T_j + \hat{\alpha}_1 + \hat{\alpha}_2 V_j + \frac{\hat{\alpha}_3 V_j}{kT_j} + \frac{\hat{\alpha}_4}{kT_j} \quad \text{und}$$

$$H = \left(1, V, \frac{V}{kT}, \frac{1}{kT} \right)^T$$

setzt. Dabei bezeichnet $t_{p-4,\gamma}$ das γ -Quantil der t -Verteilung mit $p-4$ Freiheitsgraden; s^2 ist natürlich gerade der Schätzer für die Varianz σ^2 der Fehlerterme $\ln \varepsilon_j$.

Beispiel: Auf $p = 8$ verschiedenen Belastungsstufen wurde die Lebensdauer von Tantal-Kondensatoren untersucht. Die beiden Einflüsse waren hier verschiedene Temperaturen T_j und verschiedene Spannungen V_j . In Tab. 13 sind T_j und V_j sowie n_j , r_j und t_{rj} für $j = 1, \dots, p$ und $i = 1, \dots, r_j$ angegeben.

Tab. 13: Lebensdauern von Tantal-Kondensatoren (nach Singpurwalla et al. (1975))

Belastungsstufe j	Temperatur T_j (in °K)	Spannung V_j (in Volt)	n_j	r_j	t_{1j}, \dots, t_{rj} (in Std.)
1	358	35,0	1000	4	20, 90, 700, 37000
2	358	40,6	200	4	20, 3600, 9500, 27000
3	358	46,5	50	2	800, 2800
4	358	51,5	50	5	500, 800, 2400, 10700, 10700
5	318	46,5	500	6	100, 1200, 7500, 2000, 26000, 27300
6	318	57,0	50	1	8900
7	278	46,5	175	1	1000
8	278	62,5	175	19	25, 50, 165, 500, 620, 720, 820, 910, 980, 1270, 1600, 2270, 2370, 4590, 4690, 4880, 7560, 8730, 12500

Wir wollen nun zunächst den Kleinste-Quadrate-Schätzer $\hat{\alpha}$ für den Parametervektor α bestimmen. In Tab. 14 sind die benötigten Hilfsgrößen $\hat{\lambda}_j$, $\ln(\hat{\lambda}_j/T_j)$ sowie die Matrix X angegeben.

Tab. 14: Hilfsgrößen und Designmatrix X zur Berechnung von $\hat{\alpha}$

j	$\hat{\lambda}_j$	$\ln(\hat{\lambda}_j/T_j)$	1	V_j	$\frac{V_j}{kT_j}$	$\frac{1}{kT_j}$
1	$1 \cdot 10^{-7}$	$\begin{bmatrix} -21,917685 \\ -19,983498 \end{bmatrix}$	1	35,0	1132,8547	32,367278
2	$8 \cdot 10^{-7}$	$\begin{bmatrix} -17,022395 \\ -17,406572 \end{bmatrix} = z$	1	40,6	1314,1115	32,367278
3	$145 \cdot 10^{-7}$	$\begin{bmatrix} 46,5 \\ 51,5 \end{bmatrix}$	1	46,5	1505,0784	32,367278
4	$99 \cdot 10^{-7}$	$\begin{bmatrix} 1666,9148 \\ 1694,3965 \end{bmatrix}$	1	51,5	1666,9148	32,367278
5	$4 \cdot 10^{-7}$	$\begin{bmatrix} 2077,0021 \\ 2605,1002 \end{bmatrix}$	1	46,5	1694,3965	36,438634
6	$22 \cdot 10^{-7}$	$\begin{bmatrix} 41,681603 \\ 41,681603 \end{bmatrix}$	1	57,0	2077,0021	36,438634
7	$57 \cdot 10^{-7}$	$\begin{bmatrix} 1938,1945 \\ 2605,1002 \end{bmatrix}$	1	46,5	1938,1945	41,681603
8	$95 \cdot 10^{-7}$	$\begin{bmatrix} 41,681603 \\ -17,194461 \end{bmatrix}$	1	62,5	2605,1002	41,681603

Dort ist etwa

$$\hat{\lambda}_3 = \frac{r_3}{(n_3 - r_3)t_{r_3,3} + \sum_{i=1}^{r_3} t_{i,3}} = \frac{2}{(50 - 2)2800 + 800 + 2800} = \frac{2}{138000} = 145 \cdot 10^{-7}$$

Aus Tab. 14 erhält man nun

$$X^T X = \begin{pmatrix} 8 & 368,1 & 13933,653 & 285,70959 \\ 368,1 & 19167,61 & 698958,46 & 13933,653 \\ 13933,653 & 698958,46 & 25782177 & 508666,84 \\ 285,70959 & 13933,653 & 508666,84 & 10320,823 \end{pmatrix}.$$

Somit lässt sich aus der Gleichung

$$X^T X \hat{\alpha} = X^T z$$

nun der Schätzer $\hat{\alpha}$ für den Parametervektor $\alpha = (\ln A, C, D, -B)^T$ berechnen. Es ergibt sich

$$\begin{aligned} \hat{\alpha} &= (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)^T = (X^T X)^{-1} X^T z \\ &= (-52,76327348; -0,71045739; -0,01553396; 0,74851228)^T. \end{aligned}$$

Nun kann man $\ln \lambda$ und somit λ für beliebige Werte von T und V (natürlich nur in einem Bereich, in dem sich der Ausfallmechanismus nicht verändert) aufgrund der folgenden Gleichungen schätzen:

$$\ln \lambda^* = \ln T + \hat{\alpha}_1 + \hat{\alpha}_2 V + \hat{\alpha}_3 V/(kT) + \hat{\alpha}_4/(kT) \quad \text{und} \quad \lambda^* = e^{\ln \lambda^*}.$$

Zur Überprüfung des Modells sind in Tab. 15 die direkt geschätzten Werte $\hat{\lambda}_j$ und $\ln \hat{\lambda}_j$ sowie die aufgrund der Gleichung geschätzten Werte λ_j^* und $\ln \lambda_j^*$ für $j = 1, \dots, 8$ angegeben.

Tab. 15: Direkt und mittels Gleichung geschätzte Ausfallraten

j	$\hat{\lambda}_j$	λ_j^*	$\ln \hat{\lambda}_j$	$\ln \lambda_j^*$
1	$1 \cdot 10^{-7}$	$2 \cdot 10^{-7}$	-16,118096	-15,387147
2	$8 \cdot 10^{-7}$	$7 \cdot 10^{-7}$	-14,038654	-14,224221
3	$145 \cdot 10^{-7}$	$23 \cdot 10^{-7}$	-11,141362	-12,998995
4	$99 \cdot 10^{-7}$	$64 \cdot 10^{-7}$	-11,522976	-11,960668
5	$4 \cdot 10^{-7}$	$22 \cdot 10^{-7}$	-14,731801	-13,010876
6	$22 \cdot 10^{-7}$	$102 \cdot 10^{-7}$	-13,027053	-11,494454
7	$57 \cdot 10^{-7}$	$22 \cdot 10^{-7}$	-12,075044	-13,008028
8	$95 \cdot 10^{-7}$	$61 \cdot 10^{-7}$	-11,564219	-12,000396

Diese Anpassung hat das multiple Bestimmtheitsmaß 55,4% mit $s^2 = 2,5122$. Wir wollen einmal den Schätzer für die Ausfallrate λ bei Temperatur $T = 0^\circ\text{C} = 273^\circ\text{K}$ und Spannung $V = 10$ Volt berechnen. Es ergibt sich

$$\begin{aligned} \ln \lambda^* &= \ln T + \hat{\alpha}_1 + \hat{\alpha}_2 V + \hat{\alpha}_3 V/(kT) + \hat{\alpha}_4/(kT) \\ &= \ln(273) + (-52,76327348) + (0,71045739) 10 \\ &\quad + (-0,01553396) 10/(8,63 \cdot 10^{-5} \cdot 273) + (0,74851228)/(8,63 \cdot 10^{-5} \cdot 273) \\ &= -14,872012; \end{aligned}$$

somit ist

$$\lambda^* = e^{-14,872012} = 3,5 \cdot 10^{-7}$$

der gesuchte Schätzer für die Ausfallrate λ bei $T = 273^\circ\text{K}$ und $V = 10$ Volt.

Zu bemerken ist, daß die Bestimmung der Schätzwerte $\hat{\alpha}_j$ mit obiger Matrix $X^T X$ numerisch instabil ist und daß daher mit hoher Stellenzahl gerechnet bzw. das andere in Kap. X, Abschnitt 5 vorgestellte Verfahren benutzt werden sollte.

Den zeitraffenden Prüfungen kommt neben der bisher besprochenen Aufgabe der Ermittlung von Zuverlässigkeitscharakteristika langlebiger Objekte eine weitere wichtige Aufgabe zu.

4.2. Screening-Tests, Burn-Ins

Ergebnis sich etwa für einen Gerätetyp aufgrund der in Abschnitt 2 vorgestellten Tests eine DFR-Verteilung, d. h. hat die Lebensdauerverteilung des Gerätetyps eine signifikant fallende Ausfallrate, so ist es mitunter sinnvoll, Geräte dieses Typs „vorzu-altern“, um minderwertige oder defekte Teile aufzuspüren. Der aus den Prüfkosten resultierende Preisanstieg wird durch die Erhöhung der zu erwartenden mittleren Lebensdauer (minderwertige Geräte sind ja aussortiert worden) kompensiert. Solche **Voralterungsmethoden** unter zeitraffenden Bedingungen finden vor allem in der **Sicherheits- und Raumfahrttechnik**, neuerdings aber auch vermehrt in der **Computertechnik** Verwendung.

Eine gute Übersicht über solche **Screening-Tests**, z. B. Burn-Ins, welche die Früh-ausfälle aussieben sollen, findet man bei Schäfer (1980).

Man unterscheidet etwa Screening-Tests mit

- Hochtemperaturlagerung,
- Temperaturwechsel in Luft,
- Temperaturschock in Flüssigkeit.

Solche dem eigentlichen Produktionsprozeß folgende Nachbehandlungen nennt man auch **Formier- oder Einbrennungsprozesse**. Der gewählte Screening-Test bei nachzubehandelnden Produkten hängt von der Art des Produktes, dem Einsatz-zweck und der geforderten Zuverlässigkeit ab.

Screening-Tests bereiten insofern große Schwierigkeiten, als daß man durch sie sehr leicht zusätzliche Beschädigungen von Produkten bewirkt, wenn man etwa bei der Temperaturbelastung gewisse Grenzen überschreitet; dies wirkt sich dann negativ auf die Dauer der Nutzungsphase aus.

4.3. Labor- und Einsatzbedingungen

4.3.1. Der Einfluß von Strahlungen auf den Alterungsprozeß

Einen wesentlichen Einfluß auf den Alterungsprozeß von Produkten übt z. B. die natürliche **Sonnenstrahlung** aus. Die Bestrahlungsstärke ist u.a. abhängig vom Stand der Sonne und vom Grad der Bewölkung. Untersuchungen über die Auswirkungen sind nur im Labor ohne Schwierigkeiten möglich. Die **terrestrische Strahlung** (Globalstrahlung) etwa, die die Erdoberfläche erreicht, besitzt ebenfalls einen wesentlichen Einfluß auf die Alterung von Werkstoffen. Abb. 29, vgl. Boxhammer (1981), soll einen Eindruck davon vermitteln, in welchem Maße sich der geographische Ort und die Jahreszeit auf den Alterungsprozeß von Produkten auswirken.

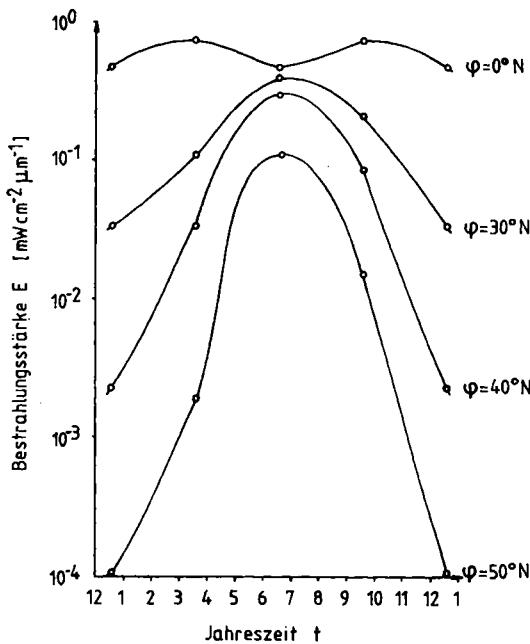


Abb. 29: Bestrahlungsstärke E als Funktion der Jahreszeit für verschiedene Breitengrade φ bei Wellenlänge 297,5 nm

Zudem wird in Abb. 29 deutlich, wie problematisch es ist, Ergebnisse von Zuverlässigkeitsprüfungen auf Orte verschiedener geographischer Breite zu übertragen.

4.3.2. Wertungsfaktoren für im Labor ermittelte Ausfallraten

Bisher haben wir ein Problem der Lebensdauerprüfungen unter Laborbedingungen noch gar nicht berücksichtigt: Elektronische Bauteile etwa, die zur Steuerung der Flugbahn einer Rakete dienen, unterliegen in jeder Hinsicht extremen Einsatzbedingungen, die im Labor kaum simuliert werden können. Insbesondere kosmische Strahlen (α -Strahlen und β -Strahlen) wirken sich negativ auf die Lebensdauer solcher Bauelemente aus. Kotowski (1961) weist im Zusammenhang mit Fragen der Flugsicherung nach, daß am Einsatzort berechnete Zuverlässigkeitsgrößen von im Labor ermittelten enorm abweichen können und zieht hieraus den wichtigen Schluß, daß Anstrengungen zur Erhöhung des Bodenaufwandes bei gleichzeitiger Einsparung bei der Bordelektronik unternommen werden müssen, um einen Zuwachs an Zuverlässigkeit zu erlangen. In den letzten Jahren wurden häufig schon spezielle Wertungsfaktoren abgeleitet, die zur Transformation der im Labor ermittelten Ausfallraten in für den Einsatzort gültige Werte dienen. Diese Wertungsfaktoren sind jedoch nur als grobe Näherung zu betrachten. Preuß (1978) etwa gibt die in Tab. 16 zusammengestellten Wertungsfaktoren für die im Labor ermittelten Ausfallraten elektronischer Bauelemente für verschiedene Einsatzorte an.

Es sei an dieser Stelle noch darauf hingewiesen, daß bei Zuverlässigkeitsermittlungen von Systemen die Zuverlässigkeitsangaben für Bauelemente durch Ausfallraten mit äußerster Vorsicht zu betrachten sind, sofern nicht detaillierte Betriebs-

Tab. 16: Wertungsfaktoren für im Labor ermittelte Ausfallraten von elektronischen Bauelementen

Umgebung	Faktor
Labor	1
Erboden	10
Schiff	20
Kraftfahrzeug	30
Eisenbahn	40
Werkbank	60
Flugzeug	150
Klassische Geschosse	400
Moderne Geschosse (Raketen, Raumschiffe)	1000

und Versuchsbedingungen angegeben werden, da sich etwa thermische, magnetische oder Strahlen-Belastungen (wie oben ausgeführt) auf die Ausfallrate auswirken. Außerdem ist zu beachten, daß sich Angaben in manchen Fällen nicht auf Vollausfälle stützen.

Die Ergebnisse aus **Tab. 17** beziehen sich auf die Vollausfälle bei einer Bauelemente-Umgebungstemperatur von 50°C, vgl. Ryerson et al. (1967), Dombrowski (1970). Es fällt auf, daß mit dem Übergang von der Germanium- zur neueren Silicium-Technik die Ausfallraten stark vermindert werden konnten.

Tab. 17: Ausfallraten (Vollausfälle) bei 50°C Umgebungstemperatur

	Halbleiter				
	Ge-Dioden	Si-Dioden	Ge-Transistoren	Si-Transistoren	Integrierte Schaltungen
Ausfallrate in Bits ($\hat{=} 10^{-8}$ Ausfälle/Std.)	5,8	2,9	200	25	25
Widerstände					Kondensatoren
	Masse-	Metall-	Keramik-	Glas-	Glimmer-
Ausfallrate in Bits ($\hat{=} 10^{-8}$ Ausfälle/Std.)	1,12	0,56	9	1,25	0,35

Es mag der Gedanke auftreten, daß die Ermüdungsphase von Objekten aus Lebensdauerprüfungen generell ausgeklammert werden kann. Das Miteinbeziehen der Ermüdungsphase in die Analyse kann jedoch wertvolle Anhaltspunkte dafür liefern, wie eine Verlängerung der Nutzungsphasen langfristig erzielt werden könnte.

Gegen zeitraffende Prüfungen spricht, daß die so gewonnenen Resultate mit dem Nachteil der Extrapolation an einen Einsatz unter Normalbedingungen erkauft

werden. Solange bei Zeitraffungstests jedoch darauf geachtet wird, daß der natürliche Ausfallmechanismus sich nicht verändert und nicht durch neue Effekte überlagert wird, sind sie jedoch vertretbare Analysemethoden.

5. Wartungs- und Erneuerungsüberlegungen

5.1. Wartung und Wartbarkeit, Erneuerung

Bei den meisten technischen Geräten besteht im Falle eines Defekts die Möglichkeit der Reparatur; durch diese wird der funktionsfähige Zustand des Gerätes zurückgewonnen. Reparaturen sowie Maßnahmen zur rechtzeitigen Erkennung und Vermeidung von Fehlern werden unter dem Oberbegriff **Wartung** zusammengefaßt. Die **Wartbarkeit** dient zur Bewertung der Eignung eines Gerätes oder einer technischen Einrichtung zur Überwachung, Instandhaltung und Instandsetzung. Dabei wird unter der **Überwachung** die Überwachung der Funktionsfähigkeit und des Zustandes verstanden; **Instandsetzung** versteht sich als Fehlerbeseitigung, bei der **Instandhaltung** hingegen muß kein Fehler vorliegen. So ist etwa eine Reifenreparatur eine Instandsetzung, die Überprüfung des Reifendrucks nur eine Instandhaltungsmaßnahme. Die Wartbarkeit selbst ist die Wahrscheinlichkeit dafür, daß eine Betrachtungseinheit unter gegebenen Bedingungen nach einem Ausfall wieder in einen funktionsfähigen Zustand gebracht werden kann. Die Wartbarkeit einer Einheit kann nur dann bestimmt werden, wenn alle Wartungsmaßnahmen erfaßt werden können.

Wartbare Einheiten sind z. B. Großrechner und Nachrichtenübertragungssysteme; dagegen sind die Wartungsmöglichkeiten für Raumschiffe im Weltraum selbst sehr stark eingeschränkt.

A. Die Wartbarkeit von Systemen

Die Wartbarkeit M (**maintainability**) im oben angegebenen Sinn ist mit der Wahrscheinlichkeit des Erfolges einer Reparatur gleichzusetzen.

Ist die Wahrscheinlichkeit für das Auftreten eines Ausfalls unabhängig von der Wahrscheinlichkeit der Beseitigung des Fehlers durch Wartung, so ist die **Ausfallwahrscheinlichkeit Q_M eines wartbaren Systems** gerade gleich dem Produkt der Ausfallwahrscheinlichkeit Q des Systems (wenn dieses nicht gewartet würde) und der Wahrscheinlichkeit $1 - M$ für eine nicht erfolgreiche Wartung:

$$Q_M = Q(1 - M) = (1 - P)(1 - M),$$

wobei P die Intaktwahrscheinlichkeit des Systems ohne Wartung angibt. Hier geht die für praktische Belange meist sinnvolle Annahme ein, daß die Wartungszeit im Verhältnis zur Betriebszeit zu vernachlässigen ist. Die **Intaktwahrscheinlichkeit eines wartbaren Systems** ist

$$P_M = 1 - Q_M = 1 - Q(1 - M) = 1 - Q + Q \cdot M = P + Q \cdot M;$$

durch Wartungsmaßnahmen läßt sich die Zuverlässigkeit eines Systems also um $Q \cdot M$ erhöhen. Eine eingehendere Bearbeitung der Wartungsproblematik findet man etwa bei Preuß (1978).

B. Die Wartung von Systemen

Zwei wesentlich verschiedene Arten der Wartung eines Systems sind die korrigie-

rende und die prophylaktische Wartung. Bei Auftreten von unvorhergesehenen Störungen (**Havarien**) in Form eines Ausfalls einer bzw. mehrerer Komponenten eines Systems, die den Systemausfall zur Folge haben, ist eine **korrigierende Wartung** nötig.

Kündigen sich die zum Ausfall eines Systems führenden Fehler zumindest zum Teil durch allmählichen **Verschleiß** (in Form einer meßbaren Funktionsminderung) an, so können mit Hilfe eines Inspektionssystems zur Überwachung und Registrierung der Zustände von Systemkomponenten oft **prophylaktische (vorbeugende) Wartungsmaßnahmen** in Gang gesetzt werden.

C. Die Erneuerung von Systemen

Bei der Wartung wird die volle Funktionsfähigkeit eines Systems durch Reparaturmaßnahmen gesichert. Bei der **Erneuerung** dagegen werden nicht wartbare, defekte Komponenten eines Systems durch neue ersetzt. Auch bei Erneuerungen unterscheidet man in gleicher Weise wie bei der Wartung zwischen **Havarieerneuerungen** und **prophylaktischen Erneuerungen**.

Für Instandhaltungsüberlegungen ist es unwesentlich, ob die Instandhaltung eines Systems durch Wartung oder durch Erneuerung erfolgt. Daher wollen wir im folgenden nur noch von Erneuerungen sprechen und Wartungsmaßnahmen in diesen Begriff miteinschließen.

5.2. Erneuerungsstrategien, Schranken der Erneuerungsfunktion

Bei der Instandhaltung können verschiedene, aus technologischen Erfordernissen resultierende **prophylaktische Erneuerungsstrategien** verwandt werden. Zu beachten ist natürlich, daß solche Strategien nur dann sinnvoll sind, wenn ein System und seine Komponenten Alterungerscheinungen unterliegen. Ob für bestimmte Systeme prophylaktische Erneuerungen sinnvoll sind, läßt sich zum Beispiel mittels der in Abschnitt 2 vorgestellten Tests zur Überprüfung der Zuordnung zu den aus Alterungskriterien resultierenden Verteilungsklassen überprüfen. Erweist sich die Planung prophylaktischer Erneuerungsstrategien als sinnvoll, so bleibt nur noch die Frage nach einer guten Strategie offen. Zwei solche Erneuerungsstrategien sollen hier vorgestellt werden.

A. Altersabhängige Erneuerungsstrategie und Gruppenerneuerungsstrategie

Wird ein Element unmittelbar nach einem Ausfall durch eine Havarieerneuerung ersetzt oder nach Erreichen eines bestimmten Alters T im Zuge einer prophylaktischen Erneuerung ausgetauscht, je nachdem, welches Ereignis zuerst auftritt, so spricht man von einer **altersabhängigen Erneuerungsstrategie**.

Wird ein Element nach Ausfall durch eine Havarieerneuerung ersetzt und findet zusätzlich zu festen Zeitpunkten T , $2T$, $3T$, ... eine prophylaktische Erneuerung statt, so spricht man von einer **Gruppen- oder Blockerneuerungsstrategie**. Die Blockerneuerungsstrategie hat gegenüber der altersabhängigen Strategie den Vorteil, daß das jeweilige Alter eines Elementes nicht registriert werden muß, und den Nachteil, daß vielfach noch relativ neue Elemente ausgetauscht werden. Beide Erneuerungsstrategien sollen hier nun näher untersucht werden.

B. Der Erneuerungsprozeß und die Erneuerungsfunktion

Dazu muß zunächst der Begriff **Erneuerungsprozeß** näher erläutert werden. Ein

Erneuerungsprozeß ist eine Folge von unabhängig identisch verteilten (mit Verteilungsfunktion F) nichtnegativen Zufallsvariablen X_1, X_2, \dots , die mit Wahrscheinlichkeit Eins nicht alle gleich Null sind.

Die Realisierung eines solchen Prozesses kann man sich etwa so vorstellen: Ein System arbeitet unendlich lange und wird nach Ausfall in einer vernachlässigbar kurzen Zeit repariert oder ersetzt; dann sind die Zufallsvariablen X_1, X_2, \dots des Erneuerungsprozesses gerade die aufeinanderfolgenden Zeitabstände zwischen zwei Ausfällen. Die Verteilungsfunktion F entspricht also gerade der Verteilungsfunktion der Intaktzeit des Systems (zwischen zwei Ausfällen). Die **Erneuerungsfunktion** $H(t)$ beschreibt nun die erwartete Anzahl von Erneuerungen im Zeitintervall $(0; t]$. Die numerische Berechnung der Erneuerungsfunktion eines Systems ist nur mit enormem Aufwand möglich; ist jedoch die Lebensdauerverteilung des Systems eine IFR-Verteilung, so läßt sich für jedes t eine untere und eine obere Schranke der Erneuerungsfunktion $H(t)$ bestimmen, vgl. etwa Beichelt (1976). Bei der Bestimmung dieser Schranken geht das sogenannte **Erneuerungstheorem**, vgl. z.B. Ross (1970), ein: Existiert der Erwartungswert μ der Verteilungsfunktion F , so gilt

$$\lim_{t \rightarrow \infty} \frac{H(t)}{t} = \frac{1}{\mu},$$

mit $\mu = \int_0^\infty \bar{F}(t) dt \left[= \int_0^\infty t f(t) dt, \text{ wenn } F \text{ Dichte } f \text{ hat} \right]$,

d.h. für große Zeiträume ist die erwartete Anzahl von Erneuerungen etwa gleich dem Verhältnis ihrer Länge t zur mittleren Intaktzeit.

C. Schranken der Erneuerungsfunktion

Sowohl für die altersabhängige als auch für die Gruppenerneuerungsstrategie ergibt sich als Abschätzung der Erneuerungsfunktion $H(t)$ nun

$$\frac{\frac{t}{\int_0^t \bar{F}(x) dx} - 1}{\frac{t \cdot F(t)}{\int_0^t \bar{F}(x) dx}} \leq H(t) \leq \frac{\frac{t \cdot F(t)}{\int_0^t \bar{F}(x) dx}}{\frac{t}{\int_0^\infty \bar{F}(x) dx}},$$

falls die Lebensdauerverteilung eines Systems eine IFR-Verteilung besitzt.

Bei Vorliegen einer NBUE-Verteilung gilt die gröbere Abschätzung

$$\frac{\frac{t}{\int_0^\infty \bar{F}(x) dx} - 1}{\frac{t}{\int_0^\infty \bar{F}(x) dx}} \leq H(t) \leq \frac{\frac{t}{\int_0^\infty \bar{F}(x) dx}}{\frac{t}{\int_0^\infty \bar{F}(x) dx}},$$

vgl. Barlow/Proschan (1978).

5.3. Die Zuverlässigkeit von Straßenverkehrssignalanlagen – Ein Beispiel

In diesem Abschnitt wollen wir – als Beispiel für die bisher in diesem Kapitel angegangenen Probleme – verschiedene Typen von Straßenverkehrssignalanlagen bzgl. ihrer Zuverlässigkeit untersuchen und vergleichen. Solche Untersuchungen sind für die Praxis nicht unwesentlich, denn die Verkehrssicherheit hängt stark von der Zuverlässigkeit eines Ampel-Systems ab. Neben einer hohen Zuverlässigkeit

spielen natürlich auch die Kosten (**Anschaffungs- und Wartungskosten**) eines Systems eine Rolle, so daß bei der Entscheidung für die Anschaffung eines bestimmten Systems Zuverlässigkeit und Preis gegeneinander abgewogen werden müssen.

5.3.1. Straßenverkehrssignalanlagen-Systeme

In den folgenden Ausführungen, die weitgehend auf Hartung/Kalin (1980) und Hartung/Jöckel (1982) zurückgehen, werden vier konkurrierende Systeme betrachtet: das System ohne Reserve, das System mit heißer Reserve, das System mit kalter Macro- und das mit kalter Micro-Reserve. Ein Ampel-System besteht hier aus N Rotlichtsignalen und das System wird aus Verkehrssicherheitsgründen abgeschaltet (gilt als defekt), sobald eines dieser Rotlichtsignale ausfällt.

Die Zuverlässigsberechnungen basieren auf den hier als exponentialverteilt angenommenen Lebensdauern der identischen Glühbirnen oder Glühdioden der N Rotlichtsignale eines Systems in einem Inspektionszeitraum I_T ; nach Ablauf von I_T werden alle Birnen bzw. Dioden ausgetauscht. Die Zuverlässigkeit eines Systems entspricht natürlich gerade der Wahrscheinlichkeit dafür, daß das System nach Ablauf von I_T noch intakt ist, vgl. Abschnitte 1 und 2.

Da man sicherlich von IFR-verteilten Lebensdauern der Glühbirnen ausgehen kann, ist es gerechtfertigt, hier mit der Exponentialverteilung zu arbeiten, wenn der Inspektionszeitraum I_T nicht zu groß gewählt wird, denn für $I_T \leq \mu$ gilt bei IFR-Verteilungen die Abschätzung

$$\bar{F}(t) \geq e^{-t/\mu} \quad \text{für } 0 \leq t \leq \mu,$$

wobei $\mu = \int_0^\infty \bar{F}(t) dt [= \int_0^\infty t f(t) dt]$ die (theoretisch) erwartete Lebensdauer der Glühbirnen ist.

Zunächst sollen nun die vier obengenannten Systeme vorgestellt, ihre Zuverlässigkeit im Inspektionszeitraum I_T , ihre Ausfallrate sowie ihre mittlere Lebenserwartung bestimmt werden. Dabei gehen wir jetzt davon aus, daß die Verteilungsfunktion der Lebensdauer einer Glühbirne bzw. eines Glühdiodens gerade

$$F(t) = 1 - e^{-\lambda t}$$

ist, d.h. ihre Überlebenswahrscheinlichkeit im Inspektionszeitraum I_T ist

$$\bar{F}(I_T) = 1 - F(I_T) = e^{-\lambda I_T}.$$

A. Das System ohne Reserve

Wir betrachten zunächst das **System ohne Reserve (SOR)**. Bei einem solchen System besteht jedes der N Rotlichtsignale aus einer einzigen Glühbirne. Um die Verkehrssicherheit nicht zu gefährden, wird es abgeschaltet, d.h. es ist defekt, wenn eine der N Glühbirnen defekt ist. Einer solchen Straßenverkehrssignalanlage entspricht somit gerade ein Seriensystem aus N Komponenten. Die Zuverlässigkeit eines Systems ohne Reserve im Inspektionszeitraum I_T berechnet sich somit als Produkt der Komponentenzuverlässigkeiten, vgl. Abschnitt 1.1, die wir alle als

$$\bar{F}(I_T) = e^{-\lambda I_T}$$

vorausgesetzt haben; d.h. die Zuverlässigkeit des Systems ist:

$$Z_{I_T}(\text{SOR}) = \prod_{j=1}^N \bar{F}(I_T) = \prod_{j=1}^N e^{-\lambda I_T} = e^{-N\lambda I_T}.$$

Die Lebensdauer des Systems ist also exponentialverteilt mit Parameter $N \cdot \lambda$, so daß sich als Ausfallrate

$$r_{\text{SOR}}(t) = \frac{f_{\text{SOR}}(t)}{Z_t(\text{SOR})} = \frac{N\lambda e^{-N\lambda t}}{e^{-N\lambda t}} = N\lambda$$

und als mittlere Lebensdauer

$$\begin{aligned}\mu_{\text{SOR}} &= \int_0^\infty t \cdot f_{\text{SOR}}(t) dt = \int_0^\infty t \cdot N \cdot \lambda \cdot e^{-N\lambda t} dt \\ &= [-t \cdot e^{-N\lambda t}]_0^\infty + \int_0^\infty e^{-N\lambda t} dt = \frac{1}{N \cdot \lambda}\end{aligned}$$

ergibt.

B. Das System mit heißer Reserve

Bei einem **System mit heißer Reserve (SHR)** ist jedes der N Rotlichtsignale mit zwei Glühbirnen bzw. Glühfäden ausgestattet, die parallel arbeiten. Das System fällt aus, wenn in einem Rotlichtsignal beide Birnen defekt sind. Betrachtet wird hier also ein System mit $2N$ gleichartigen Komponenten von denen je zwei parallelgeschaltet sind; die N Parallelschaltungen bilden dann ein Seriensystem. Die Zuverlässigkeit eines einzelnen Rotlichtsignals (**RS**) in I_T berechnet sich somit, vgl. Abschnitt 1.1, zu

$$\begin{aligned}Z_{I_T}(\text{RS}) &= 1 - (1 - \bar{F}(I_T))(1 - \bar{F}(I_T)) = 1 - (1 - e^{-\lambda I_T})(1 - e^{-\lambda I_T}) \\ &= 2 \cdot e^{-\lambda I_T} - e^{-2\lambda I_T}\end{aligned}$$

und somit ist

$$Z_{I_T}(\text{SHR}) = \prod_{j=1}^N Z(\text{RS}) = (2 \cdot e^{-\lambda I_T} - e^{-2\lambda I_T})^N$$

die Zuverlässigkeit des Systems mit heißer Reserve. Da sich die Dichte der Lebensdauerverteilung des Systems zu

$$\begin{aligned}f_{\text{SHR}}(t) &= \frac{d}{dt} (1 - Z_t(\text{SHR})) = \frac{d}{dt} (1 - (2 \cdot e^{-\lambda t} - e^{-2\lambda t})^N) \\ &= -N(2 \cdot e^{-\lambda t} - e^{-2\lambda t})^{N-1}(-2 \cdot \lambda \cdot e^{-\lambda t} + 2 \cdot \lambda \cdot e^{-2\lambda t})\end{aligned}$$

ergibt, ist

$$\begin{aligned}r_{\text{SHR}}(t) &= \frac{f_{\text{SHR}}(t)}{Z_t(\text{SHR})} = \frac{2 \cdot N \cdot \lambda (2 \cdot e^{-\lambda t} - e^{-2\lambda t})^{N-1} (e^{-\lambda t} - e^{-2\lambda t})}{(2 \cdot e^{-\lambda t} - e^{-2\lambda t})^N} \\ &= 2 \cdot N \cdot \lambda \frac{e^{-\lambda t} - e^{-2\lambda t}}{2e^{-\lambda t} - e^{-2\lambda t}}\end{aligned}$$

die Ausfallrate des Systems und

$$\mu_{\text{SHR}} = \int_0^\infty t \cdot f_{\text{SHR}}(t) dt = \sum_{i=0}^N \binom{N}{i} (-1)^{N-i} \cdot 2^i \cdot \frac{1}{\lambda(2N-i)}$$

seine mittlere Lebensdauer.

C. Systeme mit kalter Reserve

Wir kommen nun zur Beschreibung von Systemen mit kalter Reserve. Bei solchen Systemen beginnen die Reservekomponenten erst dann zu arbeiten, wenn die Erstkomponenten ausgefallen sind. Bei den Zuverlässigsberechnungen wollen wir hier die Zuverlässigkeit des Umschaltens von Erst- auf Zweitkomponenten nicht berücksichtigen, d. h. wir unterstellen für den Umschalter eine ständige Funktionsfähigkeit; man vgl. jedoch Gaede (1977) für allgemeine Berechnungen, die Ausfallmöglichkeiten für Umschalter einbeziehen.

C1. Das System mit kalter Macro-Reserve

Das **System mit kalter Macro-Reserve (SKR_I)** besteht aus N Rotlichtsignalen à zwei Glühbirnen bzw. -fäden. Zu Beginn des Inspektionszeitraumes I_T arbeitet in jedem Rotlichtsignal eine Birne. Fällt nun eine dieser N Birnen aus, so wird das gesamte System auf die Reservebirnen umgeschaltet. Das System fällt dann aus, wenn eine dieser Reservebirnen ausfällt.

Bis zum Ausfall der ersten Birne arbeiten also N in Serie geschaltete Komponenten, dann übernimmt ein zweites Seriensystem die Arbeit. Die Zuverlässigkeit des Systems SKR_I ergibt sich somit aus der Verteilung der Summe der Lebensdauern zweier identischer Seriensysteme; sie wird mittels der Faltungsformel berechnet, vgl. Kap. II, Abschnitt 7.2.B. Die Verteilungsfunktion jedes Seriensystems ist identisch mit der des Systems ohne Reserve

$$F_{SOR}(t) = 1 - e^{-N\lambda t}, \quad \text{für } t \geq 0;$$

daher ist

$$\begin{aligned} F_{SKR_I}(t) &= \int_0^t F_{SOR}(t-x) \cdot f_{SOR}(x) dx = \int_0^t (1 - e^{-N\lambda(t-x)}) N \cdot \lambda \cdot e^{-N\lambda x} dx \\ &= -e^{-N\lambda t} + 1 - t \cdot N \cdot \lambda \cdot e^{-N\lambda t} \end{aligned}$$

und

$$Z_{I_T}(SKR_I) = 1 - F_{SKR_I}(I_T) = e^{-N\lambda I_T} (1 + N \cdot \lambda \cdot I_T).$$

Die Ausfallrate des Systems mit kalter Macro-Reserve ist dann

$$r_{SKR_I}(t) = \frac{f_{SKR_I}(t)}{Z_t(SKR_I)} = \frac{(N\lambda)^2 \cdot t \cdot e^{-N\lambda t}}{e^{-N\lambda t} (1 + N\lambda t)} = N \cdot \lambda \cdot \left(\frac{1}{1 + 1/(N\lambda t)} \right)$$

und seine mittlere Lebensdauer berechnet sich zu

$$\begin{aligned} \mu_{SKR_I} &= \int_0^\infty t \cdot f_{SKR_I}(t) dt = \int_0^\infty t^2 (N\lambda)^2 \cdot e^{-N\lambda t} dt = 2 \cdot \int_0^\infty e^{-N\lambda t} dt \\ &= \frac{2}{N \cdot \lambda} \quad (= 2 \cdot \mu_{SOR}). \end{aligned}$$

Sehen wir uns die berechneten Größen einmal genauer an, so stellen wir fest, daß die Lebensdauer eines Systems mit kalter Macro-Reserve gerade **Erlang-2-verteilt** ist mit Parameter N · λ, vgl. Abschnitt 4.4 in Kap. IV. Das liegt daran, daß es sich um eine Faltung zweier unabhängiger identischer Exponentialverteilungen mit Parameter N · λ handelt.

C2. Das System mit kalter Micro-Reserve

Schließlich wollen wir noch ein **System von N Rotlichtsignalen mit kalter Micro-Reserve (SKR_{II})** betrachten. Es unterscheidet sich vom System SKR_I dadurch, daß bei Ausfall einer Erstbirne lediglich in diesem Rotsignal auf Birne 2 umgeschaltet wird. Das System fällt erst dann aus, wenn in einem Rotlichtsignal die Zweitbirne bzw. der Zweitfaden defekt ist. Im Grunde handelt es sich also um ein System, das dem mit heißer Reserve ähnelt. Der Unterschied zwischen den Systemen besteht lediglich darin, daß einmal beide Birnen eines Rotsignals gleichzeitig zu arbeiten beginnen und einmal die zweite Birne erst einsetzt, wenn die erste ausgefallen ist. Die Lebensdauerverteilung eines einzelnen Rotlichtsignals (RS) berechnet sich hier also als Faltung der Lebensdauerverteilungen zweier Birnen, d.h.

$$F_{RS}(t) = \int_0^t (1 - e^{-\lambda(t-x)}) \cdot \lambda \cdot e^{-\lambda x} dx = -e^{-\lambda t} + 1 - \lambda \cdot t \cdot e^{-\lambda t}.$$

Somit ist

$$Z_{I_T}(RS) = 1 - F_{RS}(I_T) = e^{-\lambda I_T}(1 + \lambda \cdot I_T),$$

und

$$\begin{aligned} Z_{I_T}(SKR_{II}) &= \prod_{j=1}^N Z_{I_T}(RS) = (e^{-\lambda I_T}(1 + \lambda \cdot I_T))^N \\ &= e^{-N\lambda I_T}(1 + \lambda \cdot I_T)^N \end{aligned}$$

ist die Zuverlässigkeit des Gesamtsystems. Wegen

$$f_{SKR_{II}}(t) = \frac{d}{dt} (1 - Z_t(SKR_{II})) = N \cdot \lambda^2 \cdot t \cdot e^{-N\lambda t} (1 + \lambda \cdot t)^{N-1}$$

ergibt sich die Ausfallrate des Systems SKR_{II} mit N Rotlichtsignalen zu

$$\begin{aligned} r_{SKR_{II}}(t) &= \frac{f_{SKR_{II}}(t)}{Z_t(SKR_{II})} = \frac{N \cdot \lambda^2 \cdot t \cdot e^{-N\lambda t} (1 + \lambda \cdot t)^{N-1}}{e^{-N\lambda t} (1 + \lambda \cdot t)^N} \\ &= \frac{N \cdot \lambda^2 \cdot t}{1 + \lambda \cdot t} = N \cdot \lambda \cdot \left(\frac{1}{1 + 1/(\lambda t)} \right) \end{aligned}$$

und die mittlere Lebensdauer ist

$$\mu_{SKR_{II}} = \int_0^\infty t \cdot f_{SKR_{II}}(t) dt = \frac{1}{\lambda \cdot N} \sum_{i=0}^N \frac{N!}{(N-i)!} \cdot \frac{1}{N^i}.$$

5.3.2. Vergleich der Straßenverkehrssignalanlagen-Systeme

Will man die hier vorgestellten Systeme von Straßenverkehrssignalanlagen miteinander vergleichen, so muß man natürlich die Ausfallrate λ der benutzten Glühbirnen zumindest näherungsweise kennen. Weiß man z.B. aus Erfahrung, daß in einem Zeitraum I_T etwa p Prozent der Birnen ausfallen, so ist die Ausfallrate der Birnen

$$\lambda = -\frac{1}{I_T} \ln \left(1 - \frac{p}{100} \right),$$

vgl. Abschnitt 3.1.

Die Zuverlässigkeiten der Systeme (mit N Rotlichtsignalen) sind dann durch

$$\begin{aligned} Z(\text{SOR}) &= e^{-N I_1 ((-1/I_1) \cdot \ln(1 - p/100))} = e^{N \cdot \ln(1 - p/100)} \\ &= \left(1 - \frac{p}{100}\right)^N, \end{aligned}$$

$$\begin{aligned} Z(\text{SHR}) &= (2e^{\ln(1 - p/100)} - e^{2\ln(1 - p/100)})^N \\ &= \left(2\left(1 - \frac{p}{100}\right) - \left(1 - \frac{p}{100}\right)^2\right)^N \\ &= \left(1 - \frac{p^2}{10000}\right)^N, \end{aligned}$$

$$\begin{aligned} Z(\text{SKR}_I) &= e^{N \cdot \ln(1 - p/100)} \cdot \left(1 - N \cdot \ln\left(1 - \frac{p}{100}\right)\right) \\ &= \left(1 - \frac{p}{100}\right)^N \left(1 - N \cdot \ln\left(1 - \frac{p}{100}\right)\right) \end{aligned}$$

und

$$\begin{aligned} Z(\text{SKR}_{II}) &= \left(e^{\ln(1 - p/100)} \cdot \left(1 - \ln\left(1 - \frac{p}{100}\right)\right)\right)^N \\ &= \left(1 - \frac{p}{100}\right)^N \left(1 - \ln\left(1 - \frac{p}{100}\right)\right)^N \end{aligned}$$

gegeben, wobei hier zu beachten ist, daß natürlich $p = p(I_T)$, d.h. p abhängig von I_T ist.

Betrachtet man also z.B. ein System mit kalter Macro-Reserve aus $N = 8$ Rotlichtsignalen und weiß man, daß etwa 5% der eingesetzten Birnen während eines bestimmten festgewählten Inspektionszeitraumes I_T ausfallen, so ergibt sich die Zuverlässigkeit des Systems zu

$$\begin{aligned} Z(\text{SKR}_I) &= \left(1 - \frac{5}{100}\right)^8 \left(1 - 8 \cdot \ln\left(1 - \frac{5}{100}\right)\right) \\ &= \left(\frac{95}{100}\right)^8 \left(1 - 8 \cdot \ln \frac{95}{100}\right) \\ &= 0,6634204 \cdot (1 - 8 \cdot (-0,0512933)) \\ &= 0,6634204 \cdot 1,4103464 = 0,9356525. \end{aligned}$$

In Tab. 18 sind zum Vergleich der Systeme deren Zuverlässigkeiten bei verschiedenen Größen von N und I_T bzw. p angegeben. Die sechs Inspektionszeiträume entsprechen $p = 1, 3, 5, 10, 20$ bzw. 50 Prozent ausgefallener Glühbirnen (im jeweiligen Inspektionszeitraum). In den Abb. 30–33 sind diese Ergebnisse graphisch dargestellt.

Eine Sollgrenze von beispielsweise 90% läßt sich für $N = 1$, vgl. Abb. 30, bei Systemen mit Reserve noch bei einem Inspektionszeitraum I_{T_5} (d.h. $p = 20$), dagegen beim System SOR nur noch bei einem Inspektionszeitraum I_{T_4} erreichen.

Legt man wieder einen Sollwert für die Zuverlässigkeit von 90% zugrunde, so

Tab. 18: Exemplarischer Vergleich des Systeme

p	N	Z (SOR)	Z (SHR)	Z (SKR _I)	Z (SKR _{II})
1	1	0,990	0,999	0,999	0,999
	4	0,960	0,999	0,999	0,999
	8	0,922	0,999	0,997	0,999
	16	0,851	0,998	0,988	0,999
3	1	0,970	0,999	0,999	0,999
	4	0,885	0,996	0,993	0,998
	8	0,783	0,992	0,974	0,996
	16	0,614	0,985	0,913	0,992
5	1	0,950	0,997	0,998	0,998
	4	0,814	0,990	0,981	0,994
	8	0,663	0,980	0,935	0,989
	16	0,440	0,960	0,801	0,979
10	1	0,900	0,990	0,994	0,994
	4	0,656	0,960	0,932	0,979
	8	0,430	0,922	0,793	0,959
	16	0,185	0,851	0,497	0,920
20	1	0,800	0,960	0,978	0,978
	4	0,409	0,849	0,775	0,916
	8	0,167	0,721	0,465	0,840
	16	0,028	0,520	0,128	0,706
50	1	0,500	0,750	0,846	0,846
	4	0,062	0,316	0,235	0,513
	8	0,003	0,100	0,025	0,263
	16	0,000	0,010	0,000	0,068

sind für $N = 4$, vgl. Abb. 31, bei den Systemen mit Reserve von den betrachteten Inspektionszeiträumen diejenigen bis I_{T_4} zulässig. Beim System SKR_{II} kann man sogar auf I_{T_5} übergehen, dagegen beim System SOR hier nur den Inspektionszeitraum I_{T_1} wählen.

In Abb. 32 ist ein deutlicher Abfall der Zuverlässigkeit des Systems SKR_I im Vergleich zum System SHR und zum System SKR_{II} zu beobachten. Ausgehend von einem Sollwert der Zuverlässigkeit von z. B. 90% erlauben die beiden letztgenannten Systeme einen Inspektionszeitraum I_{T_4} (d. h. $p = 10$), während das System SKR_I mindestens den Inspektionszeitraum I_{T_3} (d. h. $p = 5$) verlangt. Das System SOR verlangt nach wie vor den Inspektionszeitraum I_{T_1} (d. h. $p = 1$).

Legt man auch für $N = 16$, vgl. Abb. 33, wieder einen Sollwert der Zuverlässigkeit von 90% zu Vergleichszwecken zugrunde, so zeigt sich für die Systeme mit Reserve ein differenziertes Bild bzgl. der möglichen Wahl von Inspektionszeiträumen. Während das System mit kalter Reserve II, SKR_{II}, den Inspektionszeitraum I_{T_4} zuläßt, muß man beim System SHR auf den Inspektionszeitraum I_{T_3} und beim System SKR_I sogar auf den Inspektionszeitraum I_{T_2} zurückgehen. Das System SOR genügt dagegen bei keinem der betrachteten möglichen Inspektionszeiträume der hier gestellten Forderung, einen Sollwert der Zuverlässigkeit von mindestens 90% zu garantieren.

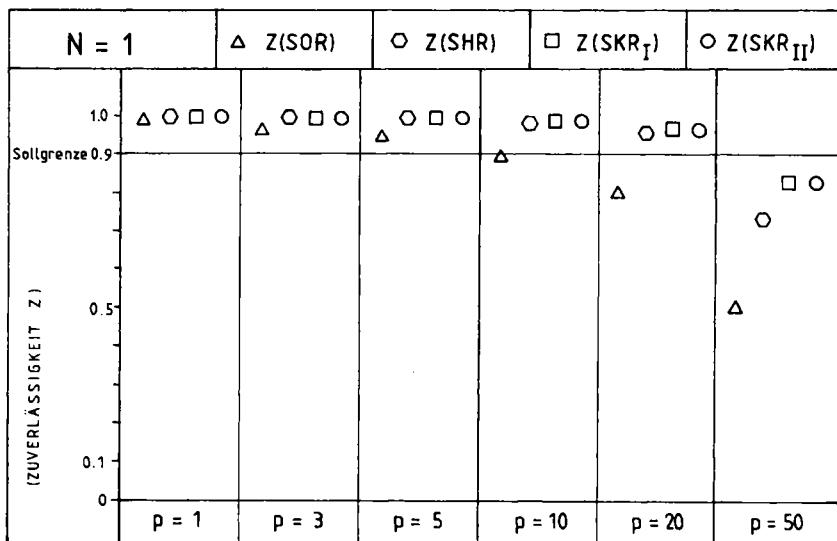


Abb. 30: Systemzuverlässigkeiten der Straßenverkehrssignalanlagen bei Systemumfang $N = 1$

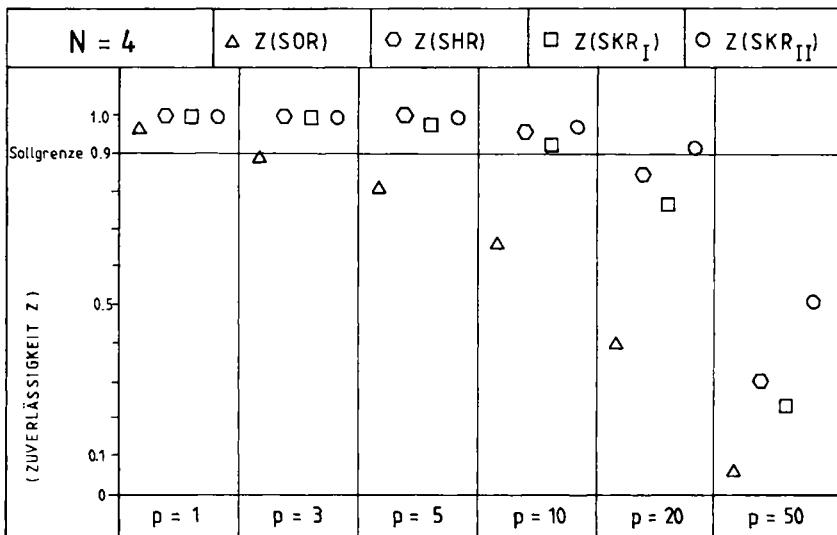
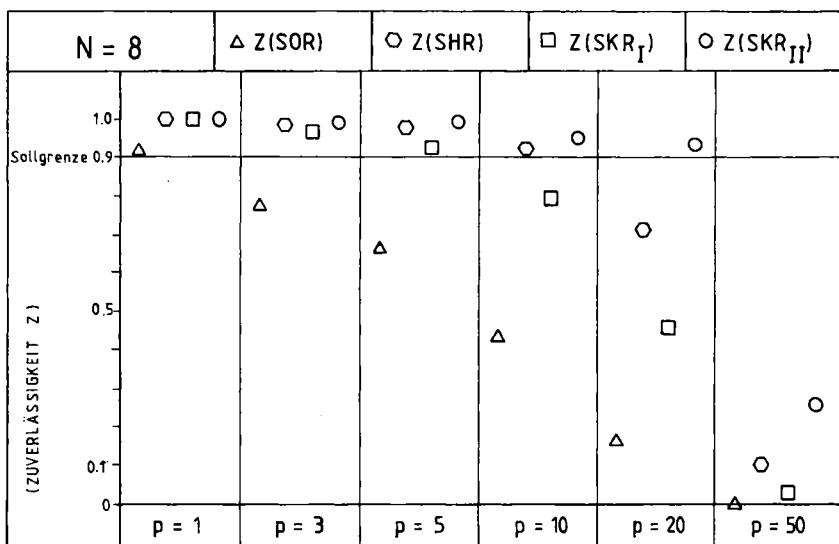
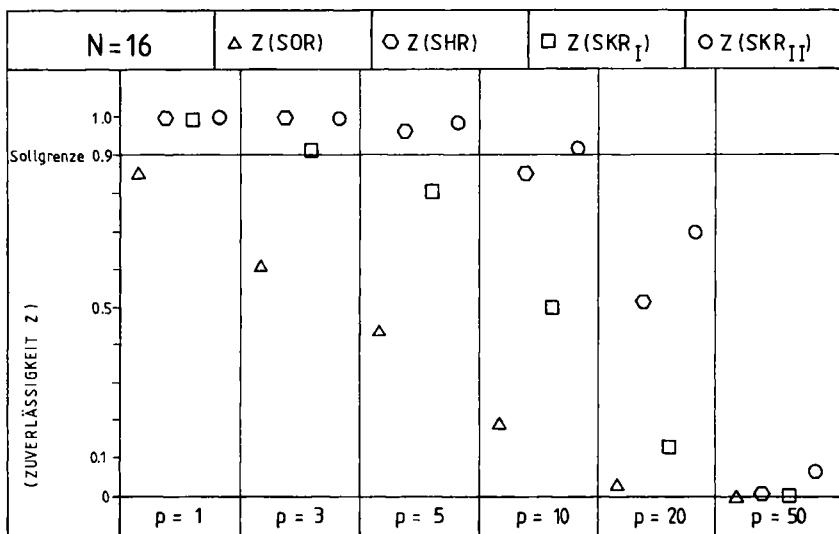


Abb. 31: Systemzuverlässigkeiten der Straßenverkehrssignalanlagen bei Systemumfang $N = 4$

5.3.3. Die Wirtschaftlichkeit der Systeme

Bei der Anschaffung eines Straßenverkehrssignalanlagen-Systems werden neben der Zuverlässigkeit auch ökonomische Überlegungen in die Entscheidung, welches System angeschafft werden soll, miteinbezogen.

Abb. 32: Systemzuverlässigkeiten der Straßenverkehrssignalanlagen bei Systemumfang $N = 8$ Abb. 33: Systemzuverlässigkeiten der Straßenverkehrssignalanlagen bei Systemumfang $N = 16$

Lassen sich etwa durch unterschiedlich lange Inspektionszeiträume bei zwei Systemen gleiche Zuverlässigkeiten erreichen, so ist zu überlegen, welches System auf Dauer gesehen kostengünstiger ist. Wir wollen einmal davon ausgehen, daß das in der Anschaffung billigere System kürzere Inspektionszeiträume erfordert als das teurere System. Können dann die durch verlängerte Inspektionszeiträume einge-

sparten Ersetzungskosten die Mehrkosten bei der Anschaffung des teuren Systems auffangen?

A. Inspektionszeiträume und Mindestzuverlässigkeit der Systeme

Möchte man eine Mindestzuverlässigkeit β eines Systems sichern, so muß man zunächst den notwendigen Inspektionszeitraum $I_T(\beta)$ bestimmen. Dieser ergibt sich als Lösung der Gleichung

$$Z_{I_T(\beta)} = \beta.$$

Möchte man etwa für ein System ohne Reserve aus acht Rotlichtsignalen eine Zuverlässigkeit $\beta = 0,9$ erreichen, so ergibt sich wegen

$$Z_{I_T(\beta)}(\text{SOR}) = e^{-N I_T(\beta) \lambda} = e^{-8 \lambda I_T(\beta)} \stackrel{!}{=} \beta = 0,9$$

gerade

$$-8 \cdot \lambda \cdot I_T(\beta) = \ln 0,9$$

und somit

$$I_T(\beta) = -\frac{1}{8 \cdot \lambda} \ln 0,9 = \frac{1}{\lambda} \cdot 0,0132,$$

d.h. der Inspektionszeitraum ist als das 0,0132-fache der erwarteten Lebensdauer der verwandten Glühbirne zu wählen.

Bei den übrigen Systemen ist $I_T(\beta)$ oft nicht so einfach zu berechnen; hier muß man dann ein iteratives Verfahren verwenden. Als Startwert kann man

$$I_T^0(\beta) = \frac{1}{\lambda} \cdot \ln 1/\beta$$

verwenden, denn alle betrachteten Systeme haben eine monoton wachsende Ausfallrate, so daß $I_T^0(\beta)$ eine Zuverlässigkeit von mindestens β garantiert, sofern

$$I_T^0(\beta) \leq \frac{1}{\lambda}$$

ist. In Tab. 19 sind die Inspektionszeiträume $I_T(\beta)$ der Systeme SHR und SKR_{II} für verschiedene Zuverlässigkeiten β und verschiedene Systemgrößen N in Einheiten der mittleren Lebensdauer $1/\lambda$ der verwandten Glühbirnen angegeben.

Tab. 19: Inspektionszeiträume $I_T(\beta)$ für SHR und SKR_{II} in Einheiten der birnenspezifischen Lebenserwartung $1/\lambda$

System	β	N		
		4	8	16
SHR	0,90	0,175	0,121	0,084
	0,95	0,119	0,083	0,058
	0,99	0,051	0,036	0,025
SKR _{II}	0,90	0,247	0,171	0,119
	0,95	0,168	0,117	0,082
	0,99	0,072	0,050	0,035

Wie diese Tabelle zu lesen ist, soll an einem Beispiel verdeutlicht werden: Um beim System SHR mit $N = 4$ Rotlichtsignalen eine Zuverlässigkeit von $\beta = 0,9$ zu erreichen, darf das Produkt $I_T \cdot \lambda$ höchstens den Wert 0,175 erreichen.

Die teilweise erheblichen Verkürzungen der Inspektionszeiträume $I_T(\beta)$ für größere N röhren daher, daß die erwarteten Systemlebensdauern mit wachsendem N abnehmen. In Tab. 20 sind die erwarteten Systemlebensdauern μ_{SHR} und $\mu_{\text{SKR}_{\text{II}}}$ in Einheiten der birnenspezifischen mittleren Lebensdauer $1/\lambda$ angegeben. Für $N = 4$ ist etwa

$$\mu_{\text{SHR}} = 0,582 \cdot \frac{1}{\lambda}.$$

Tab. 20: Erwartete Systemlebensdauer in Einheiten der erwarteten Lebensdauer $1/\lambda$ der Glühbirnen für die Systeme SHR und SKR_{II} bei verschiedenen Systemumfängen

System \ N	4	8	16
SHR	0,582	0,381	0,255
SKR _{II}	0,805	0,531	0,357

B. Systemkosten bei Mindestzuverlässigkeit der Systeme

Möchte man nun für eine konkrete Straßenverkehrssignalanlage, basierend auf standortspezifischen Kostenansätzen, eine Wirtschaftlichkeitsrechnung durchführen, so kann dies mit Hilfe einer Detailanalyse geschehen, wie sie im folgenden exemplarisch dargestellt wird.

Grundsätzlich ist es möglich, daß ein System vor Ablauf des Inspektionszeitraumes mehrmals ausfällt. Es soll jedoch vereinfachend angenommen werden, daß jedes der Systeme im Inspektionszeitraum höchstens einmal und zwar mit Wahrscheinlichkeit $1 - Z$ ausfällt, falls mit Z die Mindestzuverlässigkeit des entsprechenden Systems im Inspektionszeitraum bezeichnet wird. Daß diese Annahme nicht allzu restriktiv ist, sieht man daran, daß die Wahrscheinlichkeit für einen **mehrfachen Ausfall** des Systems vor Ablauf des Inspektionszeitraumes durch den Wert $F(I_T)^2$ beschränkt ist. Besitzt das System also eine Zuverlässigkeit von 95%, so beträgt die Wahrscheinlichkeit von mindestens zwei Sondereinsätzen höchstens $(1 - 0,95)^2 = 0,0025 = 0,25\%$, kann also im allgemeinen vernachlässigt werden.

Weiterhin wird davon ausgegangen, daß der Inspektionsplan in jedem Falle eingehalten wird.

Wir wollen nun speziell die Systeme SHR und SKR_{II} mit $N = 8$ Rotlichtsignalen betrachten. Es bezeichnen hier K_{SHR} , $K_{\text{SKR}_{\text{II}}}$ die jeweiligen Anschaffungskosten, k die artspezifischen variablen Kosten einschließlich Birnenkosten pro Einsatz, die wir zur Vereinfachung für beide Systeme als gleich ansetzen wollen. Es wird eine Mindestzuverlässigkeit von 95% angestrebt. Werden beide Systeme mit derselben Sorte von Glühbirnen bestückt, sind also die birnenspezifischen Ausfallraten λ gleich, so ergibt sich aus Tabelle 19

$$I_{T,\text{SHR}} = \frac{0,083}{\lambda}, \quad I_{T,\text{SKR}_{\text{II}}} = \frac{0,117}{\lambda}.$$

Als Gesamtkosten der beiden Systeme für T Zeiteinheiten kann nun berechnet werden

$$\begin{aligned} K_{\text{SHR}}^G &= K_{\text{SHR}} + \frac{k \cdot T}{I_{T,\text{SHR}}} + \frac{k \cdot T \cdot (1 - 0,95)}{I_{T,\text{SHR}}} \\ &= K_{\text{SHR}} + k \cdot T \cdot \lambda \cdot \left(\frac{1}{0,083} + \frac{0,05}{0,083} \right) \\ &= K_{\text{SHR}} + k \cdot T \cdot \lambda \cdot 12,65 \end{aligned}$$

und

$$\begin{aligned} K_{\text{SKR}_{\text{II}}}^G &= K_{\text{SKR}_{\text{II}}} + \frac{k \cdot T}{I_{T,\text{SKR}_{\text{II}}}} + \frac{k \cdot T \cdot (1 - 0,95)}{I_{T,\text{SKR}_{\text{II}}}} \\ &= K_{\text{SKR}_{\text{II}}} + k \cdot T \cdot \lambda \cdot \left(\frac{1}{0,117} + \frac{0,05}{0,117} \right) \\ &= K_{\text{SKR}_{\text{II}}} + k \cdot T \cdot \lambda \cdot 8,97. \end{aligned}$$

Sind die Anschaffungskosten für das System SHR größer oder gleich denen für das System SKR_{II}, so ist letzteres System natürlich in jedem Fall das kostengünstigere.

Ansonsten haben die beiden Kostenfunktionen den in Abb. 34 angegebenen Verlauf. Bis zum Kostenpunkt

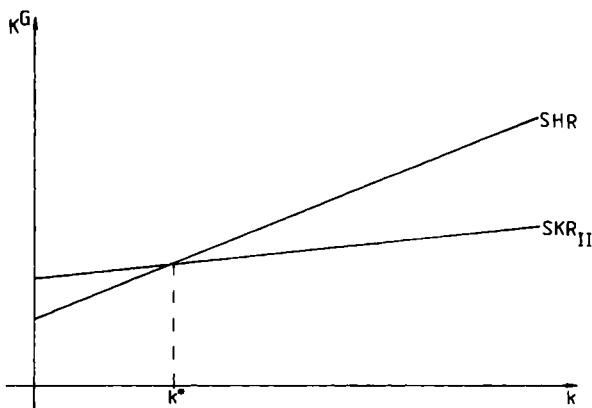


Abb. 34: Verlauf der Kostenfunktionen der Systeme SHR und SKR_{II}

$$k^* = \frac{K_{\text{SKR}_{\text{II}}} - K_{\text{SHR}}}{3,68 \cdot T \cdot \lambda}$$

ist das System SHR das kostengünstigere, darüber hinaus ist die Situation umgekehrt. Die obige Grafik zeigt, daß die Frage, welches System das kostengünstigere ist, in der Regel nicht generell beantwortet werden kann, sondern von der kosten-spezifischen Situation der Stadt abhängt (die sich in den variablen Kosten k und der Wahl der Planungsperiode T niederschlägt). Setzt man zur Demonstration die Planungsperiode einmal willkürlich gerade so fest, daß (für den Fall $K_{\text{SHR}} < K_{\text{SKR}_{\text{II}}}$)

$$k \cdot T \cdot \lambda = K_{\text{SHR}}$$

gilt, so wird in diesem speziellen Fall das System SKR_H solange den Vorzug bekommen, wie die Anschaffungskosten für das System SKR_U das 4,68-fache der Anschaffungskosten für SHR nicht übersteigen.

Natürlich sind in einer realen Bewertung die Faktoren k und λ orts- bzw. fabrikats-spezifisch vorgegeben und lediglich die Festlegung der Planungsperiode T liegt in den Händen des Entscheidungsträgers. Zu weiteren, auch allgemeineren Wirtschaftlichkeitsbetrachtungen in diesem Problemkreis sei auf Hartung/Jöckel (1982) verwiesen.

6. Verfügbarkeit von Systemen und Instandhaltungsstrategien

Die Lebensdauer eines wartbaren Systems lässt sich in Arbeitsperioden (Funktionsperioden) und Wartungsperioden einteilen. Natürlich möchte man erreichen, daß die Funktionsperioden möglichst lang und die Wartungsperioden möglichst kurz sind.

6.1. Die Verfügbarkeit von Systemen

A. Momentane Verfügbarkeit und Dauerverfügbarkeit

Die **Verfügbarkeit A** (availability) ist nun die Wahrscheinlichkeit dafür, daß ein System zu einem beliebigen Zeitpunkt funktionsfähig ist. Sie stellt somit eine Beziehung zwischen Funktions- und Wartungsperioden her. Die **momentane Verfügbarkeit A(t)** eines Systems gibt die Wahrscheinlichkeit an, mit der das System zum Zeitpunkt t funktionsfähig ist. Die **Dauerverfügbarkeit (stationäre Verfügbarkeit)**

$$A_D = \lim_{t \rightarrow \infty} A(t) \quad (\text{falls der Grenzwert existiert})$$

ist eine Kenngröße für die Verfügbarkeit eines Systems in einem längeren Zeitraum.

Diese Kenngröße lässt sich aus der **mittleren Funktionszeit τ_f** und der **mittleren Reparaturzeit τ_r** eines Systems zu

$$A_D = \frac{\tau_f}{\tau_f + \tau_r}$$

berechnen. Natürlich gilt $0 \leq A_D \leq 1$.

Im Gegensatz zur Überlebenswahrscheinlichkeit ist die Verfügbarkeit eines Systems auch gegen Ende seiner Lebensdauer hoch, denn das System muß auch dann noch seine volle Funktion erfüllen.

B. Punkt- und Intervallschätzungen, Testen von Hypothesen über die Dauerverfügbarkeit

Betrachtet man nun an einem System n_1 Funktionszeiten t_{fj} zwischen 2 Ausfällen (jeweils) und n_2 Reparaturdauern t_{ri} , so kann man die mittlere Funktionszeit τ_f des Systems durch

$$\hat{\tau}_f = \frac{1}{n_1} \sum_{j=1}^{n_1} t_{fj}$$

und die mittlere Reparaturzeit τ_r durch

$$\hat{\tau}_r = \frac{1}{n_2} \sum_{i=1}^{n_2} t_{ri}$$

schätzen, so daß sich mit

$$\hat{A}_D = \frac{\hat{\tau}_f}{\hat{\tau}_f + \hat{\tau}_r}$$

ein Schätzer für die stationäre Verfügbarkeit A_D des Systems ergibt.

Geht man nun davon aus, daß die t_{fj} einer Exponentialverteilung mit Parameter $1/\tau_f$ und die t_{ri} einer Exponentialverteilung mit Parameter $1/\tau_r$ entstammen, so kann man Tests über die stationäre Verfügbarkeit A_D durchführen und Konfidenzintervalle für A_D bestimmen, vgl. Thompson (1966).

Mit Wahrscheinlichkeit $1 - \alpha$ gilt dann

$$A_D \geq \frac{\hat{\tau}_f}{\hat{\tau}_f + \hat{\tau}_r \cdot F_{2n_1, 2n_2; 1-\alpha}},$$

und ein zweiseitiges Konfidenzintervall für A_D zum Niveau $1 - \alpha$ ist

$$\left[\frac{\hat{\tau}_f}{\hat{\tau}_f + \hat{\tau}_r \cdot F_{2n_1, 2n_2; 1-\alpha/2}} ; \frac{\hat{\tau}_f}{\hat{\tau}_f + \hat{\tau}_r \cdot F_{2n_1, 2n_2; \alpha/2}} \right];$$

die γ -Quantile $F_{2n_1, 2n_2; \gamma}$ der $F_{2n_1, 2n_2}$ -Verteilung sind dabei im Anhang vertafelt.

Will man zum Niveau α die Hypothese

$$H_0: A_D \leq A_D^0 \text{ gegen } H_1: A_D > A_D^0$$

testen, so bestimmt man den Wert der Teststatistik

$$F = \frac{\hat{\tau}_f \cdot (1 - A_D^0)}{\hat{\tau}_r \cdot A_D^0}$$

und verwirft die Hypothese H_0 , d. h. man schließt, daß die stationäre Verfügbarkeit des betrachteten Systems zum Niveau α signifikant höher ist als A_D^0 , falls gilt:

$$F > F_{2n_1, 2n_2; 1-\alpha}.$$

Weitere Tests und Konfidenzintervalle für die stationäre Verfügbarkeit A_D findet man etwa bei Gray/Schucany (1969).

Beispiel: Bei einem System wurden die $n_1 = 5$ Ausfallzwischenzeiten t_{fj} ($j = 1, \dots, 5$)

$$7,8; 9,9; 3,2; 15,7 \text{ und } 14,6$$

sowie die $n_2 = 7$ Reparaturdauern t_{ri} ($i = 1, \dots, 7$)

$$0,36; 0,52; 0,17; 0,28; 0,13; 0,40; 0,05$$

beobachtet. Der Schätzer für die mittlere Funktionszeit τ_f des Systems ist dann

$$\hat{\tau}_f = \frac{1}{5} (7,8 + 9,9 + 3,2 + 15,7 + 14,6) = \frac{51,2}{5} = 10,24$$

und ein Schätzer für die mittlere Reparaturzeit τ_r des Systems ist

$$\hat{\tau}_r = \frac{1}{7} (0,36 + 0,52 + 0,17 + 0,28 + 0,13 + 0,40 + 0,05) = 0,273.$$

Somit kann die Dauer Verfügbarkeit durch

$$\hat{A}_D = \frac{10,24}{10,24 + 0,273} = 0,974$$

geschätzt werden. Ein zweiseitiges Konfidenzintervall zum Niveau 0,90 für A_D ergibt sich hier zu

$$\begin{aligned} & \left[\frac{10,24}{10,24 + 0,273 \cdot F_{10,14; 0,95}} ; \frac{10,24}{10,24 + 0,273 \cdot F_{10,14; 0,05}} \right] \\ &= \left[\frac{10,24}{10,24 + 0,273 \cdot 2,60} ; \frac{10,24}{10,24 + 0,273 \cdot 0,35} \right] = \left[\frac{10,24}{10,95} ; \frac{10,24}{10,335} \right] \\ &= [0,935; 0,991], \end{aligned}$$

mit $F_{10,14; 0,05} = 1/F_{14,10; 0,95} = 1/2,86 = 0,35$.

Wir wollen nun noch zum Niveau $\alpha = 0,01$ die Hypothese

$$H_0: A_D \leq 0,90 \quad \text{gegen} \quad H_1: A_D > 0,90$$

testen. Da gilt

$$F = \frac{10,24 \cdot (1 - 0,9)}{0,273 \cdot 0,9} = 4,168 > 3,94 = F_{10,14; 0,99} = F_{2n_1, 2n_2; 1-\alpha},$$

muß die Nullhypothese verworfen werden, d. h. bei einer Irrtumswahrscheinlichkeit von 1% ist die stationäre Verfügbarkeit A_D des Systems signifikant größer als 0,9.

6.2. Methoden zur Erhöhung der Verfügbarkeit

Ein Nachteil des Verfügbarkeitsbegriffes ist, daß er keinerlei Information über die Konsequenz eines Ausfalls einbezieht. Beispielsweise kann für Tragseile von Seilbahnen oder Aufzügen eine Verfügbarkeit errechnet werden, die nahe bei 1 liegt (extrem lange Lebensdauer, kurze Reparaturdauer), jedoch zwingt die Möglichkeit, durch einen Ausfall Menschenleben zu gefährden, zu weiteren Maßnahmen. Hier ist eine starre Instandsetzung erforderlich, die unabhängig vom Zustand des Systems in starren Zeitabständen einen Austausch der gewissen Ermüdungserscheinungen ausgesetzten Elementen vorschreibt.

Abschließend soll kurz auf die Aktionen eingegangen werden, die zur Erhöhung der Verfügbarkeit beitragen.

A. Die Redundanzplanung

Eine Möglichkeit zur Verbesserung der Verfügbarkeit bietet die Redundanzplanung, deren Effekt der **Erhöhung der mittleren Funktionsdauer τ_r** durch Hinzuschalten redundanter Komponenten ebenso wie deren Begrenzung durch wirtschaftliche (Kostensteigerung) oder technologische (Volumen-, Gewichtssteigerung) Überlegungen schon in Abschnitt 1 dieses Kapitels besprochen wurde.

B. Die vorbeugende Instandsetzung

Eine andere Möglichkeit zur Erhöhung der Verfügbarkeit eines Systems besteht in der **Verringerung der mittleren Reparaturdauer τ_r** . Ein geeignetes Instrumentarium hierzu ist die vorbeugende Instandsetzung. Während wir Instandsetzung als Fehlerbeseitigung kennengelernt haben, wollen wir unter dem Begriff **vorbeugende Instandsetzung** die Ersetzung von noch funktionsfähigen, arbeitenden Elementen durch neue Elemente verstehen.

Natürlich erweist sich die vorbeugende Instandsetzung höchstens dann als sinnvoll, wenn die Systemelemente einer „Alterung“ unterliegen, also zumindest der NBUE-Verteilungsklasse angehören. Die meisten Untersuchungen zu diesem Problemkreis fordern sogar die IFR-Eigenschaft als Grundlage vorbeugender Instandsetzungsaktionen, vgl. Abschnitt 5.

Die Tatsache, daß die vorbeugende Instandsetzung τ_r zu verringern vermag, liegt darin begründet, daß durch einen Ausfall oft eine komplexere Instandsetzung nötig wird, da die durch den Ausfall angerichteten Schäden mitberücksichtigt werden müssen. Organisatorische Gründe (Bereitstellung von Personal und Material) ergeben einen weiteren Vorteil der vorbeugenden Instandsetzung, da andernfalls höhere Reparaturzeiten durch den Transport der Instandsetzungshilfsmittel zum ausgefäl- lenen System entstehen.

Eine Übersicht über die wesentlichen vorbeugenden Instandsetzungsaktionen findet man bei Barlow/Hunter (1960), Rinne (1972).

B1. Bereitschafts- und Präventivstrategien

Man unterscheidet etwa zwischen Bereitschaftsstrategien und Präventivstrategien. Bei **Bereitschaftsstrategien** ist der Produktzustand nicht permanent bekannt. Sie finden ihre Anwendung daher bei der vorbeugenden Instandsetzung von Systemen, die nur zu bestimmten Zeiten benötigt werden (z. B. Warnsysteme, Radargeräte, technische Sicherheitsanlagen), sowie militärischen Waffensystemen. Durch Anwendung dieser Strategie soll eine **maximale Einsatzbereitschaft** erreicht werden.

Bei **Präventivstrategien** hingegen ist der Produktzustand jederzeit bekannt. Die Quelle der Unsicherheit liegt lediglich im exakten Zeitpunkt der Funktionsminde- rung. Im Falle von Systemen, die genau zwei Zustände annehmen können, ist dies also der Ausfallzeitpunkt. Präventivstrategien für solche Systeme werden als **einstufige Strategien** bezeichnet.

B2. Periodische, sequentielle und optionale Strategien

Eine weitere Unterteilung resultiert aus den verschiedenen Möglichkeiten, die bei der Festlegung des Instandsetzungszeitraumes gegeben sind. **Periodische Strategien** sind hierbei durch eine einmalige Berechnung eines festen Instandsetzungszeit- raumes gekennzeichnet, wohingegen **sequentielle Strategien** eine Neuberechnung des Instandsetzungzeitraumes nach jeder Instandsetzungsaktion vorsehen. Eine Mischform von periodischen und sequentiellen Strategien sind die **optionalen Strategien**, die sich nach „günstigen“ Gelegenheiten wie etwa Pausen, Schichtwechsel oder anderen natürlichen Produktionsunterbrechungen des Systems richten, um konstante oder variable Instandsetzungszeiträume zu verwenden. Hierbei muß individuell entschieden werden, ob die sich bietende Gelegenheit zur Instandsetzung genutzt werden soll.

Richten wir unser Augenmerk auf die **periodischen Präventivstrategien**, so kön-

nen diese **strengperiodisch** (Auswechseln der Zündkerzen eines Kraftfahrzeugs, Auswechseln der Glühbirnenbestückung einer Straßenverkehrssignalanlage) oder aber zufällig periodisch (Beseitigung verbliebener Schmelzreste bei Hochöfen) durchgeführt werden.

Als eine einfache einstufige strengperiodische Präventivstrategie haben wir bereits die **Altersersatzstrategie** kennengelernt, vgl. Abschnitt 5. Ausgehend von dieser Strategie soll im folgenden eine **optimale Länge t_0 des Instandsetzungszeitraumes** ermittelt werden, wobei die „Optimalität“ in der Maximierung der stationären Verfügbarkeit AD des betrachteten Systems liegt. Ist F die bekannte absolut stetige Verteilungsfunktion der Arbeitszeit eines Systems, T die mittlere Arbeitszeit des Systems, T_s die mittlere Dauer einer vorbeugenden Instandsetzung und T_e die mittlere Dauer einer Ausfallinstandsetzung, so ergibt sich die optimale Länge t_0 des Instandsetzungszeitraumes im Fall $T_e > T_s$ (sonst lohnt sich eine vorbeugende Instandsetzung nicht) aus der Gleichung

$$r(t_0) \int_0^{t_0} \bar{F}(t) dt - F(t_0) - \frac{T_e}{T_s} + 1 = 0 .$$

Hierbei bezeichnet $r(t_0)$ die Ausfallrate des Systems in t_0 , also

$$r(t_0) = \frac{f(t_0)}{\bar{F}(t_0)} .$$

Die Lösung dieser Gleichung ist eindeutig, wenn $r(t)$ monoton nicht fallend ist. Die **maximale Verfügbarkeit des Systems** ist dann

$$A_{D_{\max}}^* = \frac{1}{1+r(t_0) \cdot (T_e - T_s)} .$$

7. Vergleichende Lebensdauerprüfungen: Log-Rank Test, Wilcoxon-Gehan Test, Tarone-Ware Tests und Cox's Proportional Hazards Regression Model

Aus methodischen und technischen Gründen werden vergleichende Lebensdauerprüfungen erst in den Abschnitten 5.1 und 5.6 von Kapitel XV behandelt.

8. Schätzen der [Über-] Lebensdauerverteilung: Kaplan-Meier Survival Analysis

Auch die Kaplan-Meier Survival Analysis findet man aus methodischen Gründen erst im Abschnitt 5.3 von Kapitel XV.

Kapitel XIV: Explorative Datenanalyse (EDA) und Robuste Verfahren

In diesem Kapitel werden wir uns mit Methoden der **explorativen Datenanalyse (EDA)** und der **robusten Statistik** beschäftigen.

Die **explorative Datenanalyse**, die in den letzten Jahren vielfach propagiert wurde, beinhaltet im wesentlichen Methoden der deskriptiven Statistik, die wir bereits in Kap. I kennengelernt haben. Diese Verfahren werden zum Teil direkt übernommen, zum Teil etwas verfeinert und enthalten teilweise auch neue, interessante Aspekte der deskriptiven Analyse von Daten. Warum wir hier recht ausführlich in den Abschnitten 1 bis 3 auf sie eingehen, ist darin begründet, daß sich im Rahmen der EDA viele Begriffe eingebürgert haben, die inzwischen in der Literatur zumeist ohne nähere Erklärung Verwendung finden.

Wir werden uns im **Abschnitt 1** zunächst mit explorativen Methoden für ein beobachtetes Merkmal beschäftigen. Kenngrößen für Lage und Streuung einer Beobachtungsreihe basieren in der EDA im wesentlichen auf Median und Größen, die den in Kap. I eingeführten empirischen Quantilen ähneln. Einen besseren Überblick über eine Beobachtungsreihe gewinnt man mitunter durch Symmetrisierung der Daten. Hierzu werden in der EDA speziellere Datentransformationen als die in Abschnitt 4.5 des Kap. V behandelten eingesetzt. Graphisch werden die wesentlichen Kenngrößen einer Beobachtungsreihe durch Box-Plots dargestellt und eine verfeinerte Form der Strichlisten ergibt sich in Form der Stamm-und-Blätter-Darstellungen. Genauso wie in der deskriptiven Statistik werden Histogramm und empirische Verteilungsfunktion in der EDA eingesetzt. Histogramme können auch als empirische Dichten aufgefaßt werden, jedoch sind empirische Dichten, die auf sogenannten Kernschätzungen basieren, oft besser geeignet, die Verteilung der Daten einer Beobachtungsreihe zu charakterisieren. Wurzeldiagramme ermöglichen es, unabhängig von den Häufigkeiten in den Klassen (in welche die Beobachtungswerte eingeteilt werden), die Abweichungen der Verteilung einer Beobachtungsreihe von einer theoretischen Verteilung zu beurteilen; dem selben Zweck – nämlich der Überprüfung von Verteilungsannahmen – dienen auch die sogenannten Q-Q-Plots. Anstelle von Wurzeldiagrammen und Q-Q-Plots können natürlich auch das in Kap. IV vorgestellte Wahrscheinlichkeitspapier oder Anpassungstests wie z. B. der χ^2 -Test oder der Kolmogoroff-Smirnov-Anpassungstest, vgl. Kap. IV, zur Überprüfung einer Verteilungsannahme eingesetzt werden.

Werden je zwei Merkmale an n Objekten beobachtet, so lassen sich die Daten in einem zweidimensionalen Koordinatensystem als Punkteschar darstellen. Im **Abschnitt 2** behandeln wir zunächst ein exploratives Verfahren zur Glättung einer solchen Punkteschar, das anstelle der in Abschnitt 1.3 des Kap. XII vorgestellten Glättungsverfahrens für Zeitreihen eingesetzt werden kann. Sodann wird eine explorative Alternative zur in Kap. X betrachteten Schätzung von Regressionsgeraden behandelt, und schließlich beschäftigen wir uns mit der Linearisierung zweidimensionaler Punktescharen durch Datentransformation. Bestimmt man (explorativ) dann eine Regressionsgerade für die linearisierten Daten, so kann durch Rücktransformation eine explorative Regressionsfunktion für die Ursprungsdaten bestimmt werden.

Werden mehr als zwei Merkmale beobachtet, so kommt man zu den in **Abschnitt 3** behandelten multivariaten explorativen Verfahren. Konkret stellen wir hier nur

den Scatter-Plot vor, der einen graphischen Überblick über die Daten bzw. über die Objekte gibt. Andere graphische Verfahren für mehrdimensionale Daten und Verfahren der multivariaten Statistik, die einer explorativen Interpretation zugänglich sind, werden in einem kurzen Abriß beschrieben; zur konkreten Vorgehensweise vgl. man z.B. Hartung/Elpelt (1984).

Methoden zur Parameterschätzung aus Beobachtungsdaten, die unempfindlich gegenüber Ausreißern sind, stellt die **robuste Statistik** zur Verfügung. Ein einfaches Beispiel für die robuste Schätzung der Lokation, der Lage einer Beobachtungsreihe ist der auch in der EDA und der deskriptiven Statistik vielfach eingesetzte Median.

Wir werden uns in **Abschnitt 4** zunächst mit Kenngrößen für die Robustheit und mit speziellen robusten Skalenschätzern, das sind Schätzer für die Streuung in einer Beobachtungsreihe, beschäftigen. Sodann werden drei Klassen robuster Schätzer speziell für die Lokationsschätzung einer Beobachtungsreihe – nämlich M-, L- und R-Schätzer für die Lokation – behandelt.

Der zunächst wesentlichste Unterschied zwischen üblichen statistischen Verfahren und explorativer Datenanalyse bzw. robuster Statistik ist der, daß als Kenngrößen für Lage und Streuung einer Beobachtungsreihe nicht mehr arithmetisches Mittel \bar{x} und empirische Standardabweichung s sondern gegenüber Ausreißern in den Daten unempfindliche Kenngrößen, die vielfach auf dem Median m und empirischen α -Quantilen basieren, verwandt werden. Dies ist nicht vom Ansatz sondern von der Begründung her neu. Schon immer wurden aus Gründen der leichteren Berechenbarkeit vor allem im technischen Bereich derartige Kenngrößen verwandt, vgl. Abschnitt 1.3 in Kap. IV, wo einige spezielle Kenngrößen angegeben sind. Wir wollen hier noch einige weitere wohl bekannte Maßzahlen für Lage und Streuung (Schätzer für Erwartungswert μ und Standardabweichung σ), die unter dem Namen **short-cut methods** zusammengefaßt werden, angeben, vgl. etwa Miller/Freund (1965). Bezeichnet \tilde{x}_α das empirische α -Quantil, vgl. Abschnitt 4 in Kap. I, einer Beobachtungsreihe x_1, \dots, x_n , so sind der **Median**

$$\tilde{x}_{0,5} = m$$

und das **midquartile**

$$\frac{1}{2}(\tilde{x}_{0,25} + \tilde{x}_{0,75})$$

geeignete Lagemaße, wenn die Daten in etwa eine symmetrische Verteilung aufweisen. Die entsprechenden Maßzahlen der EDA sind – wie hier – der Median bzw. der midhinge, vgl. Abschnitt 1.1. Als short-cut Streuungsmaß wird bei symmetrischer Verteilung auch

$$\frac{1}{3}(\tilde{x}_{15/16} - \tilde{x}_{1/16}) = \frac{1}{3}(\tilde{x}_{0,9375} - \tilde{x}_{0,0625})$$

verwandt, was in etwa der Schätzung für σ mittels d-spread in der EDA entspricht, vgl. Abschnitt 1.1. Weniger empfindlich gegenüber Abweichungen von der Symmetrie ist das Streuungsmaß

$$\frac{1}{4}(\tilde{x}_{0,9375} + \frac{3}{4} \cdot \tilde{x}_{0,75} - \frac{3}{4} \cdot \tilde{x}_{0,25} - \tilde{x}_{0,0625}),$$

eine Kombination aus der Schätzung für σ mittels h-spread und d-spread in der EDA.

Der weitergehend interessierte Leser sei bzgl. der explorativen Datenanalyse verwiesen auf die Bücher von Tukey (1977), Velleman/Hoaglin (1981), Chambers et al. (1983), Cleveland et al. (1983) sowie Hoaglin et al. (1983) und bzgl. robuster Verfah-

ren auf Gastwirth (1966), Gastwirth/Cohen (1970), Jurečková (1971), Koul (1971), Andrews et al. (1972), Jaeckel (1972), Huber (1973, 1981), Barnett/Lewis (1978), Schmitz (1979), Hoaglin et al. (1983), Ketellapper/Ronner (1984) sowie auf den Übersichtsartikel von Heiler (1980).

1. Verfahren für einzelne Merkmale in der EDA

In diesem Abschnitt werden wir explorative Analysemethoden für eine einzelne Beobachtungsreihe x_1, \dots, x_n für ein Merkmal behandeln.

Zunächst werden wir in den Abschnitten 1.1 und 1.2 empirische Kenngrößen der explorativen Datenanalyse für Einzeldaten und gruppierte Daten kennenlernen, die z. T. auch schon im Rahmen der Deskriptiven Statistik in Kap. I behandelt wurden.

Im Abschnitt 1.3 werden wir dann auf die im Abschnitt 4 des Kap. V behandelten Datentransformationen zurückkommen und speziell auf ihren Einsatz zur Symmetrisierung einer Beobachtungsreihe eingehen.

Der Box-Plot, der in Abschnitt 1.4 dargestellt wird, dient der graphischen Veranschaulichung einer Beobachtungsreihe und dem Vergleich mehrerer Beobachtungsreihen.

Stem-and-Leaves sowie Histogramme dienen der Veranschaulichung der empirischen Verteilung einer Beobachtungsreihe; sie werden in den Abschnitten 1.5 und 1.6 behandelt. Im Abschnitt 1.6 werden wir zudem eine Alternative zu den in Kap. I betrachteten empirischen Verteilungsfunktionen kennenlernen: Die stetige empirische Verteilungsfunktion ist die Übertragung des Histogramms, interpretiert als Dichte, in eine Verteilungsfunktion. Alternativ zum Histogramm können zur Dichteschätzung auch sogenannte Kernschätzer oder empirische Dichten, wie sie in Abschnitt 1.7 behandelt werden, verwendet werden.

Zum Vergleich einer Beobachtungsreihe mit einer theoretischen Verteilung dienen sowohl das in Abschnitt 1.8 dargestellte Wurzeldiagramm als auch die in Abschnitt 1.9 behandelten Quantile-Quantile-Plots, die eine Alternative zum Wahrscheinlichkeitspapier, vgl. Abschnitt 1.7 in Kap. IV, sind.

1.1 Empirische Kenngrößen

Im Kapitel 1 haben wir Maße zur Beurteilung von Lage und Streuung einer Gruppe von Daten x_1, \dots, x_n , die Beobachtungswerte für ein Merkmal an n Objekten $1, \dots, n$ repräsentieren, behandelt. Unter anderem haben wir dort den **Median** oder das **empirische 0,5-Quantil** $\tilde{x}_{0,5}$ kennengelernt, der im Rahmen der explorativen Datenanalyse eine herausragende Bedeutung als Schätzung für das Zentrum der Daten x_1, \dots, x_n hat und hier an die Stelle des sonst gebräuchlichsten Lagemaßes, des arithmetischen Mittels \bar{x} als Mittelwertschätzung tritt.

Der Median, den wir hier mit $m = \tilde{x}_{0,5}$ bezeichnen wollen, ist im Abschnitt 4.2.1 des Kap. I wie folgt definiert worden. Ist $x_{(1)}, \dots, x_{(n)}$ die Reihe der größtmäßig geordneten Daten x_1, \dots, x_n ,

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

so ist

$$m = \tilde{x}_{0,5} = \begin{cases} x_{((n+1)/2)}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(n/2)} + x_{((n+2)/2)}), & \text{falls } n \text{ gerade} \end{cases}$$

der Median der Beobachtungsreihe. In der EDA, wo der Median auch **summary point** genannt wird, ist eine andere hierzu äquivalente Definition üblich, die auf dem **Tiefe-Index** der Daten x_1, \dots, x_n basiert. Die Tiefe von $x_{(1)}$ und $x_{(n)}$ ist $T = 1$, die von $x_{(2)}$ und $x_{(n-1)}$ ist $T = 2$, die von $x_{(3)}$ und $x_{(n-2)}$ ist $T = 3$ usw. Bei ungeradem n ist schließlich die Tiefe von $x_{((n+1)/2)}$ gerade $T = (n+1)/2$ und bei geradem n haben $x_{(n/2)}$ und $x_{((n+2)/2)}$ die Tiefe $n/2$. Jeder Tiefe-Index T ist also genau zwei Beobachtungen zugeordnet, lediglich bei ungeradem n besitzt nur der Wert $x_{((n+1)/2)}$ den Index $T = (n+1)/2$. Um dennoch die Daten eindeutig anhand ihres Tiefe-Indexes identifizieren zu können, tritt zum Index T noch die Bezeichnung L für **lower** bzw. U für **upper**, wenn ein Wert in der geordneten Beobachtungsreihe links (unterhalb) bzw. rechts (oberhalb) der Mitte der Reihe steht. Die geordnete Beobachtungsreihe $x_{(1)}, \dots, x_{(n)}$ wird mit dem Tiefe-Index also geschrieben als

$$x_{1,L}, x_{2,L}, x_{3,L}, \dots, x_{(n-1)/2,L}, x_{(n+1)/2}, x_{(n-1)/2,U}, \dots, x_{3,U}, x_{2,U}, x_{1,U},$$

wenn n ungerade ist, und als

$$x_{1,L}, x_{2,L}, x_{3,L}, \dots, x_{n/2,L}, x_{n/2,U}, \dots, x_{3,U}, x_{2,U}, x_{1,U},$$

wenn n gerade ist. Der Median m wird nun über die Tiefe definiert als derjenige Wert mit Tiefe-Index

$$T(m) = (n+1)/2.$$

Ist n ungerade, so ist $T(m)$ eine ganze Zahl und der Median der Daten ist

$$m = x_{T(m)} = x_{(n+1)/2},$$

ist dagegen $T(m)$ keine ganze Zahl, d. h. ist n gerade, so ist

$$m = \frac{1}{2}(x_{T(m)-0.5,L} + x_{T(m)-0.5,U}) = \frac{1}{2}(x_{n/2,L} + x_{n/2,U}).$$

Ebenso werden weitere EDA-Kenngrößen definiert, die den α -Quantilen in der deskriptiven Statistik entsprechen. Die **hinges** (**Angelpunkte**) entsprechen in etwa den Quartilen. Sie sind definiert über die hinge-Tiefe

$$T(h) = ([T(m)] + 1)/2,$$

wobei $[T(m)]$ entweder $T(m)$ bedeutet, wenn $T(m)$ eine ganze Zahl ist, oder die größte ganze Zahl, die kleiner als $T(m)$ ist, wenn $T(m)$ keine ganze Zahl ist; die eckigen Klammern um $T(m)$ werden auch **Gauß-Klammern** genannt. Der **untere hinge** ist dann

$$h_L = \begin{cases} x_{T(h),L}, & \text{falls } T(h) \text{ eine ganze Zahl ist} \\ \frac{1}{2}(x_{T(h)-0.5,L} + x_{T(h)+0.5,L}), & \text{sonst} \end{cases},$$

und entsprechend ist der **obere hinge** gegeben als

$$h_U = \begin{cases} x_{T(h),U}, & \text{falls } T(h) \text{ eine ganze Zahl ist} \\ \frac{1}{2}(x_{T(h)-0.5,U} + x_{T(h)+0.5,U}), & \text{sonst} \end{cases}.$$

Die hinges sind identisch mit den Quartilen (0,25- bzw. 0,75-Quantil), falls $T(h)$ eine ganze Zahl ist und zusätzlich $n - 1$ oder $n - 2$ durch 4 teilbar ist. Ansonsten liegen sie etwas näher am Median als die Quartile.

Weiterhin werden dann in der EDA die Lagemaße e (eighthths), d (sixteenths), c (thirtyseconds) usw. analog zu den hinges berechnet. Die Tiefe der eighthths ist

$$T(e) = ([T(h)] + 1)/2,$$

die der sixteenths ist

$$T(d) = ([T(e)] + 1)/2,$$

usw. bis man schließlich zu Größen mit dem Tiefe-Index $T = 1$, also zu den Extremwerten der Beobachtungsreihe gelangt.

Die sogenannten **midsummarys** geben dann Auskunft über die **Schiefe der Verteilung** der Beobachtungsreihe, vgl. auch Abschnitt 4.5 in Kap. I. Der **Median** ist die erste midsummary, der **midhinge** $(h_L + h_U)/2$ die zweite, der **mideightths** $(e_L + e_U)/2$ die dritte, usw. bis man schließlich zum **midextreme**, dem Mittelwert der Extremwerte $x_{1,L}$ und $x_{1,U}$ gelangt. Sind die midsummarys nicht konstant, so ist die Verteilung unsymmetrisch, sind sie fallend, so ist die Verteilung linksschief und sind sie steigend, so ist die Verteilung rechtsschief. Eine entsprechende Klassifikation von Verteilungen könnten wir in Kap. I auch durch arithmetisches Mittel, Median und Modalwert vornehmen. Als Maße für die Streuung innerhalb einer Beobachtungsreihe haben wir im Abschnitt 5 des Kap. I etwa die Spannweite (range) oder den Quartilsabstand kennengelernt. Analog hierzu werden in der EDA der **hinge-Abstand (h-spread)**, der **eightths-Abstand (e-spread)** usw., also

$$h_U - h_L, e_U - e_L \text{ usw.},$$

als Streuungsmaße verwandt.

Aus diesen Abständen wird in der EDA auch die **empirische Standardabweichung** (als Schätzer für die Standardabweichung der zugrundeliegenden, theoretischen Verteilung) bestimmt. Diese empirischen Standardabweichungen basieren auf einer Normalverteilungsannahme und sind daher genau genommen auch nur bei Daten aus einer Normalverteilung zu rechtfertigen. Der hinge-Abstand bei einer Standardnormalverteilung $N(0, 1)$ ist die Differenz ihrer Quartile

$$u_{0,75} - u_{0,25} = 0,6745 - (-0,6745) = 1,349,$$

vgl. Tab. 2 des Anhangs, der eighthths-Abstand ist

$$u_{0,875} - u_{0,125} = 1,1505 - (-1,1505) = 2,301,$$

der sixteenth-Abstand ist

$$u_{0,9375} - u_{0,0625} = 1,5340 - (-1,5340) = 3,068$$

usw. Für beliebige $N(\mu, \sigma^2)$ -Verteilungen ergeben sich die entsprechenden Abstände durch Multiplikation obiger Zahlen mit σ ; es ist hier also

$$\sigma = h\text{-spread}/1,349 = e\text{-spread}/2,301 = d\text{-spread}/3,068 = \dots$$

Hiervom ausgehend wird die empirische Standardabweichung bestimmt durch die Abstände dividiert durch die entsprechenden Abstände bei $N(0, 1)$ -Verteilung.

Ist die Verteilung der Daten symmetrisch, so deuten konstante Standardabweichungen für die Daten auf eine Normalverteilung hin, sind sie fallend, so sind die „Schwänze“ der Verteilung schwächer, sind sie steigend, so sind die „Schwänze“ stärker ausgeprägt als bei einer Normalverteilung.

Beispiel: In der Tab. 35 des Abschnitts 4.1.4 von Kap. IV sind die Lebensdauern von $n = 16$ Kühlaggregaten angegeben. In der folgenden Tab. 1 ist die geordnete Beobachtungsreihe unter zusätzlicher Angabe der Tiefe-Indizes noch einmal aufgeführt.

Mit $n = 16$ ergibt sich aus Tab. 1 die Tiefe des Medians zu

Tab. 1: Lebensdauern $x_{(i)}$ von $n = 16$ Kühlaggregaten in Jahren (geordnete Beobachtungsreihe sowie Tiefe-Indizes)

i	T	$x_{(i)}$
1	1	$x_{1,L}$ 0,14
2	2	$x_{2,L}$ 0,27
3	3	$x_{3,L}$ 0,43
4	4	$x_{4,L}$ 0,68
5	5	$x_{5,L}$ 0,81
6	6	$x_{6,L}$ 1,14
7	7	$x_{7,L}$ 1,45
8	8	$x_{8,L}$ 1,82
9	8	$x_{8,U}$ 2,36
10	7	$x_{7,U}$ 2,53
11	6	$x_{6,U}$ 2,90
12	5	$x_{5,U}$ 3,45
13	4	$x_{4,U}$ 4,51
14	3	$x_{3,U}$ 5,12
15	2	$x_{2,U}$ 5,68
16	1	$x_{1,U}$ 7,84

$$T(m) = (n + 1)/2 = 17/2 = 8,5,$$

d. h. der Median der Beobachtungsreihe ist

$$m = \frac{1}{2}(x_{8,L} + x_{8,U}) = \frac{1}{2}(1,82 + 2,36) = 2,090.$$

Weiter ergibt sich für die Tiefe der hinges

$$T(h) = ([T(m)] + 1)/2 = ([8,5] + 1)/2 = (8 + 1)/2 = 4,5,$$

für die Tiefe der eighthths

$$T(e) = ([T(h)] + 1)/2 = ([4,5] + 1)/2 = (4 + 1)/2 = 2,5,$$

für die Tiefe der sixteenth

$$T(d) = ([T(e)] + 1)/2 = ([2,5] + 1)/2 = (2 + 1)/2 = 1,5$$

und für die Tiefe der thirtyseconds, die hier gleichzeitig die Extremwerte darstellen,

$$T(c) = ([T(d)] + 1)/2 = ([1,5] + 1)/2 = (1 + 1)/2 = 1.$$

In Tab. 2 sind Median, hinges, eighthths, sixteenth und Extremwerte, die zugehörig

Tab. 2: Lage- und Streuungsmaße der EDA für die Lebensdauern von $n = 16$ Kühlaggregaten aus Tab. 1

	L	U	midsummary	spread	Standard-abweichung
m	2,090		2,0900	—	—
h	0,745	3,980	2,3625	3,235	$s(h) = 2,398$
e	0,350	5,400	2,8750	5,050	$s(e) = 2,195$
d	0,205	6,760	3,4825	6,555	$s(d) = 2,137$
c	0,140	7,840	3,9900	7,700	—

gen midsummarys und spreads sowie Standardabweichungen zusammengestellt. Für die hinges ergibt sich mit $T(h) = 4,5$ etwa

$$h_L = \frac{1}{2}(x_{4,L} + x_{5,L}) = \frac{1}{2}(0,68 + 0,81) = 0,745,$$

$$h_U = \frac{1}{2}(x_{4,U} + x_{5,U}) = \frac{1}{2}(4,51 + 3,45) = 3,980,$$

so daß der midhinge gerade

$$\frac{1}{2}(0,745 + 3,980) = 2,3625,$$

der hinge-Abstand

$$h\text{-spread} = 3,980 - 0,745 = 3,235$$

und die zugehörige Standardabweichung

$$s = h\text{-spread}/1,349 = 3,235/1,349 = 2,398$$

ist. Es sei noch bemerkt, daß die Quartile, so wie sie im Abschnitt 4.2.2 des Kap. I definiert wurden, hier nicht mit den hinges übereinstimmen. Es ergibt sich für sie vielmehr wegen

$$n \cdot 0,25 = 16 \cdot 0,25 = 4 \quad \text{bzw.} \quad n \cdot 0,75 = 16 \cdot 0,75 = 12$$

hier

$$\tilde{x}_{0,25} = x_{(4)} = x_{4,L} = 0,68 \quad \text{bzw.} \quad \tilde{x}_{0,75} = x_{(12)} = x_{4,U} = 4,51.$$

Aus den midsummarys in Tab. 2 können wir ablesen, daß die Verteilung der Lebensdauern stark rechtsschief (linkssteil) ist, so daß die nahezu konstante Standardabweichung nicht mehr in Richtung auf normalverteilte Daten interpretiert werden kann.

1.2. Empirische Kenngrößen bei gruppierten Daten

Im Abschnitt 3 des Kap. I haben wir uns mit dem Begriff der **Klassenbildung** bzw. mit **gruppierten Daten** beschäftigt. Hat man n Beobachtungen in $k+2$ Klassen K_0, \dots, K_{k+1} derart eingeteilt, daß in die i -te Klasse K_i mit Klassengrenzen a_{i-1} und a_i ($a_{-1} = -\infty$, $a_{k+1} = \infty$) n_i Beobachtungen fallen ($n_0 + n_1 + \dots + n_{k+1} = n$), so kann man die in Abschnitt 1.1 vorgestellten Kenngrößen der EDA auch anhand der gruppierten Daten bestimmen. Ist z. B. $T(h)$ die Tiefe der hinges, so fällt h_L in die Klasse K_L , falls

$$\sum_{j=0}^{L-1} n_j < T(h) \quad \text{und} \quad \sum_{j=0}^L n_j \geq T(h),$$

und h_U in die Klasse K_U , falls

$$\sum_{j=U+1}^{k+1} n_j < T(h) \quad \text{und} \quad \sum_{j=U}^{k+1} n_j \geq T(h).$$

Ist nun $K_L = [a_{L-1}, a_L]$ und $K_U = [a_{U-1}, a_U]$, so ist der untere hinge der gruppierten Daten definiert als

$$h_L^* = a_{L-1} + [T(h) - (n_0 + \dots + n_{L-1}) - 0,5][a_L - a_{L-1}]/n_L$$

und der obere hinge ist definiert als

$$h_U^* = a_U - [T(h) - (n_{U+1} + \dots + n_{k+1}) - 0,5][a_U - a_{U-1}]/n_U.$$

Genauso werden eighthths e_L^* , e_U^* , sixteenths d_L^* , d_U^* usw. berechnet. Zur Berechnung

der midsummarys, spreads und Standardabweichungen werden dann natürlich diese Kenngrößen verwandt.

Beispiel: Gruppieren wir die Daten der Tab. 1 aus Abschnitt 1.1 in 10 Klassen $K_0 = [-\infty, 0]$, $K_1 = [0, 1]$, ..., $K_8 = [7, 8]$, $K_9 = [8, \infty]$, d.h. bis auf K_0 und K_9 betrachten wir Klassen der Breite 1 Jahr, so erhalten wir $n_0 = 0$, $n_1 = 5$, $n_2 = 3$, $n_3 = 3$, $n_4 = 1$, $n_5 = 1$, $n_6 = 2$, $n_7 = 0$, $n_8 = 1$ und $n_9 = 0$. Wir wollen nun Median und hinges aus den so gruppierten Daten bestimmen. Mit $T(m) = 8,5$ und $T(h) = 4,5$, vgl. Abschnitt 1.1, ergibt sich zunächst für den Median

$$n_0 + n_1 + n_2 = 8 < 8,5, \quad n_0 + n_1 + n_2 + n_3 = 11 > 8,5$$

bzw.

$$n_9 + n_8 + n_7 + n_6 + n_5 + n_4 + n_3 = 8 < 8,5, \quad 8 + n_2 = 11 > 8,5.$$

Beginnt man also mit Klasse K_0 , so würde man sagen, der Median liegt in K_3 , beginnt man hingegen mit K_9 , so kommt man zu der Entscheidung, daß der Median in K_2 liegt. In einem solchen Falle wird man die den beiden Klassen gemeinsame Grenze $a_2 = 2$ als Median m^g festlegen.

Für den unteren hinge erhalten wir mit

$$n_0 = 0 < 4,5, \quad n_0 + n_1 = 5 > 4,5,$$

daß er in Klasse $K_L = K_1$ liegt, und für den oberen hinge ergibt sich wegen

$$n_9 + n_8 + n_7 + n_6 + n_5 = 4 < 4,5,$$

$$n_9 + n_8 + n_7 + n_6 + n_5 + n_4 = 5 > 4,5$$

die Klasse $K_U = K_4$. Damit ist hier

$$h_L^g = 0 + [4,5 - n_0 - 0,5][1 - 0]/n_1 = [4,5 - 0,5] \cdot 1/5 = 4/5 = 0,8$$

und

$$h_U^g = 4 - [4,5 - 4 - 0,5][4 - 3]/n_4 = 4 - 0 \cdot 1/1 = 4.$$

Wir sehen, daß die aus den gruppierten Daten berechneten Größen m^g , h_L^g und h_U^g zwar nicht mit den in Abschnitt 1.1 berechneten m , h_L und h_U übereinstimmen, ihnen jedoch sehr nahe kommen. Je feiner die Klasseneinteilung gewählt wird, d.h. je größer die Anzahl der Klassen ist, desto besser stimmen die Kenngrößen aus ungruppiertem und gruppiertem Material überein.

1.3. Datentransformationen

Im Abschnitt 4.5 des Kap. V haben wir bereits verschiedene Typen von **Datentransformationen** kennengelernt. Ausgehend von Beobachtungswerten x_1, \dots, x_n wird dabei auf jede Einzelbeobachtung x_i , $i = 1, \dots, n$, eine monotone Funktion f angewandt, d.h., ist $x_{(1)}, \dots, x_{(n)}$ die geordnete Beobachtungsreihe, so ist auch die Reihe $f(x_{(1)}), \dots, f(x_{(n)})$ geordnet. Die Zielsetzung von Datentransformationen im Kap. V war, etwa zu erreichen, daß die transformierten Daten Voraussetzungen für bestimmte Methoden erfüllen. So erreicht man z.B. mittels der Wurzeltransformation, vgl. Abschnitt 4.5.B im Kap. V, daß Daten aus einer Poissonverteilung in etwa in solche aus einer Normalverteilung mit konstanter Varianz $1/4$ überführt werden. Begründet werden Datentransformationen durch das im Abschnitt 4.2 des Kap. V dargestellte **Fehlerfortpflanzungsgesetz**. Ist etwa X eine $Po(\lambda)$ -verteilte Zufallsvariable, so gilt für Erwartungswert und Varianz von X

$$EX = \text{Var } X = \lambda,$$

und nach dem Fehlerfortpflanzungsgesetz erhalten wir für $f(X) = \sqrt{X}$

$$\begin{aligned} E f(X) &= E \sqrt{X} \simeq \sqrt{E X} = \sqrt{\lambda} \\ \text{Var } f(X) &= \text{Var } \sqrt{X} \simeq \left[\frac{\partial f(X)}{\partial X} \Big|_{X=\lambda} \right]^2 \cdot \text{Var } X = \left[\frac{1}{2} X^{-1/2} \Big|_{X=\lambda} \right]^2 \cdot \text{Var } X \\ &= \left[\frac{1}{2} \lambda^{-1/2} \right]^2 \cdot \lambda = \frac{1}{4} \lambda^{-1} \cdot \lambda = \frac{1}{4}. \end{aligned}$$

Bei der explorativen Datenanalyse dienen Datentransformationen (**re-expressions**) dazu, das Datenmaterial übersichtlicher zu gestalten. Treten zum Beispiel sehr viele kleine Werte x_i und einige wenige sehr große Werte x_i auf, so ist eine übersichtliche Darstellung auch bei Klassenbildung kaum zu erreichen, wenn man gleiche Klassenbreiten wählen möchte. Transformationen, die in der EDA angewandt werden, sind von der Form x^p und werden zumeist in der sogenannten **ladder of power** (**Exponentenleiter**), wobei power die Größe p des Exponenten meint, aufgelistet:

p	3	2	1	$1/2$	0	$-1/2$	-1	-2
Trans-	x^3	x^2	x	$x^{1/2}$	$\ln x$	$-x^{-1/2}$	$-x^{-1}$	$-x^{-2}$
formation					$=\sqrt{x}$		$=-1/\sqrt{x}$	$=-1/x$	$=-1/x^2$	

Natürlich kann man für p auch beliebige Zwischenwerte wie z. B. 1,5, 1,6 oder -0,2 wählen. Bei $p = 1$ erfolgt keine Transformation der Daten, für $p = 0$ wird wegen $x^0 = 1$ für jedes x die Transformation $\ln x$ gewählt, die sich – betrachtet man die Varianz der resultierenden Größen – nahtlos in die Reihe der übrigen Transformationen einreihen lässt: Hat x den Erwartungswert μ und die Varianz σ^2 , so hat approximativ nach dem Fehlerfortpflanzungsgesetz $x^{1/2}$ die Varianz $\sigma^2/(4\mu)$, $\ln x$ die Varianz σ^2/μ^2 und $-x^{-1/2}$ die Varianz $\sigma^2/(4\mu^3)$.

Transformationen dieser Art sind in der EDA nur dann sinnvoll, wenn alle beobachteten Werte x_1, \dots, x_n positiv sind, da nur in diesem Falle die Ordnung der Werte nach einer Transformation erhalten bleibt. Treten sowohl positive als auch negative Werte oder nur negative Werte in der Beobachtungsreihe auf, so muß vor der Transformation eine Lokationsverschiebung derart vorgenommen werden, daß der kleinste beobachtete Wert $x_{(1)} = x_{1,L}$ positiv wird, d.h. zu den Daten wird eine Konstante c addiert, so daß gilt

$$x_{(1)} + c > 0.$$

Als Beispiel für die Anwendung von Transformationen in der EDA wollen wir hier die **Symmetrisierung einer Beobachtungsreihe** behandeln; ein weiteres Beispiel werden wir in Abschnitt 2.3 noch kennenlernen. Hat man z. B. mit den in Abschnitt 1.1 dargestellten Methoden festgestellt, daß eine Beobachtungsreihe eine rechts- oder linksschiefe Verteilung besitzt, so kann sie durch oben erwähnte **Power-Transformationen** symmetriert werden. Bei einer rechtsschiefen Verteilung muß hierzu $p < 1$, bei einer linksschiefen Verteilung $p > 1$ als Power verwandt werden. Um zu überprüfen, ob durch eine bestimmte Transformation Symmetrie erreicht wird, werden – wie in Abschnitt 1.1 beschrieben – die midsummarys der transformierten Beobachtungsreihe überprüft. Bei erreichter Symmetrie kann dann aus den resultierenden Standardabweichungen $s(h), s(e), s(d), \dots$ noch die Art der Abweichung in den „Schwänzen“ der Verteilung von einer Normalverteilung abgelesen werden.

Beispiel: Im Abschnitt 1.1 haben wir festgestellt, daß die Verteilung der Lebensdauern x_1, \dots, x_{16} von Kühlaggregaten, vgl. Tab. 1, rechtsschief ist. Wir wollen die Daten nun derart transformieren, daß Symmetrie erreicht wird. Wegen der Rechtschiefe müssen wir eine Power $p < 1$ wählen. Die mit $p = 0$, $p = 0,4$ und $p = 0,5$ transformierten Beobachtungen sind in Tab. 3 angegeben. In Tab. 4 sind Median, hinges, eightths, sixteenth, Extremwerte, daraus resultierende midsummarys, spreads und Standardabweichungen zusammengestellt.

Tab. 3: Lebensdauern x_i von 16 Kühlaggregaten, transformierte Lebensdauern $\ln x_i$ ($p = 0$), $x_i^{0,4}$ ($p = 0,4$) und $\sqrt{x_i}$ ($p = 0,5$) sowie Tiefe-Indizes T

i	T	$x_{(i)}$	$\ln x_{(i)}$	$x_{(i)}^{0,4}$	$\sqrt{x_{(i)}}$
1	1	0,14	-1,97	0,46	0,37
2	2	0,27	-1,31	0,59	0,52
3	3	0,43	-0,84	0,71	0,66
4	4	0,68	-0,39	0,86	0,82
5	5	0,81	-0,21	0,92	0,90
6	6	1,14	0,13	1,05	1,07
7	7	1,45	0,37	1,16	1,20
8	8	1,82	0,60	1,27	1,35
9	8	2,36	0,86	1,41	1,54
10	7	2,53	0,93	1,45	1,59
11	6	2,90	1,06	1,53	1,70
12	5	3,45	1,24	1,64	1,86
13	4	4,51	1,51	1,83	2,12
14	3	5,12	1,63	1,92	2,26
15	2	5,68	1,74	2,00	2,38
16	1	7,84	2,06	2,28	2,80

Tab. 4: Median, hinges, eightths, sixteenth, Extremwerte, midsummarys, spreads und Standardabweichungen für die Transformationen mit $p = 0$, $p = 0,4$, $p = 0,5$ der Lebensdauern von 16 Kühlaggregaten

	L	U	midsum.	spread	Standard-abweichung	p
m	0,730		0,7300			$p = 0$
h	-0,300	1,375	0,5375	1,675	1,242	
e	-1,075	1,685	0,3050	2,760	1,199	
d	-1,640	1,900	0,1300	3,540	1,154	
c	-1,970	2,060	0,0450	4,030		
m	1,340		1,3400			$p = 0,4$
h	0,890	1,735	1,3125	0,845	0,626	
e	0,650	1,960	1,3050	1,310	0,569	
d	0,525	2,140	1,3325	1,615	0,526	
c	0,460	2,280	1,3700	1,820		
m	1,445		1,4450			$p = 0,5$
h	0,860	1,990	1,4250	1,130	0,838	
e	0,590	2,320	1,4550	1,730	0,752	
d	0,445	2,590	1,5175	2,145	0,699	
c	0,370	2,800	1,5850	2,430		

An den midsummarys in der Tab. 4 sehen wir, daß die Transformation mit $p = 0$ zu stark ist, also aus der rechtsschiefen Verteilung eine linksschiefe macht, und daß $p = 0,5$ zu schwach ist; hier bleibt die Verteilung noch schwach rechtsschief. Für $p = 0,4$ sind alle midsummarys nahezu gleich, d. h. die Verteilung von $x_{(1)}^{0,4}, \dots, x_{(6)}^{0,4}$ ist in etwa symmetrisch; man vergleiche hierzu auch die Abb. 1 im folgenden Abschnitt 1.4. Wegen der leicht fallenden Tendenz der Standardabweichungen sind ihre „Schwänze“ etwas schwächer ausgeprägt als die einer Normalverteilung.

1.4. Box-Plots

Box-Plots (Kasten-Zeichnungen) dienen der übersichtlichen graphischen Darstellung einer Beobachtungsreihe. Entlang einer horizontalen Skala, die alle Beobachtungen umfaßt, werden zunächst die hinges durch vertikale Striche abgetragen; die Enden dieser Striche werden miteinander verbunden, so daß ein rechteckiger Kasten (die **Box**) entsteht, der 50 % aller Beobachtungswerte umfaßt. Dieser Kasten wird an der Stelle des Medians durch einen weiteren Strich geteilt. Sodann werden **innerer und äußerer Zaun (inner-fence, outer-fence)** berechnet. Die Grenzen des inner-fence sind

$$h_L - 1,5 \cdot (\text{h-spread}) \quad \text{und} \quad h_U + 1,5 \cdot (\text{h-spread}),$$

die des outer-fence

$$h_L - 3 \cdot (\text{h-spread}) \quad \text{und} \quad h_U + 3 \cdot (\text{h-spread}).$$

Werte zwischen inner- und outer-fence werden als **Ausreißer** einzeln durch Kreuze und Werte außerhalb des outer-fence werden als **krasse Ausreißer** durch Kreise markiert. Außerdem werden dann die **adjacent values** (das sind diejenigen Beobachtungen innerhalb des inner-fence, die am wenigsten von der oberen bzw. unteren Grenze des inner-fence abweichen) durch gestrichelte senkrechte Linien abgetragen. Die adjacent values werden dann durch waagerechte gestrichelte Linien mit den entsprechenden hinges verbunden. An einem solchen Box-Plot läßt sich dann die Verteilung und Struktur der Beobachtungsreihe recht gut ablesen. Ist z. B. der Box-Plot unsymmetrisch, so ist dies auch die Verteilung, hat er ein Übergewicht nach rechts, so ist die Verteilung linksschief, und umgekehrt. Mitunter werden auch noch die sogenannten **notches** (Kerben) in einem Box-Plot durch eckige Klammern angedeutet; man spricht dann auch vom **notched Box-Plot**.

Oberer und unterer notch sind gerade gegeben als

$$m \pm 1,58 \cdot (\text{h-spread})/\sqrt{n}$$

und bilden (unter Normalverteilungsannahme) in etwa ein **Konfidenzintervall** für den Median zum 90 % Niveau. Hier wird $\text{h-spread}/1,349$ als ein sehr guter Schätzer für die Standardabweichung angenommen und berücksichtigt, daß die Varianz des (empirischen) Medians bei Normalverteilung gerade das $\pi/2$ fache der Varianz des arithmetischen Mittels (als Schätzer für den Erwartungswert μ) ist. Die Grenzen eines $(1 - \alpha)$ -Konfidenzintervalls für μ bei Verwendung des arithmetischen Mittels \bar{x} sind bei Normalverteilung mit bekannter Varianz σ^2 gerade

$$\bar{x} \pm u_{1-\alpha/2} \sigma / \sqrt{n},$$

wobei σ^2/n die Varianz des Schätzers \bar{x} ist. Im Normalverteilungsfall sind nun Erwartungswert und Median theoretisch identisch, so daß μ auch durch den Me-

dian m mit der Varianz $\frac{\pi}{2} \sigma^2/n$ geschätzt werden kann. Das dann resultierende Konfidenzintervall zum Niveau $1 - \alpha$ für μ hat die Grenzen

$$m \pm u_{1-\alpha/2} \sigma \sqrt{\frac{\pi}{2}/n}.$$

Ersetzt man nun σ durch $s(h)$, so erhält man die Grenzen

$$m \pm u_{1-\alpha/2} \cdot (h\text{-spread}) \sqrt{\frac{\pi}{2}} / (1,349 \cdot \sqrt{n})$$

oder

$$m \pm u_{1-\alpha/2} \cdot 0,929 \cdot (h\text{-spread}) / \sqrt{n}.$$

Weiter ist

$$u_{1-\alpha/2} \cdot 0,929 = 1,58$$

für $\alpha = 0,09$, denn

$$u_{1-0,09/2} = u_{0,955} \approx 1,7.$$

Für $\alpha = 0,1$ erhält man $u_{1-\alpha/2} \cdot 0,929 = 1,645 \cdot 0,929 = 1,53$, jedoch bleiben wir (auch im Hinblick auf die nachfolgenden Paarvergleiche) bei dem etwas höheren Wert von 1,58, so wie in der entsprechenden Literatur üblich.

Beispiel: In der Abb. 1 sind die notched Box-Plots zu den Lebensdauern von $n = 16$ Kühlaggregaten, vgl. Tab. 1 bzw. Tab. 3, sowie zu den mit $p = 0,4$ transformierten Lebensdauern aus Tab. 3 dargestellt. Da im ersten Fall die Grenzen des inner-fence

$$0,745 - 1,5 \cdot 3,235 = -4,1075 \quad \text{und} \quad 3,980 + 1,5 \cdot 3,235 = 8,8325$$

sind, treten keine Ausreißer auf und die adjacent values sind gerade die Extremwerte der Beobachtungsreihe. Man sieht im ersten Box-Plot die starke Rechtsschiefe der Verteilung sehr schön. Mit den gleichen Überlegungen treten auch im zweiten Box-Plot, der die ungefähre Symmetrie der transformierten Daten zeigt, keine Ausreißer auf.

Mit Hilfe von Box-Plots lassen sich auch **verschiedene Beobachtungsreihen** bzgl. ihrer Struktur **vergleichen**. Zeichnet man die Box-Plots übereinander, so sind Lage- und Streuungsunterschiede direkt ablesbar. Verwendet man zusätzlich notches, so gelangt man zu Aussagen folgender Art: Bei Normalverteilungsannahme sind die notch-Intervalle zweier Box-Plots jeweils approximativ um die Mediane, sagen wir m_1 und m_2 , symmetrische 90 %-Konfidenzintervalle für die Erwartungswerte μ_1

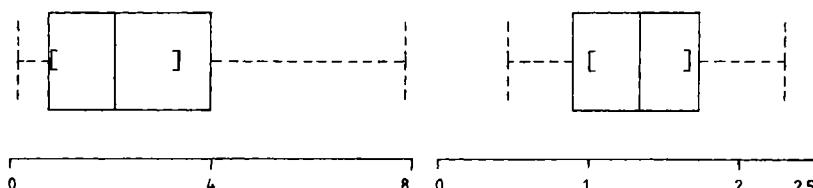


Abb. 1: Box-Plots für die Lebensdauern x_i von $n = 16$ Kühlaggregaten (links) und für die mit $p = 0,4$ transformierten Lebensdauern (rechts), $i = 1, \dots, 16$

bzw. μ_2 . Nimmt man $m_1 < m_2$ an, so sind die Wahrscheinlichkeiten, daß μ_1 größer als der zugehörige obere notch, μ_2 kleiner als der zugehörige untere notch ist, gerade 5%. Treten nun keine Überschneidungen der beiden notch-Intervalle auf, so weiß man, daß beim einseitigen Test zum Niveau 10% von $H_0: \mu_1 \geq \mu_2$ gegen $H_1: \mu_1 < \mu_2$, die Hypothese H_0 verworfen wird, d. h. unter Normalverteilungsannahme sind μ_1 und μ_2 signifikant verschieden (bzw. ist μ_1 signifikant kleiner als μ_2) zum 10%-Niveau. Für eine explorative Argumentation, die bei letzterer Aussage für ein Niveau von 5% plädiert (mit dem Wert 1,58 wie oben) vergleiche man Velleman/Hoaglin (1981). Es sei an dieser Stelle aber auch auf die exakten Verfahren zum Vergleich zweier Beobachtungsreihen in Kap. VIII hingewiesen.

Beispiel: In der Tab. 1 des Kap. X sind Größe x_i und Gewicht y_i von $n = 40$ Personen angegeben. Von diesen Personen waren die ersten zwanzig ($i = 1, 2, \dots, 20$) männlich und die übrigen Personen ($i = 21, \dots, 40$) weiblich. Sowohl für Größe als auch für Gewicht wollen wir nun die notched Box-Plots bei Männern und Frauen bestimmen und vergleichen. Die benötigten Hilfsgrößen sind in Tab. 5 zusammengestellt, die Box-Plots für die Größe sind in Abb. 2 und die für das Gewicht in Abb. 3 graphisch veranschaulicht.

Tab. 5: Hilfsgrößen zur Bestimmung der Box-Plots für Gewicht und Größe von je 20 Männern und Frauen, vgl. Tab. 1 in Kap. X

	Größe		Gewicht	
	Männer	Frauen	Männer	Frauen
m	178	163	74,5	61
h_L	173	159	69	55
h_U	181,5	168	82	65,5
h-spread	8,5	9	13	10,5
inner -fence	160,25 194,25	145,5 181,5	49,5 101,5	39,25 81,25
outer -fence	147,5 207	132 195	30 121	23,5 97
adjacent values	166 185	156 173	60 91	45 81
notches	175 181	159,8 166,2	69,9 79,1	57,3 64,7
Ausreißer	-	-	-	-

In den Abb. 2 und 3 erkennt man, daß die Verteilungen bei den Männern die umgekehrte Schiefe haben wie bei den Frauen. Sowohl für Gewicht als auch für Größe sind die Unterschiede in den Mittelwerten bei Männern und Frauen approximativ zum 10%-Niveau signifikant.

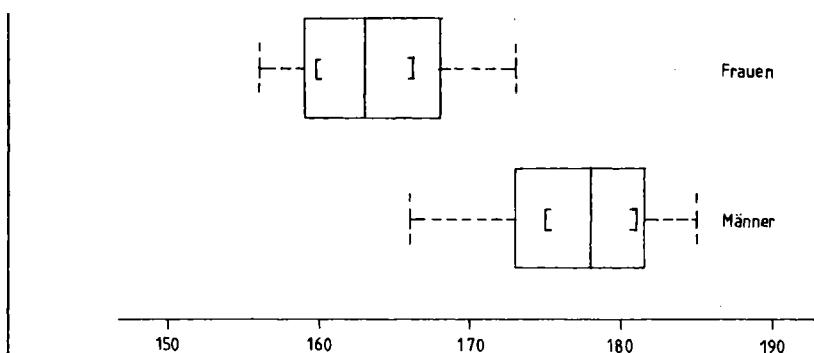


Abb. 2: Box-Plots für die Körpergröße von je 20 Männern und Frauen

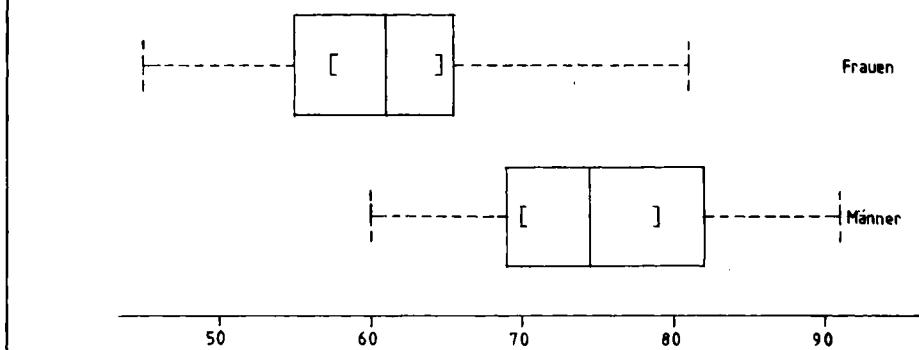


Abb. 3: Box-Plots für das Gewicht von je 20 Männern und Frauen

In etwas anderer Form sind Box-Plots in Hartung/Elpelt (1984, Kap.IX) zu finden. Wie in einigen Standard-Programmpaketen üblich ist dort z. B. neben dem Median auch noch das arithmetische Mittel im Box-Plot gekennzeichnet.

1.5. Stamm-und-Blätter-Darstellungen

Eine **Stamm-und-Blätter-** (**stem-and-leaves-**) Darstellung ähnelt einer **Strichliste** oder auch einem **Histogramm**, die wir im Abschnitt 3 des Kap.I kennengelernt haben, vgl. aber auch Abschnitt 1.6.

Betrachtet man gruppierte Daten, so entspricht die Rechteckfläche über dem zu einer bestimmten Klasse gehörigen Intervall im **Histogramm** gerade der relativen Häufigkeit des Auftretens eines Werts aus dem betreffenden Intervall innerhalb einer Beobachtungsreihe x_1, \dots, x_n .

Bei der Stamm-und-Blätter-Darstellung hingegen werden zunächst für alle Klassen untereinander die den Beobachtungswerten innerhalb einer Klasse gemeinsam ersten Ziffern notiert. Neben einem rechts von diesen Ziffern gezogenen senkrechten Strich werden sodann der Größe nach geordnet die direkt auf sie folgenden Ziffern bei den Werten der Beobachtungsreihe notiert; die verbleibenden Ziffern der Beobachtungswerte werden vernachlässigt. Sind die den Beobachtungswerten ge-

<table border="1"> <tbody> <tr><td>0</td><td>1</td><td>2</td><td>4</td><td>6</td><td>8</td></tr> <tr><td>1</td><td>1</td><td>4</td><td>8</td><td></td><td></td></tr> <tr><td>2</td><td>3</td><td>5</td><td>9</td><td></td><td></td></tr> <tr><td>3</td><td>4</td><td></td><td></td><td></td><td></td></tr> <tr><td>4</td><td>5</td><td></td><td></td><td></td><td></td></tr> <tr><td>5</td><td>1</td><td>6</td><td></td><td></td><td></td></tr> <tr><td>6</td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>7</td><td>8</td><td></td><td></td><td></td><td></td></tr> </tbody> </table>	0	1	2	4	6	8	1	1	4	8			2	3	5	9			3	4					4	5					5	1	6				6						7	8					<table border="1"> <tbody> <tr><td>0</td><td>F</td><td>4</td><td>5</td></tr> <tr><td>0</td><td>S</td><td>7</td><td></td></tr> <tr><td>0</td><td>N</td><td>8</td><td>9</td></tr> <tr><td>1</td><td>.</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>*</td><td>2</td><td></td></tr> <tr><td>1</td><td>F</td><td>4</td><td>4</td></tr> <tr><td>1</td><td>S</td><td>6</td><td></td></tr> <tr><td>1</td><td>N</td><td>8</td><td>9</td></tr> <tr><td>2</td><td>.</td><td>0</td><td></td></tr> <tr><td>2</td><td>*</td><td>2</td><td></td></tr> </tbody> </table>	0	F	4	5	0	S	7		0	N	8	9	1	.	0	1	1	*	2		1	F	4	4	1	S	6		1	N	8	9	2	.	0		2	*	2		<table border="1"> <tbody> <tr><td>0</td><td>.</td><td>4</td></tr> <tr><td>0</td><td>*</td><td>5</td><td>7</td><td>8</td><td>9</td></tr> <tr><td>1</td><td>.</td><td>0</td><td>1</td><td>2</td><td>4</td><td>4</td></tr> <tr><td>1</td><td>*</td><td>5</td><td>6</td><td>8</td><td>9</td></tr> <tr><td>2</td><td>.</td><td>0</td><td>2</td><td></td><td></td></tr> </tbody> </table>	0	.	4	0	*	5	7	8	9	1	.	0	1	2	4	4	1	*	5	6	8	9	2	.	0	2		
0	1	2	4	6	8																																																																																																																	
1	1	4	8																																																																																																																			
2	3	5	9																																																																																																																			
3	4																																																																																																																					
4	5																																																																																																																					
5	1	6																																																																																																																				
6																																																																																																																						
7	8																																																																																																																					
0	F	4	5																																																																																																																			
0	S	7																																																																																																																				
0	N	8	9																																																																																																																			
1	.	0	1																																																																																																																			
1	*	2																																																																																																																				
1	F	4	4																																																																																																																			
1	S	6																																																																																																																				
1	N	8	9																																																																																																																			
2	.	0																																																																																																																				
2	*	2																																																																																																																				
0	.	4																																																																																																																				
0	*	5	7	8	9																																																																																																																	
1	.	0	1	2	4	4																																																																																																																
1	*	5	6	8	9																																																																																																																	
2	.	0	2																																																																																																																			
*	*	0,4																																																																																																																				

Abb. 4: Stem-and-leaves-Darstellungen für die Lebensdauern von $n = 16$ Kühlaggregaten bei Klassenbreite 1 Jahr (links) sowie für die mit $p = 0,4$ transformierten Lebensdauern bei Klassenbreite 0,2 (Mitte) und Klassenbreite 0,5 (rechts)

meinsamen Ziffern mehrerer Klassen identisch, so kann das gerade angesprochene Intervall durch ein zusätzliches Zeichen markiert werden.

Beispiel: In der Abb. 4 sind drei verschiedene Stamm-und-Blätter-Darstellungen graphisch veranschaulicht. Beim linken Bild werden die Lebensdauern von $n = 16$ Kühlaggregaten, vgl. Tab. 1, in Klassen der Breite 1 Jahr eingeteilt, vgl. auch Abschnitt 1.2. Die den Lebensdauern gemeinsame Ziffer innerhalb jeder Klasse ist also die erste, d. h. die Anzahl vollendeter Lebensjahre. Hinter dem senkrechten Strich sind dann der Größe nach geordnet die ersten Nachkommastellen der insgesamt 16 Lebensdauern abgetragen. Die anderen beiden Darstellungen beziehen sich auf die symmetrierten ($p = 0,4$) Daten aus Tab. 3 im Abschnitt 1.3. Im mittleren Bild wurde die Intervallbreite 0,2 gewählt und im rechten Bild die Intervallbreite 0,5. Da hier jeweils mehrere Klassen Beobachtungswerte mit gleicher Anfangsziffer beinhalten, haben wir von einer zusätzlichen Markierung der ersten Ziffern der Daten Gebrauch gemacht. Im mittleren Bild bedeutet der Punkt, daß die erste Nachkommastelle, die hinter dem senkrechten Strich auch angegeben wird, eine Null oder eine Eins ist; der Stern bedeutet, daß sie eine Zwei oder Drei, F, daß sie eine Vier oder Fünf, S, daß sie eine Sechs oder Sieben, und N, daß sie eine Acht oder Neun ist. Im rechten Bild hingegen wurden Intervall- bzw. Klassenbreiten von 0,5 gewählt. Der Punkt hinter der ersten Ziffer besagt dann, daß die abgetragenen ersten Nachkommastellen 0,1, 2, 3 oder 4 sind; entsprechend bezieht sich der Stern auf die Nachkommastellen 5, 6, 7, 8 und 9.

1.6. Histogramm und empirische Verteilungsfunktion

Wir haben Histogramme und empirische Verteilungsfunktionen für gruppierte Daten bereits im Abschnitt 3.3 bzw. 3.4 des Kap. I kennengelernt. Trotzdem wollen wir hier noch einmal kurz auf sie eingehen.

Teilt man n Beobachtungen x_1, \dots, x_n in $k + 2$ Klassen K_0, \dots, K_{k+1} derart ein, daß alle n_i Beobachtungspunkte im Intervall $[a_{i-1}, a_i]$ zur Klasse K_i gehören, so dienen Histogramm und empirische Verteilungsfunktion zur Veranschaulichung der Verteilung der gruppierten Daten.

Im **Histogramm** werden die relativen Häufigkeiten $h_i = n_i/n$ der i-ten Klasse K_i

($i = 1, \dots, k$) durch rechteckige Flächen dargestellt. Dazu werden die Intervallgrenzen a_0, \dots, a_k ($a_{-1} = -\infty, a_{k+1} = \infty$) auf der x-Achse eines Koordinatensystems abgetragen, und die Höhe r_i des Rechtecks über der Klasse K_i bzw. dem Intervall $[a_{i-1}, a_i]$ berechnet sich zu

$$r_i = h_i/w_i, \quad w_i = a_i - a_{i-1} \quad \text{für } i = 1, \dots, k;$$

die Rechteckflächen sind also proportional zu den relativen Häufigkeiten und die Gesamtfläche unterhalb des Histogramms ist wie bei einer Dichte, vgl. Kap. II, gerade Eins, wenn wir voraussetzen, daß die Klassen K_0 und K_{k+1} leer sind, was wir hier tun und was man geschickterweise auch immer zu erreichen versuchen sollte.

Es sei noch erwähnt, daß im Rahmen der explorativen Datenanalyse oft auch die absoluten Häufigkeiten n_i durch die Histogramm-Rechtecke repräsentiert werden. Die Fläche dieses **nicht-normierten Histogramms** ist dann natürlich nicht mehr gleich Eins sondern vielmehr gleich der Anzahl n der Beobachtungen.

Die **empirische Verteilungsfunktion** $S_k(x)$ gruppierter Daten haben wir in Kap. I als Treppenfunktion definiert. Die Sprungstellen dieser Treppenfunktion konnten etwa die Klassenendpunkte a_i sein:

$$S_k(x) = \sum_{j=1}^i h_j \quad \text{für } a_i \leq x < a_{i+1};$$

hier werden die Beobachtungswerte in einer Klasse am Klassenendpunkt konzentriert. Vielfach ist es aber auch üblich, eine stetige Gleichverteilung, vgl. Abschnitt 2 in Kap. IV, innerhalb jeder Klasse anzunehmen, da auch die Histogrammdarstellung interpretiert als empirische Dichte von dieser Annahme ausgeht. Ist $n_0 = n_{k+1} = 0$, so ist dann

$$S_k(x) = 0 \quad \text{für } x \leq a_0, \quad S_k(x) = 1 \quad \text{für } x \geq a_k.$$

Im Intervall $[a_{i-1}, a_i]$, $i = 1, \dots, k$, hat $S_k(x)$ die Gestalt einer Geraden mit Steigung $r_i = h_i/w_i$, die die Punkte $(a_{i-1}, \sum_{j=1}^{i-1} h_j)$ und $(a_i, \sum_{j=1}^i h_j)$ verbindet:

$$S_k(x) = r_i x + \left(\sum_{j=1}^{i-1} r_j w_j - r_i a_{i-1} \right) = \frac{h_i}{w_i} x + \sum_{j=1}^{i-1} h_j - \frac{h_i}{w_i} a_{i-1}$$

$$\text{für } a_{i-1} \leq x \leq a_i \quad (w_i = a_i - a_{i-1}).$$

Bei dieser Definition spricht man auch von einer **stetigen empirischen Verteilungsfunktion**.

Beispiel: In Abb. 5 sind Histogramm und zugehörige stetige empirische Verteilungsfunktion für die Lebensdauern von $n = 16$ Kühlaggregaten, vgl. Tab. 1, bei Klassenbreite 1 Jahr angegeben. In Abb. 6 findet man entsprechende Darstellungen für die mit Power $p = 0,4$ transformierten Daten, vgl. Tab. 3; als Klassenbreite wurde hier $w_i = 0,5$ für $i = 1, \dots, 5$ gewählt.

1.7. Empirische Dichten

Mitunter sind Histogramme keine geeignete Darstellung für die empirische Dichte einer Beobachtungsreihe x_1, \dots, x_n . Wie in Abschnitt 1.6 erwähnt, geht man beim Histogramm von in etwa gleichverteilten Beobachtungswerten in jeder Klasse aus. Ist eine solche Annahme nicht zu rechtfertigen, so sind die im folgenden dargestellten Formen einer **empirischen Dichte** besser geeignet, die Verteilung der Beobach-

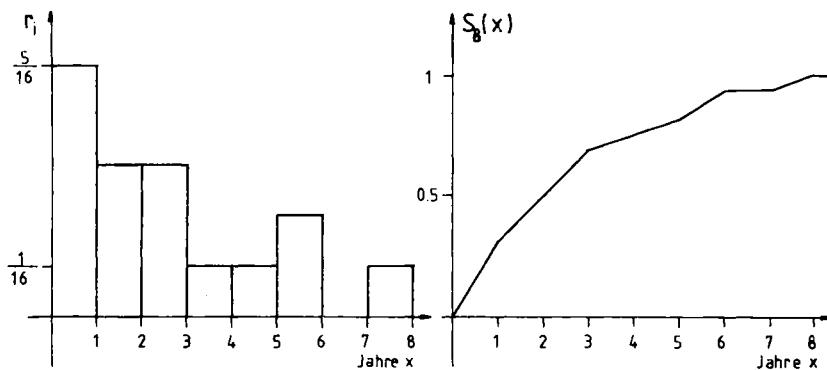


Abb. 5: Histogramm und stetige empirische Verteilungsfunktion für die Lebensdauern von $n = 16$ Kühlaggregaten bei Klassenbreite 1 Jahr

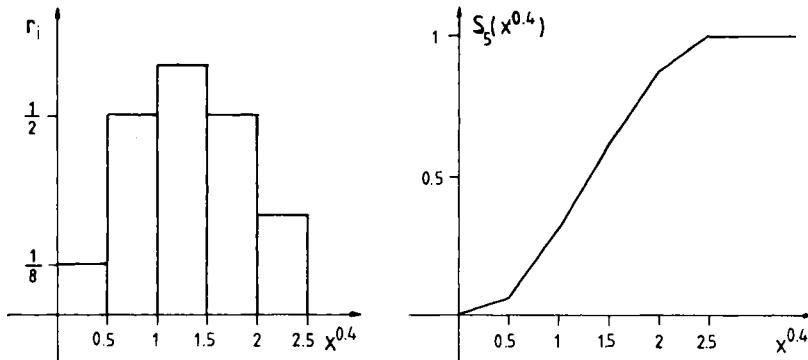


Abb. 6: Histogramm und stetige empirische Verteilungsfunktion für die mit $p = 0,4$ transformierten Lebensdauern von $n = 16$ Kühlaggregaten bei Klassenbreite 0,5

tungswerte zu repräsentieren; solche empirischen Dichten werden auch als **Dichteschätzer** oder **Kernschätzer** bezeichnet.

Bei ihnen wird zunächst ein Wertebereich, der alle Beobachtungsdaten x_1, \dots, x_n überdeckt, bestimmt. An endlich vielen, äquidistanten Stellen x dieses Wertebereichs wird dann der Wert der empirischen Dichte an der Stelle x

$$f(x) = \frac{1}{bn} \sum_{i=1}^n w\left(\frac{x - x_i}{b}\right)$$

berechnet. Hier bezeichnet b einen frei wählbaren Parameter, der bestimmt, wie breit das Intervall $[x - 0,5b, x + 0,5b]$ ist; Beobachtungswerte x_i in diesem Intervall werden zur Bestimmung des Wertes $f(x)$ herangezogen. Weiter ist $w(u)$ eine **Gewichtsfunktion**, ein **Kern** ähnlich den lag-Spektral Fenstern zur Glättung von Stichprobenspektren in der Zeitreihenanalyse, vgl. Kap. XII, Abschnitt 3.

Speziell zur Bestimmung empirischer Dichten wird in der EDA für $w(u)$ zumeist

entweder die **Boxcar-Funktion**

$$w(u) = \begin{cases} 1, & \text{falls } |u| \leq 1/2 \\ 0, & \text{sonst} \end{cases}$$

oder die **Cosinus-Funktion**

$$w(u) = \begin{cases} 1 + \cos(2\pi u), & \text{falls } |u| \leq 1/2 \\ 0, & \text{sonst} \end{cases}$$

gewählt. Im ersten Fall bestimmt sich $f(x)$ gerade als Anzahl der Beobachtungswerte im Intervall $[x - 0,5b, x + 0,5b]$ dividiert durch $b \cdot n$. In Abb. 7 sind beide Funktionen veranschaulicht.

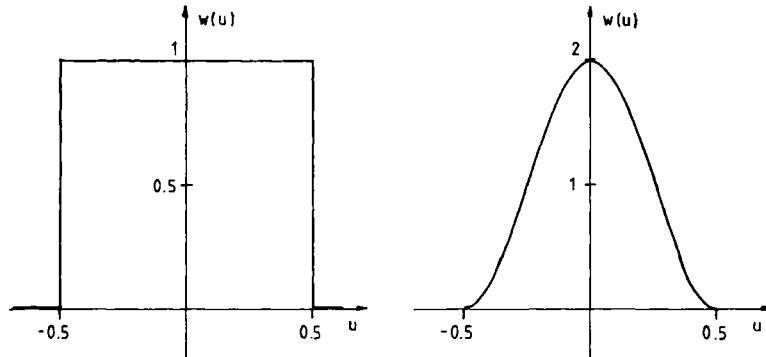


Abb. 7: Boxcar-Funktion (links) und Cosinus-Funktion (rechts)

Prinzipiell kann jede beschränkte und symmetrische Funktion $w(u) \geq 0$ als Gewichtsfunktion verwandt werden, für die gilt:

$$\int_{-\infty}^{\infty} w(u) du = 1, \quad \int_{-\infty}^{\infty} w^2(u) du < \infty$$

und $|w(u)/u| \rightarrow 0$ für $|u| \rightarrow \infty$.

Einige neben der Boxcar- und der Cosinus-Funktion (die dem Rechtecks- bzw. dem Tukey-lag-Fenster, vgl. Abschnitt 3.4 in Kap. XII, ähneln) häufig verwandte Gewichtsfunktionen seien hier noch erwähnt. Die Gewichtsfunktion

$$w(u) = \begin{cases} 1/2, & \text{falls } |u| \leq 1 \\ 0, & \text{sonst} \end{cases}$$

heißt **Rechteckskern** (auch sie ähnelt dem Rechtecks-lag-Fenster),

$$w(u) = \begin{cases} 1 - |u|, & \text{falls } |u| \leq 1 \\ 0, & \text{sonst} \end{cases}$$

heißt **Dreieckskern** (sie entspricht dem Bartlett- oder Dreiecks-lag-Fenster), und die Dichte der Standardnormalverteilung als Gewichtsfunktion verwandt:

$$w(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

heißt **Normalkern**.

Die berechneten Werte $f(x)$ werden dann als Polygonzug, in Abhängigkeit von x , verbunden dargestellt. Es ist darauf zu achten, daß der kleinste Wert x noch groß genug ist, sicherzustellen, daß im Intervall $[x, x + 0,5b]$ ein Beobachtungswert liegt, und daß der größte Wert x so klein ist, daß im Intervall $[x - 0,5b, x]$ ein Beobachtungswert liegt. Ist dies der Fall, so ist die Fläche zwischen Polygonzug und x -Achse gerade gleich Eins.

Beispiel: Wir betrachten hier die symmetrierten, mit Power $p = 0,4$ transformierten Lebensdauern von $n = 16$ Kühlaggregaten, mithin die Daten $x_{(1)}^{0,4}, \dots, x_{(16)}^{0,4}$ aus Tab. 3, deren Histogramm bei Klassenbreite 0,5 in Abb. 6 dargestellt ist, und wollen hier einmal deren empirische Dichten f_B bei Verwendung der Boxcar-Funktion und f_C bei Verwendung der Cosinus-Funktion bestimmen. Wir wählen in beiden Fällen $b = 0,5$ und bestimmen f_B sowie f_C an den Stellen $x = 0,25 + j \cdot 0,125$ für $j = 0, \dots, 18$, vgl. Tab. 6. Es ergibt sich z. B.

$$\begin{aligned} f(1) &= \frac{1}{0,5 \cdot 16} \sum_{i=1}^{16} w\left(\frac{1-x_i}{0,5}\right) \\ &= \frac{1}{8} (0 + 0 + 0 + w(0,28) + w(0,16) + w(-0,10) + w(-0,32) \\ &\quad + 0 + \dots + 0) \end{aligned}$$

und somit

$$f_B(1) = \frac{1}{8} (0 + 0 + 0 + 1 + 1 + 1 + 1 + 0 + \dots + 0) = \frac{4}{8} = 0,5$$

bzw.

$$\begin{aligned} f_C(1) &= \frac{1}{8} \left(0 + 0 + 0 + (1 + \cos(2\pi \cdot 0,28)) + (1 + \cos(2\pi \cdot 0,16)) \right. \\ &\quad \left. + (1 + \cos(-2\pi \cdot 0,10)) + (1 + \cos(-2\pi \cdot 0,32)) + 0 + \dots + 0 \right) \\ &= 0,591. \end{aligned}$$

Tab. 6: Empirische Dichten $f_B(0,25 + j \cdot 0,125)$, $f_C(0,25 + j \cdot 0,125)$ für $j = 0, \dots, 18$ zu den symmetrierten Lebensdauern $x_{(1)}^{0,4}, \dots, x_{(16)}^{0,4}$ von $n = 16$ Kühlaggregaten, vgl. Tab. 3 ($x = 0,25 + j \cdot 0,125$)

j	x	$f_B(x)$	$f_C(x)$	j	x	$f_B(x)$	$f_C(x)$
0	0,250	1/8	0,015	10	1,500	5/8	0,751
1	0,375	2/8	0,197	11	1,625	5/8	0,502
2	0,500	3/8	0,428	12	1,750	5/8	0,407
3	0,625	4/8	0,490	13	1,875	4/8	0,588
4	0,750	4/8	0,513	14	2,000	3/8	0,500
5	0,875	4/8	0,595	15	2,125	3/8	0,223
6	1,000	4/8	0,591	16	2,250	2/8	0,241
7	1,125	4/8	0,550	17	2,375	1/8	0,171
8	1,250	5/8	0,544	18	2,500	1/8	0,009
9	1,375	5/8	0,684				

In Abb. 8 sind die beiden empirischen Dichten auch als Polygonzüge dargestellt.

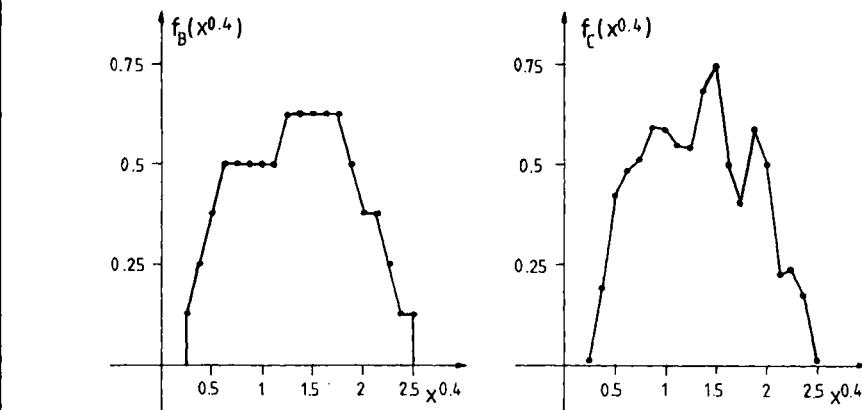


Abb. 8: Mittels Boxcar-Funktion (links) bzw. Cosinus-Funktion (rechts) bestimmte empirische Dichten zu den mit $p = 0,4$ transformierten Lebensdauern von $n = 16$ Kühlaggregaten bei Klassenbreite 0,5 und $b = 0,5$

1.8. Wurzeldiagramme

In Abschnitt 1.6 wurde erwähnt, daß im Rahmen der explorativen Datenanalyse oftmals **nicht-normierte Histogramme** zur Veranschaulichung der Verteilung gruppiertener Daten verwandt werden. Bei diesen nicht-normierten Histogrammen entspricht die Größe der Rechtecksfläche über einem Intervall $[a_{i-1}, a_i]$, d. h. über einer Klasse K_i ($i = 1, \dots, k$), gerade der (absoluten) Klassenhäufigkeit, mithin der Anzahl n_i der Beobachtungswerte aus einer Reihe x_1, \dots, x_n , die zur Klasse K_i gehören. Die Höhe r_i der Rechtecksfläche zur Klasse K_i ist dann gerade n_i/w_i mit $w_i = a_i - a_{i-1}$.

Faßt man die Häufigkeiten n_1, \dots, n_k als poissonverteilte Daten auf, so erreicht man durch eine Wurzeltransformation, daß die transformierten Daten $\sqrt{n_1}, \dots, \sqrt{n_k}$ eine Approximation an normalverteilte Größen mit Varianz $1/4$ darstellen, unabhängig vom Erwartungswert; man vgl. hierzu Abschnitt 3.4 in Kap. IV, Abschnitt 4.5.B in Kap. V sowie Abschnitt 1.3 dieses Kapitels. Das **Wurzeldiagramm (rootogram)** entspricht nun gerade dem Histogramm, wenn man die Höhen der Rechtecksflächen über den Klassen K_1, \dots, K_k gleich $\sqrt{r_1}, \dots, \sqrt{r_k}$ wählt.

Das Wurzeldiagramm wird hauptsächlich verwandt, um die Güte der Anpassung des zugehörigen Histogramms an eine Normalverteilungsdichte zu beurteilen. Dazu bestimmt man zunächst eine passende Normalverteilung. Der Mittelwert dieser Verteilung wird gleich dem midhinge

$$m(h) = \frac{1}{2}(h_L + h_U)$$

der zugrundeliegenden Daten und die Standardabweichung gleich dem Wert

$$s(h) = h \cdot \text{spread}/1,349$$

gewählt, vgl. Abschnitt 1.1 bzw. Abschnitt 1.2, falls die Ursprungsdaten nicht zur Verfügung stehen. Bezeichnet $\varphi(x)$ die Dichte der Standardnormalverteilung

$N(0, 1)$, so ist dann, vgl. Abschnitt 1 in Kap. IV,

$$f(x) = \frac{n}{s(h)} \cdot \varphi\left(\frac{x - m(h)}{s(h)}\right) = \frac{n}{s(h)} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - m(h))^2}{2s^2(h)}}$$

die an das nicht-normierte Histogramm, dessen Fläche ja gerade der Anzahl n der Beobachtungen entspricht, angepaßte Kurve. Die an das Wurzeldiagramm angepaßte Kurve ist dann natürlich $\sqrt{f(x)}$, entsteht also durch punktweises Wurzelziehen aus $f(x)$.

Nun werden die **Wurzeldiagramm-Residuen**

$$RR_i = \sqrt{r_i} - \sqrt{\hat{r}_i} \quad \text{für } i = 1, \dots, k$$

berechnet. Hierbei bestimmt man \hat{r}_i aus der Fläche \hat{n}_i zwischen $f(x)$ und x -Achse im Intervall $[a_{i-1}, a_i]$ der Breite w_i : Bezeichnet $\Phi(x)$ die Verteilungsfunktion der Standardnormalverteilung, so ist

$$\hat{r}_i = \frac{n}{w_i} \left(\Phi\left(\frac{a_i - m(h)}{s(h)}\right) - \Phi\left(\frac{a_{i-1} - m(h)}{s(h)}\right) \right) = \hat{n}_i / w_i;$$

\hat{n}_i ist also die bei dieser Normalverteilung zwischen a_{i-1} und a_i zu erwartende Anzahl von Beobachtungen.

Folglich können wir sagen, daß bei tatsächlichem Vorliegen dieser Normalverteilung approximativ gilt

$$\sqrt{n_i} - \sqrt{\hat{n}_i} \sim N(0, 1/4),$$

d. h. $\sqrt{r_i} - \sqrt{\hat{r}_i}$ hat angenähert den Erwartungswert 0 und die Standardabweichung $\sigma = 1/(2\sqrt{w_i})$. Liegt nun ein Wurzeldiagramm-Residuum außerhalb des 2σ -Intervalls $[-1/\sqrt{w_i}, 1/\sqrt{w_i}]$, so kann es als groß bezeichnet werden (2σ -Regel, vgl. auch Abschnitt 1.1 in Kap. IV), d. h. in diesem Intervall findet keine gute Anpassung an die Funktion $\sqrt{f(x)}$ statt. Aufgrund der Größe der Wurzeldiagramm-Residuen wird somit die Güte der Anpassung der „Verteilung“ der Daten bzw. des nicht-normierten Histogramms durch die Funktion $f(x)$ bewertet.

Wie man hier sieht, sind durch den Übergang vom nicht-normierten Histogramm zum Wurzeldiagramm alle Klassen (gleicher Breite) mit dem selben Wert vergleichbar, ganz gleich ob sie schwach oder stark besetzt sind. Zu erklären ist diese Tatsache dadurch, daß die Wurzeltransformation nicht nur eine angenäherte Normalverteilung liefert sondern vielmehr auch zur ungefähren Stabilisierung der Varianz auf den Wert 1/4 führt.

Prinzipiell kann diese Art der Datenanalyse natürlich nicht nur unter Verwendung der Normalverteilung sondern auch mit jeder anderen Verteilung durchgeführt werden. Die Funktion $f(x)$ ist dann gerade die n -fache Dichte dieser Verteilung.

Die Wurzeldiagramm-Residuen lassen sich auch graphisch darstellen. Dazu wird zumeist die Funktion $\sqrt{f(x)}$ an der x -Achse gespiegelt repräsentiert. Die Rechtecke des Wurzeldiagramms werden ebenfalls an dieser Achse gespiegelt (**aufgehängtes Wurzeldiagramm**) und zusätzlich noch um den Wert des zugehörigen Wurzeldiagramm-Residuums nach unten oder nach oben verschoben. Ein negatives Residu-

um bewirkt eine entsprechende Verschiebung nach unten, ein positives Residuum eine Verschiebung nach oben.

Vielfach werden, um sie besser sichtbar zu machen, auch lediglich die Residuen dargestellt. Zusätzlich werden dann waagerechte Linien auf den Höhen $1/\sqrt{w_i}$ und $-1/\sqrt{w_i}$ eingezeichnet, um extrem große Residuen direkt erkennen zu können.

Beispiel: Wir wollen die mit Power $p = 0,4$ transformierten Lebensdauern von $n = 16$ Kühlaggregaten, die nahezu symmetrische Daten $x_{(1)}^{0,4}, \dots, x_{(16)}^{0,4}$ sind, mit Hilfe des Wurzeldiagramms analysieren. Dazu bilden wir Klassen $K_1 = [0; 0,5], \dots, K_5 = [2; 2,5]$ der Breite $w_i = 0,5$ und passen an das nicht-normierte Histogramm die n -fache Dichte $f(x)$ einer Normalverteilung mit Mittelwert

$$m(h) = \frac{1}{2}(h_L + h_U) = \frac{1}{2}(0,890 + 1,735) = 1,3125$$

und Standardabweichung

$$\begin{aligned} s(h) &= h\text{-spread}/1,349 = (h_U - h_L)/1,349 = (1,735 - 0,890)/1,349 \\ &= 0,626 \end{aligned}$$

an, vgl. Tab. 4 in Abschnitt 1.3.

In Tab. 7 sind die Hilfsgrößen zur Bestimmung des Wurzeldiagramms sowie die

Tab. 7: Hilfsgrößen zur Bestimmung des Wurzeldiagramms und Wurzeldiagramm-Residuen zu den mit $p = 0,4$ transformierten Lebensdauern von $n = 16$ Kühlaggregaten

i	$[a_{i-1}; a_i]$	w_i	n_i	r_i	$\sqrt{r_i}$	\hat{r}_i	$\sqrt{\hat{r}_i}$	$\sqrt{r_i} - \sqrt{\hat{r}_i}$
1	$[0; 0,5]$	0,5	1	2	1,414	2,528	1,590	-0,176
2	$[0,5; 1]$	0,5	4	8	2,828	6,784	2,605	0,223
3	$[1; 1,5]$	0,5	5	10	3,162	9,888	3,145	0,017
4	$[1,5; 2]$	0,5	4	8	2,828	7,872	2,806	0,022
5	$[2; 2,5]$	0,5	2	4	2,000	3,424	1,850	0,150

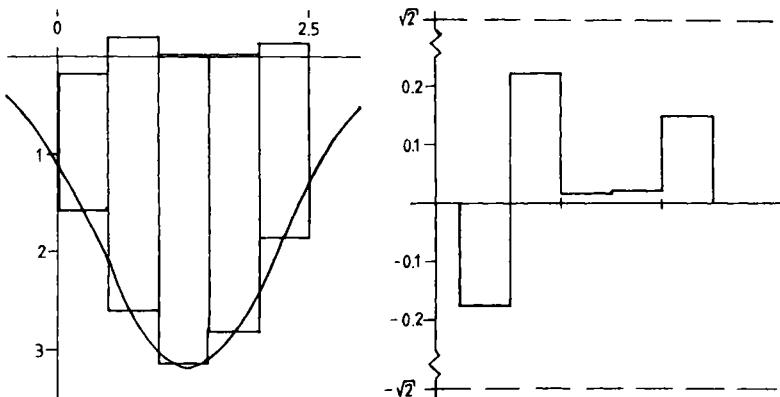


Abb. 9: Aufgehängtes Wurzeldiagramm und zugehörige Funktion $\sqrt{f(x)}$ (links) sowie Wurzeldiagramm-Residuen bei eingezeichnetem 2σ -Bereich (rechts) für die mit $p = 0,4$ transformierten Lebensdauern von $n = 16$ Kühlaggregaten bei Klassenbreite 0,5

Wurzeldiagramm-Residuen zusammengestellt. In Abb. 9 sind das aufgehängte Wurzeldiagramm sowie die Vergleichskurve $\sqrt{f(x)}$ graphisch veranschaulicht. Im rechten Teil der Abbildung sind zudem noch einmal nur die Residuen abgetragen. Die Linien bei $1/\sqrt{w_i}$, $-1/\sqrt{w_i}$, die weit weg von den Residuen liegen:

$$1/\sqrt{w_i} = 1/\sqrt{0,5} = \sqrt{2} = 1,414, \quad -1/\sqrt{w_i} = -1,414 \quad \text{für } i = 1, \dots, 5,$$

sind in Abb. 9 ebenso eingezeichnet. Da diese Residuen alle betragsmäßig sehr klein sind, ist die Anpassung des Histogramms durch die Normalverteilung sehr gut.

Es sei hier noch erwähnt, daß anstelle der Wurzeldiagramm-Residuen mitunter auch die sogenannten **Double-Root-Residuen**

$$\text{DRR}_i = \begin{cases} \sqrt{2 + 4n_i} - \sqrt{1 + 4\hat{n}_i}, & \text{falls } n_i \neq 0 \\ 1 - \sqrt{1 + 4\hat{n}_i}, & \text{falls } n_i = 0 \end{cases}, \quad i = 1, \dots, k,$$

zur Bewertung der Güte der Anpassung des nicht-normierten Histogramms durch die Funktion $f(x)$ herangezogen werden. Diese Double-Root-Residuen sind approximativ $N(0, 1)$ -verteilt, so daß man sie nach der 2σ -Regel als groß bezeichnen kann, wenn sie außerhalb des Intervalls $[-2; 2]$ liegen.

Natürlich können anstelle der hier vorgestellten explorativen Methoden zur Überprüfung von Verteilungsannahmen auch die gewöhnlichen Anpassungstests (χ^2 -Test, Kolmogoroff-Smirnov-Test) verwendet werden, die in Kap. IV vorgestellt wurden.

1.9. Q-Q-Plots zur Überprüfung von Verteilungsannahmen

Um zu überprüfen, ob die Daten einer Beobachtungsreihe x_1, \dots, x_n Realisationen aus einer bestimmten Verteilung darstellen, werden sogenannte empirische Quantile-Quantile-Plots (Q-Q-Plots) verwendet.

Der i -te Wert $x_{(i)}$ der geordneten Beobachtungsreihe $x_{(1)}, \dots, x_{(n)}$ entspricht gerade dem empirischen $i/(n+1)$ -Quantil der Beobachtungsreihe. Trägt man nun diese empirischen Quantile gegen die entsprechenden theoretischen Quantile $\xi_{i/(n+1)}$ der interessierenden Verteilung in einem Koordinatensystem ab, so liegen die Punkte $(\xi_{i/(n+1)}, x_{(i)})$ für $i = 1, \dots, n$ etwa auf der Ursprungsgeraden mit Steigung 1, falls die Daten aus dieser Verteilung stammen. Starke Abweichungen von der Geraden signalisieren ein Nichtzutreffen der Verteilung, und starke Abweichungen nur der letzten Punkte von der Geraden (links nach unten, rechts nach oben) können als Ausreißer interpretiert werden.

Q-Q-Plots sind Alternativen zum Wahrscheinlichkeitspapier, wie es im Abschnitt 1.7 des Kap. IV etwa zur Überprüfung einer Normalverteilungsannahme vorgestellt wurde. Sie lassen sich auch zur Überprüfung von Verteilungsannahmen für mehrdimensionale Daten verwenden; man vgl. hierzu Kap. IX und X in Hartung/Elpelt (1984).

Beispiel: In der Tab. 3 sind geordnete Lebensdauern $x_{(1)}, \dots, x_{(16)}$ von $n = 16$ Kühlaggregaten sowie die transformierten Lebensdauern $x_{(1)}^{0,4}, \dots, x_{(16)}^{0,4}$ angegeben. In der Abb. 10 sind Q-Q-Plots für die Lebensdauern dargestellt, wobei zum einen

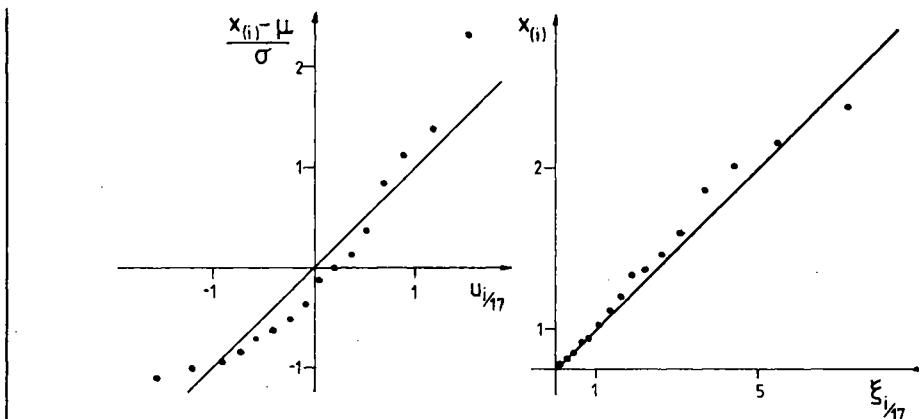


Abb. 10: Q-Q-Plot für die Lebensdauern von $n = 16$ Kühlaggregaten (geordnete Werte $x_{(i)}$) zum Vergleich mit der Normalverteilung (links) und mit der Exponentialverteilung (rechts)

eine Normalverteilung mit Mittelwert bzw. Standardabweichung

$$\mu = m = 2,571 \quad \text{bzw.} \quad \sigma = s(h) = 2,235$$

und zum anderen eine Exponentialverteilung mit Parameter

$$\lambda = \frac{1}{m} = 0,39$$

zugrundegelegt wurde; man vgl. auch Tab. 2. In **Abb. 11** ist der Q-Q-Plot für die

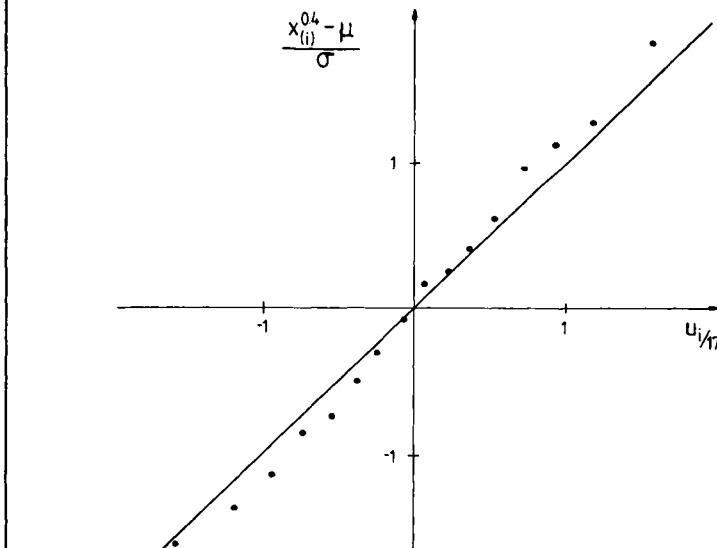


Abb. 11: Q-Q-Plot für die mit $p = 0,4$ transformierten Lebensdauern von $n = 16$ Kühlaggregaten zum Vergleich mit der Normalverteilung

transformierten Daten bei Zugrundelegung einer Normalverteilung mit

$$\mu = m = 1,3175 \quad \text{und} \quad \sigma = s(h) = 0,5350,$$

vgl. Tab. 4, angegeben. Die zur Anfertigung der Plots benötigten Größen sind in Tab. 8 zusammengestellt.

In Abb. 10 erkennt man die Abweichung von einer Normalverteilung sehr deutlich, wohingegen die Anpassung der Daten an eine $\text{Ex}(0,39)$ -Verteilung recht gut ist. Die transformierten Lebensdauern zeigen, vgl. auch Abschnitt 1.8, kaum Abweichungen von einer Normalverteilung. Die Punkte $(u_{i/17}, (x_{(i)}^{0,4} - 1,3175)/0,5350)$ liegen nahezu auf einer Geraden durch den Ursprung, vgl. Abb. 11. Die Steigung dieser Punkte-Geraden ist zwar etwas größer als 1, d.h. liegt oberhalb der eingezeichneten Geraden, was sich jedoch durch Wahl einer etwas größeren Standardabweichung leicht beheben lässt.

Tab. 8: Lebensdauern $x_{(i)}$, standardisierte Lebensdauern, standardisierte transformierte Lebensdauern sowie Quantile $u_{i/17}$ der Standardnormalverteilung und $\xi_{i/17} = \ln(1 - i/17)/0,39$

i	$x_{(i)}$	$\frac{x_{(i)} - 2,571}{2,235}$	$\frac{x_{(i)}^{0,4} - 1,3175}{0,5350}$	$u_{i/17}$	$\xi_{i/17} = \ln(1 - i/17)/0,39$
1	0,14	-1,088	-1,603	-1,565	0,155
2	0,27	-1,030	-1,360	-1,186	0,321
3	0,43	-0,958	-1,136	-0,929	0,498
4	0,68	-0,846	-0,855	-0,723	0,688
5	0,81	-0,788	-0,743	-0,542	0,893
6	1,14	-0,640	-0,500	-0,378	1,116
7	1,45	-0,502	-0,294	-0,228	1,361
8	1,82	-0,336	-0,089	-0,073	1,631
9	2,36	-0,094	0,173	0,073	1,933
10	2,53	-0,018	0,248	0,228	2,275
11	2,90	0,147	0,397	0,378	2,670
12	3,45	0,393	0,603	0,542	3,138
13	4,51	0,868	0,958	0,723	3,710
14	5,12	1,140	1,126	0,929	4,448
15	5,68	1,391	1,276	1,186	5,487
16	7,84	2,357	1,799	1,565	7,265

2. Verfahren für zwei Merkmale in der EDA

Wir werden uns in diesem Abschnitt mit einigen explorativen Methoden für zwei Merkmale beschäftigen, d.h. wir werden Verfahren für zweidimensionale Punktescharen $(x_1, y_1), \dots, (x_n, y_n)$ behandeln.

Zunächst wird im Abschnitt 2.1 die **Glättung (smoothing)** einer solchen Punkteschar mittels sogenannter gleitender Medianen, die in etwa den in Kap. XII behandelten gleitenden Durchschnitten entsprechen, dargestellt.

Im Abschnitt 2.2 werden wir dann ein Verfahren zur explorativen Bestimmung einer Regressionsgeraden für zweidimensionale, lineare Punktescharen kennenlernen, und im Abschnitt 2.3 schließlich beschäftigen wir uns mit der Linearisierung

zweidimensionaler Punktescharen. Hierzu werden die Transformationen aus der in Abschnitt 1.3 angegebenen „ladder of power“ verwandt.

2.1. Glätten zweidimensionaler Punktescharen

Wir haben uns im Abschnitt 1 des Kap. XII mit der Glättung bzw. Filterung von Zeitreihen beschäftigt. Dort wurden zu diesem Zwecke etwa gleitende Durchschnitte eingesetzt. Wurde für $t = 1, \dots, n$ der Wert y_t der Zeitreihe beobachtet, so ergab sich die Reihe der gleitenden Durchschnitte der Ordnung 5 z. B. zu

$$y_t^* = \frac{1}{5}(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}) \quad \text{für } t = 3, \dots, n-2.$$

In der explorativen Datenanalyse werden zu Glättungszwecken häufig **gleitende Mediane (running medians)** eingesetzt. Werden etwa zu Zeitpunkten $t = 1, \dots, n$ Werte y_1, \dots, y_n beobachtet, so besteht die Reihe der **gleitenden Mediane der ungeraden Ordnung $2k + 1$** aus den Werten

$$y_t^* = m(y_{t-k}, \dots, y_t, \dots, y_{t+k}) \quad \text{für } t = k+1, \dots, n-k,$$

wobei $m(y_{t-k}, \dots, y_t, \dots, y_{t+k})$ den Median der $2k + 1$ Werte $y_{t-k}, \dots, y_t, \dots, y_{t+k}$ meint. Die Reihe der **gleitenden Mediane der geraden Ordnung $2k$** hingegen ist gegeben durch

$$y_t^* = \frac{1}{2}(m(y_{t-k}, \dots, y_t, \dots, y_{t+k-1}), \dots, y_t, \dots, y_{t+k}) \quad \text{für } t = k+1, \dots, n-k.$$

Häufig werden auch mehrere gleitende Mediane nacheinander angewandt. So versteht man zum Beispiel unter einem **gleitenden Median 53**, daß zunächst die Reihe der gleitenden Mediane der Ordnung 5 und ausgehend von diesen dann die Reihe der gleitenden Mediane der Ordnung 3 bestimmt wird. Fügt man hinter die letzte Ordnung noch ein R an (z. B. also 53R), so meint man, daß gleitende Mediane der Ordnung 3 nicht nur einmal gebildet werden, sondern solange bis sich die resultierende Reihe nicht mehr verändert.

Bei der Bildung eines gleitenden Medians der Ordnung $2k + 1$ oder $2k$ tritt wie bei gleitenden Durchschnitten das Problem auf, daß für die ersten und letzten k Werte kein Wert für die geglättete Reihe mehr vorhanden ist. Hier kann man sich so behelfen, daß für die Randwerte gleitende Mediane geringerer Ordnung verwandt werden. Bei gleitenden Medianen der Ordnung 5 wird man etwa

$$y_1^* = y_1, \quad y_n^* = y_n, \quad y_2^* = m(y_1, y_2, y_3), \quad y_{n-1}^* = m(y_{n-2}, y_{n-1}, y_n) \\ \text{setzen.}$$

Das bisher beschriebene Vorgehen bezieht sich auf reine Zeitreihen, d. h. zu jedem Zeitpunkt t wird ein Wert beobachtet und die Abstände zwischen den Zeitpunkten sind äquidistant. Beobachtet man hingegen zu vorgegebenen Werten x_1, \dots, x_n die eigentlich interessierenden Größen y_1, \dots, y_n , wobei die x_i nicht notwendig alle verschieden sind, so ist das Verfahren der gleitenden Mediane (für die Werte y_1, \dots, y_n) nicht mehr direkt anwendbar und auch nicht mehr exakt. Man kann sich dann so behelfen, daß man die Punkte $(x_1, y_1), \dots, (x_n, y_n)$ nach der Größe der x_i -Koordinate ordnet (vom kleinsten zum größten), dann zunächst Mittelwerte über diejenigen Werte y_i mit gleichem Wert x_i bildet und auf diese Reihe z_1, \dots, z_{n^*} gleitende Mediane anwendet.

Hier ist n^* die Anzahl unterschiedlicher Werte in x_1, \dots, x_n und für $j = 1, \dots, n^*$ ist z_j das arithmetische Mittel derjenigen Werte y_i mit gleichem, j -t-großtem Wert x_i .

Tab. 9: Körpergröße x_i , Gewicht y_i von $n = 40$ Personen, mittlere Gewichte z_j , gleitende Mediane 5 (z_j^*) und 53 (z_j^{**}); vgl. auch Tab. 1 in Kap. X

i	x_i	y_i	j	z_j	z_j^*	z_j^{**}
1	156	60	1	55	55	55
2	156	50	2	47	55	55
3	157	47	3	57,5	55	55
4	158	54	4	58,5	57,5	57,5
5	158	61	5	51	58,5	58,5
6	160	72	6	63	63	63
7	160	45	7	66,5	64	64
8	161	51	8	81	64	64
9	162	63	9	64	64	64
10	163	76	10	62	64	64
11	163	57				
12	164	81				
13	166	67				
14	166	61				
15	167	62				
16	168	56				
17	168	57				
18	168	66	11	58,5	62	62
19	168	60				
20	170	81	12	75	62	62
21	170	69	13	56	67,5	67,5
22	172	62				
23	173	77				
24	173	70	14	67,5	73	68
25	173	64				
26	173	65				
27	175	75				
28	175	73	15	73	68	73
29	175	70				
30	178	79	16	75,5	73	73
31	178	72	17	68	75,5	75,5
32	179	68				
33	180	82				
34	180	64	18	82	81	81
35	180	91				
36	183	82	19	82	82	82
37	184	74	20	81	82	82
38	184	88				
39	185	86				
40	185	84	21	85	85	85

Beispiel: In der Tab. 1 des Kap. X sind Größe x_i und Gewicht y_i von $n = 40$ Personen angegeben. Für die zweidimensionale Punkteschar (x_i, y_i) , $i = 1, \dots, 40$, soll nun ein gleitender Median 53 berechnet werden. In Tab. 9 sind die nach der KörpergröÙe geordneten Daten x_i und y_i und die Mittelwerte z_1, \dots, z_{21} der y_i für die $n^* = 21$ verschiedenen Werte unter den x_i angegeben. Weiterhin findet man in der Tabelle die gleitenden Mediane der Ordnung 5 (z_j^*) und die resultierenden gleitenden Mediane 53 (z_j^{**}). In Abb. 12 sind die originäre Punkteschar (durch Punkte) und die Glättung mit dem gleitenden Median 53 (durch Kreuze) verbunden zum

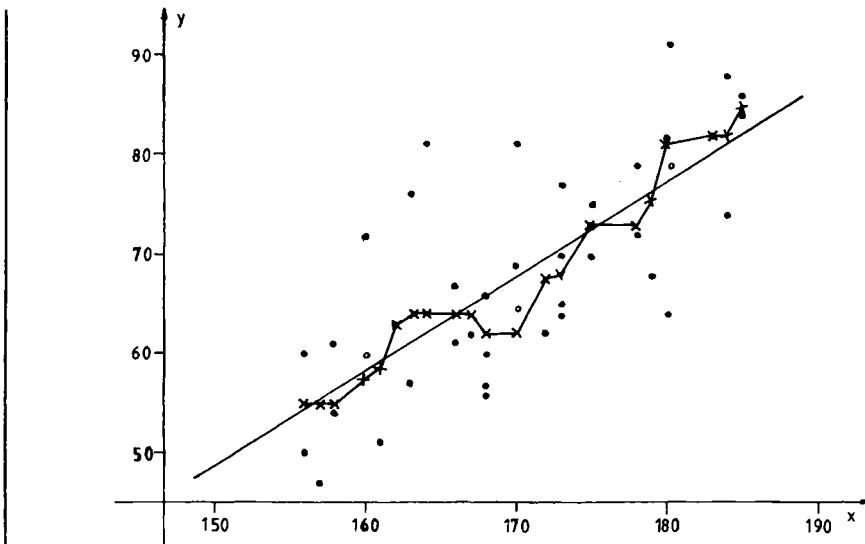


Abb. 12: Körpergröße x und Gewicht y für $n = 40$ Personen (Punkte), gleitende Mediane 53 (Kreuze), Medianpunkte (Kreise) und explorative Regressionsgerade

Polygonzug dargestellt. Auf die ebenfalls eingezeichnete Gerade und die Kreise gehen wir erst im Abschnitt 2.2 ein.

2.2. Explorative Regressionsgeraden

Im Abschnitt 1 des Kap. X haben wir uns mit der Schätzung von Regressionsgeraden nach der Methode der kleinsten Quadrate beschäftigt. Ist die Anzahl n der zur Verfügung stehenden Daten sehr groß, so ist die Bestimmung von Steigungsparameter und Absolutglied sehr aufwendig. Sind zudem in den Daten Ausreißer vorhanden, so beeinflussen diese die Parameterschätzungen.

Ein schnelles und gegen Ausreißer robustes Verfahren zur **explorativen Bestimmung einer Regressionsgeraden (resistant line regression)**

$$y = \alpha + \beta x$$

wollen wir in diesem Abschnitt behandeln.

Werden zu n Werten x_1, \dots, x_n eines Regressors die Regressandenwerte y_1, \dots, y_n beobachtet, so lassen sich die n Punkte $(x_1, y_1), \dots, (x_n, y_n)$ in einem zweidimensionalen Koordinatensystem abtragen. Nun wird die x-Achse in drei Teile unterteilt, derart daß im linken, mittleren und rechten Teil gleich viele der Werte x_1, \dots, x_n liegen. Dies ist natürlich nur dann möglich, wenn n durch 3 teilbar ist. Ist dies nicht der Fall, so verlangt man, daß im Falle $n + 1$ teilbar durch 3 im linken und rechten Teil $\frac{n+1}{3}$ und im mittleren Teil $\frac{n+1}{3} - 1$ der Werte x_1, \dots, x_n liegen und daß im

Fall $n - 1$ teilbar durch 3 im linken und rechten Teil $\frac{n-1}{3}$ und im mittleren Teil $\frac{n-1}{3} + 1$ der Werte x_1, \dots, x_n liegen.

Durch die Dreiteilung der x -Achse wird das gesamte Koordinatensystem in drei nebeneinanderliegende Abschnitte untergliedert. Nun werden in jedem Teil getrennt die Mediane der dort befindlichen Regressorwerte und Regressandenwerte bestimmt. Diese Mediane seien im linken Teil mit x_L, y_L , im mittleren Teil mit x_M, y_M und im rechten Teil mit x_R, y_R bezeichnet. Die Punkte $(x_L, y_L), (x_M, y_M), (x_R, y_R)$ heißen auch **Medianpunkte (summary points)** des Koordinatensystems.

Nun wird zunächst als Anfangsschätzung für den Steigungsparameter β der Wert

$$b_1 = (y_R - y_L)/(x_R - x_L)$$

und für das Absolutglied α der Wert

$$a_1 = \frac{1}{3}(y_L + y_M + y_R - b_1(x_L + x_M + x_R))$$

berechnet. Mit

$$\hat{y}_{L,1} = a_1 + b_1 x_L, \quad \hat{y}_{M,1} = a_1 + b_1 x_M, \quad \hat{y}_{R,1} = a_1 + b_1 x_R$$

und

$$e_{L,1} = y_L - \hat{y}_{L,1}, \quad e_{M,1} = y_M - \hat{y}_{M,1}, \quad e_{R,1} = y_R - \hat{y}_{R,1}$$

bestimmt man dann

$$b'_1 = (e_{R,1} - e_{L,1})/(x_R - x_L)$$

und daraus als neue Schätzung für β den Wert

$$b_2 = b_1 + b'_1.$$

Unter Verwendung von b_2 berechnet man sodann $\hat{y}_{L,2}, \hat{y}_{M,2}, \hat{y}_{R,2}, e_{L,2}, e_{M,2}$ und $e_{R,2}$, bestimmt

$$b'_2 = (e_{R,2} - e_{L,2})/(x_R - x_L)$$

und daraus als neue Schätzung für β

$$b_3 = b_2 - b'_2(b_2 - b_1)/(b'_2 - b'_1)$$

usw. bis im i -ten Schritt $b_i = 0$ gilt. Der zugehörige Wert $b_i = b$ wird dann als endgültige Schätzung für β verwandt und als resultierende Schätzung für das Absolutglied α erhält man

$$a = a_1 + \frac{1}{3}(e_{L,i-1} + e_{M,i-1} + e_{R,i-1}) \quad \text{für } i > 1$$

$$\text{und } a = a_1 \quad \text{für } i = 1,$$

d.h. die geschätzte explorative Regressionsgerade ist

$$\hat{y} = a + bx.$$

Beispiel: In Tab. 9 sind für $n = 40$ Personen Körpergröße x_i und Gewicht y_i zusammengestellt und in Abb. 12 sind die Punkte (x_i, y_i) für $i = 1, \dots, 40$ in einem Koordinatensystem abgetragen. Wir wollen nun für diese Daten eine explorative Regressionsgerade bestimmen. Mit $n = 40$, d.h. $n - 1 = 39$ ist teilbar durch 3, und $x_1 \leq x_2 \leq \dots \leq x_{40}$ liegen die ersten 13 Punkte ($i = 1, \dots, 13$) im linken, die mittleren 14 Punkte ($i = 14, \dots, 27$) im mittleren und die letzten 13 Punkte ($i = 28, \dots, 40$) im rechten Teil des Koordinatensystems. Aus den Daten der Tab. 9 berechnen sich dann die in Abb. 12 durch Kreise dargestellten Medianpunkte zu

$$(x_L, y_L) = (x_7, y_1) = (160; 60),$$

$$(x_M, y_M) = (\frac{1}{2}(x_{20} + x_{21}), \frac{1}{2}(y_{25} + y_{26})) = (170; 64,5) \quad \text{und}$$

$$(x_R, y_R) = (x_{34}, y_{30}) = (180; 79).$$

Damit erhalten wir

$$\begin{aligned} b_1 &= (79 - 60)/(180 - 160) = 19/20 = 0,95 \\ a_1 &= \frac{1}{3}(60 + 64,5 + 79 - 0,95(160 + 170 + 180)) = \frac{1}{3}(203,5 - 0,95 \cdot 510) \\ &= -93\frac{2}{3} \end{aligned}$$

und somit

$$\begin{aligned} \hat{y}_{L,1} &= -93\frac{2}{3} + 0,95 \cdot 160 = 58\frac{1}{3}, \quad e_{L,1} = 60 - 58\frac{1}{3} = 1\frac{2}{3}, \\ \hat{y}_{M,1} &= 67\frac{5}{6}, \quad e_{M,1} = -3\frac{1}{3}, \quad \hat{y}_{R,1} = 77\frac{1}{3}, \quad e_{R,1} = 1\frac{2}{3}. \end{aligned}$$

Es ist also

$$b'_1 = (e_{R,1} - e_{L,1})/(x_R - x_L) = (1\frac{2}{3} - 1\frac{2}{3})/20 = 0$$

und somit

$$b = b_1 \quad \text{und} \quad a = a_1 \quad \text{wegen} \quad i = 1.$$

Die geschätzte Regressionsgerade

$$\hat{y} = a + bx = -93\frac{2}{3} + 0,95x$$

ist bereits in Abb. 12 eingezeichnet.

2.3. Linearisieren zweidimensionaler Punktescharen

Im Abschnitt 1.3 haben wir uns mit Datentransformationen beschäftigt und dabei etwa zur Symmetrisierung einer Beobachtungsreihe x_1, \dots, x_n eine Transformation der Stärke p aus der „ladder of power“ verwandt.

Diese „ladder of power“ kann auch zur Linearisierung einer zweidimensionalen Punkteschar $(x_1, y_1), \dots, (x_n, y_n)$ verwandt werden. Welche Transformationen auf x_1, \dots, x_n bzw. y_1, \dots, y_n angewandt werden müssen, um Linearität zu erreichen, überprüft man unter Verwendung der Medianpunkte (x_L, y_L) , (x_M, y_M) und (x_R, y_R) , die – wie im Abschnitt 2.2 beschrieben – bestimmt werden. Aus ihnen berechnet man die Halbsteigungen (half-slopes)

$$b_L = \frac{y_M - y_L}{x_M - x_L} \quad \text{und} \quad b_R = \frac{y_R - y_M}{x_R - x_M}.$$

Ist der Quotient

$$q = b_R/b_L$$

gleich Eins, so ist die Punkteschar bereits linear und es ist keine Transformation vonnöten. Haben b_R und b_L unterschiedliche Vorzeichen, d.h. ist q negativ, so ist durch Power-Transformationen keine Linearisierung der Punkteschar möglich. Sind b_R und b_L positiv und $q < 1$, so erreicht man Linearität durch Transformationen von x_1, \dots, x_n mit Power $p_x < 1$ und/oder von y_1, \dots, y_n mit Power $p_y > 1$; bei positiven b_R und b_L mit $q > 1$ ist $p_x > 1$ und/oder $p_y < 1$ zu wählen, bei negativen b_R und b_L mit $q < 1$ ist $p_x < 1$ und/oder $p_y < 1$ und bei negativen b_R und b_L mit $q > 1$ ist $p_x > 1$ und/oder $p_y > 1$ zu wählen. Diese Regeln für die Wahl von p_x und p_y sind in Abb. 13 auch graphisch veranschaulicht. Wie p_x und p_y konkret zu wählen sind, um Linearität zu erreichen, kann im wesentlichen nur durch Ausprobieren festgestellt werden. Der Aufwand hierbei ist jedoch nicht sehr groß, da immer nur die

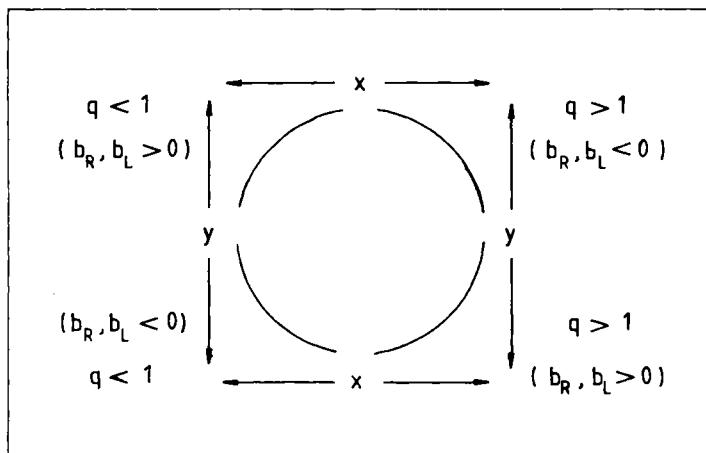


Abb. 13: Wahl der Transformationen zur Linearisierung zweidimensionaler Punktescharen in Abhängigkeit von den Halbsteigungen b_L , b_R und vom Quotienten q

Medianpunkte transformiert werden müssen, und die Überprüfung auf Linearität durch Bestimmung der Halbsteigungen der transformierten Medianpunkte erfolgen kann.

Hat man Power-Transformationen der Stärke p_x bzw. p_y derart bestimmt, daß die transformierte Punkteschar $(x_1^{p_x}, y_1^{p_y}), \dots, (x_n^{p_x}, y_n^{p_y})$ linear ist, so kann für sie mit den Methoden aus Abschnitt 2.2 eine explorative Regressionsgerade unter Verwendung der transformierten Medianpunkte

$$(x_L^{p_x}, y_L^{p_y}), (x_M^{p_x}, y_M^{p_y}) \text{ und } (x_R^{p_x}, y_R^{p_y})$$

bestimmt werden.

Beispiel: In der Tab. 10 sind die Gewichte y_i von Kükenembryonen in Gramm nach $x_i = i + 5$ Tagen für $i = 1, \dots, 11$ angegeben; für sie ergeben sich die Medianpunkte

$$(x_L, y_L) = (7,5; 0,0655), \quad (x_M, y_M) = (11; 0,2610)$$

$$(x_R, y_R) = (14,5; 1,5060),$$

so daß wir die Halbsteigungen

$$b_L = \frac{0,2610 - 0,0655}{11 - 7,5} = 0,056, \quad b_R = \frac{1,5060 - 0,2610}{14,5 - 11} = 0,356$$

und damit

$$q = b_R/b_L = 6,357$$

Tab. 10: Gewichte y_i von Kükenembryonen in Gramm nach $x_i = i + 5$ Tagen für $i = 1, \dots, 11$ (vgl. Snedecor, G. W./Cochran, W. G. (1967): Statistical methods. Iowa, p. 461)

i	1	2	3	4	5	6	7	8	9	10	11
x_i	6	7	8	9	10	11	12	13	14	15	16
y_i	0,029	0,052	0,079	0,125	0,181	0,261	0,425	0,738	1,130	1,882	2,812

erhalten. Da $q > 1$ und $b_L, b_R > 0$ sind, haben wir Transformationen mit Power $p_x > 1$ und/oder $p_y < 1$ zu wählen, um Linearität zu erreichen, vgl. Abb. 13.

In der Tab. 11 sind für verschiedene p_x und p_y die transformierten Medianpunkte, resultierende Halbsteigungen und Quotienten angegeben.

Tab. 11: Medianpunkte, Halbsteigungen und Quotienten für Transformationen mit verschiedenen Stärken p_x und p_y für die Daten aus Tab. 10

p_x	p_y	Transfor- ma- tionen	Medianpunkte					
			links	Mitte	rechts	b_L	b_R	q
1	1	(x, y)	(7,5; 0,0655)	(11; 0,2610)	(14,5; 1,506)	0,056	0,356	6,357
1	0,5	(x, \sqrt{y})	(7,5; 0,2546)	(11; 0,5109)	(14,5; 1,2174)	0,073	0,202	2,767
2	0,5	(x^2, \sqrt{y})	(56,5; 0,2546)	(121; 0,5109)	(210,5; 1,2174)	0,004	0,008	2,000
1	0	$(x, \ln y)$	(7,5; -2,7474)	(11; -1,3432)	(14,5; 0,3773)	0,401	0,492	1,227
2	0	$(x^2, \ln y)$	(56,5; -2,7474)	(121; -1,3432)	(210,5; 0,3773)	0,022	0,019	0,864
1	-0,5	$(x, -1/\sqrt{y})$	(7,5; -3,9716)	(11; -1,9574)	(14,5; -0,8348)	0,575	0,321	0,575
1	-0,1	$(x, -1/y^{0,1})$	(7,5; -1,3165)	(11; -1,1438)	(14,5; -0,9633)	0,0493	0,0516	1,047

Wir sehen in der Tab. 11, daß für $p_x = 1$ und $p_y = -0,1$ in etwa Linearität der Punkteschar $(x_1, -1/y_1^{0,1}), \dots, (x_n, -1/y_n^{0,1})$ erreicht wird. Für die so transformierten Daten wollen wir nun eine explorative Regressionsgerade bestimmen. Die Medianpunkte sind in Tab. 11 bereits angegeben, und wir erhalten

$$b_1 = \frac{-0,9633 - (-1,3165)}{14,5 - 7,5} = 0,0504,$$

$$a_1 = \frac{1}{3}(-3,424 - 0,0504 \cdot 33) = -1,6957.$$

Da sich mit

$$\begin{aligned} -\widehat{1/y_1^{0,1}} &= a_1 + b_1 x_L = -1,3177, \quad -\widehat{1/y_M^{0,1}} = -1,1413, \\ -\widehat{1/y_R^{0,1}} &= -0,9649 \end{aligned}$$

gerade

$$e_{L,1} = 0,0012, \quad e_{M,1} = -0,0027, \quad e_{R,1} = 0,0014$$

und somit

$$b'_1 = \frac{0,0014 - 0,0013}{14,5 - 7,5} = 0,00003 \simeq 0$$

ergibt, wollen wir

$$-\widehat{1/y^{0,1}} = a + bx = a_1 + b_1 x$$

als explorative Regressionsgerade verwenden. Als Schätzung für das Gewicht y selbst ergibt sich wegen

$$\widehat{y^{0,1}} = -1/(a + bx)$$

dann gerade

$$\hat{y} = (-1/(a + bx))^{10} = \frac{1}{(a + bx)^{10}}.$$

In der Tab. 12 sind neben x_i und y_i auch $-1/y_i^{0,1}$, \hat{y}_i und $\widehat{-1/y_i^{0,1}}$ angegeben. In der Abb. 14 sind die Punktescharen (links original und rechts transformiert) sowie die explorativen Regressionsfunktionen auch graphisch dargestellt.

Tab. 12: Gewichte y_i von Kükenembryonen am Tag $x_i = i + 5$, transformierte Gewichte $-1/y_i^{0,1}$, aus der explorativen Regressionsgeraden geschätzte Größen $\hat{y}_i = 1/(a + bx_i)^{10}$ für $i = 1, \dots, 11$ (mit $a = 0,0504$, $b = -1,6957$)

i	x_i	y_i	$-1/y_i^{0,1}$	$\widehat{-1/y_i^{0,1}}$	\hat{y}_i
1	6	0,029	-1,425	-1,3933	0,0363
2	7	0,052	-1,344	-1,3429	0,0524
3	8	0,079	-1,289	-1,2925	0,0769
4	9	0,125	-1,231	-1,2421	0,1144
5	10	0,181	-1,186	-1,1917	0,1731
6	11	0,261	-1,144	-1,1413	0,2667
7	12	0,425	-1,089	-1,0909	0,4189
8	13	0,738	-1,031	-1,0405	0,6723
9	14	1,130	-0,988	-0,9901	1,1046
10	15	1,882	-0,939	-0,9397	1,8625
11	16	2,812	-0,902	-0,8893	3,2323

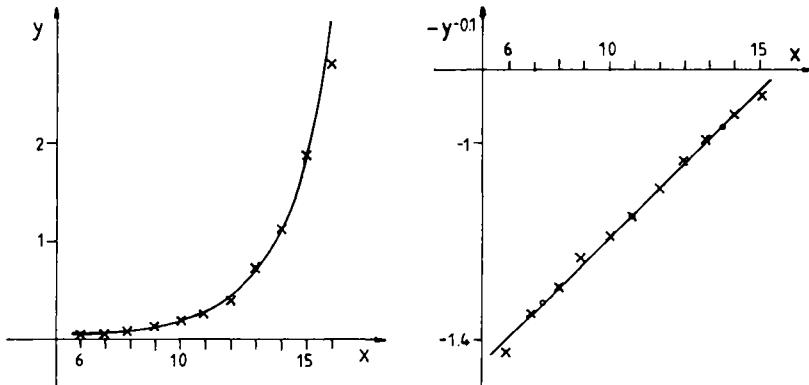


Abb. 14: Gewicht y von Kükenembryonen am Tag x und zugehörige explorative Regressionsfunktion (links); mit $p_y = -0,1$ transformierte Gewichte von Kükenembryonen am Tag x und zugehörige explorative Regressionsgerade (rechts)

3. Verfahren für mehrdimensionale Daten in der EDA

Wir sind bisher davon ausgegangen, daß an n Objekten ein oder zwei Merkmale beobachtet werden. Daraus ergaben sich eine Beobachtungsreihe x_1, \dots, x_n bzw. eine Reihe zweidimensionaler Daten $(x_1, y_1), \dots, (x_n, y_n)$. Werden $p \geq 2$ Merkmale

X_1, \dots, X_p an n Objekten erhoben, so spricht man allgemein von **mehrdimensionalem Datenmaterial**. Für das i -te Objekt ($i = 1, \dots, n$) erhält man dann ein p -Tupel (x_{i1}, \dots, x_{ip}) von Beobachtungswerten zu den p Merkmalen. Diese insgesamt $n \cdot p$ Tupel werden häufig in einer $n \times p$ -**Datenmatrix**

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

zusammengefaßt; die i -te Zeile von X enthält dann die am i -ten Objekt ($i = 1, \dots, n$) beobachteten Merkmalswerte und die j -te Spalte von X enthält alle an den n Objekten beobachteten Merkmalswerte zum Merkmal X_j ($j = 1, \dots, p$). Das Element x_{ij} der Datenmatrix X gibt somit den Merkmalswert zum j -ten Merkmal X_j an, der am i -ten Objekt beobachtet wurde.

Wir wollen hier in Abschnitt 3.1 den **Scatter-Plot**, ein schnelles Hilfsmittel zur graphischen Veranschaulichung mehrdimensionaler Daten vorstellen. Im Abschnitt 3.2 werden dann einige weitere explorative Verfahren für mehrdimensionale Daten beschrieben; für eine genaue und ausführliche Verfahrensdarstellung vergleiche man Hartung/Elpelt (1984).

3.1. Der Scatter-Plot

Werden p quantitative Merkmale X_1, \dots, X_p jeweils an n Objekten aus einer Grundgesamtheit beobachtet, so können die insgesamt $n \cdot p$ Beobachtungswerte in einer $n \times p$ -Datenmatrix X zusammengefaßt werden. Je größer p ist, desto unübersichtlicher ist eine solche Datenmatrix.

Eine schnelle Möglichkeit zur graphischen Veranschaulichung der Datenmatrix X und somit der Beziehungen zwischen den Merkmalen bzw. den Objekten bietet der **Scatter-Plot** (**Scatter-Plot-Matrix**, **Draftsman-Display**), insbesondere wenn er durch Computer-Einsatz unterstützt wird. Beim Scatter-Plot wird jedes der p Merkmale gegen jedes andere abgetragen, d. h. für jedes Paar (X_j, X_k) , $j \neq k$, von Merkmalen werden die n Beobachtungspunkte $(x_{1j}, x_{1k}), \dots, (x_{nj}, x_{nk})$ in einem zweidimensionalen Koordinatensystem abgetragen. Die insgesamt $p(p-1)$ Koordinatensysteme werden wie eine Matrix angeordnet, d. h. jedes Koordinatensystem entspricht einer quadratischen Fläche; die Quadrate in der Diagonale des Scatter-Plots bleiben leer.

Beispiel: Aufgrund einer Umfrage bei 396 amerikanischen Autobesitzern über Reparaturanfälligkeiten verschiedener Teile ihrer Wagen, die von 5 verschiedenen Herstellern stammen (1 ≡ American Motors, 2 ≡ Chrysler, 3 ≡ Ford, 4 ≡ General Motors, 5 ≡ bzgl. des US-Markts ausländische Hersteller) ergab sich eine 5×7 -Datenmatrix

$$X = \begin{pmatrix} 0,452 & -0,394 & -0,752 & -0,241 & -0,355 & 0,472 & 0,150 \\ 0,424 & -0,339 & -0,515 & -0,638 & 0,643 & -0,354 & 0,026 \\ 0,063 & -0,045 & -0,008 & -0,270 & -0,414 & -0,289 & -0,504 \\ -0,766 & -0,307 & 0,358 & 0,873 & 0,260 & -0,360 & -0,287 \\ 0,658 & 1,524 & 0,518 & 0,454 & -0,703 & 1,357 & 1,164 \end{pmatrix}$$

für die 5 Hersteller. In diese Matrix, die – wie in Hartung/Elpelt (1984, Kap. V, Abschnitt 3) detailliert beschrieben – mittels Skalierungsverfahren in Kontingenztabellen aus den Umfrageergebnissen gewonnen wurde, gehen sieben verschiedene Reparaturanfälligkeitssmerkmale ein:

- X_1 – äußere Karosserie,
- X_2 – innere Karosserie,
- X_3 – Eisenteile der Karosserie,
- X_4 – Bremsen,
- X_5 – Automatikgetriebe,
- X_6 – Stoßdämpfer,
- X_7 – Vorderradaufhängung.

In der Abb. 15 ist der Scatter-Plot zu dieser Datenmatrix zu sehen. Im zweituntersten Quadrat der dritten Spalte etwa sind die Daten (x_{i3}, x_{i2}) für $i = 1, \dots, 5$ abgetragen.

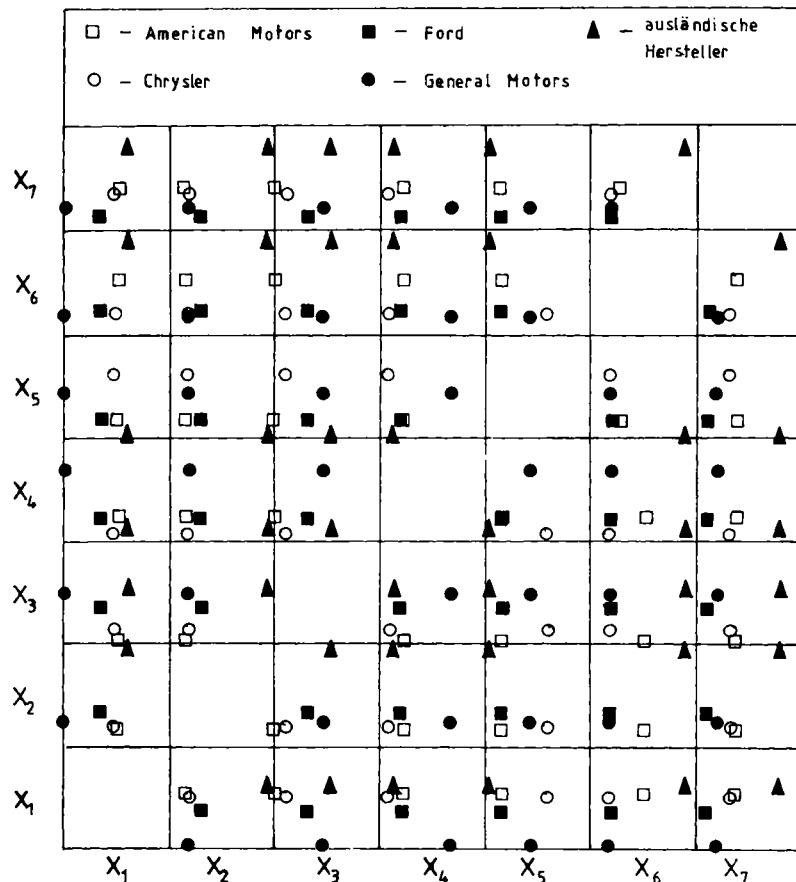


Abb. 15: Scatter-Plot für sieben Merkmale, die an 5 Auto-Herstellern erhoben wurden

Bei großen Datenmengen lassen sich die Möglichkeiten eines Scatter-Plots durch die Verwendung von Computern, auch hinsichtlich des Zeitaufwandes, verbessern. Um etwa Gruppen einander ähnlicher Objekte zu finden, lässt man jeweils einige Objektpunkte in jedem Quadrat aufblinken, um zu überprüfen, ob die zugehörigen Objekte stets eine Gruppe bilden (d. h. eng zusammen liegen) oder nicht. Diese und ähnliche Vorgehensweisen werden auch als „**Bürsten eines Scatter-Plots**“ bezeichnet.

3.2. Weitere explorative Verfahren für mehrdimensionale Daten

Wir haben in Abschnitt 3.1 den Scatter-Plot als ein graphisches Verfahren zur Veranschaulichung mehrdimensionaler oder multivariater Daten kennengelernt. Hier wollen wir nun kurz auf weitere multivariate Verfahren der explorativen Datenanalyse eingehen. Für eine ausführliche Darstellung vgl. man Hartung/Elpelt (1984).

Die **Multidimensionale Skalierung (MDS)** dient der Darstellung von Objekten in einem niedrigdimensionalen Repräsentationsraum. Dabei geht man von den Ähnlichkeiten der Objekte aus, die entweder aus den p an ihnen beobachteten Merkmalswerten berechnet oder direkt beobachtet werden, und sucht eine Konfiguration der Objekte im Repräsentationsraum derart, daß die Ähnlichkeiten durch die euklidischen Abstände möglichst gut repräsentiert werden; ist die gewählte Dimension q des Repräsentationsraumes kleiner oder gleich drei, so ermöglicht die MDS eine graphische Darstellung der Objekte in einem q -dimensionalen Koordinatensystem. Zu den verschiedenen Verfahren der MDS vergleiche man z. B. Kap. VI in Hartung/Elpelt (1984).

Die **Clusteranalyse**, vgl. etwa Kap. VII in Hartung/Elpelt (1984), ermöglicht es (aufgrund mehrdimensionaler Daten oder aufgrund von Objektähnlichkeiten), Gruppen ähnlicher Objekte zu bestimmen. Sie liefert also eine Einteilung der Objekte in Klassen, wobei die Objekte innerhalb einer Klasse möglichst homogen und die Klassen untereinander möglichst heterogen sind.

Die **Faktorenanalyse** versucht, ausgehend von Korrelationen der an n Objekten beobachteten p Merkmale, die Merkmale durch wenige, sie charakterisierende Faktoren zu beschreiben; man vgl. auch Kap. VIII in Hartung/Elpelt (1984). Sind adäquate Faktoren gefunden, so lassen sich die Objekte dann auch im Raum dieser Faktoren repräsentieren.

Ein wichtiges Gebiet der explorativen Analyse multivariater Daten sind **graphische Verfahren**, wie sie in Kap. IX von Hartung/Elpelt (1984) beschrieben werden. Insbesondere gibt es graphische Verfahren, die jedes Objekt einzeln aufgrund der an ihm beobachteten Merkmalswerte darstellen, und so einen visuellen Vergleich der Objekte ermöglichen. Dabei werden die beobachteten **Merkmalswerte durch Streichen** (z. B. Profile, Polygonzüge, Sterne, Sonnen, Glyphs) oder auch z. B. **durch Winkel** (Diamanten) repräsentiert. Eine ganz andere Darstellungsform sind **Gesichter**, wobei jedem Merkmal ein Gesichtsteil (z. B. Nase, Mund, Auge) zugeordnet wird, dessen Form dann in Abhängigkeit vom jeweiligen Merkmalswert variiert wird (Chernoff-faces, Flury-Riedwyl-faces). Auch **trigonometrische Funktionen** (Andrews-Plots, Blumen) können zur Repräsentation von Objekten verwandt werden. Komplexere Darstellungen von Objekten ergeben sich, wenn man die **Ähnlichkeiten** der an ihnen beobachteten **Merkmale** in die Darstellung miteinbezieht; Beispiele für derartige graphische Verfahren sind **Quader**, **Burgen** und **Bäume**. Die sogenannten **Facetten** berücksichtigen die **Diskriminationsgüte** der beobachteten Merkmale, d. h. die Güte, mit der die Merkmale zwischen den Objekten unterscheiden.

Zum Schluß sei hier noch der sogenannte **Bi-Plot**, vgl. Hartung/Elpelt (1984, Kap. IX), erwähnt. Hierbei werden Objekte und Merkmale gemeinsam in einem zweidimensionalen Koordinatensystem dargestellt. Die euklidischen Abstände zwischen den Objektpunkten repräsentieren dann ähnlich wie bei der multidimensionalen Skalierung die Ähnlichkeiten der Objekte. Die Merkmale werden durch Pfeile ausgehend vom Ursprung (0, 0) des Koordinatensystems dargestellt und zwar derart, daß die Winkel zwischen je zwei Pfeilen die Korrelationen zwischen den beiden zugehörigen Merkmalen beschreiben. Verwendet man das System der Pfeile (die Winkel) und trägt auf ihnen (für jedes Objekt) die zugehörigen Merkmalswerte als Strecken ab, so erhält man schließlich die **Bi-Plot-Sonnen**.

Zu erwähnen bleibt, daß viele Verfahren und graphische Darstellungsformen sich nur anwenden lassen, wenn die beobachteten Merkmale quantitativer Natur sind. Mittels einer **Datenaufbereitung** durch **Skalierung kategorialer Merkmale**, vgl. auch Abschnitt 7 in Kap. V, können nun auch qualitative oder gemischt quantitativ-qualitative Daten einer derartigen statistischen Analyse bzw. graphischer Darstellung zugänglich gemacht werden. Eine ausführliche Abhandlung dieser Vorgehensweise findet sich in Kap. V von Hartung/Elpelt (1984).

4. Robuste Schätzungen

Sind in einer Stichprobe Ausreißer vorhanden, so können viele Schätzfunktionen wie z. B. das arithmetische Mittel unbefriedigende Werte liefern, wenn diese Ausreißer nicht vorher eliminiert werden.

Beispiel: Eine Latte der Länge 5 m wird in 5 Stücke der Länge 1 m zersägt. Beim Nachmessen der 5 Stücke ergeben sich folgende Abweichungen in mm von der Sollänge 1 m:

$$-0,5; 1,5; -1,0; 0,5; -9,5.$$

Es fällt auf, daß das letzte Stück ungewöhnlich stark vom Sollwert 1 m abweicht. Hier sind nämlich beim Durchsägen der Latte jedes Mal ca. 2 mm durch Sägespäne verlorengegangen. Das arithmetische Mittel $\bar{x}_5 = -1,8$ aller 5 Meßwerte gibt deshalb ein falsches Bild von dem Justagezustand der Sägemaschine. Eine adäquate Beschreibung der Situation wird dagegen durch das Mittel $\bar{x}_4 = 0,125$ der ersten 4 Meßwerte gegeben.

Nicht immer ist wie im obigen Beispiel eine schnelle und sichere Identifizierung der Ausreißer möglich. Deshalb betrachtet man sogenannte **robuste Schätzfunktionen**, die von Ausreißern garnicht oder zumindest wenig beeinflußt werden. Eine solche Schätzfunktion ist z. B. der schon im Abschnitt 4 des Kap. I und im Abschnitt 1.1 dieses Kapitels beschriebene Median.

Beispiel: Für obiges Beispiel ergibt sich für den Median der Wert $m = \tilde{x}_{0,5} = -0,5$, wodurch der Justagezustand der Sägemaschine zufriedenstellend charakterisiert wird.

Die Robustheit des Medians gegenüber Ausreißern wird allerdings beim Vorliegen normalverteilter Daten mit einer im Vergleich zum arithmetischen Mittel reduzierten Güte erkauft. Wir werden uns in diesem Abschnitt mit verschiedenen robusten Schätzfunktionen beschäftigen. Zunächst betrachten wir jedoch in Abschnitt 4.1 einige Kenngrößen bzw. Kennfunktionen, die zur Charakterisierung der Robustheitseigenschaften solcher Schätzfunktionen dienen.

In Abschnitt 4.2 gehen wir kurz auf zwei spezielle robuste **Skalenschätzer**, also auf Kenngrößen für die Streuung eines Merkmals ein. Diese Kenngrößen sind der Median der absoluten Abweichungen vom Median und der Quartilsabstand, die in den nachfolgenden Abschnitten benötigt werden.

In den Abschnitten 4.3–4.5 beschäftigen wir uns mit drei speziellen Klassen robuster Schätzfunktionen, nämlich den M-, L- und R-Schätzern. Schätzfunktionen dieser Typen sind grundsätzlich bei beliebigen Problemstellungen einsetzbar. Wir werden uns hier jedoch auf den Fall von **Lokationsschätzern**, d. h. Schätzern für die Lage von Merkmalswerten, beschränken.

4.1. Charakterisierung von Robustheitseigenschaften

Kenngrößen, welche die **Robustheitseigenschaften einer Schätzfunktion** beschreiben, sind die Sensitivitätskurve, die Einflußkurve und der Bruchpunkt. Wir werden im folgenden jede Schätzfunktion allgemein mit $T_n = T_n(x_1, \dots, x_n)$ bezeichnen, wenn sie auf einer Stichprobe x_1, \dots, x_n vom Umfang n basiert. So ist zum Beispiel das arithmetische Mittel der Beobachtungswerte x_1, \dots, x_n eine Schätzfunktion

$$T_n = T_n(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n = \bar{x}$$

für das Zentrum der zugrundeliegenden Verteilung, die natürlich von den konkreten Beobachtungswerten abhängt. Betrachtet man das Grenzverhalten einer Schätzfunktion T_n für wachsenden Stichprobenumfang n ($n \rightarrow \infty$), so bezeichnen wir den Grenzwert der Schätzfunktion mit T .

4.1.1. Die Sensitivitätskurve

Ist $T_n = T_n(x_1, \dots, x_n)$ eine Schätzfunktion basierend auf n Beobachtungswerten, so nennt man die Funktion

$$SC(x; x_1, \dots, x_{n-1}, T_n) = n \cdot (T_n(x_1, \dots, x_{n-1}, x) - T_{n-1}(x_1, \dots, x_{n-1}))$$

die **Sensitivitätskurve (sensitivity curve)** der Schätzfunktion. Die Sensitivitätskurve beschreibt also die Reaktion einer Schätzfunktion, wenn zu einer Stichprobe x_1, \dots, x_{n-1} noch ein Wert x ergänzt wird. Ihr Wert ist gerade die n -fache Differenz von T_n und T_{n-1} in Abhängigkeit vom Wert x .

Beispiel: Mit den Daten $x_1 = -0,5$, $x_2 = 1,5$, $x_3 = -1$ und $x_4 = 0,5$ ($n - 1 = 4$) ergibt sich die Sensitivitätskurve des arithmetischen Mittels $T_5 = \bar{x} = \bar{x}_5$ zu

$$\begin{aligned} SC(x, x_1, \dots, x_4, \bar{x}_5) &= 5 \cdot (\frac{1}{5}(-0,5 + 1,5 - 1 + 0,5 + x) - \frac{1}{4}(-0,5 + 1,5 - 1 + 0,5)) \\ &= 5 \cdot (0,1 + \frac{1}{5}x - 0,125) = x - 0,125. \end{aligned}$$

Entsprechend erhält man für den Median $T_5 = m = m_5$, vgl. z. B. Abschnitt 1.1 dieses Kapitels,

$$SC(x; x_1, \dots, x_4, m_5) = \begin{cases} -2,5, & x < -0,5 \\ 5x, & \text{falls } -0,5 \leq x \leq 0,5 \\ 2,5, & 0,5 < x \end{cases}$$

In Abb. 16 sind diese beiden Sensitivitätskurven auch graphisch dargestellt. Es wird hier deutlich, daß das arithmetische Mittel im Gegensatz zum Median durch Hinzunahme eines einzelnen Beobachtungswertes x in beliebiger Weise verändert

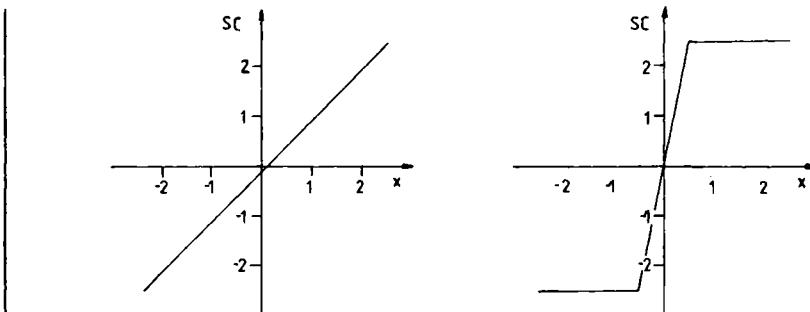


Abb. 16: Sensitivitätskurven $SC(x)$ für das arithmetische Mittel (links) und den Median (rechts)

werden kann. Liegt – bei Vorliegen obiger Stichprobe x_1, \dots, x_4 – der neue Wert x zwischen $-0,5$ und $0,5$, so reagiert allerdings der Median sehr viel heftiger als das arithmetische Mittel.

4.1.2. Die Einflußkurve

Die Sensitivitätskurve ist zur allgemeinen Beschreibung von Robustheitseigenschaften einer Schätzfunktion nur bedingt geeignet, da sie nicht nur von der Schätzfunktion selbst, sondern auch noch von den konkreten Beobachtungswerten und deren Anzahl abhängig ist. Deshalb geht man zumeist über zur **Einflußkurve (influence curve)**, die einer Folge von Stichprobenergebnissen x_1, x_2, \dots aus einer Verteilung mit Verteilungsfunktion F den Grenzwert

$$IC(x; F, T) = \lim_{n \rightarrow \infty} SC(x; x_1, \dots, x_{n-1}, T_n)$$

zuordnet.

Beispiel: Ist F die Verteilungsfunktion einer Normalverteilung mit Mittelwert μ und Varianz σ^2 , dann ist die Einflußkurve für das arithmetische Mittel \bar{x} gegeben durch

$$IC(x; F, \bar{x}) = x - \mu,$$

und im Falle des Medians m erhalten wir

$$IC(x; F, m) = \frac{1}{2} \cdot \sqrt{2\pi} \cdot \sigma \cdot \text{sgn}(x - \mu) = \begin{cases} -\sqrt{\pi}/2 \cdot \sigma, & x - \mu < 0 \\ 0, & \text{falls } x - \mu = 0 \\ \sqrt{\pi}/2 \cdot \sigma, & x - \mu > 0 \end{cases}$$

Beide Einflußkurven sind in Abb. 17 für $\mu = 0$ und $\sigma = 1$ dargestellt.

Die **Einflußkurve** beschreibt die Reaktion der Schätzfunktion auf einen zusätzlichen Wert in der Stichprobe, wenn von einer sehr großen Stichprobe aus der zugrundeliegenden Verteilung ausgegangen wird, ist also eine **lokale Kenngröße** der Robustheit.

Der maximale Effekt, den eine einzelne Beobachtung auf den Schätzwert haben kann, wird durch den maximalen Wert des Betrages der Einflußkurve beschrieben. Man nennt diese Größe deshalb auch **Ausreißerempfindlichkeit (gross-error sensitivity)**. Das arithmetische Mittel hat hier im Gegensatz zum Median einen unendlichen Wert.

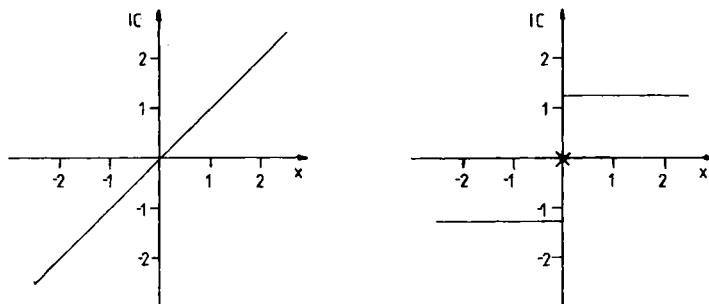


Abb. 17: Einflußkurven $IC(x)$ für das arithmetische Mittel (links) und den Median (rechts)

Ein anderer Aspekt ist das Verhalten der Schätzfunktion bei kleinen Änderungen der Stichprobenwerte, die sich etwa durch Rundung oder Klassenbildung ergeben können; hierfür ist die Steigung der Einflußkurve maßgebend. Je größer der maximale Wert des Absolutbetrages der Steigung ist, desto stärker können solche Änderungen auf den Schätzwert durchschlagen. Man spricht hier von **Rundungsempfindlichkeit (local-shift sensitivity)**. Beim Median ergibt sich hier ein unendlicher Wert, während zum arithmetischen Mittel der endliche Wert 1 gehört.

Eine Eigenschaft mancher robuster Schätzfunktionen, die weder beim Median noch beim arithmetischen Mittel auftritt, ist das **vollständige Ignorieren** weit außerhalb liegender Stichprobenwerte. Dabei wird die Einflußkurve identisch Null, wenn der Abstand vom zu schätzenden Parameter eine gewisse Grenze überschreitet. Als Beispiel für solche Schätzfunktionen werden wir später etwa Andrews' wave und Tukey's biweight kennenlernen.

4.1.3. Der Bruchpunkt

Der **Bruchpunkt (breakdown point)** ε^* gibt diejenige Grenze an, bis zu welcher der Anteil von Ausreißern in einer Stichprobe steigen darf, ohne daß sich dadurch der Schätzwert unbeschränkt verändern kann. Im Gegensatz zur Einflußkurve ist der Bruchpunkt eine **globale Kenngröße** der Robustheit. Der maximal erreichbare Wert beträgt $\varepsilon^* = 0,5$.

Beispiel: Beim arithmetischen Mittel reicht es schon aus, einen einzigen Wert in einer Stichprobe zu verändern, um den Schätzwert über alle Grenzen wachsen zu lassen. Der Bruchpunkt des arithmetischen Mittels ist deshalb $\varepsilon^* = 0$.

Eine beliebige Veränderung des Medians durch Ausreißer in der Stichprobe ist dagegen nicht möglich, wenn ihr Anteil weniger als 50 % beträgt. Der Median hat daher als Bruchpunkt den höchst möglichen Wert $\varepsilon^* = 0,5$.

4.2. Robuste Skalenschätzer

Wir werden uns später zwar im wesentlichen nur noch mit robusten Lokations-schätzern beschäftigen, benötigen dazu aber teilweise Skalenschätzer, die die Streuung eines Merkmals beschreiben. Deshalb sollen hier nun einige gebräuchliche Schätzungen für Skalenparameter vorgestellt und deren Robustheitseigenschaften beschrieben werden.

Der gebräuchlichste Skalenschätzer ist die **Standardabweichung**, vgl. z. B. Ab-

schnitt 5 in Kap.I, einer Stichprobe x_1, \dots, x_n :

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Sie kann, wie das arithmetische Mittel, durch die Veränderung eines einzigen Stichprobenwertes über alle Grenzen wachsen, d. h. der Bruchpunkt der Standardabweichung s ist $\varepsilon^* = 0$. Somit ist s im Gegensatz zu den in den Abschnitten 4.2.1 und 4.2.2 vorgestellten Schätzern, kein robuster Skalenschätzer.

4.2.1. Der Median der absoluten Abweichungen vom Median

Der **Median der absoluten Abweichungen vom Median MAD (median absolute deviation)** einer Stichprobe x_1, \dots, x_n berechnet sich als Median m_y der Größen

$$y_i = |x_i - m_x| \quad \text{für } i = 1, \dots, n,$$

wobei m_x den Median von x_1, \dots, x_n bezeichnet.

Der MAD wird mitunter noch durch Multiplikation mit einer Konstanten korrigiert. Sind x_1, \dots, x_n etwa Realisationen aus einer $N(\mu, \sigma^2)$ -Verteilung, so ist der MAD ein Schätzer für

$$u_{0,75} \cdot \sigma = 0,6745 \cdot \sigma,$$

wobei $u_{0,75}$ das 0,75-Quantil der Standardnormalverteilung bezeichnet. Zur Schätzung von σ selbst wird in diesem Fall daher $MAD/0,6745 = 1,4828$ MAD verwandt. Wie der Median hat der MAD oder auch ein korrigierter MAD den maximalen Bruchpunkt $\varepsilon^* = 0,5$ und seine Einflußkurve ist beschränkt, so daß ein einzelner Beobachtungswert auch nur beschränkten Einfluß auf den Schätzwert haben kann.

Beispiel: Wir wollen den MAD der 5 Beobachtungswerte $x_1 = -0,5$, $x_2 = 1,5$, $x_3 = -1$, $x_4 = 0,5$ und $x_5 = -9,5$ bestimmen. Dazu berechnen wir zunächst den Median $m_x = -0,5$ von x_1, \dots, x_5 und erhalten

$$\begin{aligned} y_1 &= |x_1 - m_x| = |-0,5 - (-0,5)| = |-0,5 + 0,5| = 0, \\ y_2 &= 2, \quad y_3 = 0,5, \quad y_4 = 1 \quad \text{und} \quad y_5 = 9. \end{aligned}$$

Es ist also

$$\text{MAD} = m_y = 1.$$

Würde man hier $x_5 = -9,5$ als „Ausreißer“ nicht berücksichtigen, so hätte man mit $m_x = \frac{1}{2}(-0,5 + 0,5) = 0$, d.h. $y_i = |x_i|$ für $i = 1, \dots, 4$, den Wert $\text{MAD} = \frac{1}{2}(0,5 + 1) = 0,75$. Der Wert der Sensitivitätskurve an der Stelle $x = -9,5$ würde sich somit zu

$$\text{SC}(-9,5; x_1, \dots, x_4, \text{MAD}_5) = 5(0,75 - 1) = -1,25$$

ergeben. Im Vergleich dazu ergibt sich bei Berücksichtigung von x_1, \dots, x_5 eine Standardabweichung von $s = s_5 = 4,41$, wohingegen sich für x_1, \dots, x_4 der Wert $s = s_4 = 1,11$ berechnen würde, d.h. es ist

$$\text{SC}(-9,5; x_1, \dots, x_4, s_5) = 5(4,41 - 1,11) = 16,5.$$

4.2.2. Der Quartilsabstand

Im Abschnitt 4 des Kap. I haben wir oberes und unteres Quartil (0,25- bzw. 0,75-Quantil) kennengelernt. Die Differenz dieser beiden Größen

$$IQR = \tilde{x}_{0,75} - \tilde{x}_{0,25}$$

bezeichnet man auch als **Quartilsabstand (interquartile range)** einer Stichprobe x_1, \dots, x_n .

Bei symmetrischen Verteilungen ist 0,5 IQR asymptotisch äquivalent zum MAD und hat deshalb auch die gleiche Einflußkurve. Der Bruchpunkt beträgt jedoch nur $\varepsilon^* = 0,25$. Dieser gegenüber dem Bruchpunkt des MAD halb so große Wert kommt allerdings nur dann zum Tragen, wenn Ausreißer überwiegend an einem Ende der Stichprobe auftreten.

Beispiel: Hat man $x_1 = -0,5$, $x_2 = 1,5$, $x_3 = -1$, $x_4 = 0,5$ und $x_5 = -9,5$ beobachtet, so ergibt sich mit $\tilde{x}_{0,25} = -1$ und $\tilde{x}_{0,75} = 0,5$, vgl. Abschnitt 4 in Kap. I, der Quartilsabstand zu

$$IQR = \tilde{x}_{0,75} - \tilde{x}_{0,25} = 0,5 - (-1) = 1,5.$$

Kommt zu den obigen Daten noch der Ausreißer $x_6 = 10,5$ hinzu, so berechnet man

$$IQR = \tilde{x}_{0,75} - \tilde{x}_{0,25} = 1,5 - (-1) = 2,5,$$

erhält also einen Wert in der gleichen Größenordnung wie oben. Tauscht man $x_6 = 10,5$ gegen $x_6 = -10,5$ aus, so ergibt sich

$$IQR = \tilde{x}_{0,75} - \tilde{x}_{0,25} = 0,5 - (-9,5) = 10.$$

Diese starke Veränderung des Schätzwertes wird dadurch verursacht, daß die beiden Ausreißer $x_6 = -10,5$ und $x_5 = -9,5$ am gleichen Ende der Stichprobe liegen und ihr Anteil an der Stichprobe $2/6 = 0,33$ beträgt, also größer ist als der Bruchpunkt $\varepsilon^* = 0,25$.

Zum Vergleich wollen wir noch den MAD für alle 3 Datensituationen angeben. Für die Ausgangswerte wurde schon im Beispiel des Abschnitts 4.2.1 $MAD = 1$ berechnet. Mit dem Ausreißer $x_6 = 10,5$ erhält man $MAD = 1,25$ und mit dem Ausreißer $x_6 = -10,5$ den Wert $MAD = 1,75$.

4.3. M-Schätzer für die Lokation

Ein **M-Schätzer** ergibt sich generell als Lösung einer Minimierungsaufgabe. Die Gestalt dieser Minimierungsaufgabe ist bei Lokationsschätzern abhängig davon, ob diese skaleninvariant sein sollen, d. h. ob sie von der Wahl der Meßskala für die Stichprobenwerte x_1, \dots, x_n unabhängig sind, oder keine Skaleninvarianz erreicht werden soll.

Nicht-skaleninvariante M-Schätzer T_n für die Lokation werden vermittels einer reellwertigen Funktion $\varrho(x)$ aus einer Stichprobe x_1, \dots, x_n bestimmt, indem die **Minimierungsaufgabe**

$$\text{minimiere } \sum_{i=1}^n \varrho(x_i - t) \quad \text{über alle reellen Zahlen } t$$

gelöst wird. Der Wert t , für den $\sum_{i=1}^n \varrho(x_i - t)$ minimal wird, ist dann gerade der

Schätzwert T_n basierend auf der Stichprobe x_1, \dots, x_n . Ist die Funktion $\varphi(x)$ differenzierbar und bezeichnet man die Ableitung mit $\psi(x)$, so gilt also

$$\sum_{i=1}^n \psi(x_i - T_n) = 0;$$

bei streng monotonem $\psi(x)$ ergibt sich T_n somit als (eindeutige) Lösung dieser Gleichung. Oft werden nicht-skaleninvariante M-Schätzer auch direkt durch die Angabe einer Funktion $\psi(x)$ definiert.

Beispiel: Für die Funktion $\varphi(x) = x^2/2$, d.h. $\psi(x) = x$, ergibt sich aus der Gleichung

$$0 = \sum_{i=1}^n \psi(x_i - T_n) = \sum_{i=1}^n (x_i - T_n) = \sum_{i=1}^n x_i - nT_n$$

gerade

$$nT_n = \sum_{i=1}^n x_i \quad \text{bzw.} \quad T_n = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x},$$

d.h. der Schätzer T_n für die Lokation ist gerade das arithmetische Mittel \bar{x} .

Die **Einflußkurve** eines nicht-skaleninvarianten M-Schätzers unterscheidet sich nur durch eine von der zugrundegelegten Verteilung abhängigen Konstante $C(F)$ von der Funktion $\psi(x)$. Ist $f(x)$ die Dichte dieser Verteilung mit Verteilungsfunktion $F(x)$ und $E_F T$ der Erwartungswert der Schätzfunktion T unter dieser Verteilung, so gilt

$$IC(x; F, T) = C(F) \cdot \psi(x - E_F T),$$

wobei die Konstante $C(F)$ gegeben ist durch

$$C(F) = (\int \psi'(x - E_F T) f(x) dx)^{-1};$$

$\psi'(x)$ bezeichnet den Wert der Ableitung von ψ an der Stelle x .

Beispiel: Wir haben gesehen, daß die ψ -Funktion $\psi(x)$ des arithmetischen Mittels \bar{x} gerade $\psi(x) = x$ ist. Damit erhalten wir die Ableitung

$$\psi'(x) = 1$$

und somit

$$C(F) = (\int 1 \cdot f(x) dx)^{-1} = (\int f(x) dx)^{-1} = 1^{-1} = 1.$$

Die Einflußkurve des arithmetischen Mittels \bar{x} ergibt sich also zu

$$IC(x, F; \bar{x}) = x - E_F \bar{x}.$$

Ist F z.B. die Verteilungsfunktion einer Verteilung mit Erwartungswert μ , so ist – man vgl. auch Abschnitt 4.1.2 –

$$IC(x, F; \bar{x}) = x - \mu.$$

Aus der Gestalt der Einflußkurve von nicht-skaleninvarianten M-Schätzern erkennt man, daß sich zu vorgegebener Form der Einflußkurve leicht der entsprechende M-Schätzer ergibt. Man muß lediglich die Funktion $\psi(x)$ so festlegen, daß sie die gleiche Form besitzt wie die gewünschte Einflußkurve. Bei den M-Schätzern ist somit die Form der Einflußkurve stets unabhängig von der zugrundeliegenden Verteilung.

Außerdem lässt sich der **Bruchpunkt** ε^* eines nicht-skaleninvarianten M-Schätzers leicht bestimmen, wenn $\psi(x)$ eine monotone, beschränkte Funktion ist, die sowohl positive als auch negative Werte annimmt. Es gilt dann nämlich

$$\varepsilon^* = \frac{1}{1 + \eta} \quad \text{mit} \quad \eta = \min\left(-\frac{\psi(-\infty)}{\psi(+\infty)}, -\frac{\psi(+\infty)}{\psi(-\infty)}\right).$$

Ist also $-\psi(-\infty) = \psi(+\infty)$, so ergibt sich stets $\eta = 1$ und damit der maximale Bruchpunkt $\varepsilon^* = 0,5$. Ist die Funktion $\psi(x)$ unbeschränkt, so ist der Bruchpunkt des M-Schätzers stets $\varepsilon^* = 0$.

Beispiel: Der Median $T_n = m$ ist ein M-Schätzer mit $\varrho(x) = |x|$ und

$$\psi(x) = \text{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{falls } x = 0 \\ -1, & x < 0 \end{cases}$$

Da $\psi(x)$ monoton und beschränkt ist mit $-\psi(-\infty) = \psi(+\infty) = 1$, ergibt sich der Bruchpunkt des Medians zu $\varepsilon^* = 0,5$.

Das arithmetische Mittel $T_n = \bar{x}$ besitzt die unbeschränkte ψ -Funktion $\psi(x) = x$ und somit den Bruchpunkt $\varepsilon^* = 0$.

In den Beispielen haben wir bisher das arithmetische Mittel \bar{x} und den Median m als M-Schätzer kennengelernt. Diese M-Schätzer sind automatisch skaleninvariant. Für allgemeine Funktionen $\varrho(x)$ und $\psi(x)$, wie weiter oben beschrieben, kann nicht immer Skaleninvarianz erreicht werden, wenn $\sum_{i=1}^n \varrho(x_i - t)$ minimiert wird. Ein Beispiel hierzu werden wir im Abschnitt 4.3.1 kennenlernen.

Um generell **Skaleninvarianz** bei M-Schätzern zu erreichen, wird zunächst eine Skalenschätzung S_n z.B. mit Hilfe des MAD, vgl. Abschnitt 4.2, vorgenommen und dann ausgehend von einer Stichprobe x_1, \dots, x_n

$$\sum_{i=1}^n \varrho\left(\frac{x_i - t}{cS_n}\right) \quad \text{bzgl. } t \text{ minimiert,}$$

wobei c eine positive Streuungskonstante bezeichnet. Der Wert t , für den obige Summenfunktion minimal wird, ist dann ein skaleninvarianter M-Schätzer $T_n = t$ für die Lokation. Ist die Funktion $\varrho(x)$ differenzierbar, so gilt also für die Ableitung $\psi(x)$

$$\sum_{i=1}^n \psi\left(\frac{x_i - T_n}{cS_n}\right) = 0.$$

Die **Einflußkurve** des skaleninvarianten M-Schätzers ist gerade

$$\text{IC}(x; F, T) = C(F) \cdot \psi\left(\frac{x - E_F T}{cE_F S}\right)$$

mit

$$C(F) = cE_F S \left(\int \psi'\left(\frac{x - E_F T}{cE_F S}\right) f(x) dx \right)^{-1},$$

wobei $E_F S$ bzw. $E_F T$ die Erwartungswerte von S und T unter der zugrundeliegenden Verteilung mit Dichte $f(x)$ und Verteilungsfunktion $F(x)$ bezeichnet ($S_n \rightarrow S$, $T_n \rightarrow T$ für $n \rightarrow \infty$).

Der **Bruchpunkt** eines skaleninvarianten M-Schätzers lässt sich wie folgt bestimmen. Ist ε_M^* der Bruchpunkt des nicht-skaleninvarianten M-Schätzers mit gleicher ϱ -bzw. ψ -Funktion und ε_S^* der Bruchpunkt des verwandten Skalenschätzers, so ergibt sich der Bruchpunkt ε_{MS}^* des skaleninvarianten M-Schätzers gerade als Minimum dieser beiden Werte:

$$\varepsilon_{MS}^* = \min(\varepsilon_M^*, \varepsilon_S^*).$$

Wir werden nun in den Abschnitten 4.3.1 und 4.3.2 spezielle M-Schätzer betrachten, im Abschnitt 4.3.3 Berechnungsmethoden für M-Schätzer vorstellen und im Abschnitt 4.3.4 die verschiedenen M-Schätzer anhand von Beispielen miteinander vergleichen.

4.3.1. Huber-k-Schätzer

Huber-k-Schätzer sind eine Klasse von M-Schätzern, in der als Grenzfälle das arithmetische Mittel und der Median enthalten sind.

Der Huber-k-Schätzer wird durch die ϱ -Funktion

$$\varrho(x) = \begin{cases} x^2/2, & \text{falls } |x| \leq k \\ k \cdot |x| - k^2/2, & \text{falls } |x| > k \end{cases}$$

bzw. die ψ -Funktion

$$\psi(x) = \begin{cases} x, & \text{falls } |x| \leq k \\ k \cdot \text{sgn}(x), & \text{falls } |x| > k \end{cases}$$

festgelegt, wobei k eine positive, reelle Konstante ist. Hier bezeichnet wieder $\text{sgn}(x)$ die Vorzeichenfunktion Signum, d. h.

$$\text{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{falls } x = 0. \\ -1, & x < 0 \end{cases}$$

Ist k so klein, daß sämtliche Stichprobenwerte x_1, \dots, x_n , die nicht identisch mit dem Median m der Stichprobe sind, einen Abstand größer als k vom Median haben, so ist der Huber-k-Schätzer identisch mit dem Median. Ist dagegen k größer als die halbe Spannweite der Stichprobenwerte, vgl. Abschnitt 5 in Kap. I, so erhält man als Huber-k-Schätzer das arithmetische Mittel \bar{x} . In diesen beiden Fällen ist der Huber-k-Schätzer dann natürlich automatisch skaleninvariant, was sonst nicht der Fall ist.

Wir werden uns nun zunächst mit **nicht-skaleninvarianten** Huber-k-Schätzern beschäftigen. Der nicht-skaleninvariante Huber-k-Schätzer für die Lokation ergibt sich bei Vorliegen einer Stichprobe x_1, \dots, x_n als eindeutige Lösung der Gleichung

$$\sum_{i=1}^n \psi(x_i - T_n) = 0.$$

Außer mit dieser Gleichung lässt sich auch wie folgt überprüfen, ob eine Schätzung T_n tatsächlich Huber-k-Schätzer ist. Bezeichnet $x_{(1)}, \dots, x_{(n)}$ die geordnete Reihe der Stichprobenwerte, d. h. $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, so sucht man zunächst das kleinste Intervall $[x_{(s)}, x_{(n-s+1)}]$, das alle Stichprobenwerte aus $[T_n - k, T_n + k]$ enthält. In diesem Intervall werden nun alle Stichprobenwerte, die größer als $T_n + k$ sind, durch $T_n + k$ bzw. alle Stichprobenwerte, die kleiner als $T_n - k$ sind, durch T_n

– k ersetzt. Ist T_n ein Huber- k -Schätzer, so muß nun das arithmetische Mittel der im Intervall $[T_n - k, T_n + k]$ liegenden, zum Teil modifizierten Stichprobenwerte identisch mit T_n sein.

Beispiel: Ausgehend von den $n = 6$ der Größe nach geordneten Stichprobenwerten $x_{(1)} = -1,5$, $x_{(2)} = 0,5$, $x_{(3)} = 1,1$, $x_{(4)} = 1,9$, $x_{(5)} = 2,3$ und $x_{(6)} = 11,5$, die in mm gemessen wurden, wollen wir zeigen, daß der Huber-4-Schätzer den Wert $T_6 = 1,66$ besitzt.

Das kleinste Intervall der Form $[x_{(s)}, x_{(6-s+1)}]$, das alle Werte aus

$$[T_6 - 4, T_6 + 4] = [-2,34; 5,66]$$

enthält, ist

$$[x_{(1)}, x_{(6)}] = [-1,5; 11,5].$$

Wir ersetzen nun den oberhalb von $T_6 + 4$ liegenden Stichprobenwert $x_{(6)} = 11,5$ durch 5,66 und erhalten als arithmetisches Mittel der modifizierten, im Intervall $[-2,34; 5,66]$ liegenden Stichprobenwerte

$$\bar{x} = \frac{1}{6}(-1,5 + 0,5 + 1,1 + 1,9 + 2,3 + 5,66) = 1,66.$$

Damit ist $T_6 = 1,66$ tatsächlich der Huber-4-Schätzer.

Die **Einflußkurve** des nicht-skaleninvarianten Huber- k -Schätzers ergibt sich zu

$$IC(x; F, T) = (F(k) - F(-k))\psi(x - E_F T).$$

In Abb. 18 sind im Falle der Standardnormalverteilung für verschiedene k die Einflußkurven graphisch dargestellt. Hier sieht man deutlich, daß bei kleinem k eine Annäherung an die Einflußkurve des Medians und bei großem k eine Annäherung an die Einflußkurve des Mittelwerts erfolgt.

Da beim Huber- k -Schätzer immer $-\psi(-\infty) = \psi(+\infty)$ gilt, hat der **Bruchpunkt** für jeden Wert $k > 0$ den maximalen Wert $\varepsilon^* = 0,5$.

Anhand des obigen Beispiels wollen wir nun zeigen, daß die bisher betrachteten Huber- k -Schätzer nicht automatisch skaleninvariant sind.

Beispiel: Werden die Werte aus dem vorhergehenden Beispiel nicht in mm, sondern in cm angegeben, so ergibt sich als Huber-4-Schätzer statt der erwarteten 1,66 mm = 0,166 cm der Wert 0,2633 cm. Es sind nämlich sämtliche Stichprobenwerte weniger als 4 cm vom Schätzwert entfernt, so daß der Huber-4-Schätzer mit dem arithmetischen Mittel übereinstimmt. Wählt man dagegen $k = 0,4$, so erhält man als Huber-0,4-Schätzer für die cm-Werte 0,166 cm = 1,66 mm, d. h. der Huber-0,4-Schätzer für die cm-Werte ist identisch mit dem Huber-4-Schätzer für die mm-Werte. Das liegt daran, daß 4 mm gerade 0,4 cm entspricht.

Macht man, wie weiter vorne in diesem Abschnitt 4.3 beschrieben, den Huber- k -Schätzer mittels einer Skalenschätzung S_n invariant, d. h. berechnet man den Schätzwert T_n aus der Gleichung

$$\sum_{i=1}^n \psi\left(\frac{x_i - T_n}{cS_n}\right) = 0,$$

wobei c eine Skalenkonstante ist, so ist dieser **skaleninvariante Huber- k -Schätzer** identisch mit dem nicht-skaleninvarianten Huber- k^* -Schätzer mit $k^* = cS_n k$.

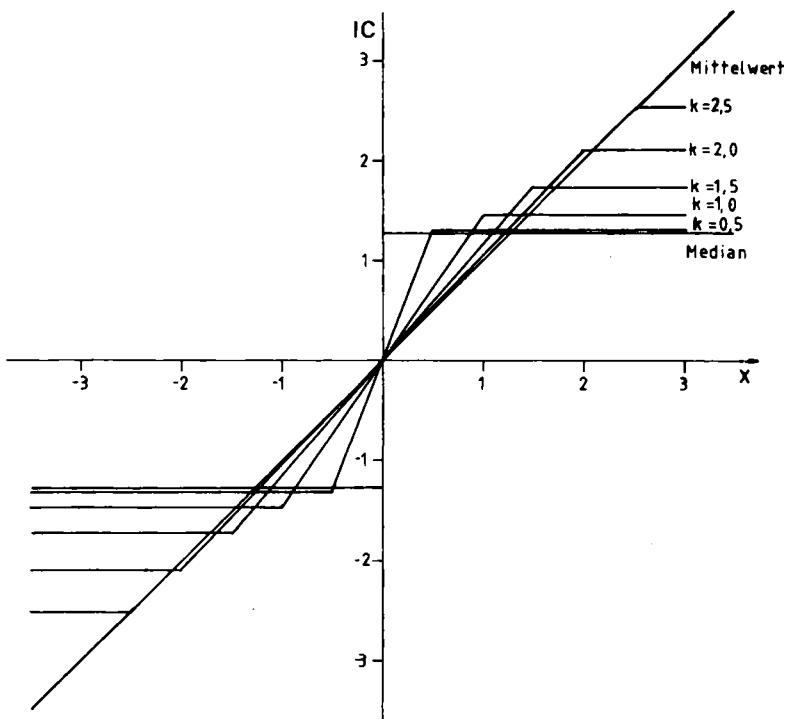


Abb. 18: Einflußkurven der Huber-k-Schätzer für $k = 2,5; 2; 1,5; 1; 0,5$ sowie des arithmetischen Mittels und des Medians

Beispiel: Für unser Beispiel soll der skaleninvariante Huber-3-Schätzer bestimmt werden, wobei wir als Skalenschätzer

$$S_6 = 1,4826 \text{ MAD}$$

verwenden und $c = 1$ setzen wollen.

Geht man von den Angaben in mm aus, so ergeben sich die absoluten Abweichungen vom Median

$$m = \frac{1}{2}(x_{(3)} + x_{(4)}) = \frac{1}{2}(1,1 + 1,9) = 1,5$$

zu $y_1 = |1,5 - (-1,5)| = 3$, $y_2 = 1$, $y_3 = 0,4$, $y_4 = 0,4$, $y_5 = 0,8$ und $y_6 = 10$, d.h. es ist

$$\text{MAD} = \frac{1}{2}(y_5 + y_2) = \frac{1}{2}(0,8 + 1) = 0,9.$$

Damit ist

$$k^* = c \cdot S_6 \cdot k = 1 \cdot (1,4826 \cdot 0,9) \cdot 3 = 4$$

und der skaleninvariante Huber-3-Schätzer stimmt mit dem nicht-skaleninvarianten Huber-4-Schätzer überein, d.h. es ist $T_6 = 1,66 \text{ mm}$.

Geht man nicht von mm-Angaben sondern von cm-Angaben aus, so ergibt sich $\text{MAD} = 0,09$ und der skaleninvariante Huber-3-Schätzer ist identisch mit dem

nicht-skaleninvarianten Huber-0,4-Schätzer ($k^* = 1,4826 \cdot 0,09 \cdot 3 = 0,4$), d. h. es ist $T_6 = 0,166 \text{ cm} = 1,66 \text{ mm}$.

4.3.2. Andrews' wave und Tukey's biweight

Bei der bisher vorgestellten Klasse von M-Schätzern, den Huber-k-Schätzern, ist ein Einfluß beliebig großer Stichprobenwerte immer, wenn auch in beschränktem Maße, vorhanden. Das ist nicht mehr der Fall bei M-Schätzern, deren ψ -Funktion außerhalb eines endlichen Bereiches verschwindet (**redescending M-Estimators**).

Man kann z. B. die ψ -Funktion der Huber-k-Schätzer wieder linear auf Null abfallen lassen und erhält die mit vielen Parametern in ihren Eigenschaften steuerbaren **Hampel-Schätzer**, vgl. z. B. Andrews et al. (1972), Huber (1981), die hier jedoch nicht weiter besprochen werden sollen.

Im Prinzip ähnliche ψ -Funktionen haben die M-Schätzer von Andrews und Tukey. **Andrews' wave** ist festgelegt durch

$$\varrho(x) = \begin{cases} \frac{1}{\pi^2} \cdot (1 - \cos \pi x), & \text{falls } |x| \leq 1 \\ \frac{2}{\pi^2}, & \text{falls } |x| > 1 \end{cases}$$

vgl. Abb. 19, bzw., durch die ψ -Funktion, vgl. Abb. 20,

$$\psi(x) = \begin{cases} \frac{1}{\pi} \cdot \sin \pi x, & \text{falls } |x| \leq 1 \\ 0, & \text{falls } |x| > 1 \end{cases}.$$

Für **Tukey's biweight** hat man

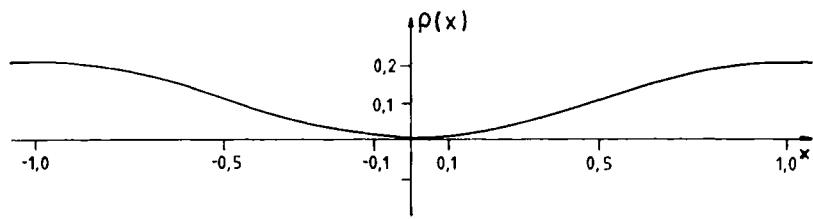
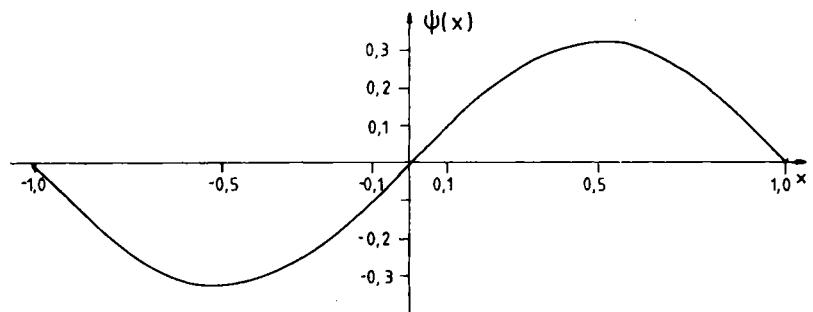
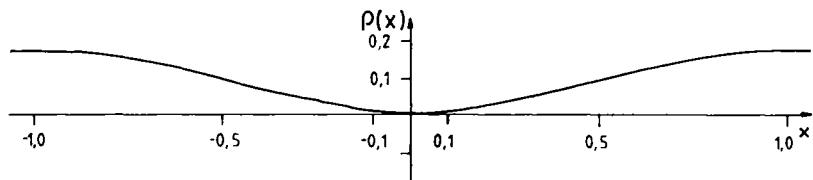
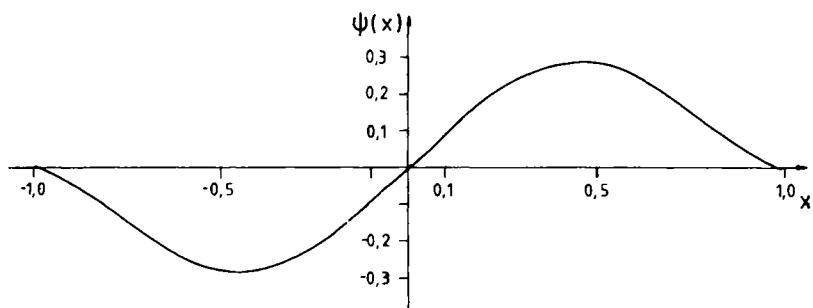
$$\varrho(x) = \begin{cases} \frac{1}{6} \cdot (1 - (1 - x^2)^3), & \text{falls } |x| \leq 1 \\ \frac{1}{6}, & \text{falls } |x| > 1 \end{cases},$$

vgl. Abb. 21, bzw.

$$\psi(x) = \begin{cases} x(1 - x^2)^2, & \text{falls } |x| \leq 1 \\ 0, & \text{falls } |x| > 1 \end{cases},$$

vgl. Abb. 22. Es werden bei beiden Schätzern praktisch nur die **skaleninvarianten Versionen** benutzt. Dabei nimmt man als Skalenschätzung S_n oft den MAD, um einen Bruchpunkt von $\varepsilon^* = 0,5$ zu erreichen. Einfacher berechnen läßt sich jedoch der Quartilsabstand IQR, der allerdings bei beiden Schätzern den Bruchpunkt $\varepsilon^* = 0,25$ zur Folge hat.

Bei der Wahl der Konstanten c muß man sich darüber im Klaren sein, daß Stichprobenwerte, die einen größeren Abstand als $c \cdot S_n$ vom Schätzwert haben, überhaupt nicht berücksichtigt werden. Eine zu kleine Wahl von c kann außerdem sehr

Abb. 19: ρ -Funktion von Andrews' waveAbb. 20: ψ -Funktion von Andrews' waveAbb. 21: ρ -Funktion von Tukey's biweightAbb. 22: ψ -Funktion von Tukey's biweight

viele Lösungen t der Gleichung

$$\sum_{i=1}^n \psi\left(\frac{x_i - t}{cS_n}\right) = 0$$

zur Folge haben, wodurch sich bei der Berechnung eines sinnvollen Schätzwertes Schwierigkeiten ergeben können.

4.3.3. Die Berechnung von M-Schätzern

Sieht man einmal von arithmetischem Mittel und Median ab, so erfordert die Berechnung von M-Schätzern in der Regel den Einsatz iterativer Verfahren. Wir werden hier zwei Methoden vorstellen, die Lösungen $T_n = t$ der Gleichung

$$\sum_{i=1}^n \psi\left(\frac{x_i - t}{cS_n}\right) = 0,$$

also skaleninvariante M-Schätzer liefern. Zur Bestimmung entsprechender nicht-skaleninvarianter Schätzer hat man hier lediglich cS_n durch den Wert 1 zu ersetzen.

Da obige Gleichung unter Umständen mehrere Lösungen besitzt, ist die Wahl eines geeigneten Startwerts T_n^0 zur Bestimmung des gesuchten M-Schätzers T_n für die Lokation vermittels der iterativen Verfahren sehr wichtig. Es hat sich gezeigt, daß in den meisten Fällen der Median m ein geeigneter Startwert ist, d. h. $T_n^0 = m$. In der Regel liegt dann bereits nach einem einzigen Iterationsschritt ein zufriedstellendes Ergebnis vor.

Als ein mögliches Iterationsverfahren wollen wir hier zunächst das **Newton-Raphson-Verfahren** vorstellen. Mit einem Startwert T_n^0 für t bzw. T_n und der Ableitung ψ' von ψ berechnet man hier bei Vorliegen der Stichprobe x_1, \dots, x_n im p -ten Schritt ($p = 1, 2, \dots$)

$$T_n^p = T_n^{p-1} + cS_n \left(\sum_{i=1}^n \psi(v_i^{p-1}) \right) / \left(\sum_{i=1}^n \psi'(v_i^{p-1}) \right),$$

wobei für $i = 1, \dots, n$

$$v_i^{p-1} = (x_i - T_n^{p-1}) / (cS_n).$$

Stabilisieren sich die Schätzwerte, d. h. ist für ein p

$$|T_n^{p-1} - T_n^p| < \varepsilon,$$

wobei $\varepsilon > 0$ eine kleine, frei wählbare Zahl ist, so wird das Verfahren abgebrochen und als M-Schätzer wird

$$T_n = T_n^p$$

verwandt.

Beispiel: Die ψ -Funktion des Huber-k-Schätzers ist bis auf die Stellen $-k$ und k überall differenzierbar. An diesen beiden Stellen setzen wir als Ableitung den Wert 1 fest und erhalten

$$\psi'(x) = \begin{cases} 1, & \text{falls } |x| \leq k \\ 0, & \text{falls } |x| > k \end{cases}$$

Ist keine Skaleninvarianz gefordert, so berechnet man im p -ten Iterationsschritt

$$T_n^p = T_n^{p-1} + \frac{\sum_{i=1}^n \psi(x_i - T_n^{p-1})}{\sum_{i=1}^n \psi'(x_i - T_n^{p-1})};$$

dabei gibt die Summe im Nenner des Bruches gerade die Anzahl der Stichprobenwerte an, die von T_n^{p-1} höchstens den Abstand k haben.

Mit den Daten $x_1 = -1,5$, $x_2 = 0,5$, $x_3 = 1,1$, $x_4 = 1,9$, $x_5 = 2,3$ und $x_6 = 4,1$ und dem Median $T_6^0 = m = 1,5$ als Startwert ergibt sich damit z. B. für den nicht-skaleninvarianten Huber-1-Schätzer ($k = 1$) im 1. Iterationsschritt

$$\begin{aligned} T_6^1 &= 1,5 + \frac{-1 + (0,5 - 1,5) + (1,1 - 1,5) + (1,9 - 1,5) + (2,3 - 1,5) + 1}{4} \\ &= 1,5 + \frac{-0,2}{4} = 1,45. \end{aligned}$$

Im 2. Iterationsschritt erhält man

$$\begin{aligned} T_6^2 &= 1,45 + \frac{-1 + (0,5 - 1,45) + (1,1 - 1,45) + (1,9 - 1,45) + (2,3 - 1,45) + 1}{4} \\ &= 1,45 + 0 = 1,45. \end{aligned}$$

Der gesuchte Huber-1-Schätzer hat also den Wert $T_6 = 1,45$, der sich wegen der stückweisen Linearität der ψ -Funktion schon nach endlich vielen Iterationsschritten exakt ergibt.

Ein M-Schätzer lässt sich auch als gewichtetes Mittel der Stichprobenwerte darstellen, wobei i. a. allerdings die Gewichte vom zu schätzenden Wert abhängig sind. Deshalb führt die Ausnutzung dieser Darstellung zur Berechnung des Schätzwertes T_n in der Regel ebenfalls zu einem iterativen Verfahren.

Mit der **Gewichtsfunktion**

$$w(x) = \psi(x)/x$$

und einem Startwert T_n^0 für T_n berechnet man bei dieser Methode im p -ten Schritt ($p = 1, 2, \dots$)

$$T_n^p = \frac{\sum_{i=1}^n w(v_i^{p-1}) x_i}{\sum_{i=1}^n w(v_i^{p-1})}$$

mit

$$v_i^{p-1} = (x_i - T_n^{p-1})/(cS_n) \quad \text{für } i = 1, \dots, n.$$

Auch dieses Verfahren wird abgebrochen, wenn sich die Ergebnisse stabilisieren, d. h. wenn in einem Schritt p der Abstand zwischen T_n^p und T_n^{p-1} kleiner ist als eine kleine, wählbare Zahl $\epsilon > 0$. Es ist dann

$$T_n = T_n^p$$

der gesuchte M-Schätzer zur Stichprobe x_1, \dots, x_n .

Beispiel: Unabhängig vom Startwert T_n^0 erhält man das arithmetische Mittel $T_n = \bar{x}$ bereits nach einem Iterationsschritt, da mit $\psi(x) = x$ die Gewichtsfunktion $w(x)$ konstant gleich 1 ist und somit gilt

$$T_n^1 = \sum_{i=1}^n 1 \cdot x_i / \sum_{i=1}^n 1 = \sum_{i=1}^n x_i / n = \bar{x}.$$

Für alle übrigen in den Abschnitten 4.3.1 und 4.3.2 vorgestellten M-Schätzer gilt dies nicht, d. h. ihre Gewichtsfunktionen sind nicht konstant. Zur Veranschaulichung sind in Abb. 23 die Gewichtsfunktion $w(x)$ der Huber-1- und Huber-2-Schätzer und in den Abb. 24 und 25 diejenigen von Andrews' wave und Tukey's biweight dargestellt.

Es sei hier noch erwähnt, daß bei der Anwendung des zuletzt dargestellten Iterationsverfahrens auf Andrews' wave Probleme auftreten können. Die Gewichts-

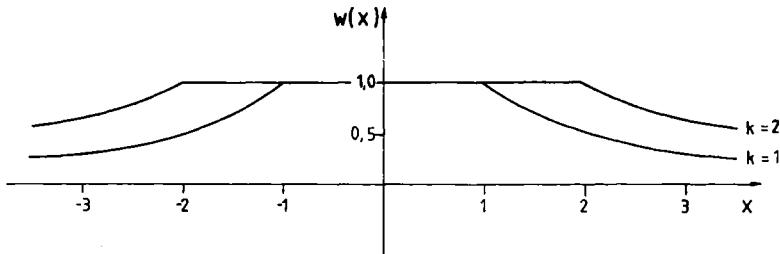


Abb. 23: Gewichtsfunktionen $w(x)$ der Huber- k -Schätzer für $k = 1$ und $k = 2$

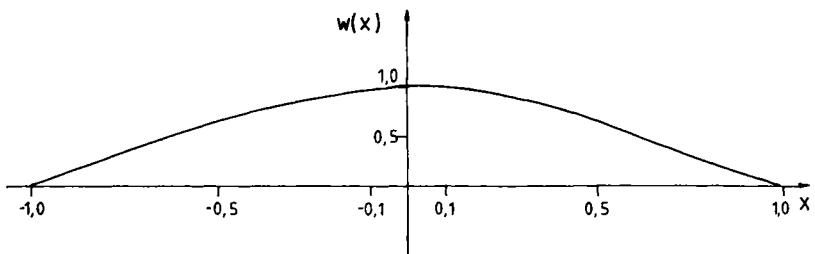


Abb. 24: Gewichtsfunktion $w(x)$ von Andrews' wave

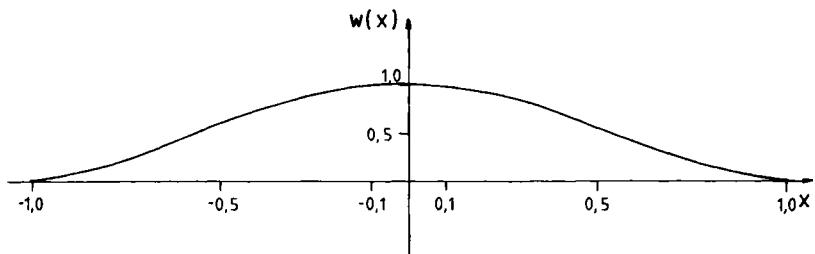


Abb. 25: Gewichtsfunktion $w(x)$ von Tukey's biweight

Tab. 13: Daten x_i , Größen $v_i^{p-1} = (x_i - T_6^{p-1})/(6 \cdot 0,9)$, Werte $w(v_i^{p-1})$ der Gewichtsfunktion für $i = 1, \dots, 6$ sowie Summen $\Sigma = \sum_{i=1}^6 w(v_i^{p-1})$ für $p = 1, \dots, 5$

i	1	2	3	4	5	6	Σ
x_i	-1,5	0,5	1,1	1,9	2,3	7,0	
v_i^0	-0,566	-0,185	-0,074	0,074	0,148	1,019	
$w(v_i^0)$	0,478	0,933	0,989	0,989	0,957	0	4,346
v_i^1	-0,487	-0,117	-0,006	0,142	0,216	1,087	
$w(v_i^1)$	0,581	0,973	1,000	0,960	0,909	0	4,423
v_i^2	-0,472	-0,101	0,010	0,158	0,232	1,102	
$w(v_i^2)$	0,605	0,980	1,000	0,951	0,985	0	4,431
v_i^3	-0,468	-0,097	0,014	1,162	0,236	1,106	
$w(v_i^3)$	0,610	0,981	1,000	0,948	0,892	0	4,431
$v_i^4 = v_i^5$	-0,467	-0,097	0,014	0,163	0,237	1,107	
$w(v_i^4) = w(v_i^5)$	0,611	0,981	1,000	0,948	0,891	0	4,431

funktion ist hier

$$w(x) = \begin{cases} \frac{1}{\pi x} \cdot \sin \pi x, & \text{falls } |x| \leq 1 \\ 0, & \text{falls } |x| > 1 \end{cases}$$

so daß $w(x)$ für $x = 0$ nicht definiert ist.

Beispiel: Mit Hilfe des zuletzt vorgestellten Iterationsverfahrens wollen wir zu den Daten $x_1 = -1,5$, $x_2 = 0,5$, $x_3 = 1,1$, $x_4 = 1,9$, $x_5 = 2,3$ und $x_6 = 7,0$ hier Tukey's biweight bestimmen. Dabei wollen wir als Startwert T_6^0 den Median $m = 1,5$ der Daten, als Skalenschätzer den MAD, d.h. $S_6 = \text{MAD} = 0,9$, und als Konstante $c = 6$ verwenden. Das Verfahren soll abgebrochen werden, wenn in einem Schritt p gilt: $T_6^p = T_6^{p-1}$.

Mit der Gewichtsfunktion

$$w(x) = \psi(x)/x = \begin{cases} (1 - x^2)^2, & \text{falls } |x| \leq 1 \\ 0, & \text{falls } |x| > 1 \end{cases}$$

von Tukeys biweight erhalten wir, vgl. auch **Tab. 13**, im ersten Iterationsschritt

$$\begin{aligned} T_6^1 &= \frac{1}{4,346} (0,478 \cdot (-1,5) + 0,933 \cdot 0,5 + 0,989 \cdot 1,1 \\ &\quad + 0,989 \cdot 1,9 + 0,957 \cdot 2,3 + 0 \cdot 7,0) \\ &= 4,918/4,346 = 1,132. \end{aligned}$$

In der Tab. 13 sind für $i = 1, \dots, 6$ neben den Daten x_i die Größen v_i^{p-1} und $w(v_i^{p-1})$ sowie die Summen $\sum_{i=1}^6 w(v_i^{p-1})$ für $p = 1, \dots, 5$ zusammengestellt, aus denen sich dann weiter

$T_6^2 = 1,047$, $T_6^3 = 1,026$, $T_6^4 = 1,022$, $T_6^5 = 1,021$, $T_6^6 = 1,021$
ergibt, so daß der gesuchte M-Schätzer gerade

$$T_6 = T_6^6 = 1,021$$

ist.

4.3.4. M-Schätzer für einige typische Beispiele

Wir wollen hier die Unterschiede zwischen den vorgestellten M-Schätzern anhand von 8 verschiedenen Datensätzen x_1, \dots, x_6 , vgl. Tab. 14, demonstrieren. In dieser Tabelle sind arithmetisches Mittel \bar{x} , Standardabweichung s, Median m und Median der absoluten Abweichungen vom Median MAD für jeden Datensatz bereits angegeben.

Die ersten 4 Datensätze enthalten jeweils einen „Ausreißer“ (x_6), der sich immer mehr von den übrigen Werten x_1, \dots, x_5 , die stets identisch sind, entfernt. Bei den Datensätzen 5 und 6 sind sowohl x_5 als auch x_6 Ausreißer, die unterschiedlich stark von den übrigen Werten x_1, \dots, x_4 abweichen. Die Werte x_1, x_2, x_3 und x_4, x_5, x_6 bilden in den Datensätzen 7 und 8 zwei Blöcke ähnlicher Daten.

Tab. 14: 8 typische Datensätze x_1, \dots, x_6 , Mittelwerte \bar{x} , Standardabweichungen s, Median m und MAD

Nummer des Datensatzes	i	1	2	3	4	5	6				
	x_i	x_1	x_2	x_3	x_4	x_5	x_6	\bar{x}	s	m	MAD
1		-1,5	0,5	1,1	1,9	2,3	4,1	1,400	1,879	1,500	0,900
2		-1,5	0,5	1,1	1,9	2,3	7,0	1,883	2,840	1,500	0,900
3		-1,5	0,5	1,1	1,9	2,3	11,5	2,633	4,544	1,500	0,900
4		-1,5	0,5	1,1	1,9	2,3	15,3	3,267	6,044	1,500	0,900
5		-1,5	0,5	1,1	1,9	4,7	5,3	2,000	2,588	1,500	2,000
6		-1,5	0,5	1,1	1,9	11,5	15,3	4,800	6,862	1,500	2,000
7		-1,5	0,5	1,1	4,9	6,3	9,6	3,483	4,166	3,000	2,900
8		-1,5	0,5	1,1	11,7	13,2	14,6	6,600	7,303	6,400	6,350

In der Tab. 15 sind nun Huber-k-Schätzer, Andrews' waves und Tukey's biweights zu allen 8 Datensätzen zusammengestellt, die mit Hilfe des zweiten im Abschnitt 4.3.3 vorgestellten Iterationsverfahrens bestimmt wurden. Als Startwert T_6^0 für die M-Schätzer wurde hier stets der Median gewählt und die Ergebnisse T_6^1 der ersten Iterationsschritte sind in Tab. 15 (kursiv, jeweils zweite Zeile) ebenfalls aufgelistet. Bei den nicht-skaleninvarianten Huber-k-Schätzern wurde $k = 1$, $k = 2$, $k = 4$ und $k = 8$ gewählt, bei den invarianten Huber-k-Schätzern $k = 1$, $k = 1,5$ und $k = 2$; als Skalenschätzer wurde hier $S_6 = 1,4826 \cdot \text{MAD}$ und als Konstante $c = 1$ gewählt. Tukey's biweights und Andrews' waves wurden mit dem Skalenschätzer $S_6 = \text{MAD}$ berechnet, wobei jeweils die Konstanten $c = 6$, $c = 9$ und $c = 12$ und bei Andrews' waves zudem $c = 1,8\pi$, $c = 2,1\pi$ sowie $c = 2,4\pi$ verwandt wurden.

Tab. 15: M-Schätzer T_6 für die Lokation von 8 Datensätzen 1, ..., 8 aus Tab. 14; Ergebnisse bei Verwendung des zweiten Iterationsverfahrens aus Abschnitt 4.3.3 mit dem Median m, vgl. Tab. 14, als Startwert und Resultate T_6^* des ersten Iterationsschritts (kursiv)

Nr.	Huber-k-Schätzer (nicht-skaleninvariant)			Huber-k-Schätzer ($S_6 = 1,4826 \text{ MAD}$, $c = 1$)			Tukeys biweight ($S_6 = \text{MAD}$)			Andrews' wave ($S_6 = \text{MAD}$)		
	$k = 1$	$k = 2$	$k = 4$	$k = 8$	$k = 1$	$k = 1,5$	$k = 2$	$c = 6$	$c = 9$	$c = 12$	$c = 9$	$c = 12$
1	1,450	1,450	1,400	1,400	1,450	1,450	1,446	1,478	1,428	1,414	1,458	1,422
	1,458	1,463	1,400	1,400	1,460	1,463	1,454	1,486	1,440	1,423	1,471	1,433
2	1,450	1,450	1,660	1,883	1,450	1,450	1,450	1,021	1,259	1,564	0,983	1,399
	1,456	1,460	1,640	1,883	1,457	1,460	1,463	1,131	1,335	1,550	1,081	1,427
3	1,450	1,450	1,660	2,460	1,450	1,450	1,450	1,021	0,922	0,896	0,983	0,909
	1,455	1,459	1,648	2,328	1,456	1,456	1,459	1,131	0,984	0,974	1,081	0,961
4	1,450	1,450	1,660	2,460	1,450	1,450	1,450	1,021	0,922	0,893	0,983	0,909
	1,455	1,458	1,651	2,360	1,456	1,458	1,461	1,131	0,984	0,930	1,081	0,961
5	1,500	1,833	2,000	2,000	2,050	2,000	2,000	1,974	1,990	1,995	1,980	1,992
	1,500	1,708	2,000	2,000	1,845	2,000	2,000	1,900	1,956	1,975	1,919	1,964
6	1,500	1,833	2,500	4,300	2,155	2,724	3,465	0,609	2,742	3,942	0,869	3,350
	1,500	1,749	2,353	3,731	1,937	2,527	3,065	0,820	2,219	3,288	1,063	2,673
7	3,000	3,000	3,200	3,483	3,200	3,483	3,483	3,361	3,435	3,457	3,388	3,445
	3,000	3,000	3,146	3,483	3,143	3,460	3,483	3,292	3,398	3,435	3,328	3,414
8	6,400	6,400	6,400	6,620	6,600	6,600	6,600	6,598	6,599	6,600	6,598	6,599
	6,400	6,400	6,400	6,567	6,600	6,600	6,600	6,573	6,588	6,593	6,578	6,595

4.4. L-Schätzer für die Lokation

Ein **L-Schätzer** T_n für die Lokation basierend auf einer Stichprobe x_1, \dots, x_n ist eine gewichtete Summe bzw. **Linearkombination der geordneten Stichprobe** $x_{(1)}, \dots, x_{(n)}$ ($x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$):

$$T_n = \sum_{i=1}^n a_i x_{(i)},$$

wobei a_1, \dots, a_n feste Gewichte sind. Im Gegensatz zu den M-Schätzern sind L-Schätzer stets **skaleninvariant**, und ihre Berechnung ist ohne aufwendige Iterationsverfahren möglich. Hingegen läßt sich die Einflußkurve eines L-Schätzers nicht so leicht bestimmen.

Zwei spezielle L-Schätzer, nämlich **Median** und **midquartile**, haben wir schon in der Einleitung zu diesem Kapitel als sogenannte short-cut methods kennengelernt. Der Median ist z. B. definiert als

$$m = \tilde{x}_{0,5} = \begin{cases} x_{((n+1)/2)}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(n/2)} + x_{((n+2)/2)}), & \text{falls } n \text{ gerade} \end{cases},$$

so daß er – aufgefaßt als L-Schätzer – durch die Gewichte

$$a_i = \begin{cases} 1, & \text{für } i = (n+1)/2 \\ 0, & \text{sonst} \end{cases}$$

für ungerades n bzw.

$$a_i = \begin{cases} 1/2, & \text{für } i = n/2, (n+2)/2 \\ 0, & \text{sonst} \end{cases}$$

für gerades n bestimmt ist. Genauso lassen sich die Quartile $\tilde{x}_{0,25}$ und $\tilde{x}_{0,75}$ vermittels der geordneten Stichprobe $x_{(1)}, \dots, x_{(n)}$ darstellen, so daß das midquartile und ebenso auch die im Abschnitt 1 betrachteten midsummaries als L-Schätzer mit speziellen Gewichten a_1, \dots, a_n darstellbar sind.

Wir wollen hier nun noch drei spezielle Klassen von L-Schätzern betrachten, nämlich die α -getrimmten Mittel, die α -winsorisierten Mittel und die α -Gastwirth-Cohen-Mittel.

4.4.1. Das α -getrimmte Mittel

Die **Gewichte** a_1, \dots, a_n des α -getrimmten Mittels, das bei den M-Schätzern ein Analogon in den Huber-k-Schätzern hat, werden wie folgt festgelegt. Zu vorgegebenem α zwischen 0 und 0,5 bestimmt man zunächst $n_1 = [\alpha n]$, die größte ganze Zahl kleiner oder gleich αn und setzt

$$a_i = \begin{cases} 0, & \text{falls } i \leq n_1 \text{ oder } i \geq n - n_1 + 1 \\ \frac{n_1 + 1 - \alpha n}{(1 - 2\alpha)n}, & \text{falls } i = n_1 + 1 \text{ oder } i = n - n_1. \\ \frac{1}{(1 - 2\alpha)n}, & \text{falls } n_1 + 1 < i < n - n_1 \end{cases}$$

Es sei hier noch erwähnt, daß die Gewichte a_{n_1+1} und a_{n-n_1} oftmals auch 0 oder 1 gesetzt werden, was bei großem Stichprobenumfang n praktisch keinen Einfluß auf den Schätzwert T_n hat.

Das α -getrimmte Mittel hat die gleiche **Einflußkurve** wie der nicht-skaleninvariante Huber-k-Schätzer, vgl. Abschnitt 4.3.1, falls k so gewählt ist, daß für die Verteilungsfunktion F der zugrundeliegenden zentralen Zufallsvariablen gilt: $F(-k) = \alpha$. Der **Bruchpunkt** des α -getrimmten Mittels ist gerade $\varepsilon^* = \alpha$ und als Grenzfälle ergeben sich für $\alpha = 0$ das arithmetische Mittel, für $\alpha = 0,5$ der Median als Schätzwerte T_n .

Beispiel: Für die geordnete Stichprobe $x_{(1)} = -1,5, x_{(2)} = 0,5, x_{(3)} = 1,1, x_{(4)} = 1,9, x_{(5)} = 2,3$ und $x_{(6)} = 7,0$ wollen wir das α -getrimmte Mittel für $\alpha = 0,1$ und $\alpha = 0,2$ berechnen.

Für $\alpha = 0,1$ ergibt sich

$$n_1 = [\alpha n] = [0,1 \cdot 6] = [0,6] = 0,$$

so daß wir die Gewichte

$$a_i = \begin{cases} \frac{0,4}{0,8 \cdot 6} = \frac{0,4}{4,8} = \frac{1}{12} & \text{für } i = 1, 6 \\ \frac{1}{0,8 \cdot 6} = \frac{1}{4,8} = \frac{5}{24} & \text{für } i = 2, \dots, 5 \end{cases}$$

erhalten. Damit ergibt sich das 0,1-getrimmte Mittel zu

$$\begin{aligned} T_6 &= \sum_{i=1}^6 a_i x_{(i)} = \frac{1}{12}(-1,5) + \frac{5}{24}(0,5 + 1,1 + 1,9 + 2,3) + \frac{1}{12} \cdot 7 \\ &= \frac{1}{12} \cdot 5,5 + \frac{5}{24} \cdot 5,8 = \frac{1}{24}(11 + 29) = \frac{40}{24} = \frac{5}{3} \\ &= 1,667. \end{aligned}$$

Für das 0,2-getrimmte Mittel erhalten wir mit

$$n_1 = [\alpha n] = [0,2 \cdot 6] = [1,2] = 1,$$

d.h.

$$a_i = \begin{cases} 0 & \text{für } i = 1, 6 \\ \frac{0,8}{0,6 \cdot 6} = \frac{0,8}{3,6} = \frac{2}{9} & \text{für } i = 2, 5, \\ \frac{1}{0,6 \cdot 6} = \frac{1}{3,6} = \frac{5}{18} & \text{für } i = 3, 4 \end{cases}$$

dann gerade

$$\begin{aligned} T_6 &= \frac{2}{9} \cdot 0,5 + \frac{5}{18}(1,1 + 1,9) + \frac{2}{9} \cdot 2,3 = \frac{2}{9} \cdot 2,8 + \frac{5}{18} \cdot 3,0 \\ &= \frac{1}{18}(11,2 + 15) = \frac{26,2}{18} = 1,456. \end{aligned}$$

4.4.2. Das α -winsorisierte Mittel

Beim α -getrimmten Mittel haben die $[n_1] = n_1$ ersten und letzten Werte in der geordneten Stichprobe $x_{(1)}, \dots, x_{(n)}$ keinen Einfluß auf den Schätzwert T_n , da ihre Gewichte a_i gerade 0 gesetzt werden. Beim α -winsorisierten Mittel hingegen werden

mit $[\alpha n] = n_1$ die Beobachtungen $x_{(1)}, \dots, x_{(n_1)}$ gleich $x_{(n_1+1)}$ und die Beobachtungen $x_{(n-n_1+1)}, \dots, x_{(n)}$ gleich $x_{(n-n_1)}$ gesetzt, und dann das arithmetische Mittel der modifizierten Stichprobe berechnet. Dadurch ändert sich der **Bruchpunkt** nicht, d.h. $\varepsilon^* = \alpha$, und die **Gewichte** sind

$$a_i = \begin{cases} \frac{1}{n} \cdot \frac{x_{(n_1+1)}}{x_{(i)}}, & \text{falls } i \leq n_1 \\ \frac{1}{n}, & \text{falls } n_1 + 1 \leq i \leq n - n_1 \\ \frac{1}{n} \cdot \frac{x_{(n-n_1)}}{x_{(i)}}, & \text{falls } i \geq n - n_1 + 1 \end{cases}$$

Beispiel: Zum Datensatz $x_{(1)}, \dots, x_{(6)}$ aus dem Beispiel des Abschnitts 4.4.1 ergibt sich das 0,1-winsorisierte Mittel wegen

$$n_1 = [\alpha n] = [0,6] = 0$$

gerade als arithmetisches Mittel

$$T_6 = \sum_{i=1}^n a_i x_{(i)} = \frac{1}{n} \sum_{i=1}^n x_{(i)} = \bar{x} = \frac{11,3}{6} = 1,883$$

der Daten, und das 0,2-winsorisierte Mittel erhält man wegen

$$n_1 = [\alpha n] = [1,2] = 1$$

unter Verwendung der Gewichte

$$a_i = \begin{cases} \frac{1}{6} \cdot \frac{0,5}{-1,5} & \text{für } i = 1 \\ \frac{1}{6} & \text{für } i = 2, \dots, 5, \\ \frac{1}{6} \cdot \frac{2,3}{7,0} & \text{für } i = 6 \end{cases}$$

d.h. es ist

$$\begin{aligned} T_6 &= \frac{1}{6} \cdot \frac{0,5}{-1,5} \cdot (-1,5) + \frac{1}{6} (0,5 + 1,1 + 1,9 + 2,3) + \frac{1}{6} \cdot \frac{2,3}{7,0} \cdot 7,0 \\ &= \frac{1}{6} (0,5 + 0,5 + 1,1 + 1,9 + 2,3 + 2,3) = \frac{8,6}{6} = 1,433. \end{aligned}$$

4.4.3. Das α -Gastwirth-Cohen-Mittel

Zu einem vorgegebenem Wert α zwischen 0 und 0,5 bestimmt man beim **α -Gastwirth-Cohen-Mittel**, vgl. Gastwirth/Cohen (1970), den L-Schätzer für die Lokation der (geordneten) Stichprobe $x_{(1)}, \dots, x_{(n)}$ als

$$T_n = \gamma(x_{(n_1+1)} + x_{(n-n_1)}) + (1 - 2\gamma)\tilde{x}_{0,5},$$

wobei $n_1 = [\alpha n]$ die größte ganze Zahl kleiner oder gleich αn ist, $\tilde{x}_{0,5}$ den Median der Beobachtungsreihe bezeichnet und $\gamma \in (0, 0,5)$ ein frei wählbarer Parameter ist.

Ist die Anzahl n der Beobachtungen ungerade, so sind die Gewichte dieses L-Schätzers also gerade

$$a_i = \begin{cases} \gamma & \text{für } i = n_1 + 1, n - n_1 \\ 1 - 2\gamma & \text{für } i = (n+1)/2, \\ 0 & \text{sonst} \end{cases}$$

ist n gerade, so ergeben sie sich zu

$$a_i = \begin{cases} \gamma & \text{für } i = n_1 + 1, n - n_1 \\ 1/2 - \gamma & \text{für } i = n/2, (n+2)/2. \\ 0 & \text{sonst} \end{cases}$$

Spezialfälle des α -Gastwirth-Cohen-Mittels sind mit $\alpha = 1/3$, d.h. $n_1 = [n/3]$, und $\gamma = 0,3$ das **Gastwirth-Mittel**

$$T_n = 0,3(x_{(n_1+1)} + x_{(n-n_1)}) + 0,4\tilde{x}_{0,5}$$

und mit $\alpha = 1/4 = 0,25$, d.h. $n_1 = [0,25n]$, und $\gamma = 1/4$ der **trimean**

$$T_n = \frac{1}{4}(x_{(n_1+1)} + x_{(n-n_1)}) + \frac{1}{2}\tilde{x}_{0,5}.$$

Beispiel: Wir wollen hier Gastwirth-Mittel und trimean für die Daten $x_{(1)}, \dots, x_{(6)}$ des Beispiels aus Abschnitt 4.4.1 berechnen. Da $n = 6$ gerade ist, ergibt sich zunächst der Median der Daten zu

$$\tilde{x}_{0,5} = \frac{1}{2}(x_{(3)} + x_{(4)}) = \frac{1}{2}(1,1 + 1,9) = 1,5,$$

so daß wir als Gastwirth-Mittel mit $n_1 = [1/3 \cdot n] = [6/3] = 2$ gerade

$$T_n = 0,3(x_{(3)} + x_{(4)}) + 0,4\tilde{x}_{0,5} = 0,3(1,1 + 1,9) + 0,4 \cdot 1,5 = 1,5,$$

also den Median, erhalten. Für den trimean ergibt sich mit $n_1 = [0,25n] = [0,25 \cdot 6] = [1,5] = 1$ hingegen

$$T_n = 0,25(x_{(2)} + x_{(5)}) + 0,5\tilde{x}_{0,5} = 0,25(0,5 + 2,3) + 0,5 \cdot 1,5 = 1,45.$$

4.5. R-Schätzer für die Lokation

R-Schätzer für die Lokation werden mit Hilfe von **Zweistichproben-Rangtests** für Lokationsvergleiche gewonnen. Betrachten wir einmal den Fall, daß die Stichprobenumfänge in der ersten Stichprobe x_1, \dots, x_n und in der zweiten Stichprobe y_1, \dots, y_n jeweils gleich n sind und daß keine Bindungen (alle $x_1, \dots, x_n, y_1, \dots, y_n$ haben verschiedene Werte) auftreten. Ordnet man nun der kombinierten Stichprobe gemeinsam Rangzahlen $R(x_i)$, $R(y_j)$ für $i = 1, \dots, n$ zu, insgesamt tauchen dann die Rangzahlen $1, 2, \dots, 2n$ auf, so läßt sich die Teststatistik jedes Zweistichproben-Rangtests in der Form

$$S_{n,n} = \frac{1}{n} \sum_{i=1}^n a(R(x_i))$$

schreiben. Dabei sind die Größen $a(j)$ für $j = 1, \dots, 2n$ die sogenannten **Scores** des jeweiligen Tests. Die Scores werden dabei mittels einer **Scorefunktion** $J(t)$ z.B. durch

$$a(j) = 2n \cdot \int_{\frac{j-1}{2n}}^{\frac{j}{2n}} J(t) dt$$

definiert. Ist $J(t)$ derart gewählt, daß gilt $\sum_{j=1}^{2n} a(j) = 0$, so ist unter der Hypothese gleicher Lokation in beiden Stichproben der Wert der Teststatistik $S_{n,n}$ gerade Null.

Beispiel: Wir haben im Kap. VIII, Abschnitt 1.2.1 den Wilcoxon-Rangsummentest kennengelernt und dort die Teststatistik

$$W_{n,n} = \sum_{i=1}^n R(x_i)$$

verwandt, die unter der Hypothese gleicher Lokation den Erwartungswert

$$\frac{1}{2} \cdot n(2n + 1) = n^2 + \frac{n}{2}$$

besitzt. Wollen wir den Wilcoxon-Rangsummentest also, wie oben beschrieben, mittels Scores definieren, so müssen wir anstelle von $W_{n,n}$ eine Teststatistik

$$S_{n,n} = \frac{1}{cn} \left(\sum_{i=1}^n R(x_i) - n^2 - \frac{n}{2} \right) = \frac{1}{c} \left(\frac{1}{n} \sum_{i=1}^n R(x_i) - n - \frac{1}{2} \right)$$

betrachten, wobei c eine Konstante ist.

Mit der Scorefunktion

$$J(t) = t - \frac{1}{2}$$

ergeben sich die Scores

$$\begin{aligned} a(j) &= 2n \int_{\frac{j-1}{2n}}^{\frac{j}{2n}} \left(t - \frac{1}{2} \right) dt = 2n \left[\frac{1}{2} t^2 - \frac{1}{2} t \right]_{\frac{j-1}{2n}}^{\frac{j}{2n}} = n [t^2 - t]_{\frac{j-1}{2n}}^{\frac{j}{2n}} \\ &= n \left(\frac{j^2}{4n^2} - \frac{j}{2n} - \frac{(j-1)^2}{4n^2} + \frac{j-1}{2n} \right) = \frac{j^2 - j^2 + 2j - 1}{4n} - \frac{j - j + 1}{2} \\ &= \frac{2j - 1}{4n} - \frac{1}{2} = \frac{1}{4n} (2j - 1 - 2n), \end{aligned}$$

und es ist

$$\begin{aligned} S_{n,n} &= \frac{1}{n} \sum_{i=1}^n a(R(x_i)) = \frac{1}{n} \sum_{i=1}^n \frac{1}{4n} (2R(x_i) - 1 - 2n) \\ &= \frac{1}{4n^2} \left(2 \sum_{i=1}^n R(x_i) - n - 2n^2 \right) = \frac{1}{2n} \left(\frac{1}{n} \sum_{i=1}^n R(x_i) - \frac{1}{2} - n \right). \end{aligned}$$

Die Scorefunktion $J(t) = t - \frac{1}{2}$ führt also auf den Wilcoxon-Rangsummentest mit der Konstanten $c = 2n$.

Will man nun einen **R-Schätzer** T_n für die Lokation basierend auf einem Zwei-Stichproben-Rangsummentest aus der Stichprobe x_1, \dots, x_n gewinnen, so bestimmt man ihn derart, daß mit $y_1 = 2T_n - x_1, \dots, y_n = 2T_n - x_n$ ungefähr gilt:

$$S_{n,n} = 0.$$

Speziell der **auf dem Wilcoxon-Rangsummentest basierende R-Schätzer** T_n zu einer

Stichprobe x_1, \dots, x_n lässt sich als Median der n^2 Werte $\frac{1}{2}(x_i + x_j)$ ($i, j = 1, \dots, n$) bestimmen. Er hat den Bruchpunkt $\varepsilon^* = 1 - 1/\sqrt{2} = 0,293$.

Beispiel: Wir wollen zu den Daten $x_1 = 0,5, x_2 = 1,1, x_3 = 1,9$ und $x_4 = 7,3$ den auf dem Wilcoxon-Rangsummentest basierenden R-Schätzer T_n bestimmen. Er ergibt sich als Median der in Tab. 16 angegebenen $n^2 = 16$ Werte, d.h. es ist

$$T_n = \frac{1}{2}(1,5 + 1,9) = 1,7.$$

Tab. 16: Werte $\frac{1}{2}(x_i + x_j)$ für $i, j = 1, \dots, 4$ zur Berechnung des auf dem Wilcoxon-Rangsummentests basierenden R-Schätzers

		x_j	0,5	1,1	1,9	7,3
x_i		0,5	0,8	1,2	3,9	
0,5		0,5	0,8	1,2	3,9	
1,1		0,8	1,1	1,5	4,2	
1,9		1,2	1,5	1,9	4,6	
7,3		3,9	4,2	4,6	7,3	

Wir wollen nun noch konkret aufzeigen, daß bei Vorliegen dieser Stichprobe x_1, \dots, x_4 durch Spiegelung an $T_n = 1,7$, d.h. $y_i = 2 \cdot 1,7 - x_i = 3,4 - x_i$ für $i = 1, \dots, 4$, der Wert der Teststatistik $S_{n,n}$ des Wilcoxon-Rangsummentests 0 ist. Dazu betrachten wir die kombinierte Stichprobe $x_1 = 0,5, x_2 = 1,1, x_3 = 1,9, x_4 = 7,3, y_1 = 3,4 - x_1 = 2,0, y_2 = 3,4 - x_2 = 2,3, y_3 = 3,4 - x_3 = 1,5$ und $y_4 = 3,4 - x_4 = -3,9$. Die Rangzahlen ergeben sich hier zu $R(x_1) = 2, R(x_2) = 3, R(x_3) = 5, R(x_4) = 8, R(y_1) = 7, R(y_2) = 6, R(y_3) = 4$ und $R(y_4) = 1$, so daß wir

$$\begin{aligned} S_{4,4} &= \frac{1}{2 \cdot 4} \left(\frac{1}{4} \sum_{i=1}^4 R(x_i) - \frac{1}{2} - 4 \right) = \frac{1}{8} \left(\frac{1}{4} (2 + 3 + 5 + 8) - \frac{1}{2} - 4 \right) \\ &= \frac{1}{8} \left(\frac{18}{4} - \frac{1}{2} - 4 \right) = \frac{1}{8} \cdot \frac{18 - 2 - 16}{4} = \frac{1}{8} \cdot 0 = 0 \end{aligned}$$

erhalten. Natürlich kann $S_{n,n}$ hier auch direkt mittels der Scores

$$a(j) = \frac{1}{4^n} (2j - 1 - 2n) = \frac{1}{16} (2j - 1 - 8) = \frac{1}{16} (2j - 9)$$

berechnet werden:

$$\begin{aligned} S_{4,4} &= \frac{1}{4} \sum_{i=1}^4 a(R(x_i)) = \frac{1}{4} (a(2) + a(3) + a(5) + a(8)) \\ &= \frac{1}{4} \cdot \frac{1}{16} [(4 - 9) + (6 - 9) + (10 - 9) + (16 - 9)] \\ &= \frac{1}{64} (-5 - 3 + 1 + 7) = \frac{0}{64} = 0. \end{aligned}$$

Es sei an dieser Stelle noch eine andere, manchmal verwandte Scorefunktion erwähnt, nämlich

$$J(t) = \Phi^{-1}(t),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist. Die zu dieser Scorefunktion gehörigen Scores sind die sogenannten **Normal-Scores**. Der **Bruchpunkt** des aus den Normal-Scores resultierenden R-Schätzers für die Lokation ist

$$\varepsilon^* = 2\Phi(-\sqrt{\ln 4}) \simeq 0,239$$

und somit etwas niedriger als der des zuvor betrachteten R-Schätzers zum Wilcoxon-Rangsummentest. Mit asymptotisch zu den Normal-Scores äquivalenten, den sogenannten **van der Waerden-Scores** arbeitet als zugehöriger Rangtest der **van der Waerden-Test**, auch X-Test genannt, vgl. van der Waerden (1971).

Zu erwähnen ist noch, daß obige Scorefunktion zum Wilcoxon-Rangsummentest die Umkehrfunktion der Verteilungsfunktion der stetigen Gleichverteilung $U(-\frac{1}{2}, \frac{1}{2})$ auf dem Intervall von $-\frac{1}{2}$ bis $\frac{1}{2}$ ist, vgl. Abschnitt 2 in Kap. IV.

Kapitel XV: Meta-Analyse zur Kombination von Studien, Experimenten und Prognosen

Dieses Kapitel beschäftigt sich mit der Kombination von Wissen aus unterschiedlichen Quellen. Hierzu dienliche Methoden sind von zunehmender Bedeutung. Dies konstatierte zum Beispiel schon 1992 das US National Research Council, wie deren Stellungnahme und ausgesprochenen Empfehlungen in Draper et al. (1992) zu entnehmen ist.

1. Kombination von Wissen – Combining Information

Draper et al. (1992) nennen ausdrücklich die folgenden Bereiche:

Archäologie, Astronomie, Chemie, Energie, Erziehung, Geodäsie, Geologie, Medizin, Militär, Physik, Psychologie, Staatsstatistik, Technik, Umwelt und Wirtschaft,

in denen Methoden zur Kombination von Erkenntnissen oder Vermutungen zum Einsatz gelangen. Der Begriff **Meta-Analyse** wurde von Glass (1976) in den Sozial- und Verhaltenswissenschaften eingeführt und hat sich inzwischen auch in anderen Bereichen für jede Art der Erkenntnisgewinnung aus Daten zweiter Hand, d. h. bereits verarbeiteten Daten, eingebürgert. Neben diesen Bereichen ist obige Liste auf jeden Fall zu ergänzen um die Bereiche Landwirtschaft, Genetik, Biologie, Pharmazie sowie Biometrie, Epidemiologie, Ökonometrie, Technometrie und Informationsverarbeitung.

Bereits der folgende beispielhafte Ausschnitt aus dem Anwendungsspektrum zeigt die praktische Vielfalt und Bedeutung der Meta-Analyse auf.

1.1. Astronomie

In Abschnitt 4 von Kap. V sind in der dortigen Tab. 8 Messungen von 15 Observatorien, vorgenommen in den Jahren zwischen 1895 und 1961, der Astronomischen Einheit (AE) [mittlerer Abstand Sonne - Erde] angegeben. Eingetragen sind ebenfalls die sogenannten Messbereiche, die jedes Observatorium angibt mit dem Hinweis, daß in diesem Bereich der wahre Wert auf jeden Fall zu finden sei: Wir haben diese Angaben in **Abb. 1** veranschaulicht. Da im Beobachtungszeitraum die AE als konstant anzusehen ist, sind die deutlich erkennbaren Abweichungen, einige Meßbereiche sind sogar disjunkt, nur auf Meßfehler zurückzuführen. Wir verzichten auf die Angabe des Ergebnisses einer Meta-Analyse, um der möglichen Kritik auszuweichen, die Meßtechnik habe sich über die Jahre hinweg verbessert und das Zustandekommen der Meßbereiche sei nicht nachvollziehbar, so daß sie folglich auch untereinander nicht vergleichbar seien. Damit haben wir auch gleich einige der Schwierigkeiten kennen gelernt, aus vorliegenden Informationen zur wahren Erkenntnis zu gelangen.

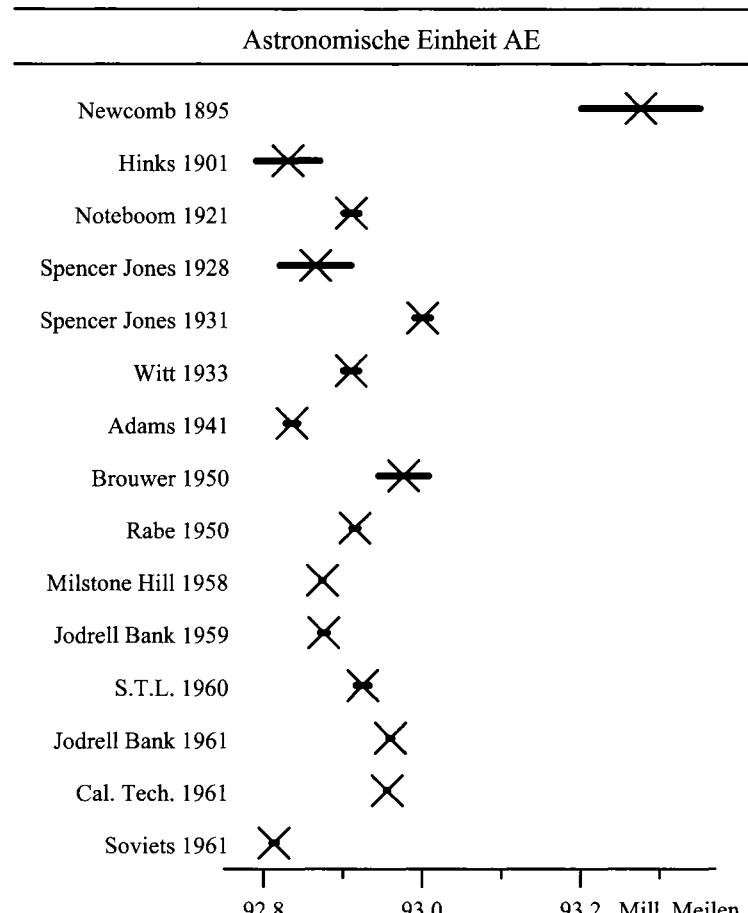


Abb. 1: Messungen der AE und sichere Meßbereiche

1.2. Evidence Based Medicine

Die Evidenz Basierte Medizin [EBM] beruht auf der Annahme, daß die Wirkung eines Medikamentes, falls vorhanden, auch in Klinischen Prüfungen, möglichst in mehreren, nachweisbar ist. Dies heißt natürlich umgekehrt, zeigen die Klinischen Studien keine nachweisbare, d. h. signifikante Wirkung, so kann das Medikament als unwirksam eingestuft werden. Der Nachweis der Wirksamkeit hängt somit direkt von der **richtigen Meta-Analyse** der durchgeföhrten Klinischen Prüfungen ab.

In **Abb. 2** sind die Ergebnisse von 13 Studien, veröffentlicht zwischen 1984 und 1997, zur Wirksamkeit des Medikamentes Cisapride im Vergleich zu Placebo bei der Behandlung der Funktionellen Dyspepsie veranschaulicht. Ein gezeichnet sind zudem die 95%-Konfidenzintervalle des Behandlungseffekts jeder Studie und der Meta-Analyse zur Erlangung einer Gesamtbeurteilung.

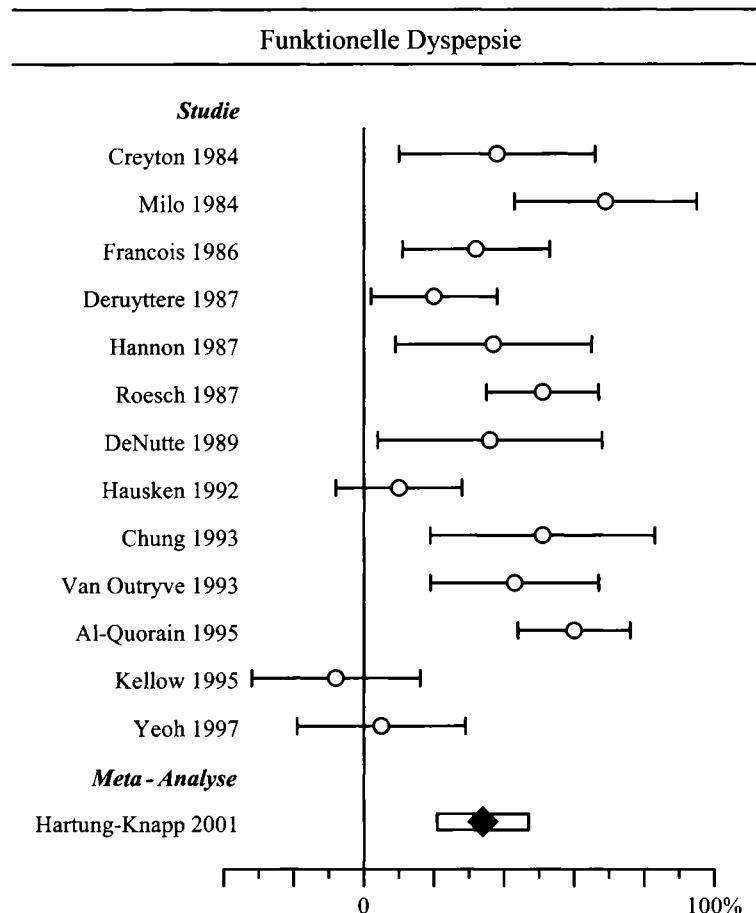


Abb. 2: Cisapride – Steigerung der Rekonvaleszensquote

Trotz eines stärkeren und zweier schwächeren nicht signifikanten Einzelergebnisse herrscht doch ein recht überzeugender Gesamteindruck vor, der sich dann auch in der Meta-Analyse wieder findet und einen deutlichen Gesamteffekt für das Medikament bescheinigt, vgl. Abschnitt 4.4.

Eine Veranschaulichung der Versuchsergebnisse wie in Abb. 2 wird gelegentlich auch als **Forest Plot** bezeichnet. Die damit verbundene Art der **Abbildungsgestaltung** erweist sich für dieses Kapitel XV als ausgesprochen zweckmäßig, so daß wir sie hier auch fortan beibehalten.

Eine Kombination mit den kompletten **Box Plots** gemäß Abschnitt 1.4 in Kapitel XIV für jede Studie führt zu einer verfeinerten Darstellung, die mehr Einzelinformationen zeigt.

In vergleichenden Studien zu Überlebenszeiten dient der Quotient der Hazardraten zum Vergleich der Überlebenszeiten, hier der Kontrollgruppe ge-

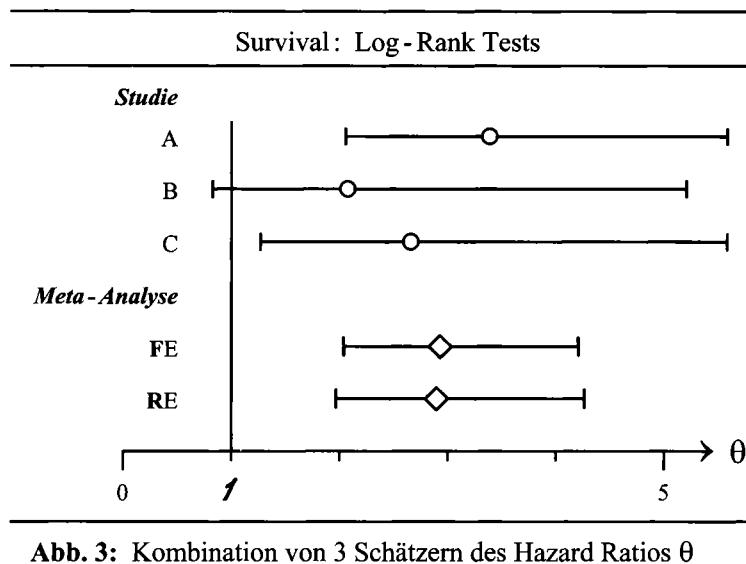


Abb. 3: Kombination von 3 Schätzern des Hazard Ratios θ

genüber der Behandlungsgruppe, so daß letztere eine längere Überlebenszeit aufweist, wenn dieser Quotient, auch Hazard Ratio genannt, einen Wert größer als 1 aufweist. In **Abb. 3** sind drei solcher Studien an Hand ihrer Hazard Ratios veranschaulicht. Die ebenfalls eingezeichneten 95%-Konfidenzintervalle sind unsymmetrisch bzgl. der eingetragenen Punktschätzer, da die zugrunde liegenden Analysen für die logarithmierten Hazard Ratios durchgeführt werden, d. h. über die Teststatistiken zu den Log-Rank Tests, vgl. Abschnitt 5.2.

Meta-Analysen werden im Wesentlichen in den beiden Modellen: **Fixed Effects Model (FE)** und **Random Effects Model (RE)** durchgeführt.

Im **FE-Modell** wird der Gesamteffekt als gewogenes Mittel der Studieneffekte berechnet, wobei die Gewichte reziprok proportional zur Varianz der Studieneffekte sind. Das **RE-Modell** berücksichtigt außerdem in der Varianz der Studieneffekte noch eine zusätzliche Zwischen-Studienvarianz, die auf eine **Wechselwirkung** zwischen Behandlung und Studien zurückzuführen ist, vgl. auch Abb. 20 und Abb. 21 in Abschnitt 5 von Kap. V.

Ist folglich keine Wechselwirkung zwischen Behandlung und Studien vorhanden, so spricht man von homogenen Studien und befindet sich im FE-Modell. Andernfalls hat man inhomogene oder heterogene Studien und legt das RE-Modell zugrunde.

In Abb. 3 sind auch die Ergebnisse der Meta-Analysen in den beiden Modellen veranschaulicht. Dargestellt sind die jeweils kombinierten Hazard Ratios der drei Überlebenszeitstudien mit zugehörigen 95%-Konfidenzintervallen. Beide Intervalle liegen deutlich oberhalb von 1, wobei das RE-Konfidenzintervall erwartungsgemäß etwas breiter ausfällt. Die Behandlung führt also zu signifikant längerem Überleben. Deutliche Wechselwirkungen zwischen Behandlung und Studien sind nicht erkennbar.

1.3. Verhaltensforschung – Kombination unabhängiger p-Werte und Tests

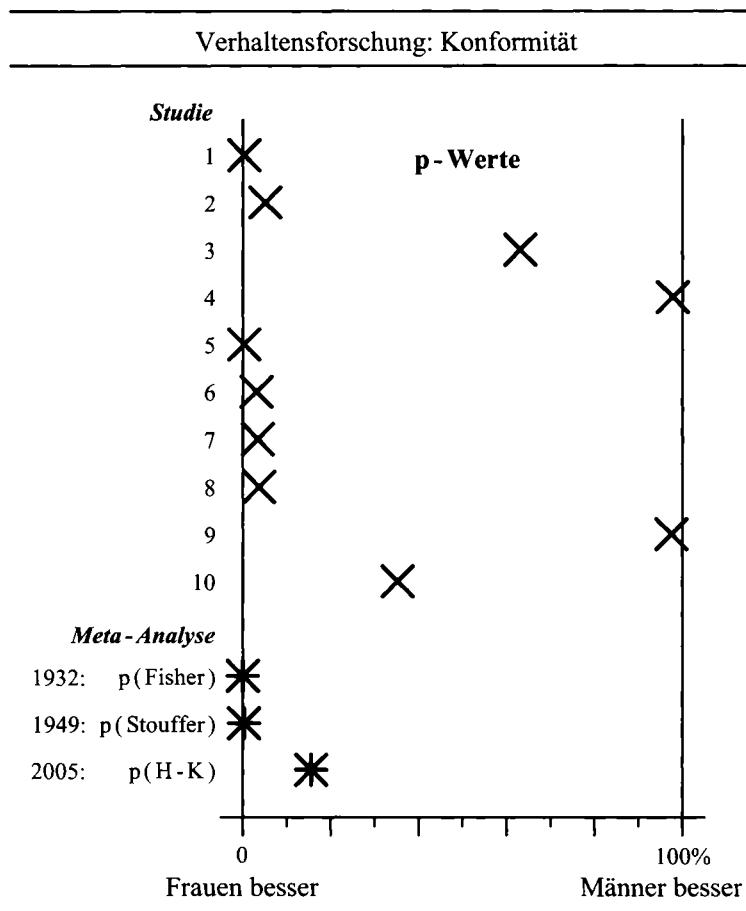


Abb. 4: Nullhypothese H_0 : Konformität (F) \leq Konformität (M)

Geradezu widersprüchliche, konträr signifikante Studienergebnisse sind in Abb. 4 veranschaulicht in der Form von p-Werten zur Nullhypothese H_0 : Grad der Konformität ist bei Frauen geringer als bei Männern. Die meisten der 10 Studien zeigen kleine p-Werte und somit einen signifikant oder nahezu signifikant höheren Grad an Konformität für Frauen als für Männer.

Aber zwei der Studien zeigen auch signifikant das Gegenteil. Die klassischen Kombinationsmethoden von Fisher und Stouffer reagieren darauf aber nicht. Deshalb schlagen Hartung/Knapp (2005a) für solche Fälle von extremer Heterogenität der Studien eine Modifikation vor, die in Abschnitt 9 dargestellt wird. Die drei Kombinationsergebnisse sind ebenfalls in Abb. 4 eingetragen.

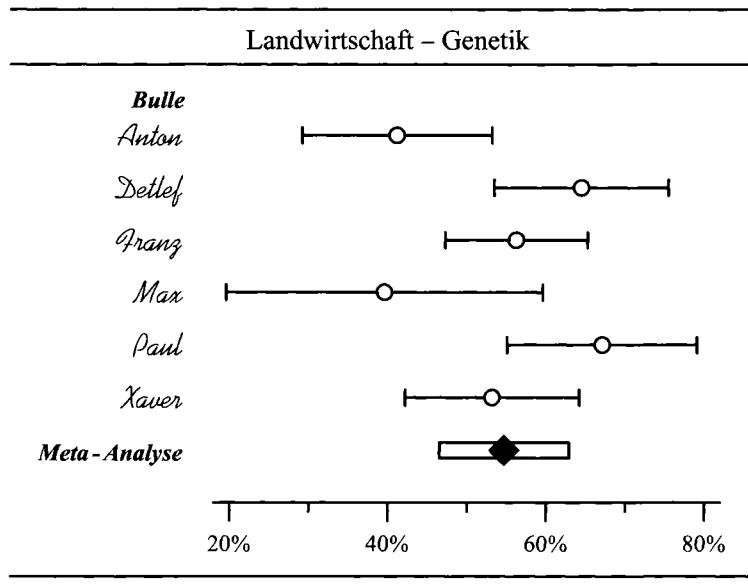


Abb. 5: Besamungssicherheit von Bullen einer Zucht

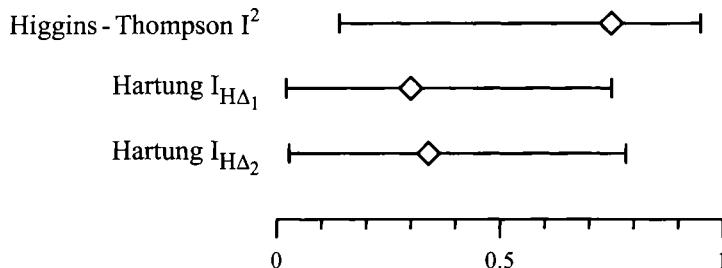
1.4. Landwirtschaft - Genetik - Heterogenitätsmaße

Eine gänzlich andere Fragestellung veranschaulicht Abb. 5. Sechs Bullen eines Zuchttamms sind in der künstlichen Befruchtung von Kühen eingesetzt. Da nicht jede Besamung erfolgreich verläuft, ist es für die Tierzüchter wichtig, die Besamungssicherheit [in %] ihrer Zucht zu kennen, die in der landwirtschaftlichen Praxis negativ als komplementär zu 100% verwandt wird und dann **Umbullquote** heißt. Hier besorgt die Meta-Analyse die korrekte Beurteilung des Zuchttamms in der Vermehrungssicherheit, vgl. Abschnitt 3.4.

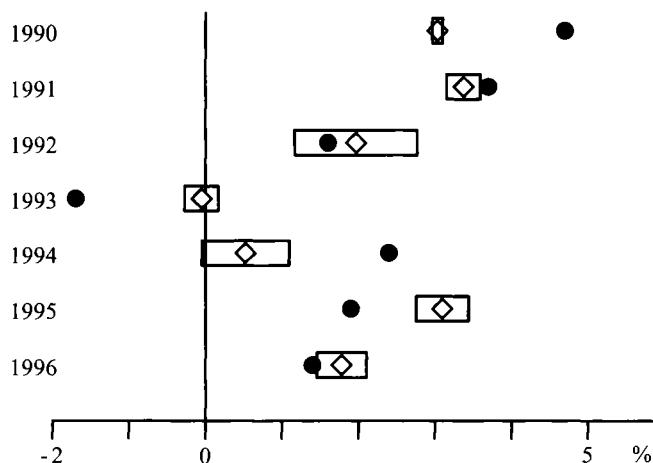
In Anlehnung an den **Intraklass-Korrelationskoeffizienten** der Varianzanalyse in Abschnitt 4.1 von Kap. XI, der Werte zwischen 0 und 1 annimmt, ist man bestrebt, ein entsprechendes Heterogenitäts- oder Inhomogenitätsmaß für die an einer Meta-Analyse beteiligten Studien, Experimente oder Prognosen zu definieren.

Drei solche Maße sind für die Bullen-Daten aus Abb. 5 in Abb. 6 veranschaulicht, nebst zugehörigen 95%-Konfidenzintervallen. Das erste Maß ist im Sinne von Higgins/Thompson (2002) gebildet und die beiden anderen Maße werden hier entwickelt, wie auch die Konfidenzintervalle für alle drei Maße, vgl. Abschnitt 13. Die letzteren Maße lassen sich im vorliegenden Anwendungsbereich auch als **Verallgemeinerte Heritabilitätskoeffizienten** auffassen, vgl. Abschnitt 4.1 in Kap. XI. Dies sind Kennzahlen der Genetik, welche die **Erblichkeit** eines Merkmals, hier die Besamungsfähigkeit der Bullen des Zuchttamms, quantifizieren.

Inhomogenitätsmaße
Genetik: Verallgemeinerter Heritabilitätskoeffizient

**Abb. 6:** Erblichkeit der Besamungsfähigkeit

Vorhersagen des Bruttoinlandsprodukts
BIP ● Kombinierte Prognosen ◊

**Abb. 7:** Gegenüberstellung der wahren BIP - Werte und der kombinierten BIP - Prognosen von sieben Wirtschaftsinstituten aus dem Vorjahr

1.5. Gesamtwirtschaftliche Vorhersagen

Jährlich werden von Wirtschaftsinstituten Vorhersagen zur zukünftigen wirtschaftlichen Entwicklung veröffentlicht. Über Jahre hinweg bleibt die Zusammensetzung der Institute gleich, so daß man rückwirkend gut feststellen kann,

wie gut ein Institut die am Jahresende bekannte Zielgröße vorhergesagt hat. Daraus lassen sich Scoringpunkte für die Institute ermitteln, die dann bei der Kombination der Prognosen als Gewichtung dienen. Daß die kombinierten Vorhersagen auch nicht immer genau den wahren Wert treffen, zeigt **Abb. 7**.

Die BIP-Vorhersagen von sieben Wirtschaftsinstituten wurden jährlich kombiniert, vgl. Hartung/Argac (2009). Die kombinierten Prognosen und zugehörige 95%-Konfidenzintervalle sind in Abb. 7 den wahren Steigerungs-raten des Brutto-Inlands-Produktes gegenübergestellt. Politische Entwick-lungen sind mit der kombinierten Prognose zwar tendenziell gut erfassbar, aber natürlich nicht vollständig, vgl. Abschnitt 7.

Der Vorteil kombinierter Prognosen wird z. B. auch bestätigt von folgen-dem Zitat von **Chatfield** (2003): „Empirical results are very encouraging in that the average of several forecasts typically gives better forecasts than the individual forecasts“.

1.6. Betriebswirtschaftliche Prognosen

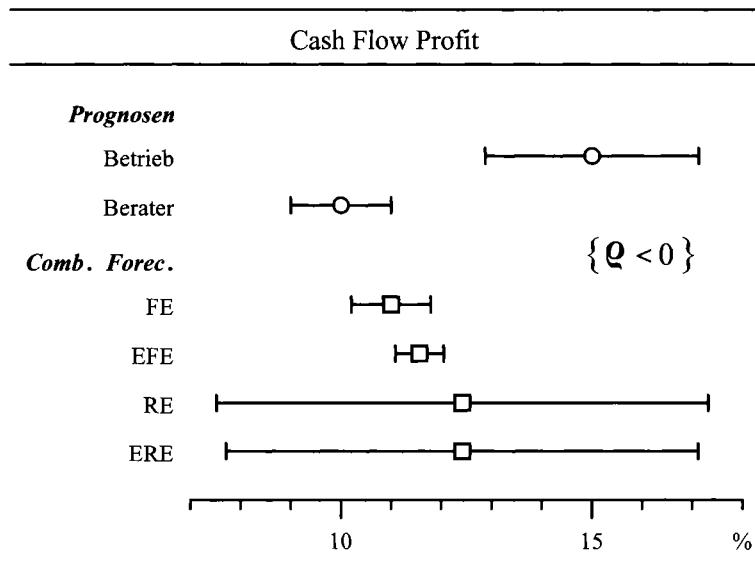


Abb. 8: Combined Forecasts in der Investitionsberatung

Etwa mittels Methoden der Regressionsanalyse aus Kap. X oder der Zeitrei-henanalyse aus Kap. XII lassen sich Prognosen erstellen. Wir betrachten hier, vgl. etwa Pflaumer (2009), Pflaumer/Kohler (2003), den im Rahmen einer Investitionsentscheidung typischen Fall, daß ein Betrieb vor einer geplanten Investition eine Vorhersage der zu erwartenden Steigerung des Cash Flows [**Cash Flow Profit**] berechnet und das Risiko durch eine weitere Vorhersage seitens eines Unternehmensberaters mindern möchte. In **Abb. 8** haben wir einen solchen Fall veranschaulicht, wobei wir zu den beiden Prognosen auch zugehörige 95%-Konfidenzintervalle eingezeichnet haben, die sich allerdings

noch nicht einmal überlappen. Zudem haben wir noch eine stark negative Korrelation zu konstatieren, was in der Praxis häufiger zu beobachten ist.

Im FE-Modell haben wir die reziprok zur Varianz gewichteten Einzelprognosen kombiniert und im erweiterten FE-Modell EFE auch die Korrelation berücksichtigt, was durch die negative Korrelation hier sogar ein schmaleres 95%-Konfidenzintervall für die zudem auch noch höhere kombinierte Prognose zur Folge hat [„Borrowing of Strength“], vgl. Abschnitt 8.

Ein solches Vorgehen zur Kombination von Prognosen oder **Combined Forecasting** im wirtschaftlichen Bereich wird gründlich diskutiert etwa von Clemen/Winkler (1986), Diebold (1988), Makridakis/Weelwright/Hyndman (1998) und dort zitierten Referenzen, s. auch Klapper (2000), Troschke (2002).

Die Idee, Inhomogenitäten zwischen den Prognosen durch einen zusätzlichen Parameter zu modellieren, scheint hier noch nicht verwirklicht worden zu sein. Wir haben dies ohne und mit Berücksichtigung der Korrelation getan. Die entsprechenden Ergebnisse für unser Anwendungsbeispiel im RE-Modell und im erweiterten RE-Modell ERE zur Gewinnung einer kombinierten Prognose sind ebenfalls in Abb. 8 eingezeichnet. Hier im Fall von nur zwei Prognosen bewirkt die Schätzung eines zusätzlichen Parameters eine kräftige Verbreiterung des Konfidenzintervalls.

1.7. Stationarität eines kreuz-korrierten Zeitreihenpanels - Kombination abhängiger p-Werte

Die **Abb. 9** entnehmen wir Abschnitt 1.4 in Kap. X von Hartung/Elpelt (2007) bzw. dem FernKurs Hartung/Elpelt/Schwaiger (1991), wo für ein derartiges Panel von Zeitreihen die multivariate Profilanalyse dargestellt wird, mit der u. a. auf Parallelität der Reihen im zeitlichen Verlauf getestet werden kann.

Hier interessieren wir uns für die Stationarität der zu einem Panel kombinierten Zeitreihen bzw. mittels Differenzenbildungen daraus abgeleiteten Zeitreihen, wobei zugelassen wird, daß diese kreuz-korriert sind, vgl. Abschnitt 11. Wesentlich ist dabei die Kombination nicht unabhängiger p-Werte aus den **Dickey-Fuller-Tests** gemäß Hartung (1999b), vgl. auch Abschnitt 10.

1.8. Umwelt - Physik - Technik

Die Feststellung einer durch Asbest-Fasern verursachten Belastung eines Objekts, hier einer Ostseefähre, ist eine komplizierte Angelegenheit, vgl. Weber/Argac/Hallier/Hartung (2001), Hartung/Knapp (1998) und wird in Abschnitt 4.7 ausführlich geschildert und beispielhaft demonstriert.

Zunächst ist ein hoher physikalisch-technischer Aufwand nötig, um auf einem Luft-Meßfilter neben einer Reihe ähnlich aussehender Fasern die interessierenden Asbestfasern zu identifizieren und deren Anzahl flächenbezogen zu ermitteln. Dies geschieht mit Hilfe eines Elektronenrastermikroskops mit 2500-facher Vergrößerung und unter Einsatz energiedispersiver Röntgenanalytik für jede einzelne verdächtige Faser. Hierfür werden zufällig kleinste Flächenstückchen auf dem Filter ausgewählt, deren gesamte Flächeninhalte z. B. 1 mm² ergeben. In **Abb. 10** ist ein solcher Vorgang schematisch

 Kombination von Zeitreihen

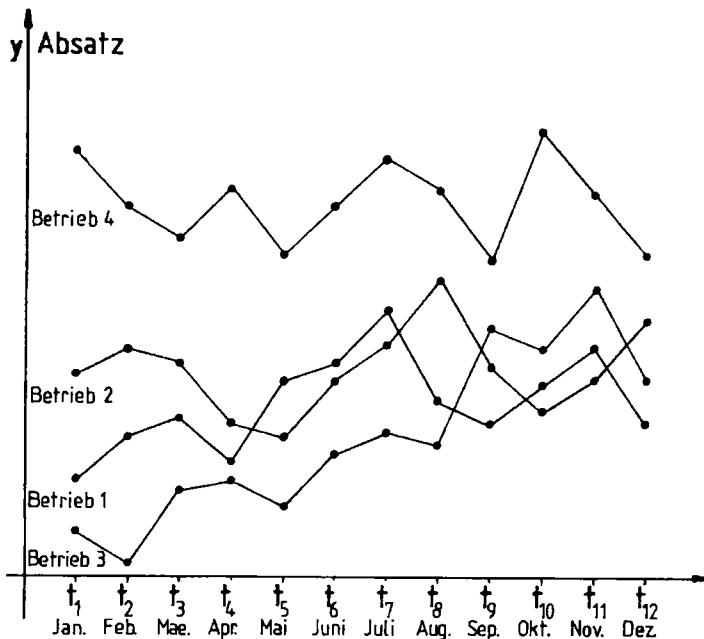


Abb. 9: [In -] Stationarität eines Panels

veranschaulicht.

Den erhobenen Zählergebnissen läßt sich eine Poisson-Verteilung mit Parameter λ [$Po(\lambda)$] zuordnen. Für diesen Parameter ist ein erlaubter Grenzwert λ_G , z. B. als: $\lambda_G = 5$, festgelegt. Ab diesem Grenzwert gilt das betroffene Objekt als belastet und muß zunächst gesperrt werden, um gegebenenfalls saniert werden zu können.

Allerdings ist die Feststellung einer [signifikanten] Überschreitung des Grenzwertes wiederum eine komplexe Angelegenheit und nicht allein durch extreme Unterschiede zwischen den beteiligten Meß-Laboren kompliziert.

In Abb. 11 sind für ein reales Beispiel die Meßergebnisse $\widehat{\lambda}_i$ dreier konkurrierender Labore, vom selben Schiff, zusammen mit jeweiligem 95%-Konfidenzintervall veranschaulicht, vgl. Abschnitt 4.7. Ebenfalls angegeben sind die Ergebnisse der Meta-Analysen im Fixed Effects (FE) und im Random Effects (RE) Model durch Angabe des jeweils kombinierten Gesamtschätzers für λ nebst zugehörigem 95%-Konfidenzintervall. Die beiden Gesamtergebnisse unterscheiden sich hier besonders deutlich in den Konfidenzintervallen, was bedingt wird durch eine merkliche **Wechselwirkung** zwischen Zählergebnis und Laboren.

Erst die **Meta-Analyse** erreicht also durch Kombination konkurrierender Labore eine **gerechte Beurteilung** der Asbest-Belastung auf der in

 Physik – Technik

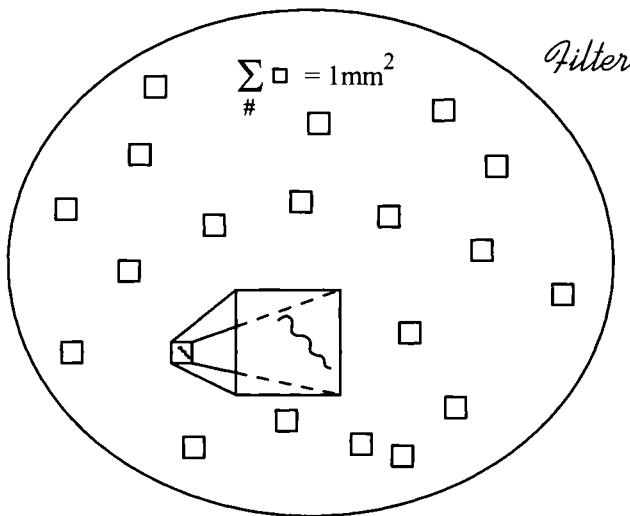


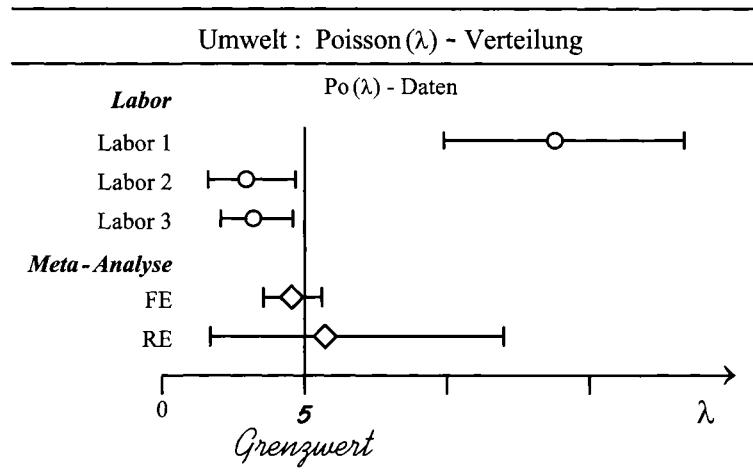
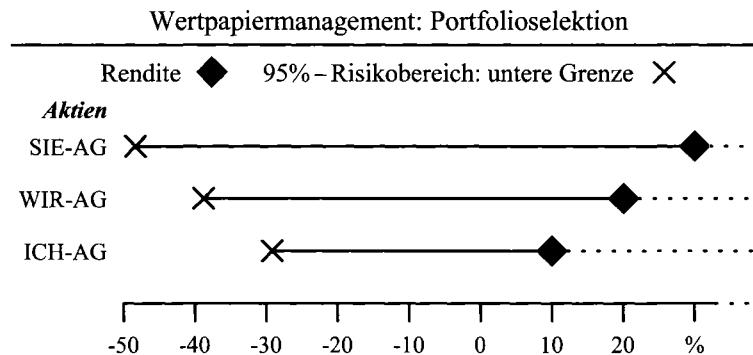
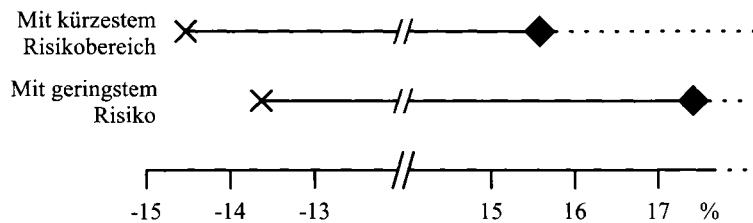
Abb. 10: Poissonverteilung $Po(\lambda)$: Asbest - Faser Zählung mit dem Elektronenrastermikroskop

Verdacht geratenen Ostseefähre.

1.9. Pareto-Optimale Portfolioselektion im Wertpapiermanagement

Während die Meta-Analyse zur Kombination von Studien, Experimenten, Erhebungen und Prognosen sich in den bisherigen Anwendungsbeispielen zumeist eher retrospektiv orientiert, ist sie in der Portfolioselektion tendenziell prospektiv ausgerichtet. Das Konzept der Portfolioselektion behandelt die optimale Zusammensetzung eines Wertpapier-Portfolios nach den Kriterien mittlere **Rendite** und deren Schwankung bzw. Standardabweichung, die spezifisch berechnet wird und dann **Volatilität** genannt wird. Dabei möchte man eine möglichst hohe Rendite und zugleich eine möglichst geringe Volatilität erreichen. Da die beiden Zielgrößen i. d. R. gegenläufig sind, entsteht ein typisches Problem der **Vektor- oder Pareto-Optimierung**, so daß eine zusätzliche Präferenzbildung erforderlich wird, z. B. Hartung (1972a, 1974a). In Abschnitt 16 gehen wir näher darauf ein.

Zwei sinnvolle derartige **Präferenzen**, die beide zu Pareto-optimalen Anlagen führen, werden in Abb. 12 an Hand eines Portfolios bestehend aus Anteilen von drei verschiedenen Aktienpaketen illustrativ vorgeführt. Dargestellt sind die historisch berechneten oder prognostisch erwarteten mittleren Renditen und zugehörige 95%-Konfidenzintervalle, bzw. deren untere Grenzen, die hier als **Risikobereiche** interpretiert werden können. Die beiden Anlagestrategien: „**Stabilste Anlage**“ bzw. „**Schmalster Risikobe-**

**Abb. 11:** Asbest - Belastung einer Ostseefähre**Meta - Analyse:** Portfolioselektion aus SIE-, WIR- und ICH-AG**Abb. 12:** Pareto – Optimale Portfolioselektionen nach den Präferenzen: Stabilste Anlagestrategie und Anlage mit größter unterer Grenze des 95% – Risikobereiches

reich“ und „**Minimales Risiko**“ bzw. „Größte untere Konfidenzgrenze“ sind Pareto-optimal in dem Sinne, daß es jeweils keine gleichmäßig bessere Aktienzusammenstellung mit sowohl höherer Rendite als auch zugleich geringerer Volatilität gibt.

Die Konzeption der Portfolioauswahl, z. B. Ryan (1978) als eine frühe Referenz, hat die finanztheoretische Literatur nachhaltig beeinflußt und wird auch zu allgemeinen Diversifikationsentscheidungen herangezogen, z. B. Lee (2008). Eine der Gründe für die **Diversifikation** ist die **Risikostreuung** und die damit verbundene langfristige **Stabilisierung der Erträge**.

Begründer der Portfolioauswahl ist **Harry Markowitz**, der für seine „pioneering work in portfolio theory“ im Jahr 1990 den **Nobel-Preis** für Wirtschaft erhielt. Eine Sammlung seiner Beiträge wird von ihm unter Markowitz (2009) veröffentlicht.

1.10. Normalverteilungs-Annahmen

In der Einführung des Fixed Effects Model (FE) in folgendem Abschnitt 2 und des Random Effects Model (RE) in Abschnitt 3 wird für die das Gesamtergebnis etwa in der i-ten Studie beschreibende Zufallsvariable y_i eine Normalverteilung mit Erwartungswert ψ und Varianz σ_i^2 , d. h. $y_i \sim \mathcal{N}(\psi, \sigma_i^2)$, angenommen. In den Anwendungen ist y_i zumeist eine Funktion weiterer, die Einzeldaten der Studienteilnehmer zusammenfassender Statistiken. Ist y_i etwa die Differenz von Mittelwerten, so sind diese Mittelwerte und damit y_i auf Grund des **Zentralen Grenzwertsatzes**, vgl. Abschnitt 9 in Kap. II, praktisch auch dann als normalverteilt einzustufen, wenn die Einzeldaten nicht als normalverteilt angesehen werden können.

Schon im 16. Jahrhundert rechnete der Astronom Tycho Brahe mit den Mittelwerten von jeweils 2 Beobachtungen, vgl. Abschnitt 5 der „Einführung und Grundlagen“ zu diesem Buch. Im Feldversuchswesen etwa verrechneten die „alten Biometer“ die originalen Versuchsdaten nur über eine vorgesetzte Mittelbildung von 3 bis 5 ursprünglichen Werten. Die Idee war hierbei stets, „gut normalverteilt“ Ausgangsdaten für weitere Verrechnungen zu sichern.

Des Weiteren wird in den FE- und RE-Modellen, wenn nicht anders zugelassen, die Existenz einer möglichst annähernd unverzerrten Schätzung $\hat{\sigma}_i^2$ für σ_i^2 angenommen. Auch hier ist $\hat{\sigma}_i^2$ zumeist eine Funktion der Varianzschätzungen in den Studiengruppen. Die übliche Varianzschätzung s^2 für die Varianz in einer Gruppe bleibt unverzerrt, wenn die Verteilung nicht die Normalverteilung ist.

Die **Meta-Analyse** ist folglich **robust** in Hinsicht auf Abweichungen von der Normalverteilung in den zugrunde liegenden Basisdaten.

Weitere hierzu nützliche Überlegungen werden in den Abschnitten 4.5 und 5 von Kap. V sowie in Hotelling (1960) dargelegt.

2. Fixed Effects Model (FE) der Meta-Analyse

Die Bezeichnung Fixed Effects Model (FE) hat sich eingebürgert für Situationen, in denen wir in verschiedenen unabhängigen Studien bzw. Erhebungen,

Experimenten oder Prognosen den selben Parameter oder Effekt, bezeichnet mit ψ , geschätzt oder prognostiziert haben, aber aufgrund der unterschiedlichen Umgebungssituationen keine gleichen Varianzen oder Zuverlässigkeit vorliegen haben. Ansonsten liegt keine Variabilität zwischen den Studien, Experimenten oder Prognosen vor, was gegebenenfalls mit einem Test auf Homogenität der Studien etc. überprüft werden kann. Eine synonyme Bezeichnung für dieses homogene Fixed Effects Model ist **Common Mean Problem (CM)**.

2.1. Darstellung des Modells und Gauß-Markov Schätzer

In $K \geq 2$ Studien, Experimenten oder Prognosen liegen zumindest genähert normalverteilte Beobachtungen y_i mit gemeinsamem Erwartungswert $E y_i = \psi$ und den möglicherweise unterschiedlichen Varianzen $\text{Var } y_i = \sigma_i^2$ vor. Wir haben also das [homogene] Fixed Effects Model (FE):

$$y_i = \psi + e_i, \quad e_i \sim \mathcal{N}(0, \sigma_i^2), \quad i = 1, \dots, K,$$

wobei die „Fehler“ e_1, \dots, e_K als paarweise unabhängig angenommen werden.

Der unverzerrte Schätzer minimaler Varianz für ψ , der Gauß-Markov Schätzer $\hat{\psi}_{GM}$, ist bei bekannten Varianzen gegeben als:

$$\hat{\psi}_{GM} = \sum_{i=1}^K \delta_i y_i, \quad \text{wobei } \delta_i = \left(1/\sigma_i^2\right) / \sum_{j=1}^K 1/\sigma_j^2,$$

und die Varianz des Gauß-Markov Schätzers ist der K-te Teil des **harmonischen Mittels** von $\sigma_1^2, \dots, \sigma_K^2$, vgl. Abschnitt 4.4 in Kap. I, d. h.

$$\text{Var } \hat{\psi}_{GM} = \frac{1}{K} \bar{\sigma}_h^2 \quad \text{mit } \bar{\sigma}_h^2 = \frac{K}{1/\sigma_1^2 + \dots + 1/\sigma_K^2}.$$

[Der interessierte Leser findet einige formale Ausführungen im letzten Abschnitt 17.5.]

2.2. Kombinierte Schätzung des Gesamteffektes

Nun sind in der Regel die Varianzen nicht bekannt, sondern nur ihre Schätzungen $\hat{\sigma}_i^2$, die als unabhängig von den y_i angenommen werden. Mit diesen Schätzungen an Stelle der wahren Varianzen wird entsprechend zum Gauß-Markov Schätzer dann ψ ebenfalls unverzerrt [vgl. Abschnitt 17.5] im FE-Modell geschätzt durch:

$$\hat{\psi}_{FE} = \sum_{i=1}^K d_i y_i, \quad d_i = \frac{v_i}{v_\Sigma}, \quad v_i = \frac{1}{\hat{\sigma}_i^2}, \quad v_\Sigma = \sum_{j=1}^K v_j, \quad \hat{\sigma}_i^2 \text{ gegeben.}$$

Seine Varianz wird entsprechend, nach Cochran (1937, 1954), Yates/Cochran (1938), geschätzt mit

$$\widehat{\text{Var}}_C \hat{\psi}_{FE} = 1/v_\Sigma.$$

Da diese Schätzung die wahre Varianz von $\hat{\psi}_{FE}$ im Mittel unterschätzt [vgl. Meier (1953), Li/Shi/Roth (1994), Böckenhoff/Hartung (1998) oder Abschnitt 17.5] schlägt Hartung (1999a), ausgehend von Cochrans Quadratform Q_C , vgl. Cochran (1937), den folgenden Schätzer für die Varianz von $\hat{\psi}_{FE}$ vor:

$$\widehat{\text{Var}}_H \hat{\psi}_{FE} = \frac{1}{(K-1)v_{\Sigma}} Q_C, \text{ mit } Q_C = \sum_{i=1}^K v_i (y_i - \hat{\psi}_{FE})^2.$$

2.3. Konfidenzintervalle für den Gesamteffekt

Mit den $(1 - \alpha/2)$ -Quantilen $u_{1-\alpha/2}$ der Standardnormalverteilung $\mathcal{N}(0, 1)$, $0 < \alpha < 1$, ergibt sich das klassische $(1 - \alpha)$ -Konfidenzintervall für den gemeinsamen Effekt ψ zu:

$$KI_{FE}(\psi)_C = \hat{\psi}_{FE} \pm u_{1-\alpha/2} \sqrt{1/v_{\Sigma}}.$$

Da die Quadratform Q_C genähert einer χ^2 -Verteilung mit $(K-1)$ Freiheitsgraden entstammt, ist $\hat{\psi}_{FE}/(\widehat{\text{Var}}_H \hat{\psi}_{FE})^{1/2}$ genähert t-verteilt mit $(K-1)$ Freiheitsgraden, so daß mit dem $(1 - \alpha/2)$ -Quantil $t_{K-1;1-\alpha/2}$ der t_{K-1} -Verteilung und der zweiten Varianzschätzung für $\hat{\psi}_{FE}$ ein $(1 - \alpha)$ -Konfidenzintervall für ψ gegeben ist als:

$$KI_{FE}(\psi)_H = \hat{\psi}_{FE} \pm t_{K-1;1-\alpha/2} \sqrt{\frac{Q_C}{(K-1)v_{\Sigma}}}.$$

Simulationsstudien zeigen, daß dieses Konfidenzintervall das nominelle Konfidenzniveau praktisch einhält, vgl. z. B. Hartung (1999a), Hartung/Knapp (2000a; 2001a, b, c). Da das Intervall aufwendiger zu berechnen ist, werden wir es nicht in jedem Beispiel mitberechnen, s. auch den nächsten Abschnitt 3.

2.4. Homogenitätstest von Cochran

Geht man zunächst von einem **inhomogenen** Modell aus, in dem jede Studie oder Prognose einen eigenen Erwartungswert ψ_i besitzt, so prüft der Homogenitätstest von Cochran (1937) die Nullhypothese $H_{0,hom} : \psi_1 = \dots = \psi_K$ gegen die Alternative, daß wenigstens zwei der ψ_i verschieden sind. Zum Niveau α :

$$\text{verwerfe } H_{0,hom} \text{ falls } Q_C > \chi_{K-1;1-\alpha}^2,$$

mit Cochrans Quadratform Q_C von oben und dem $(1 - \alpha)$ -Quantil $\chi_{K-1;1-\alpha}^2$ der χ^2 -Verteilung mit $(K-1)$ Freiheitsgraden.

Wird die Homogenitätshypothese verworfen, so sollte die Meta-Analyse auf jeden Fall im Random Effects Model vorgenommen werden, das im nächsten Abschnitt 3 vorgestellt wird.

Tab. 1: Progression der Behinderung von Patienten mit Multipler Sklerose (MS)

Studie i	Azathioprine (E)			Placebo (P)		
	n_{E_i}	\bar{x}_{E_i}	$s_{E_i}^2$	n_{C_i}	\bar{x}_{C_i}	$s_{C_i}^2$
1	162	3.0	159	175	4.2	165
2	15	1.7	81	20	8.3	96
3	30	2.0	121	32	4.5	125
4	27	1.7	190	25	4.2	185

2.5. Beispiel: Klinische Prüfungen zu Multipler Sklerose MS [Zwei-armige Studien]

Wir betrachten $K = 4$ Klinische Studien, die alle mit der selben Disability Scale durchgeführt wurden, zum Fortschreiten der Behinderung von Patienten mit Multipler Sklerose (MS), vgl. Yudkin et al. (1991). Mit dem Medikament Azathioprine behandelte experimentelle Gruppen E werden jeweils mit Placebo-Kontrollgruppen C verglichen. In **Tab. 1** sind die zugehörigen Fallzahlen n_{E_i}, n_{C_i} , die beobachteten Mittelwerte $\bar{x}_{E_i}, \bar{x}_{C_i}$, sowie die Stichprobenvarianzen $s_{E_i}^2$ und $s_{C_i}^2$ aufgeführt.

Den Gruppen liegen Erwartungswerte μ_E und μ_C zu Grunde, die in der i -ten Studie jeweils geschätzt werden mittels \bar{x}_{E_i} und \bar{x}_{C_i} . Interessiert sind wir am Behandlungseffekt $\mu_C - \mu_E$. Wir setzen:

$$\psi = \mu_C - \mu_E, \quad y_i = \bar{x}_{C_i} - \bar{x}_{E_i} \text{ und } \hat{\sigma}_i^2 = \frac{s_{C_i}^2}{n_{C_i}} + \frac{s_{E_i}^2}{n_{E_i}},$$

und wenden obige Methoden an. Die Berechnung der benötigten Größen lässt sich der **Tab. 2** entnehmen. Wir erhalten:

$$\widehat{\psi}_{FE} = 2.16333, \quad v_\Sigma = 0.813 \text{ und } Q_C = 2.43427,$$

so daß

$$\widehat{\text{Var}}_C \widehat{\psi}_{FE} = 1.23 \text{ und } \widehat{\text{Var}}_H \widehat{\psi}_{FE} = 0.99806.$$

Mit den Quantilen $u_{0.975} = 1.96$ und $t_{3;0.975} = 3.182$ ergeben sich die beiden 95%-Konfidenzintervalle für $\psi = \mu_C - \mu_E$ zu:

$$KI_{FE}(\psi)_C = [-0.010, 4.337] \text{ und } KI_{FE}(\psi)_H = [-1.016, 5.342].$$

In **Abb. 13** sind die Studienergebnisse y_i mit zugehörigen 95%-Konfidenzintervallen graphisch veranschaulicht. Eingetragen sind ebenfalls $\widehat{\psi}_{FE}$ und das 95%-Konfidenzintervall $KI_{FE}(\psi)_C$, sowie bereits die entsprechenden Meta-Analyse Ergebnisse im RE-Modell aus Abschnitt 3.3. Im Homogenitätstest haben wir $Q_C \leq \chi^2_{3;1-\alpha}$ für α mit Werten von 0 bis 0.48, so daß also die Homogenität bis zu diesem Niveau nicht signifikant widerlegt ist. Dennoch

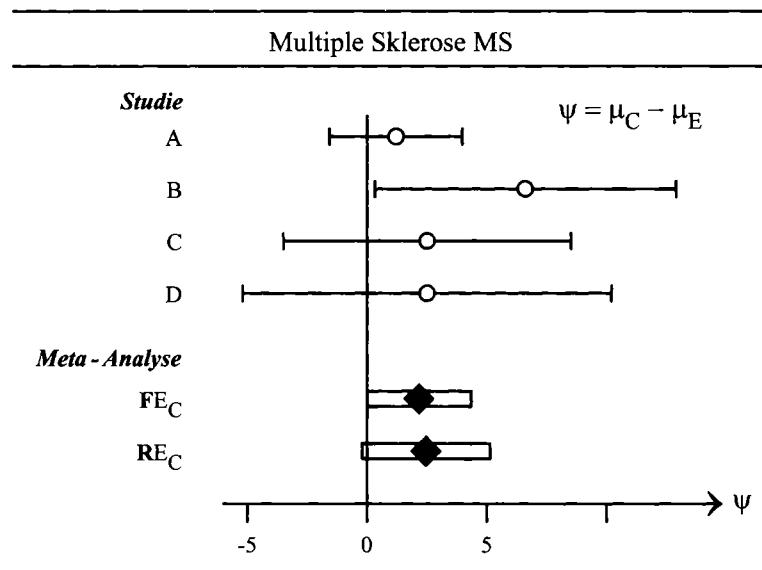


Abb. 13: Azathioprine – Verzögerung der Behinderung

Tab. 2: Arbeitstabelle (FE) zu MS-Studien

i	y_i	$\hat{\sigma}_i^2$	$v_i = 1/\hat{\sigma}_i^2$	$d_i = v_i/v_{\Sigma}$	$d_i y_i$	$v_i(y_i - \sum d_j y_j)^2$
1	1.2	1.924	0.51975	0.63930	0.76716	0.48233
2	6.6	10.200	0.09804	0.12059	0.79589	1.92982
3	2.5	7.940	0.12594	0.15491	0.38728	0.01427
4	2.5	14.437	0.06927	0.08520	0.21300	0.00785
$\Sigma :$	12.8	34.501	0.81300	1	2.16333	2.43427

mag eine Auswertung im Random Effects Model sinnvoll sein, s. Abschnitt 3.3.

Da die Beispiele in diesem Kapitel XV zumeist eher den Charakter von **Fallstudien** aufweisen, wählen wir im Folgenden als **Hervorhebung der Beispiele** i. d. R. die Zuweisung eigener Unterabschnitte, wie hier im Beispiel der MS-Studien, so daß wir auf die bisherige zusätzliche Kennung durch Seitenstriche verzichten können.

2.6. Korrelationen

Im Jahre 1904 veröffentlichte **Karl Pearson** eine Abhandlung über den Zusammenhang zwischen Sterblichkeit und einer Impfung gegen Typhus. Weil die einzelnen diesbezüglichen Untersuchungen ihm in den jeweiligen Fallzahlen als zu klein erschienen, hat er die Korrelationen von fünf Studien zusammengefaßt, um eine sichere Basis für seine Schlußfolgerungen zu erreichen.

Diese Begebenheit zählt zu den Anfängen der Meta-Analyse.

2.6.1. Kombination von Korrelationskoeffizienten

In Abschnitt 1 von Kapitel IX wird gründlich der Pearsonsche Korrelationskoeffizient r_{XY} zweier normalverteilter Merkmale X und Y behandelt. In der dortigen Tab. 2 sind die Fishersche Z-Transformation und ihre Umkehrung:

$$z = \operatorname{arctanh} r_{XY} \text{ und } r_{XY} = \operatorname{arctanh}^{-1} z,$$

ausführlich vertafelt. Bezeichnet ρ die zugrunde liegende, theoretische Korrelation von X und Y, so gilt, wenn der Stichprobenumfang n bei der Rechnung von r_{XY} nicht allzu klein ist, s. Kap. IX für Genaues, in guter Näherung:

$$E z = \operatorname{arctanh} \rho = \zeta \text{ und } \operatorname{Var} z = \frac{1}{n - 3}.$$

Ist der Zusammenhang zwischen den Merkmalen X und Y in K Studien mit den Umfängen n_1, \dots, n_K untersucht worden, so haben wir als Ergebnisse die K Korrelationskoeffizienten r_1, \dots, r_K vorliegen. Führen wir nun, am besten mittels Tab. 2 in Kap. IX, die Fishersche Z-Transformation für alle r_i durch, so erhalten wir:

$$z_1 = \operatorname{arctanh} r_1, \dots, z_K = \operatorname{arctanh} r_K,$$

so daß in guter Näherung gilt:

$$z_i \sim \mathcal{N}\left(\zeta, \frac{1}{n_i - 3}\right), \text{ für } i = 1, \dots, K.$$

Setzen wir nun formal:

$$y_i = z_i, \psi = \zeta \text{ und } \sigma_i^2 = \frac{1}{n_i - 3},$$

so können wir die eben vorgestellten Methoden direkt anwenden.

2.6.2. Beispiel: Korrelation zwischen Rückenspeckdicke und Fleisch-Fett-Verhältnis im Schlachtkörper von Schweinen

In vier Studien wurde die Korrelation zwischen Rückenspeckdicke und Fleisch-Fett-Verhältnis im Schlachtkörper von „leichten“ Schweinen untersucht, vgl. auch das Beispiel von Haiger zur multiplen Korrelation in Abschnitt 5 von Kapitel IX.

In **Tab. 3** sind die Daten der 4 Studien sowie die Auswertung im Fixed Effects Modell zusammengestellt. Wir erhalten daraus den gemeinsamen Schätzer $\hat{\zeta}$ für ζ und das zugehörige 95%-Konfidenzintervall für ζ wie folgt:

$$\hat{\zeta} = 0.311 \text{ und } \text{KI}(\zeta) = \hat{\zeta} \pm 1.96 \cdot \sqrt{1/43} = [0.012, 0.610].$$

Tab. 3: Korrelation Rückenspeckdicke und Fleisch-Fett-Verhältnis, Daten, Z-Transformierte und Auswertung im FE-Modell

Studie i	n _i	r _i	z _i	σ_i^2	v _i = 1/ σ_i^2	d _i = v _i /v _Σ	d _i z _i
1	11	0.37	0.388	1/8	8	0.186	0.072
2	15	0.26	0.266	1/12	12	0.279	0.074
3	7	0.19	0.192	1/4	4	0.093	0.018
4	22	0.32	0.332	1/19	19	0.442	0.147
$\Sigma :$	-	-	1.178	0.511	43	1	0.311

Die Cochransche Homogenitätsstatistik ergibt sich mit

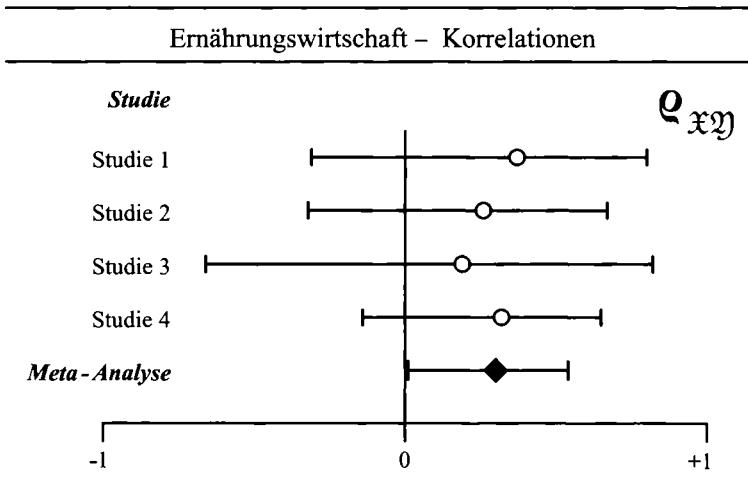
$$Q_C = \sum_{i=1}^4 (n_i - 3) z_i^2 - \frac{\left(\sum_{j=1}^4 (n_j - 3) z_j \right)^2}{\sum_{k=1}^4 n_k - 12} = 4.295 - \frac{13.372^2}{43} = 0.137$$

als sehr klein, so daß wir eine Auswertung im Fixed Effects Modell rechtfer tigen können.

Zurückgerechnet von z auf Korrelationen liefert Tab. 2 in Kap. IX den kom binierten Korrelationskoeffizienten \hat{r} mit zugehörigem 95%-Konfidenzintervall für ρ wie folgt:

$$\hat{r} = 0.3 \text{ und } KI(\rho) = [0.01, 0.54].$$

In Abb. 14 sind die Korrelationen r_i der vier Studien und \hat{r} der Meta-

**Abb. 14:** Rückenspeckdicke \mathfrak{X} – Fleischfettverhältnis \mathfrak{Y}

Analyse sowie die jeweils zugehörigen 95%-Konfidenzintervalle graphisch ver anschaulicht.

Analog können etwa die **Assoziationsmaße** aus Kap. VII kombiniert werden, da zu allen Maßen dort die zugehörigen Varianzen ausführlich hergeleitet sind.

3. Random Effects Model (RE) der Meta-Analyse

Liegt zwischen den Ergebnissen y_1, \dots, y_K von K unabhängigen Studien, Experimenten oder Prognosen eine zusätzliche Variation vor, die durch eine zusätzliche zufällige Abweichung a_i des i-ten Ergebnisses y_i vom gemeinsamen Mittel ψ erklärbar ist, so erhält man das Random Effects Model (RE):

$$y_i = \psi + a_i + e_i, \quad a_i \sim \mathcal{N}(0, \tau^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2), \quad \hat{\sigma}_i^2 \text{ gegeben},$$

wobei die a_i und e_i sowohl untereinander als auch miteinander als unabhängig angenommen werden. Die Größen a_i sind zufällige Interaktionen oder **Wechselwirkungen**, vgl. Abschnitt 5 im Kap. V, und deren Varianz τ^2 wird auch Heterogenitätsparameter genannt.

3.1. Schätzung des Heterogenitätsparameters

Zunächst wird eine Schätzung für die zusätzliche Varianzkomponente τ^2 benötigt. Bezeichnet \bar{y} das arithmetische Mittel der y_i , so gilt für die ungewichteten quadratischen Abweichungen vom Mittel im Erwartungswert:

$$E \sum_{i=1}^K (y_i - \bar{y})^2 = \tau^2(K - 1) + \frac{K-1}{K} \sum_{i=1}^K \sigma_i^2.$$

Setzt man hier für σ_i^2 die gegebenen Schätzungen ein und wendet die Momentenmethode an, vgl. Abschnitt 1.A in Kap. III, so ergibt sich der von Hedges (1983) vorgeschlagene Schätzer:

$$\hat{\tau}_{HED}^2 = \frac{1}{K-1} \sum_{i=1}^K (y_i - \bar{y})^2 - \frac{1}{K} \sum_{i=1}^K \hat{\sigma}_i^2, \quad \text{mit } \bar{y} = \frac{1}{K} \sum_{i=1}^K y_i,$$

der bei 0 gestutzt wird, falls er negativ ausfällt. Gewichtet man die y_i mit $d_i = v_i/v_\Sigma$ und legt die Cochran'sche Quadratform Q_C zu Grunde, vgl. Abschnitt 2, so ergibt die Momentenmethode den von DerSimonian/Laird (1986) vorgeschlagenen Schätzer:

$$\begin{aligned} \hat{\tau}_{DSL}^2 &= \frac{v_\Sigma [Q_C - (K-1)]}{v_\Sigma^2 - \sum_{i=1}^K v_i^2}, \quad \text{wobei} \\ v_i &= 1/\hat{\sigma}_i^2, \quad v_\Sigma = \sum_{i=1}^K v_i \quad \text{und} \quad Q_C = \sum_{i=1}^K v_i \left(y_i - \sum_{j=1}^K v_j y_j / v_\Sigma \right)^2. \end{aligned}$$

Der Schätzer wird auch bei 0 gestutzt, wenn er negativ wird. Ebenfalls von der Cochran'schen Quadratform ausgehend ist ein nicht-negativer minimalverzerrter Schätzer nach Hartung (1981) mit einer Verzerrungskorrektur

gemäß Hartung (1999c, 2003) gegeben als:

$$\hat{\tau}_{\text{HAR}}^2 = \frac{Q_C^2}{\left(v_{\Sigma} - \sum_{i=1}^K v_i^2/v_{\Sigma}\right) [Q_C + 2(K - 1)]}.$$

Je nach Anwendungsgebiet wird der Hedges- oder der DSL-Schätzer bevorzugt. Der positive Schätzer ist auf jeden Fall dann zu empfehlen, wenn die anderen negativ ausfallen. Verzerrung und Varianz des positiven Schätzers $\hat{\tau}_{\text{HAR}}^2$ sind praktisch identisch wie beim DSL-Schätzer, wenn der bei 0 gestutzt wird, vgl. Hartung/Makambi (2002a). Neben diesen explizit gegebenen Schätzern sind in verschiedenen Anwendungen implizit oder iterativ definierte Schätzer gebräuchlich, auf die wir jedoch erst später in Abschnitt 15 eingehen werden.

Beispiel: Fortsetzung MS-Studien. Aus der Tab. 2 ermitteln wir:

$$\bar{y} = 3.2, \quad \sum(y_i - \bar{y})^2 = 16.54, \quad \sum\hat{\sigma}_i^2 = 34.501, \text{ so daß:}$$

$$\hat{\tau}_{\text{HED}}^2 = 16.54/3 - 34.501/4 = -3.11.$$

Weiterhin haben wir mit den dortigen Werten:

$$v_{\Sigma} = 0.813, \quad \sum v_i^2 = 0.30041 \text{ und } Q_C = 2.43427, \text{ so daß:}$$

$$\hat{\tau}_{\text{DSL}}^2 = -1.28 \text{ und } \hat{\tau}_{\text{HAR}}^2 = 1.58418.$$

Mit letzterer Schätzung werden wir das Beispiel fortsetzen.

3.2. Gesamtergebnis und Konfidenzintervalle

Wie im Fixed Effects Model wird das gemeinsame Mittel ψ bei bekannten Varianzen durch den Gauß-Markov Schätzer als unverzerrten Schätzer kleinster Varianz geschätzt. Mit geschätzten Varianzen erhalten wir den kombinierten Schätzer für ψ als:

$$\hat{\psi}_{\text{RE}} = \sum_{i=1}^K c_i y_i, \quad c_i = \frac{w_i}{w_{\Sigma}}, \quad w_i = \frac{1}{\hat{\tau}^2 + \hat{\sigma}_i^2}, \quad w_{\Sigma} = \sum_{j=1}^K w_j,$$

wobei für $\hat{\tau}^2$ hier einer der obigen Schätzer eingesetzt werden kann.

Der klassische Varianzschätzer für $\hat{\psi}_{\text{RE}}$ ist

$$\hat{\sigma}^2(\hat{\psi}_{\text{RE}})_C = 1/w_{\Sigma},$$

so daß mit dem $(1 - \alpha/2)$ -Quantil $u_{1-\alpha/2}$ der $\mathcal{N}(0, 1)$ -Verteilung das entsprechende $(1 - \alpha)$ -Konfidenzintervall für ψ gegeben ist als:

$$\text{KI}_{\text{RE}}(\psi)_C = \hat{\psi}_{\text{RE}} \pm \sqrt{1/w_{\Sigma}} \cdot u_{1-\alpha/2}.$$

Da dieses Intervall das Konfidenzniveau nicht immer einhält, befaßt sich Hartung (1999a) mit den Quadratformen

$$Q_H = \sum_{i=1}^K w_i (y_i - \hat{\psi}_{RE})^2 \text{ und } q_H = \frac{1}{K-1} \sum_{i=1}^K c_i (y_i - \hat{\psi}_{RE})^2$$

und zeigt, daß Q_H einer χ^2_{K-1} -Verteilung genügt und unabhängig von $\hat{\psi}_{RE}$ ist, wenn die geschätzten Varianzen durch die wahren ersetzt werden. Folglich ergibt sich eine weitere Varianzschätzung für den Gesamteffekt,

$$\hat{\sigma}^2(\hat{\psi}_{RE})_H = q_H,$$

so daß mit dem $(1 - \alpha/2)$ -Quantil $t_{K-1,1-\alpha/2}$ der t_{K-1} -Verteilung das zugehörige $(1 - \alpha)$ -Konfidenzintervall für ψ gegeben ist als:

$$KI_{RE}(\psi)_H = \hat{\psi}_{RE} \pm \sqrt{q_H} \cdot t_{K-1,1-\alpha/2}.$$

Simulationen sowohl für normalverteilte als auch für binäre Versuchsauskommen zeigen, daß dieses Intervall das Konfidenzniveau gut einhält, vgl. Hartung (1999a), Hartung/Knapp (2000a; 2001a, b, c; 2003a). Beruhigend ist ebenfalls, daß dieses Intervall wenig empfindlich ist bezüglich recht unterschiedlich ausfallender Schätzungen für τ^2 . Breite und Lage des Intervalls variieren nur geringfügig. Anzumerken ist, daß dieses Konfidenzintervall erneut vorgeschlagen wurde von Sidik/Jonkman (2002). Weiterhin sei erwähnt, daß Hartung (1999a) auch eine **Interpolation** zwischen beiden Konfidenzintervallen vorschlägt, die wir hier jedoch wegen des damit verbundenen Aufwandes nicht näher diskutieren. Auch werden wir nicht für jedes Beispiel beide Konfidenzintervalle berechnen.

In spezifischen Konstellationen von Versuchsdaten ist es möglich, daß der Schätzer $q_H = Q_H/(K-1)w_\Sigma$ wesentlich kleiner als der Varianzschätzer $1/v_\Sigma$ aus dem Fixed Effects Model in Abschnitt 2 ausfällt [vgl. Tab. 29 in Abschnitt 14.2.4]. In der Praxis kann man dann q_H ersetzen durch den **adjustierten** Varianzschätzer:

$$q_{H_{adj}} = \max\{q_H, 1/v_\Sigma\}.$$

Damit genügt man dem Prinzip der **ordnungserhaltenden** Varianzschätzung, vgl. Hartung (1999a, c, d; 2003). Entsprechend kann man auch mit der Varianzschätzung $Q_C/(K-1)v_\Sigma$ im Fixed Effects Model verfahren.

3.3. Beispiel: MS-Studien im RE-Modell

Von Abschnitt 3.1 übernehmen wir die positive Schätzung $\hat{\tau}_{HAR}^2 = 1.58418$ für τ^2 . In **Tab. 4** sind die benötigten Berechnungen zusammengestellt. Daraus erhalten wir:

$$\hat{\psi}_{RE} = 2.45790, \text{ sowie}$$

$$\hat{\sigma}^2(\hat{\psi}_{FE})_C = 1/w_\Sigma = 1.86105 \text{ und } \hat{\sigma}^2(\hat{\psi}_{FE})_H = q_H = 1.18303.$$

Tab. 4: Arbeitstabelle (RE) zu MS-Studien, vgl. auch Tab. 2, mit $\hat{\tau}_{\text{HAR}}^2 = 1.58418$

Nr. i	w _i	c _i	c _i y _i	c _i (y _i - $\hat{\psi}_{\text{RE}}$) ²
1	0.28505	0.53050	0.63660	0.83942
2	0.08486	0.15793	1.04234	2.70960
3	0.10500	0.19541	0.48853	0.0000346
4	0.06242	0.11617	0.29043	0.0000206
$\Sigma:$	0.53733	1	2.45790	3.54908

Alle Größen haben höhere Werte als im FE-Modell. Die 95%-Konfidenzintervalle für ψ ergeben sich mit $u_{0.975} = 1.96$ und $t_{3;0.975} = 3.182$ im RE-Modell zu:

$$\text{KI}_{\text{RE}}(\psi)_C = [-0.216, 5.132] \text{ und } \text{KI}_{\text{RE}}(\psi)_H = [-1.003, 5.919].$$

Sie sind erwartungsgemäß etwas breiter als die entsprechenden Intervalle im FE-Modell. In **Abb. 13** von Abschnitt 2.5 sind die Studienergebnisse nebst zugehörigen 95%-Konfidenzintervallen und die Ergebnisse der Meta-Analyse in beiden Modellen für die Cochranischen 95%-Konfidenzintervalle graphisch veranschaulicht.

3.4. Beispiel: Landwirtschaft – Besamungsquote von Bullen [Einarmige Studien]

Für $K = 6$ Bullen eines Zuchttamms, deren Samen zur künstlichen Befruchtung in jeweils n_i Kuhherden eingesetzt war, wurde die Besamungssicherheit X [in %] mit Erwartungswert μ_X untersucht, vgl. Snedecor/Cochran (1967, p. 290). In **Tab. 5** sind die Stichprobenumfänge n_i , die beobachteten Mittelwerte \bar{x}_i , die Standardabweichungen s_i sowie die versuchsintern geschätzten Varianzen s_i^2/n_i von \bar{x}_i angegeben. Wir setzen:

$$\psi = \mu_X, \quad y_i = \bar{x}_i \quad \text{und} \quad \hat{\sigma}_i^2 = s_i^2/n_i,$$

so daß wir im RE-Modell

$$y_i = \psi + a_i + e_i, \quad a_i \sim \mathcal{N}(0, \tau^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2)$$

für τ^2 die folgenden expliziten Schätzungen gemäß Abschnitt 3.1 erhalten:

$$\hat{\tau}_{\text{HED}}^2 = 89.68, \quad \hat{\tau}_{\text{DSL}}^2 = 66.80 \text{ und } \hat{\tau}_{\text{HAR}}^2 = 58.44.$$

Etwa mit der DSL-Schätzung für τ^2 erhalten wir dann für den gemeinsamen Mittelwert $\psi = \mu_X$ die Schätzungen:

$$\hat{\psi}_{\text{RE}} = 54.70, \quad \text{mit}$$

$$\hat{\sigma}^2(\hat{\psi}_{\text{RE}})_C = 1/w_{\Sigma} = 16.953, \quad \text{und} \quad \hat{\sigma}^2(\hat{\psi}_{\text{RE}})_H = q_H = 19.489.$$

Tab. 5: Besamungsquote von sechs Bullen einer Zucht in der künstlichen Befruchtung

Bulle	n_i	\bar{x}_i	s_i	$\hat{\sigma}_i^2 = s_i^2/n_i$
Anton	5	41.2	13.26	35.14
Detlef	2	64.5	7.78	30.25
Franz	7	56.3	11.53	18.99
Max	5	39.6	22.48	101.06
Paul	7	67.1	16.45	38.64
Xaver	9	53.2	15.79	27.72

Mit den Quantilen $u_{0.975} = 1.96$ und $t_{5,0.975} = 2.571$ ergeben sich daraus die 95%-Konfidenzintervalle für $\psi = \mu_X$ zu:

$$KI_{RE}(\psi)_C = [46.62, 62.77] \text{ und } KI_{RE}(\psi)_H = [43.35, 66.05].$$

Im einleitenden Abschnitt 1.4 sind in der dortigen **Abb. 5** die Besamungsquoten der sechs Bullen sowie $\hat{\psi}_{RE}$ und zugehörige 95%-Konfidenzintervalle graphisch veranschaulicht.

3.5. Studien-Cluster [Hartung-Knapp]

Oft müssen Ergebnisse von Studien zusammengefaßt werden, etwa im Rahmen der Evidence Based Medicine [EBM], die in verschiedenen Zeitperioden durchgeführt wurden. In der Zwischenzeit haben sich z.B. die Beurteilungsskalen oder auch die medizinischen Responderdefinitionen geändert. Durch die Studien mit jeweils identischer Skala oder derselben Responderfestlegung entstehen Cluster innerhalb der K Studien, welche meta-analytisch zusammengefaßt werden sollen.

Witte/Victor (2004) zeigen für Klinische Prüfungen auf, daß durch unterschiedliche Auswertungspopulationen, etwa Intent-To-Treat Population, Full Analysis Set, Per Protocol Population, verschiedene Cluster entstehen, wobei Witte und Victor sich mit der Zusammenfassung von Noninferiority Studien beschäftigen.

Allgemein entstehen durch **Klassenbildung** von einflußnehmenden **Ko-variablen** hier betrachtete Studien-Cluster, s. auch das Tuberkulosebeispiel zur Meta-Regression in Abschnitt 17.3.

Innerhalb der K Studien haben wir $L < K$ Cluster von Studien, wobei im Cluster j sich $k_j - k_{j-1}$ Studien befinden, mit $k_0 = 0$ und $k_L = K$, so daß wir im Cluster j das Modell wie folgt gegeben haben:

$$y_i \sim \mathcal{N}(\psi_j, \tau^2 + \sigma_i^2), \quad i = k_{j-1} + 1, \dots, k_j, \quad \hat{\sigma}_i^2 \text{ gegeben},$$

dabei steht ψ_j für den gemeinsamen Effekt der Studien im Cluster j. Dieses Modell stellt eine **Mischung** aus einem **Inhomogenen Fixed Effects Modell** und dem **Random Effects Modell** dar, vgl. Hartung/Knapp (2005a).

Für die gemeinsame Beurteilung aller Studien werden Gewichte λ_j für Cluster j etwa nach Clusterumfang oder Alter der Studien im j -ten Cluster festgelegt, mit $\lambda_j > 0$ und $\sum_{j=1}^L \lambda_j = 1$, oder einfach $\lambda_j = 1/L$ gesetzt.

Der Einfachheit halber bestimmen wir zunächst einen ungewichteten verallgemeinerten Hedges Schätzer für τ^2 gemäß:

$$\hat{\tau}_{\text{HHK}}^2 = \frac{1}{K-L} \sum_{j=1}^L \sum_{i=k_{j-1}+1}^{k_j} \left\{ y_i - \sum_{h=k_{j-1}+1}^{k_j} \frac{y_h}{k_j - k_{j-1}} \right\}^2 - \frac{1}{K-L} \sum_{j=1}^K \left\{ 1 - \frac{1}{k_j - k_{j-1}} \right\} \sum_{i=k_{j-1}+1}^{k_j} \hat{\sigma}_i^2.$$

Damit berechnen wir unsere Gewichte:

$$c_i = w_i / w_\Sigma, \quad w_i = 1 / (\hat{\tau}_{\text{HHK}}^2 + \hat{\sigma}_i^2), \quad w_\Sigma = \sum_{i=1}^K w_i.$$

Der aus allen Studien kombinierte Schätzer für den über die Cluster gemittelten gemeinsamen Effekt $\psi = \sum_{j=1}^L \lambda_j \psi_j$:

$$\hat{\psi} = \sum_{j=1}^L \lambda_j \sum_{i=k_{j-1}+1}^{k_j} c_i y_i / c_{\Sigma(j)}, \quad \text{mit } c_{\Sigma(j)} = \sum_{i=k_{j-1}+1}^{k_j} c_i,$$

besitzt die geschätzte Varianz:

$$\widehat{\text{Var}} \hat{\psi} = \frac{1}{K-L} \left\{ \sum_{j=1}^L \sum_{i=k_{j-1}+1}^{k_j} c_i \left(y_i - \sum_{h=k_{j-1}+1}^{k_j} c_h y_h / c_{\Sigma(j)} \right)^2 \right\} \cdot \left\{ \sum_{j=1}^L \lambda_j^2 / c_{\Sigma(j)} \right\},$$

so daß ein $(1-\alpha)$ -Konfidenzintervall für ψ gegeben ist als:

$$\text{KI}(\psi)_{\text{HK}} = \hat{\psi} \pm \sqrt{\widehat{\text{Var}} \hat{\psi}} \cdot t_{K-L; 1-\alpha/2},$$

worin wir bei nicht so strikter Auffassung das t_{K-L} -Quantil auch gegen das $\mathcal{N}(0, 1)$ -Quantil $u_{1-\alpha/2}$ austauschen können.

Die Schätzer $\hat{\psi}$ und $\widehat{\text{Var}} \hat{\psi}$ liefern unabhängige unverzerrte Schätzungen für ψ und $\text{Var} \hat{\psi}$, wenn die wahren Varianzkomponenten σ_i^2 und τ^2 anstelle ihrer Schätzungen eingesetzt würden. Es bleibt noch anzumerken, daß für $L = 1$ der Schätzer $\hat{\tau}_{\text{HHK}}^2$ wieder mit dem Hedges Schätzer $\hat{\tau}_{\text{HED}}^2$ aus Abschnitt 3.1 zusammenfällt.

3.6. Beispiel: Fortsetzung Landwirtschaft

Wir setzen das Beispiel zur Erhebung der Besamungssicherheit von $K = 6$ Bullen aus dem vorletzten Abschnitt 3.4 fort. Hier nehmen wir einmal an: Bei Durchsicht der Erfassungsbögen fällt einem findigen Tierarzt auf, daß ausgerechnet die 3 Bullen mit den niedrigsten Besamungserfolgen alle im Süden im Einsatz waren, während die 3 Bullen mit den besseren Erfolgen alle im Norden wirksam waren. Also wird ein Teil der Variation der Besamungsergebnisse durch diese externe, geographische Begleitvariable erklärt und deren geeignete Berücksichtigung in der Auswertung der Erhebungsdaten muß zu einem genaueren Gesamtergebnis führen, d. h. zu einem schmäleren 95%-Konfidenzintervall für $\psi = \mu_X$ als in Abschnitt 3.4 in Unkenntnis dieser Kovariablen berechnet. Wir betrachten folglich die $L = 2$ **Erhebungscluster**:

$$\text{Cluster(I)} = \{\text{Anton, Max, Xaver}\} \text{ und}$$

$$\text{Cluster(II)} = \{\text{Detlef, Franz, Paul}\},$$

so daß wir gemäß der Formeln im letzten Abschnitt 3.5 vorgehen können. Zunächst berechnen wir den verallgemeinerten Hedges Schätzer:

$$\hat{\tau}_{\text{HHK}}^2 = \frac{1}{4} \{110.507 + 63.547\} - \frac{1}{4} \left\{ \frac{2}{3} \cdot 163.92 + \frac{2}{3} \cdot 87.79 \right\} = 1.56$$

und damit $c_i = w_i / w_{\Sigma}$ mit $w_i = (1.56 + \hat{\sigma}_i^2)^{-1}$, $w_{\Sigma} = \sum_{j=1}^6 w_j$, so daß sich die c_i -Werte im Cluster (I) und deren Summe $c_{\Sigma(I)}$ ergeben zu:

$$0.15454, 0.05526, 0.19366 \text{ und } c_{\Sigma(I)} = 0.40346.$$

Entsprechend ergeben sich für Cluster (II) die c_i -Werte:

$$0.17828, 0.27717, 0.14107 \text{ und } c_{\Sigma(II)} = 0.59652.$$

Mit den Clustergewichten $\lambda_1 = \lambda_2 = 1/2$ erhalten wir das kombinierte Gesamtergebnis:

$$\hat{\psi} = \frac{1}{2} \cdot 46.738 + \frac{1}{2} \cdot 61.305 = 54.02.$$

Die zugehörige Varianz schätzen wir zu:

$$\widehat{\text{Var}} \hat{\psi} = \frac{1}{4} \{15.70 + 13.50\} \cdot \left\{ \frac{1}{4 \cdot 0.40346} + \frac{1}{4 \cdot 0.59652} \right\} = 7.58278,$$

und erhalten das 95%-Konfidenzintervall für $\psi = \mu_X$ mittels des Quantils $t_{4;0.975} = 2.776$ als:

$$\text{KI}(\psi)_{\text{HK}} = 54.02 \pm 2.776 \cdot \sqrt{7.58278} = 54.02 \pm 7.644 = [46.376, 61.66],$$

dessen Breite mit 15.28 tatsächlich geringer ist als die Breiten 16.15 bzw. 22.7 der 95%-Konfidenzintervalle für ψ in Abschnitt 3.4, welche ohne die Berücksichtigung der zur Clusterbildung führenden Kovariablen berechnet sind. In **Abb. 15** wird der Sachverhalt durch Angabe der jeweiligen Schätzer mit zugehörigem 95%-Konfidenzintervall graphisch veranschaulicht.

Landwirtschaft: Erhebungs - Cluster
Besamungssicherheit der Bullen

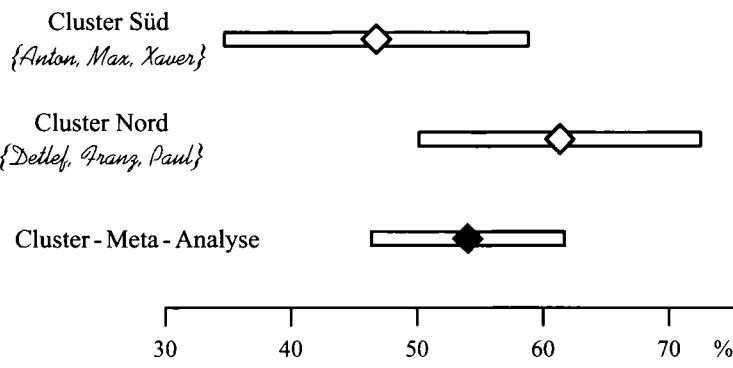


Abb. 15: Meta – Analysen in den Clustern
und für die Cluster – Kombination

4. Diskrete Beobachtungen

Häufig werden etwa in Klinischen Studien Patienten an Hand einer ordinalen Skala beurteilt, die in verschiedenen Studien unterschiedlich viele Ausprägungen haben mag. Für die wesentliche Analyse eines Behandlungserfolges wird i. d. R. dann das Skalenniveau gesenkt auf ein nominales mit nur noch zwei Ausprägungen, indem man in den ordinalen Skalen jeweils die Kategorien festlegt, die einen „Responder“ der Behandlung, z. B. „als geheilt“, beschreiben, die übrigen Kategorien stehen dann für „Nonresponder“. Bei einer Gruppe von n Patienten wird das Auskommen des j -ten Patienten dann beschrieben durch eine 0 oder 1 wertige Bernoulli-verteilte Zufallsvariable X_j mit Parameter p , d. h. $X_j \sim B(1, p)$, vgl. Abschnitt 3.1 in Kap. IV, wobei $p = P(X_j = 1)$ für die Wahrscheinlichkeit steht, ein Responder zu sein. Das Ergebnis der Gruppe wird nun beschrieben durch die mit den Parametern n und p binomialverteilte Zufallsvariable $X = \sum_{j=1}^n X_j \sim B(n, p)$ und p wird unverzerrt geschätzt durch $\hat{p} = m/n$, wobei $m = \sum_{j=1}^n X_j$ für die Anzahl der Responder in der Gruppe steht. Bei vergleichenden Studien erfolgt die Auswertung dann über Normalverteilungsapproximationen, s. folgende Abschnitte 4.1 - 5.

Wird das Auskommen einer einzelnen Versuchseinheit j eines Experiments oder eines einzelnen Patienten j bereits durch eine $B(L, p)$ -verteilte Zufallsvariable X_j beschrieben, etwa wenn im Rahmen einer „Quality of Life“-Beurteilung bei einem Patienten gezählt wird, wieviele von L unabhängigen und gleichwertigen Tests er besteht, so wird X_j gemäß der **Arcus-Sinus-Transformation**:

$$Y_j = g(X_j) = \sqrt{L + 0.5} \cdot \arcsin \sqrt{\frac{X_j + 0.375}{L + 0.75}}$$

in eine praktisch normalverteilte Zufallsvariable Y_j überführt, für welche

dann gilt: $Y_j \sim \mathcal{N}(g(Lp), 1/4)$. Bei n Versuchseinheiten wird das Gesamtergebnis dann beschrieben durch

$$y = \sum_{j=1}^n Y_j / n \sim \mathcal{N}(g(Lp), 1/4n),$$

so daß wir in einer Meta-Analyse mehrerer solcher Studien oder Experimente die Verfahren aus den vorstehenden Abschnitten 2 und 3 direkt auf die derart transformierten Daten anwenden können, vgl. Abschnitt 4.5E in Kap. V, in dem die Transformation auch für ein umfangreiches Beispiel demonstriert wird, sowie Anscombe (1948).

Sind L groß und p klein ($L \geq 30, p < 0.1$), so ist X_j genähert poissonverteilt mit Parameter λ , d. h. $X_j \sim Po(\lambda)$, wobei $\lambda \approx Lp$, vgl. Abschnitt 3 in Kap. V. Wir haben es dann mit „selteneren Ereignissen“ zu tun. Uns bekannte Anwendungen im klinischen Bereich sind etwa Studien mit der $Po(\lambda)$ -verteilten Haupt-Zielgröße: Anzahl des Auftretens von Harnwegsinfektionen bei Patienten mit regelmäßigen, aber reversiblen derartigen Infektionen, oder allgemein als Neben-Zielgrößen im Bereich der Nebenwirkungen. Wir werden in Abschnitt 4.7 als eine Anwendung der Poissonverteilung im Bereich Umwelt bzw. Technik eine Meta-Analyse zur Ermittlung der Belastung durch Asbest-Fasern behandeln.

Zunächst wenden wir uns jedoch der anfangs geschilderten Situation zu, in welcher der einzelne Patient schließlich durch eine Bernoulli-verteilte 0-1-Variable beschrieben wird, lediglich der Parameter p kann in den Vergleichsgruppen unterschiedlich ausfallen.

Wir betrachten $i = 1, \dots, K \geq 2$ unabhängige Studien mit jeweils unabhängigen Vergleichsgruppen: Experimentelle Gruppe E und Kontrollgruppe C. Von den n_{E_i} und n_{C_i} Beobachtungen liegen jeweils m_{E_i} und m_{C_i} Responder vor, so daß

$$\hat{p}_{E_i} = m_{E_i} / n_{E_i} \text{ und } \hat{p}_{C_i} = m_{C_i} / n_{C_i}$$

die jeweiligen Erfolgswahrscheinlichkeiten p_E und p_C schätzen, die es dann zu vergleichen gilt. Zu diesem Zwecke betrachten wir Effektmaße, die folglich die Größen ψ , y_i und $\hat{\sigma}_i^2$ in den oben eingeführten FE- und RE-Modellen definieren.

4.1. Risiko Differenz RD

Es ist: $RD = p_E - p_C$ und wir setzen: $\psi = RD$, so daß in der i -ten Studie diese Differenz geschätzt wird durch:

$$y_i = \frac{m_{E_i}}{n_{E_i}} - \frac{m_{C_i}}{n_{C_i}} = \widehat{RD}_i.$$

Wir schätzen die Varianz von y_i versuchsintern erwartungstreu gemäß:

$$\hat{\sigma}_i^2 = \frac{m_{E_i}(n_{E_i} - m_{E_i})}{n_{E_i}^2(n_{E_i} - 1)} + \frac{m_{C_i}(n_{C_i} - m_{C_i})}{n_{C_i}^2(n_{C_i} - 1)}, \quad n_{E_i} \geq 2, \quad n_{C_i} \geq 2.$$

Falls z. B. $m_{E_i} = 0$ oder $m_{E_i} = n_{E_i}$, dann sollte man m_{E_i} ersetzen durch 0.5 bzw. $n_{E_i} - 0.5$.

4.2. Odds Ratio OR

Statt des Odds Ratio OR, wobei, vgl. Abschnitt 3.1 in Kap. VII,

$$OR = \frac{p_E}{1 - p_E} / \frac{p_C}{1 - p_C}$$

ist, betrachten wir den logarithmierten (natürlichen ln) Log OR, so daß wir setzen: $\psi = \ln OR$, der in der i-ten Studie geschätzt wird mittels:

$$y_i = \ln \left\{ \frac{(m_{E_i} + 0.5)(n_{C_i} - m_{C_i} + 0.5)}{(n_{E_i} - m_{E_i} + 0.5)(m_{C_i} + 0.5)} \right\} = \ln \widehat{OR}_i.$$

Versuchsintern wird die Varianz von y_i nahezu unverzerrt geschätzt mittels:

$$\widehat{\sigma}_i^2 = \frac{1}{m_{E_i} + 0.5} + \frac{1}{n_{E_i} - m_{E_i} + 0.5} + \frac{1}{m_{C_i} + 0.5} + \frac{1}{n_{C_i} - m_{C_i} + 0.5}.$$

4.3. Relatives Risiko RR und Noninferiority

Das Relative Risiko RR ist definiert als $RR = p_E/p_C$ und wird ebenfalls als Log RR analysiert. Man setzt dann: $\psi = \ln RR$, welches in der i-ten Studie geschätzt wird durch:

$$y_i = \ln \left\{ \frac{(m_{E_i} + 0.5)/(n_{E_i} + 0.5)}{(m_{C_i} + 0.5)/(n_{C_i} + 0.5)} \right\} = \ln \widehat{RR}_i.$$

Die Varianz von y_i wird versuchsintern nahezu unverzerrt geschätzt mittels:

$$\widehat{\sigma}_i^2 = \frac{1}{m_{E_i} + 0.5} - \frac{1}{n_{E_i} + 0.5} + \frac{1}{m_{C_i} + 0.5} - \frac{1}{n_{C_i} + 0.5}.$$

Falls $m_{E_i} = n_{E_i}$ oder $m_{C_i} = n_{C_i}$, dann wird der Hilfswert 0.5 nicht zu m_{E_i} bzw. m_{C_i} addiert.

Sogenannte **Noninferiority** Studien dienen dem Nachweis der Nichtunterlegenheit einer neuen experimentellen Behandlung E im Vergleich zu einer Standard-Kontrollbehandlung C. Wird wie hier das Versuchsauskommen durch Erfolgswahrscheinlichkeiten p_E und p_C beschrieben, so ist es üblich, den Noninferiority Parameter $\Delta > 0$ relativ z. B. als $\Delta \cdot 100$ in % anzugeben, d. h. wir testen zum Niveau $\alpha/2$ die Nullhypothese

$$H_0 : p_E \leq p_C - \Delta \cdot p_C \text{ gegen } H_1 : p_E > p_C - \Delta \cdot p_C.$$

Dazu äquivalent ist $RR = p_E/p_C > 1 - \Delta$ nachzuweisen, indem das $(1 - \alpha)$ -Konfidenzintervall für RR oberhalb von $1 - \Delta$ liegt, bzw. das $(1 - \alpha)$ -Konfidenzintervall für $\psi = \ln RR$ oberhalb von $\ln(1 - \Delta)$.

Beispiel: Wir betrachten die beiden für das Medikament Cisapride (E) ungünstigsten Studien von Kellow und Yeoh aus **Tab. 6** zur Behandlung der Funktionellen Dyspepsie im Vergleich zu einer Standard-Placebobehandlung (C), vgl. nachfolgendes Beispiel. Uns interessiert hier, ob Cisapride in diesen

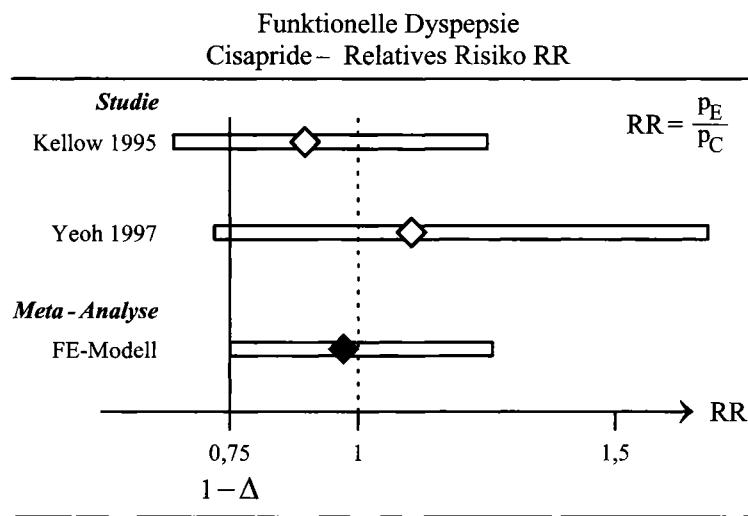


Abb. 16: Noninferiority Studien mit $\Delta = 25\%$.
Schätzer und 95% – Konfidenzintervalle für RR

beiden Studien wenigstens die Nichtunterlegenheits-Schranke von $-\Delta \cdot p_C = -0.25 \cdot p_C$ erreicht, wobei wir wie ebenfalls üblich $1 - \alpha = 0.95$ wählen.

Die Versuchsdaten in Tab. 6 ergeben für die erstgenannte Studie mit $m_{E_1}/n_{E_1} = 19/29$ und $m_{C_1}/n_{C_1} = 22/30$ für $\ln \widehat{RR}_1$:

$$y_1 = -0.10976 \text{ und } \widehat{\sigma}_1^2 = 0.029.$$

Daraus ergibt sich in der ersten Studie das 95%-Konfidenzintervall für $\psi = \ln RR$ zu:

$$KI(\psi)_1 = y_1 \pm 1.96 \cdot \sqrt{\widehat{\sigma}_1^2} = [-0.44378, 0.22426] = [a, b],$$

bzw. mittels $[e^a, e^b]$ und $e^{y_1} = \widehat{RR}_1$ für RR:

$$KI(RR)_1 = [0.64, 1.25] \text{ und } \widehat{RR}_1 = 0.896.$$

Entsprechend erhalten wir mit $m_{E_2}/n_{E_2} = 21/38$ und $m_{C_2}/n_{C_2} = 19/38$ in der zweitgenannten Studie von Yeoh für $\ln \widehat{RR}_2$:

$$y_2 = 0.09764 \text{ und } \widehat{\sigma}_2^2 = 0.046, \text{ so daß:}$$

$$KI(\psi)_2 = y_2 \pm 1.96 \cdot \sqrt{\widehat{\sigma}_2^2} = [-0.32205, 0.51311] \text{ bzw. für RR :}$$

$$KI(RR)_2 = [0.72, 1.68] \text{ und } \widehat{RR}_2 = 1.103.$$

Für die Meta-Analyse zur Kombination der beiden Studien berechnen wir $d_i = v_i/v_\Sigma$ mit $v_i = 1/\widehat{\sigma}_i^2$ und $v_\Sigma = v_1 + v_2$ zu:

$$d_1 = 0.612, d_2 = 0.388 \text{ und } v_\Sigma = 56.25.$$

Damit berechnen wir den kombinierten Schätzer und das 95%-Konfidenzintervall für $\psi = \ln RR$ zu:

$$\hat{\psi}_{FE} = d_1 \cdot y_1 + d_2 \cdot y_2 = -0.029 \text{ sowie}$$

$$KI_{FE}(\psi) = \hat{\psi}_{FE} \pm 1.96 \cdot \sqrt{1/v_\Sigma} = -0.029 \pm 0.26 = [-0.289, 0.231]$$

und für RR :

$$KI_{FE}(RR) = [0.75, 1.26] \text{ und } \widehat{RR}_{FE} = 0.971 = e^{\hat{\psi}_{FE}}.$$

Also wird in der Kombination der beiden Studien von $RR = p_E/p_C$ die zugehörige Noninferiority Schranke $1 - \Delta = 0.75$ signifikant knapp erreicht. Obige Schätzer und 95%-Konfidenzintervalle für RR sind in **Abb. 16** auch graphisch veranschaulicht mit Bezug auf die Noninferiority Schranke.

4.4. Beispiel: Klinische Prüfungen zu Funktioneller Dyspepsie [Zweiarmige Studien]

Funktionelle Dyspepsie beschreibt nicht notwendigerweise organisch bedingte Beschwerden im oberen Bauchbereich. Näheres entnehme man den Abhandlungen von Allescher et al. (1999) und Allescher et al. (2001), die eine Meta-Analyse von diesbezüglichen Studien enthält. Der Vergleich von unterschiedlichen Medikamentengruppen bezüglich ihrer Behandlungserfolge in jeweiligen placebokontrollierten Studien steht dabei im Vordergrund, vgl. auch Allescher/Knapp/Hartung (2002).

Im Bereich der Gastropunktika interessierte insbesondere das Medikament Cisapride. In **Tab. 6** sind die Daten nebst Hilfsgrößen von 13 Studien mit jeweiligem Vergleich von Cisapride und Placebo aufgeführt, vgl. auch Hartung/Knapp (2001b). Beurteilt wurde das Responderverhalten auf Grund einer allen 13 Studien gemeinsamen Beschwerdenskala. Wir weisen darauf hin, daß in Tab. 6 die Standardabweichungen und nicht die Varianzen aufgeführt sind. Wir beschränken uns auf die Effektmaße: Risiko Differenz (RD) und Log Odds Ratio (Log OR). Für das Maß RD erhalten wir gemäß Abschnitt 3.1 die expliziten Schätzungen für τ^2 als:

$$\hat{\tau}_{HED}^2 = 0.0340, \quad \hat{\tau}_{DSL}^2 = 0.0391 \text{ und } \hat{\tau}_{HAR}^2 = 0.0348,$$

folglich ein recht gleichmäßiges Bild, so daß wir im Folgenden etwa mit dem DSL-Schätzer weiterarbeiten, der sich für das Maß Log OR ergibt zu: $\hat{\tau}_{DSL}^2 = 0.8086$. Der Homogenitätstest aus Abschnitt 2.4, falls von Interesse, muß für beide Maße zum 5%-Niveau abgelehnt werden. Somit ist eine Meta-Analyse der 13 Studien im Random Effects Modell zusätzlich gerechtfertigt. In **Tab. 7** sind die wesentlichen Ergebnisse der Meta-Analyse für beide Effektmaße zusammengestellt. Es zeigt sich ein deutlich signifikantes Bild in allen aufgeführten 95%-Konfidenzintervallen für den jeweils meta-analytisch kombinierten Gesamteffekt von Cisapride. Im einleitenden Abschnitt 1.2 sind in der dort befindlichen **Abb. 2** die Risiko Differenzen Schätzer \widehat{RD}_i der 13 Studien und \widehat{RD}_{RE} der diesbezüglichen Meta-Analyse im RE-Modell sowie zugehörige 95%-Konfidenzintervalle graphisch veranschaulicht.

Bzgl. Literaturhinweisen zu obigen Effektmaßen sei auf Hartung/Knapp (2001b) und Hartung/Knapp/Sinha (2008) hingewiesen.

Tab. 6: Wirksamkeitsprüfungen von Cisapride (E) und Placebo (C) bei Patienten mit Funktioneller Dyspepsie

Studie ^(*)	\widehat{p}_{E_i} m_{E_i}/n_{E_i}	\widehat{p}_{C_i} m_{C_i}/n_{C_i}	\widehat{RD}_i	$\widehat{\sigma}_i$	$\widehat{\ln OR}_i$	$\widehat{\sigma}_{\widehat{\ln OR}_i}$
Creyton 1984	15/16	9/16	0.3750	0.1425	2.4567	1.1492
Milo 1984	12/16	1/16	0.6875	0.1281	3.8067	1.1832
Francois 1986	29/34	18/34	0.3235	0.1065	1.6401	0.5937
Deruyttere 1987	42/56	31/56	0.1964	0.0889	0.8835	0.4093
Hannon 1987	14/22	6/22	0.3636	0.1431	1.5404	0.6524
Roesch 1987	44/54	17/55	0.5057	0.0825	2.2860	0.4559
De Nutt 1989	14/17	7/15	0.3569	0.1639	1.6740	0.8201
Hausken 1992	29/58	23/58	0.1034	0.0927	0.4200	0.3755
Chung 1993	10/14	3/15	0.5143	0.1647	2.3026	0.8756
v. Outryve 1993	17/26	6/27	0.4316	0.1253	1.8888	0.6199
Al-Quorain 1995	38/44	12/45	0.5970	0.0848	2.8574	0.5537
Kellow 1995	19/29	22/30	-0.0782	0.1217	-0.3697	0.5684
Yeoh 1997	21/38	19/38	0.0526	0.1159	0.2113	0.4601

(*): Vollständige Literaturangaben in Hartung/Knapp (2001b)

4.5. Odds Ratio Schätzer von Mantel-Haenszel und Peto oder Yusuf-Peto

Etwa in der **Epidemiologie** beliebt ist der Odds Ratio Schätzer von Mantel/Haenszel (1959), vgl. z. B. Kreienbrock/Schach (2005). Die K unabhängigen Studien entsprechen da oft K Schichten einer untersuchten Grundgesamtheit, z. B. K = 2 Schichten, wenn getrennt nach Geschlechtern untersucht wird. Für jede Schicht i kann dann eine (2×2) -Kontingenztafel aufgestellt werden, sortiert nach Gruppe E oder C und in die andere Richtung nach Responder JA oder NEIN, vgl. Abschnitt 1 in Kap. VII, und der i-te Schätzer berechnet werden. Die hier vorgestellten Schätzer werden bei kleinen Stichprobenumfängen n_i , d. h. bei dünn besetzten Kontingenztafeln, sogenannten „Sparse Data“, empfohlen, vgl. Breslow (1981), Sweeting/Sutton/Lambert (2004), Bradburn et al. (2007). Die Schätzer setzen, auch für die Schichten, ein homogenes Fixed Effects Modell voraus. Wie bisher werden die Erfolgswahrscheinlichkeiten in der i-ten Studie oder Schicht geschätzt durch:

$$\widehat{p}_{E_i} = m_{E_i}/n_{E_i} \text{ und } \widehat{p}_{C_i} = m_{C_i}/n_{C_i}, \text{ wobei } n_i = n_{E_i} + n_{C_i}$$

den Gesamtstichprobenumfang der i-ten Schicht oder Studie bezeichnet. Mit:

$$\widehat{OR}_i = \frac{m_{E_i}(n_{C_i} - m_{C_i})}{m_{C_i}(n_{E_i} - m_{E_i})} \text{ und } \omega_i = \frac{m_{C_i}(n_{E_i} - m_{E_i})}{n_i}$$

Tab. 7: Funktionelle Dyspepsie: Schätzung des Gesamteffektes von Cisapride und zugehörige 95%-Konfidenzintervalle

Resultate	Risiko Differenz	Log Odds Ratio
$\hat{\psi}_{RE}$	0.3375	1.4911
$KI_{RE}(\psi)_C$	[0.2119, 0.4632]	[0.8881, 2.0941]
$KI_{RE}(\psi)_H$	[0.2021, 0.4730]	[0.8206, 2.1617]
$\widehat{\tau^2_{DSL}}$	0.0391	0.8086
Q_C	49.44	41.67
		$\chi^2_{12;0.95} = 21.03$

definieren **Mantel/Haenszel** (1959) als kombinierten Schätzer für das gemeinsame Odds Ratio OR:

$$\widehat{OR}_{MH} = \sum_{i=1}^K \frac{\omega_i}{\sum_{j=1}^K \omega_j} \widehat{OR}_i.$$

Setzen wir:

$$\begin{aligned} R_i &= m_{E_i}(n_{C_i} - m_{C_i})/n_i, & S_i &= m_{C_i}(n_{E_i} - m_{E_i})/n_i, \\ T_i &= (m_{E_i} - m_{C_i} + n_{C_i})/n_i, & U_i &= (m_{C_i} - m_{E_i} + n_{E_i})/n_i, \end{aligned}$$

so erhalten wir, vgl. Robins/Breslow/Greenland (1986), Silcocks (2005), als Schätzer für die Varianz von $\ln \widehat{OR}_{MH}$:

$$\begin{aligned} \widehat{\text{Var}} \ln \widehat{OR}_{MH} &= \frac{\sum_{i=1}^K T_i R_i}{2 \left(\sum_{j=1}^K R_j \right)^2} + \frac{\sum_{i=1}^K (T_i S_i + U_i R_i)}{2 \sum_{j=1}^K R_j \sum_{k=1}^K S_k} \\ &\quad + \frac{\sum_{i=1}^K U_i S_i}{2 \left(\sum_{j=1}^K S_j \right)^2}. \end{aligned}$$

Mit dem $(1 - \alpha/2)$ -Quantil $u_{1-\alpha/2}$ der $\mathcal{N}(0, 1)$ -Verteilung ist ein $(1 - \alpha)$ -Konfidenzintervall für $\ln OR$ dann gegeben als:

$$KI(\ln OR)_{MH} = \ln \widehat{OR}_{MH} \pm u_{1-\alpha/2} \sqrt{\widehat{\text{Var}} \ln \widehat{OR}_{MH}}.$$

Der **Peto Schätzer**, auch Yusuf-Peto Schätzer genannt, vgl. Yusuf et al. (1985), für das Log Odds Ratio in der i-ten Schicht oder Studie lautet:

$$\begin{aligned} [\ln \widehat{OR}]_{i,Peto} &= \frac{m_{E_i} - e_i}{V_i}, \text{ mit } e_i = \frac{n_{E_i}(m_{E_i} + m_{C_i})}{n_i} \text{ und} \\ V_i &= \frac{n_{E_i} n_{C_i} (m_{E_i} + m_{C_i})(n_i - m_{E_i} - m_{C_i})}{(n_i - 1)n_i^2}, \quad n_i = n_{E_i} + n_{C_i}. \end{aligned}$$

In der i-ten Schicht wird seine Varianz geschätzt als:

$$\widehat{\text{Var}}[\ln \widehat{\text{OR}}]_{i,\text{Peto}} = \frac{1}{V_i}.$$

Der aus allen Schichten oder Studien kombinierte Schätzer für das gemeinsame Log Odds Ratio lautet:

$$[\ln \widehat{\text{OR}}]_{\text{Peto}} = \sum_{i=1}^K \frac{V_i}{\sum_{j=1}^K V_j} [\ln \widehat{\text{OR}}]_{i,\text{Peto}},$$

und seine Varianz wird geschätzt durch:

$$\widehat{\text{Var}}[\ln \widehat{\text{OR}}]_{\text{Peto}} = \frac{1}{\sum_{i=1}^K V_i},$$

also ganz analog zum Vorgehen im Fixed Effects Modell in Abschnitt 2. Ein $(1 - \alpha)$ -Konfidenzintervall für $\ln \text{OR}$ ist dann gegeben als:

$$\text{KI}(\ln \text{OR})_{\text{Peto}} = [\ln \widehat{\text{OR}}]_{\text{Peto}} \pm u_{1-\alpha/2} \sqrt{\widehat{\text{Var}}[\ln \widehat{\text{OR}}]_{\text{Peto}}}.$$

Greenland/Robins (1985) leiten Schätzer vom Mantel-Haenszel Typ für die Risiko Differenz und das Relative Risiko her.

4.6. Poissonverteilte Daten

Sei X eine poissonverteilte Zufallsvariable mit Parameter λ , $X \sim \text{Po}(\lambda)$, dann ist: $E X = \lambda$ und $\text{Var } X = \lambda$. Daher wird in Abschnitt 4.5.B von Kap. V die folgende Wurzeltransformation nahegelegt, vgl. Anscombe (1948):

$$Y = \sqrt{X + 3/8}.$$

Es gilt dann praktisch stets:

$$Y \sim \mathcal{N}(\mu, 0.25) \text{ mit } \mu = \sqrt{\lambda + 1/8}.$$

Es gilt dann auch der **Steinersche Verschiebungssatz**, vgl. Abschnitt 8.2 A in Kap. II, für Y :

$$\text{Var } Y = E(Y^2) - (EY)^2 \text{ bzw. } \frac{1}{4} = \left(\lambda + \frac{3}{8} \right) - \left(\lambda + \frac{1}{8} \right).$$

Beobachten wir in der i-ten Studie zur selben $\text{Po}(\lambda)$ -Verteilung X_{ij} bei $j = 1, \dots, n_i$ Versuchsstoffen, so erhalten wir als Ergebnis der i-ten Studie den Mittelwert der transformierten Beobachtungen:

$$y_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \sim \mathcal{N}(\mu, 1/4n_i) \text{ mit } Y_{ij} = \sqrt{X_{ij} + 3/8}.$$

Setzen wir noch $\psi = \mu = (\lambda + 1/8)^{1/2}$ und $\sigma_i^2 = 1/4n_i$, so haben wir formal die Form, für welche wir die Fixed Effects und Random Effects Modelle in den Abschnitten 2 und 3 dargelegt haben.

4.7. Beispiel Umwelt-Technik: Asbest-Faser Zählung [Einarmige Studien]

Zur Feststellung einer durch Asbest-Fasern verursachten Belastung werden Meßfilter aufgestellt und nach einer einheitlichen Richtlinie ausgezählt, vgl. Weber/Argac/Hallier/Hartung (2001) und Hartung/Knapp (1998).

Mit einem Elektronenrastermikroskop werden auf dem Filter zufällig eine große, aber fixierte Anzahl von Flächenstückchen ausgewählt, die zusammen einen vorgeschriebenen Flächeninhalt ergeben. Auf diesen Flächenstückchen wird dann geschaut, ob sich eine Asbest-Faser, neben einer Reihe anderer Faserarten, identifizieren läßt, vgl. auch den einführenden Abschnitt 1.8 und die dortige **Abb. 10** für eine schematische Veranschaulichung der physikalisch-technischen Vorgänge. Meistens findet man keine Asbest-Faser, aber ab einer geringen Anzahl dabei insgesamt identifizierbarer Asbest-Fasern gilt das Objekt als belastet. Die Poissonverteilung beschreibt diese Zählsituation recht präzise, vgl. Hartung/Knapp (1998).

Das Ergebnis hängt von einer Vielzahl an wenig beeinflußbaren Faktoren ab: Vom Raum, in dem der Filter aufgestellt wird, von der gewählten Stelle im Raum, von den klimatischen Verhältnissen an dieser Stelle, etwa Luftstrom oder ruhige Ecke, von der Aufstellzeit, dann vom Labor und im Labor von der Erfahrung desjenigen, der auszählt.

Deshalb betrachten wir hier einmal in einem kleinen Ausschnitt eine für viele Situationen typische Begebenheit: Eine Hafenbehörde beauftragt ein ausgewiesenes Labor, die Asbestbelastung auf einer im Hafen liegenden Ostsseefähre zu ermitteln. Auf Grund des Ergebnisses legt sie das Schiff still. Der Schiffseigner beauftragt ein anderes Labor, welches zu einem für ihn günstigeren Resultat kommt. Daraufhin beauftragt ein Gericht ein drittes Labor, eine klärende Asbestanalyse für die Fähre vorzunehmen. Es erklärt sich fast schon von selbst, daß jedes Labor nur eine geringe Stückzahl an Filtern aufstellt und auswertet, da jede Analyse aufwendig und somit kostspielig ist. Wir haben es also auch in dieser Hinsicht mit einem „Problem der kleinen Zahlen“ zu tun, vgl. das Tod-durch-Hufschlag-Beispiel von Bortkiewicz aus dem Jahr 1898 in Abschnitt 3.4 von Kapitel IV. In **Tab. 8** sind beispielsweise Zählergebnisse für Asbest-Fasern der drei Labore angegeben.

Der erlaubte Grenzwert lag, zumindest hier, bei $\lambda = 5$. Bei $\hat{\lambda} = 5$ oder mehr mußte das betroffene Objekt geschlossen bzw. stillgelegt werden. In unserem Beispiel ergeben sich die folgenden Mittelwerte:

$$\bar{X}_1 = \frac{1}{3} \sum_{j=1}^3 X_{1j} = 15.33, \quad \bar{X}_2 = \frac{1}{5} \sum_{j=1}^5 X_{2j} = 4.2, \quad \bar{X}_3 = \frac{1}{8} \sum_{j=1}^8 X_{3j} = 5,$$

die recht gegensätzliche Konsequenzen zur Folge haben. Jedes Laborergebnis hat eine andere Genauigkeit, die nicht spezifiziert wird.

In **Tab. 9** sind die transformierten Zählergebnisse Y_{ij} sowie deren Mittelwerte y_i zusammengestellt: $Y_{ij} = \sqrt{X_{ij} + 3/8}$ und $y_i = (1/n_i) \sum_{j=1}^{n_i} Y_{ij}$.

Jede transformierte Variable Y_{ij} entstammt praktisch einer Normalverteilung mit konstanter Varianz $1/4$ und Erwartungswert μ_i : $Y_{ij} \sim \mathcal{N}(\mu_i, 1/4)$,

Tab. 8: Zählergebnisse X_{ij} für Asbest-Fasern von $i = 1, 2, 3$ Laboren an Hand jeweils verschiedener Filter $ij, j = 1, \dots, n_i$ mit $n_1 = 3, n_2 = 5$ und $n_3 = 8$

Labor 1	Labor 2	Labor 3
6	0	1
31	7	6
9	0	0
	13	1
	1	0
	23	
	0	
	9	

Tab. 9: Asbest-Fasern: Transformierte Zählergebnisse, vgl. Tab 8

Y _{ij}	Labor i			
	j	1	2	3
1	2.52488	0.61237	1.17260	
2	5.60134	2.71570	2.52488	
3	3.06186	0.61237	0.61237	
4		3.65718	1.17260	
5		1.17260	0.61237	
6			4.83477	
7			0.61237	
8			3.06186	
y _i	3.72936	1.75404	1.82548	

mit $\mu_i = \sqrt{\lambda_i + 1/8}$, wobei λ_i der Parameter der $Po(\lambda_i)$ -Verteilung ist, die das i-te Labor beobachtet hat. Damit erhalten wir für unsere drei Labore:

$$y_1 \sim \mathcal{N}(\mu_1, 1/12), \quad y_2 \sim \mathcal{N}(\mu_2, 1/20) \text{ und } y_3 \sim \mathcal{N}(\mu_3, 1/32),$$

und somit für die zugehörigen μ_i die 95%-Konfidenzintervalle, vgl. Tab. 9,

$$KI(\mu_1) = 3.72936 \pm 1.96\sqrt{1/12} = [3.16356, 4.29516],$$

$$KI(\mu_2) = 1.75404 \pm 1.96\sqrt{1/20} = [1.31577, 2.19231] \text{ und}$$

$$KI(\mu_3) = 1.82548 \pm 1.96\sqrt{1/32} = [1.47900, 2.17196].$$

Durch Rücktransformieren gemäß: $\lambda_i = \mu_i^2 - 1/8$, ergeben sich daraus die 95%-Konfidenzintervalle für λ_i zu:

$$KI(\lambda_1) = [9.88, 18.32], \quad KI(\lambda_2) = [1.61, 4.68], \quad KI(\lambda_3) = [2.06, 4.59].$$

Tab. 10: Asbest-Fasern: Meta-Analyse im Fixed Effects Modell, vgl. Tab. 9

Labor					
i	y_i	$v_i = 1/\sigma_i^2$	$d_i = v_i/v_{\Sigma}$	$d_i y_i$	$d_i y_i^2$
1	3.72936	12	0.1875	0.69926	2.60777
2	1.75404	20	0.3125	0.54814	0.96146
3	1.82548	32	0.5	0.91274	1.66619
$\sum:$	7.30888	64	1	2.16014	5.23542
$\hat{\mu}_{FE} = 2.16014, v_{\Sigma} = 64, Q_C = \sum v_i y_i^2 - v_{\Sigma} \hat{\mu}^2 = 36.42977, \sum v_i^2 = 1568$					

Das erste Labor liegt deutlich signifikant über dem Grenzwert 5, während die beiden anderen Labore signifikant unter dem Grenzwert bleiben. Während λ_2 und λ_3 als homogen angesehen werden können, ist λ_1 doch deutlich verschieden von den beiden. Die beste Erklärung dafür wäre, wenn die Labore auf verschiedenen Schiffen gemessen hätten, aber dem war nicht so, sondern die Ursachen liegen woanders, wie oben näher erläutert.

Müssen wir die zugrunde liegenden Parameter doch als identisch betrachten, da sie zum selben Schiff gehören, so können wir eine Meta-Analyse im Fixed Effects Modell ($K = 3$) durchführen.

In **Tab. 10** sind die dafür benötigten Berechnungen zusammengestellt. Wir erhalten als kombiniertes 95%-Konfidenzintervall für das gemeinsame μ :

$$KI(\mu) = 2.16014 \pm 1.96 \sqrt{1/64} = [1.91566, 2.40514],$$

und daraus das kombinierte 95%-Konfidenzintervall für den gemeinsamen Parameter λ der dem Schiff eigenen Poissonverteilung:

$$KI(\lambda) = [3.55, 5.60],$$

welches nicht unterhalb des Grenzwertes 5 liegt und somit das Schiff nicht signifikant entlastet.

Die bereits aufgefallene Diskrepanz zwischen den Laboren zeigt sich auch im Cochran'schen Homogenitätstest, vgl. Tab. 10 sowie Tab. 4 im Anhang, $Q_C = 36.42 > 5.991 = \chi^2_{2,0.95}$. Dies legt insgesamt nahe, die Inhomogenität zwischen den Laboren durch einen zusätzlichen Modellparameter zu berücksichtigen, d. h. eine Meta-Analyse im Random Effects Modell durchzuführen. Der für solche Fälle vorgesehene Inhomogenitätsparameter τ^2 wird nach Abschnitt 3.1 stets positiv geschätzt mittels des Schätzers, vgl. Tab. 10,

$$\hat{\tau}_{HAR}^2 = \frac{v_{\Sigma} Q_C^2}{(v_{\Sigma}^2 - \sum v_i^2) [Q_C + 2(K - 1)]} = 0.831.$$

Die für eine Auswertung im RE-Modell benötigten Berechnungen sind in **Tab. 11** zusammengestellt. Die kombinierte Schätzung für μ fällt größer

Tab. 11: Asbest-Fasern: Meta-Analyse im Random Effects Modell, vgl. Tab. 10, mit $\hat{\tau}^2 = \hat{\tau}_{\text{HAR}}^2 = 0.831$

Labor					
i	$\hat{\tau}^2 + \sigma_i^2$	$w_i = (\hat{\tau}^2 + \sigma_i^2)^{-1}$	$c_i = w_i/w_{\Sigma}$	$c_i y_i$	$c_i y_i^2$
1	0.91433	1.09369	0.32276	1.20370	4.48903
2	0.881	1.13507	0.33498	0.58756	1.03060
3	0.8625	1.15976	0.34226	0.62479	1.14054
$\sum:$	-	3.38852	1	2.41605	6.66017
$\hat{\mu}_{\text{RE}} = 2.41605, w_{\Sigma} = 3.38852, Q_H = \sum w_i y_i^2 - w_{\Sigma} \hat{\mu}_{\text{RE}}^2 = 2.78832$					

aus als im FE-Modell und damit auch für den gemeinsamen Parameter $\lambda = \mu^2 - 1/8$:

$$\hat{\mu}_{\text{RE}} = 2.41605 \text{ und } \hat{\lambda}_{\text{RE}} = 5.712.$$

Die kombinierten 95%-Konfidenzintervalle für μ , vgl. Abschnitt 3.2, ergeben sich zu:

$$\begin{aligned} \text{KI}_{\text{RE}}(\mu)_C &= \hat{\mu}_{\text{RE}} \pm 1.96 \sqrt{1/w_{\Sigma}} = [1.35129, 3.48081] \text{ und} \\ \text{KI}_{\text{RE}}(\mu)_H &= \hat{\mu}_{\text{RE}} \pm 1.96 \sqrt{Q_H/(K-1)w_{\Sigma}} = [1.15884, 3.67326], \end{aligned}$$

wobei wir hier der praktischen Relevanz wegen das t-Quantil gegen das Normal-Quantil ausgetauscht haben. Mittels Rücktransformation erhalten wir schließlich die kombinierten 95%-Konfidenzintervalle für den gemeinsamen Parameter λ der schiffseigenen $\text{Po}(\lambda)$ -Verteilung wie folgt:

$$\text{KI}_{\text{RE}}(\lambda)_C = [1.701, 11.991] \text{ und } \text{KI}_{\text{RE}}(\lambda)_H = [1.218, 13.368],$$

was angesichts der extremen Zählergebnisse eine durchaus realistische Einschätzung der Asbestverhältnisse auf dem verdächtigen Schiff widerspiegelt. Im einführenden Abschnitt 1.8 sind in der dort befindlichen **Abb. 11** die Schätzer und 95%-Konfidenzintervalle für λ sowohl der drei Labore als auch der Meta-Analyse [KI_C] in beiden Modellen graphisch veranschaulicht.

5. Überlebenszeiten und Time-to-Event-Daten

Beschreibt für eine Person oder ein Objekt die Zufallsvariable X die Lebensdauer oder die Zeit bis zum Eintreten eines interessierenden Ereignisses bzw. Events und ist $F_X(t) = P(X \leq t)$ die zugehörige Verteilungsfunktion mit $F_X(0) = 0$, so ist $F_X(t)$ berechenbar, wenn man die zugehörige Ausfallrate $r(t)$ oder Hazardrate $h(t)$ kennt, wobei

$$h(t) = r(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t < X \leq t + \Delta t \mid X > t)$$

in etwa die Wahrscheinlichkeit angibt, im nächsten Zeitabschnitt $\Delta t = 1$ das interessierende Ereignis zu beobachten, vorausgesetzt, es ist bis jetzt zur Zeit t noch nicht eingetreten. Zu Beginn des Abschnitts 4 in Kap. IV wird ausführlich die Beziehung zwischen Hazardrate und Verteilungsfunktion dargelegt, wie auch in Abschnitt 2 von Kap. XIII, wo speziell noch $\bar{F}_X(t) = 1 - F_X(t)$ zur Beschreibung der **Überlebenswahrscheinlichkeit** betrachtet wird, die speziell auch **Survival-Function** genannt wird und entsprechend geschrieben wird als:

$$S(t) = S_X(t) = P(X > t), \text{ wobei } S(0) = 1 \text{ und } S(\infty) = 0.$$

5.1. Tests auf Gleichheit zweier Survival-Funktionen

In zwei unabhängigen Gruppen: Experimentell E und Kontrolle C, werde die Wahrscheinlichkeit, das Eintreten des interessierenden Ereignisses später als t zu erleben, beschrieben durch die Überlebensfunktionen $S_E(t)$ und $S_C(t)$. Wir testen die Hypothesen:

$$H_0 : S_E(t) = S_C(t) \text{ gegen } H_1 : S_E(t) \neq S_C(t).$$

Bezeichnen $h_E(t)$ und $h_C(t)$ die zugehörigen Hazardraten, dann gibt es eine Funktion $\theta(t)$ derart, daß etwa gilt:

$$h_C(t) = \theta(t)h_E(t).$$

Die Funktion $\theta(t)$ wird Hazard Ratio genannt und die nachfolgend aufgeführten Tests sind am wirkungsvollsten, aber keinesfalls darauf beschränkt, wenn dieser Hazard Ratio konstant ist, d. h. wenn gilt:

$$\theta(t) = \frac{h_C(t)}{h_E(t)} = \theta.$$

In diesem Fall logarithmieren wir θ und setzen, auch zur Meta-Analyse solcher Studien,

$$\psi = \ln \theta.$$

Konfidenzintervalle werden für ψ hergeleitet, die dann durch Rücktransformation der Grenzen in Konfidenzintervalle für θ umgewandelt werden können. Mit den Ausführungen in Abschnitt 4 von Kap. IV gilt hier:

$$S_C(t) = S_E(t)^\theta,$$

so daß für $S_E(t) < 1$ und $\theta > 1$ oder $\psi > 0$ folgt:

$$S_C(t) < S_E(t),$$

womit die folgenden Tests hier auch einseitig angewandt einen Sinn ergeben.

Tab. 12: Daten zum Log-Rank Test im i-ten Intervall I_i

$I_i = [t_{i-1}, t_i)$		
bis t_{i-1} überlebt:	in I_i	in I_i verschwunden:
at risk	verstorben	zensiert
r_{iE}	d_{iE}	c_{iE}
r_{iC}	d_{iC}	c_{iC}
$\sum: r_i$	$\sum: d_i$	$\sum: c_i$

5.1.1. Der Log-Rank Test

Der gebräuchlichste Test auf Gleichheit zweier Survival-Funktionen ist der auf Mantel (1966) und Peto/Peto (1972) zurückgehende sogenannte Log-Rank Test. Wir erklären ihn mittels der Begriffe aus der Überlebenszeitanalyse, obwohl er bei allen zeitlichen Events eingesetzt wird.

Werden von der Zeit $t = 0$, Behandlungsbeginn, an bis zum Studienende, maximal nach einer Zeit T seit Studienbeginn, die Patienten beobachtet und sind zum Zeitpunkt der Studienauswertung d Patienten verstorben, so können wir im Nachhinein die Zeitspanne $[0, T]$ derart in d disjunkte Intervalle zerlegen, daß jedes Intervall genau einen Todeszeitpunkt enthält, falls diese alle genau bekannt und verschieden sind (keine „ties“). Sonst erhalten wir Intervalle mit mehreren Todesfällen zum selben Zeitpunkt.

Andererseits gibt es Anwendungen, s. Beispiel 5.1.3, in denen die Intervalleinteilung schon dadurch vorgegeben ist, daß nur an bestimmten Untersuchungsterminen das Vorliegen des interessierenden Ereignisses festgestellt werden kann und so zwangsläufig mit mehreren Ereignissen zu diesem Termin zu rechnen ist. Wir gehen also von k Intervallen I_i aus, die jeweils d_i Todeszeitpunkte enthalten, sonst werden welche vereinigt, so daß $d = \sum_{i=1}^k d_i$ ist und gilt:

$$[0, T] \subset \{I_1 \cup \dots \cup I_k\}, \quad I_i \cap I_j = \emptyset \text{ für } i \neq j.$$

Im i-ten Intervall $I_i = [t_{i-1}, t_i)$ kommen r_i Patienten lebend an, stehen also „at risk“, davon r_{iE} in der Gruppe E und r_{iC} in der Gruppe C. Davon sterben im Intervall I_i d_i Patienten, aufgeteilt in d_{iE} in E und d_{iC} in C. Es fehlen dann noch als Differenz zwischen r_i und d_i die sogenannten zensierten Patienten c_i , aufgeteilt nach c_{iE} in E und c_{iC} in C, die aus der Studie verschwunden sind und von denen wir nicht wissen, ob sie noch leben oder nicht. In **Tab. 12** haben wir diesen Sachverhalt kurz zusammengestellt.

Dabei gehen wir davon aus, daß die Zensierungen zufällig sind und nicht in irgend einem Zusammenhang mit der Behandlung stehen, d. h. die Häufigkeiten der zensierten Beobachtungen korrelieren nicht mit deren Zugehörigkeit zu Gruppe E oder C.

Beschreibt nun etwa die Zufallsvariable X , mit Realisation d_{iC} , die Anzahl an Todesfällen in der Gruppe C im Intervall I_i , so ist X hypergeometrisch verteilt, d. h. $X \sim H(N, M, n)$, wobei im Urnenmodell mit $N = r_i$ Kugeln $M =$

r_{iC} Kugeln rot sind und wir beim $n = d_i$ -maligen Ziehen ohne Zurücklegen die Anzahl X der dabei gezogenen roten Kugeln beobachten. Aus Abschnitt 3.2 in Kap. IV entnehmen wir, leicht umgeformt, für den Erwartungswert und die Varianz von X :

$$EX = \frac{Mn}{N} \quad \text{und} \quad \text{Var } X = \frac{N-n}{N-1}n \left(\frac{N-M}{N} \right) \frac{M}{N}.$$

Setzen wir hier nun unsere Ersatzgrößen ein und beachten, daß $r_i - r_{iC} = r_{iE}$ ist, so erhalten wir, mit später gebrauchten Bezeichnungen,

$$\begin{aligned} e_i &= \frac{r_{iC}d_i}{r_i} = EX, \quad \text{und} \\ g_i &= \frac{r_{iE}r_{iC}d_i(r_i - d_i)}{(r_i - 1)r_i^2} = \text{Var } X. \end{aligned}$$

Hiermit setzen wir

$$z_i = d_{iC} - e_i \quad \text{und} \quad G = \sum_{i=1}^k g_i$$

und erhalten damit die **Log-Rank Statistik**:

$$Z_{LR} = \sum_{i=1}^k \frac{z_i}{\sqrt{G}},$$

die bei Gültigkeit von H_0 , d.h. wenn $S_E(t) = S_C(t)$ ist, genähert nach $\mathcal{N}(0, 1)$ verteilt ist, so daß mit deren $(1 - \alpha/2)$ -Quantil $u_{1-\alpha/2}$ wir zum Niveau α :

$$H_0 \text{ verwerfen, wenn } |Z_{LR}| > u_{1-\alpha/2}, \quad \text{bzw. } Z_{LR}^2 > \chi^2_{1,1-\alpha},$$

wobei $\chi^2_{1,1-\alpha}$ das $(1 - \alpha)$ -Quantil der χ^2 -Verteilung mit einem Freiheitsgrad und $F_{\chi^2_1}$ deren Verteilungsfunktion bezeichnen. Der „zweiseitige“ p-Wert oder Level attained, vgl. Abschnitt 4 in Kap. III und den noch folgenden Abschnitt 9, von Z_{LR} ist dann $P = 1 - F_{\chi^2_1}(Z_{LR}^2)$, d. h. $\chi^2_{1,1-P} = Z_{LR}^2$.

Ab jetzt nehmen wir einen **konstanten Hazardquotienten θ** an. Dann erhalten wir mit $\psi = \ln \theta$ approximativ:

$$Z_{LR} \sim \mathcal{N}(\psi\sqrt{G}, 1), \quad \text{bzw.}$$

$$Z_{LR}^* = \sum_{i=1}^k \frac{z_i}{G} \sim \mathcal{N}(\psi, 1/G),$$

woraus sich ein $(1 - \alpha)$ -Konfidenzintervall und Schätzer $\hat{\psi}$ für ψ ergeben zu:

$$KI(\psi) = Z_{LR}^* \pm u_{1-\alpha/2}\sqrt{1/G} \quad \text{und} \quad \hat{\psi} = Z_{LR}^*.$$

Durch Entlogarithmieren, d. h. $\exp(\cdot)$, der Grenzen erhält man ein entsprechendes Intervall nebst Schätzer für den Hazardquotienten θ .

5.1.2. Der Wilcoxon-Gehan Test und Tarone-Ware Tests

Führt man Gewichte ω_i ein, so kommt man zum **Weighted Log-Rank Test** mittels der Teststatistik

$$Z_\omega = \frac{\sum_{i=1}^k \omega_i z_i}{\left(\sum_{i=1}^k \omega_i^2 g_i \right)^{1/2}}.$$

Mit den Gewichten

$$\omega_i = r_i = r_{iE} + r_{iC}$$

erhält man die von Gehan (1965) vorgeschlagene Erweiterung des Wilcoxon Tests auf den Fall zensierter Daten, d. h. durch zensierte Beobachtungen unvollständige Daten in der Auswertung.

Im Vergleich zum üblichen Log-Rank Test bewertet der Wilcoxon-Gehan Test frühe Unterschiede im Ausfallverhalten der beiden Gruppen E und C stärker.

Mit anderen Gewichten als $\omega_i = 1$, das ist gerade wieder der übliche Log-Rank Test, oder als $\omega_i = r_i$ erhält man Tests nach Tarone/Ware (1977), vgl. auch Klein/Moeschberger (1997), Fleming/Harrington (2005).

5.1.3. Beispiel: Progressionsfreies Überleben bei Tumorerkrankung

An einem Tumor erkrankte Patienten, vgl. etwa Wagner et al. (2006), erhalten zunächst radiologische und chemische Behandlungen bis die Weiterentwicklung des Tumors gestoppt ist. Mit einem neuen Präparat E und einem Standardpräparat C wird medikamentös weiterbehandelt. Neben der zweiten Zielgröße, der Überlebenszeit, interessiert als erste Zielvariable die progressionsfreie Zeit nach Beginn der medikamentösen Behandlung: **Progression Free Survival PFS**.

Da die Untersuchungen zur Ermittlung einer gegebenenfalls vorliegenden Progression des Tumors aufwendig ist und zudem vom Patienten als unangenehm wahrgenommen wird, sind nur zwei Untersuchungstermine geplant, nach 3 Monaten und nach 6 Monaten, vom jeweiligen Beginn der medikamentösen Behandlung an gerechnet. So auf etwa ± 1 Woche können die beiden Termine auch eingehalten werden. Mit den genauen Zeiten zu rechnen gibt wenig Sinn, denn eine im 3. Monat erkannte Progression kann schon nach einem Monat vorgelegen haben, aber zu diesem Zeitpunkt fand keine Untersuchung statt. Wir legen hier also zwei Intervalle an,

$$I_1 = [0 \text{ bis } 4 \text{ Monate}) \text{ und } I_2 = [4 \text{ bis } 7 \text{ Monate}),$$

so daß die Untersuchungen nach 3 und nach 6 Monaten jeweils in eines der Intervalle fallen.

In **Tab. 13** sind zum Einen die Ergebnisse der Studie in Bezug auf PFS angegeben und zum Anderen auch alle Hilfsgrößen zur Berechnung der Log-Rank Statistik Z_{LR} sowie der Wilcoxon-Gehan Statistik $Z_{WG} = Z_\omega$ mit $\omega_i =$

Tab. 13: Progressionsfreies Überleben PFS, Daten zum Beispiel und Hilfsgrößen zum Log-Rank und Wilcoxon-Gehan Test

PFS	I ₁			I ₂		
	0 bis 4 Monate			4 bis 7 Monate		
Variable	r ₁	d ₁	c ₁	r ₂	d ₂	c ₂
E	100	23	7	70	54	10
C	100	77	3	20	6	8
Σ	200	100	10	90	60	18
e _i	$100 \times 100/200 = 50$			$20 \times 60/90 = 13.33$		
z _i	$77 - 50 = 27$			$6 - 13.33 = -7.33$		
g _i	$\frac{100 \times 100 \times 100(200-100)}{199 \times 200^2} = 12.56281$			$\frac{70 \times 20 \times 60(90-60)}{89 \times 90^2} = 3.49563$		
r _i z _i	$200 \times 27 = 5400$			$-7.33 \times 90 = -660$		
r _i ² g _i	502512.4			28314.603		

r_i aufgeführt. Wir berechnen daraus:

$$\sum z_i = 19.66 \quad \text{und} \quad G = \sum g_i = 16.05844, \text{ sowie}$$

$$\sum r_i z_i = 4740 \quad \text{und} \quad \sum r_i^2 g_i = 530827,$$

so daß wir den Log-Rank Testwert erhalten als:

$$Z_{LR} = \frac{\sum z_i}{\sqrt{G}} = 4.906,$$

und der Wilcoxon-Gehan Testwert ergibt sich als:

$$Z_{WG} = \frac{\sum r_i z_i}{\sqrt{\sum r_i^2 g_i}} = 6.506.$$

Beide Werte übertreffen natürlich erheblich den kritischen Wert $1.96 = u_{0.975}$, so daß zum geplanten Niveau von 5% die Nullhypothese H_0 verworfen wird. Der große Unterschied in den festgestellten Progressionen zwischen E und C zum ersten Zeitpunkt erklärt den hohen Wert der Wilcoxon-Gehan Statistik.

Der „zweiseitige“ p-Wert von Z_{LR} etwa ergibt sich hier zu:

$$P = 1 - F_{\chi^2_1}(4.906^2) = 1 - [2 \cdot \Phi(4.906) - 1] = 0.00000093,$$

wobei die Verteilungsfunktion $\Phi(x)$ der $\mathcal{N}(0, 1)$ -Verteilung in Tab. 1 im Anhang bis $x = 3.09$ vertafelt ist und sonst mit der dort angegebenen Approximation nach Hastings berechnet werden kann.

Mit

$$Z_{LR}^* = \frac{\sum z_i}{G} = 1.22428 = \hat{\psi}$$

erhalten wir das 95%-Konfidenzintervall für den logarithmierten Hazardquotienten $\psi = \ln \theta$ als:

$$KI(\psi) = 1.22428 \pm 1.96\sqrt{1/G} = [0.73492, 1.71364],$$

woraus wir das 95%-Konfidenzintervall für θ errechnen als:

$$KI(\theta) = [2.08532, 5.54912], \text{ sowie den Schätzer:}$$

$$\hat{\theta} = e^{1.22428} = 3.40172.$$

Im Sinne der stark größeren PF-Survival-Funktion ist das neue Medikament also sehr erfolgreich. Allerdings sind nach 6 Monaten sowohl in E als auch in C nur noch 6 Patienten progressionsfrei.

In Abschnitt 5.3 berechnen wir die Kaplan-Meier Kurven als Schätzungen für die Survival-Funktionen in diesem Beispiel und kommen zu weiteren Einblicken.

5.2. Kombination von vergleichenden Überlebenszeitstudien

Hier gehen wir davon aus, daß in $j = 1, \dots, K$ unabhängigen Time-to-Event Studien jeweils die beiden unabhängigen Gruppen E und C miteinander verglichen werden.

Liegen nur Angaben zu den p-Werten der jeweiligen Vergleichstests vor oder können wir absolut nicht von einem konstanten Hazardquotienten ausgehen, so werden die K „zweiseitigen“ p-Werte P_1, \dots, P_K mit den Methoden aus Abschnitt 9 kombiniert, etwa mit der Stouffer Methode zum kombinierten p-Wert P_{COMB} , gegeben als:

$$P_{COMB} = 1 - \Phi \left\{ \frac{1}{\sqrt{K}} \sum_{j=1}^K \Phi^{-1}(1 - P_j) \right\}.$$

Liegt dieser unter einem geforderten Signifikanzniveau, so haben wir also durch die Kombination der Studien ein signifikantes Gesamtereignis erzielt.

Liegt ein konstanter Hazardquotient vor, so kombinieren wir dessen K Schätzungen.

5.2.1. Kombination der Hazard Ratios

Liegen wie in Abschnitt 5.1.1 transformierte Log-Rank Statistiken $Z_{LR,j}^*$ für jede Studie vor, bzw. können berechnet werden, so setzen wir:

$$y_j = Z_{LR,j}^* = Z_{LR,j} / \sqrt{G_j},$$

und erhalten bei konstantem Hazardquotienten θ mit $\psi = \ln \theta$ das approximative Modell:

$$y_j \sim \mathcal{N}(\psi, \tau^2 + \sigma_j^2) \text{ mit } \hat{\sigma}_j^2 = 1/G_j.$$

Setzt man $\tau^2 = 0$, so können die Methoden des Fixed Effects Modells aus Abschnitt 2 angewandt werden. Andernfalls geschieht die Meta-Analyse im Random Effects Modell gemäß Abschnitt 3.

5.2.2. Beispiel Progressionsfreies Überleben PFS [3 Studien]: Kombiniertes Konfidenzintervall für den Hazard Ratio

Von $K = 3$ jeweils die Gruppen E und C vergleichenden PF-Überlebenszeitstudien sind die Werte der Log-Rank Statistiken $Z_{LR,j}$ und der Varianzschätzungen G_j wie folgt gegeben:

$$\begin{aligned} Z_{LR,1} &= 4.880 \text{ und } G_1 = 16, \\ Z_{LR,2} &= 1.587 \text{ und } G_2 = 4.726, \\ Z_{LR,3} &= 2.648 \text{ und } G_3 = 7.306. \end{aligned}$$

Daraus ergeben sich y_j und $\hat{\sigma}_j^2$ zu:

$$\begin{aligned} y_1 &= 1.22 \text{ und } \hat{\sigma}_1^2 = 0.0625, \\ y_2 &= 0.73 \text{ und } \hat{\sigma}_2^2 = 0.2116, \\ y_3 &= 0.98 \text{ und } \hat{\sigma}_3^2 = 0.1369. \end{aligned}$$

In Tab. 14 sind einige Rechen-Hilfsgrößen zusammengestellt. Im Fixed Effects Modell erhalten wir daraus den gemeinsamen Schätzer für $\psi = \ln \theta$ als:

$$\hat{\psi}_{FE} = 1.0746.$$

Das zugehörige 95%-Konfidenzintervall ist:

$$KI_{FE}(\psi) = \hat{\psi}_{FE} \pm 1.96 \sqrt{1/v_\Sigma} = [0.712, 1.437].$$

Mit $\bar{y} = 0.97667$, $\sum(y_j - \bar{y})^2/2 = 0.0604$, $\sum \hat{\sigma}_j^2 = 0.411$ erhalten wir für den Hedges Schätzer von τ^2 den Wert, vgl. Abschnitt 3.1,

$$\hat{\tau}_{HED}^2 = 0.0604 - 0.411/3 = -0.1974.$$

Tab. 14: Beispiel kombinierter Hazard Ratio, Hilfsgrößen im FE-Modell

Studie j	$v_j = 1/\hat{\sigma}_j^2$	$d_j^* = v_j/v_\Sigma$	$d_j^* y_j$	$d_j^* y_j^2$
1	16	0.57	0.6954	23.8144
2	4.726	0.169	0.1234	2.5185
3	7.306	0.261	0.2558	7.0157
$\Sigma:$	28.031	1	1.0746	33.3486

Mit $Q_C = \sum v_j y_j^2 - (\sum v_j y_j)^2 / v_\Sigma = 33.3486 - 907.34/28.031 = 0.97667$ und $\sum v_j^2 = 331.698$ ergibt sich für den DerSimonian-Laird Schätzer der Wert:

$$\hat{\tau}_{DSL}^2 = (Q_C - 2) \cdot \frac{v_\Sigma}{v_\Sigma^2 - \sum v_j^2} = (Q_C - 2) \cdot 0.062 = -0.0612.$$

Der positive Schätzer für τ^2 ergibt sich zu:

$$\hat{\tau}_{HAR}^2 = \frac{v_\Sigma \cdot Q_C^2}{(v_\Sigma^2 - \sum v_j^2) [Q_C + 4]} = 0.01156.$$

Damit berechnen wir die neuen Gewichte:

$$w_j = 1/(\hat{\tau}_{HAR}^2 + \hat{\sigma}_j^2), \quad w_\Sigma = \sum w_j = 24.71992, \quad c_j^* = w_j/w_\Sigma,$$

und im Random Effects Modell erhalten wir den gemeinsamen Schätzer für ψ als:

$$\hat{\psi}_{RE} = \sum c_j^* y_j = 1.06355,$$

mit dem zugehörigen 95%-Konfidenzintervall für $\psi = \ln \theta$:

$$KI_{RE}(\psi) = \hat{\psi}_{RE} \pm 1.96 \sqrt{1/w_\Sigma} = [0.677, 1.450],$$

das erwartungsgemäß etwas breiter ausfällt. Im einleitenden Abschnitt 1.2 sind in der dort befindlichen **Abb. 3** die Hazard Ratio Schätzer $\hat{\theta}_i = \exp(y_i)$ der drei Studien, $\hat{\theta}_{FE} = \exp(\hat{\psi}_{FE})$ sowie $\hat{\theta}_{RE} = \exp(\hat{\psi}_{RE})$ der Meta-Analyse in den beiden Modellen und jeweils zugehörige 95%-Konfidenzintervalle grafisch veranschaulicht.

5.3. Schätzen der Überlebenswahrscheinlichkeiten und Konfidenzbereiche

Hat man eine Studie zu Überlebenszeiten durchgeführt, so interessieren, auch im Vergleich zu anderen Studien, die Überlebenswahrscheinlichkeiten zu markanten Zeitpunkten, etwa nach 3, 6 oder 12 Monaten seit Behandlungsbeginn, und zur Einschätzung der Zuverlässigkeit wird dazu die Angabe eines Konfidenzbereichs benötigt.

5.3.1. Kaplan-Meier Schätzer der Survival-Funktion und Bestimmung eines Konfidenzstreifens

Wir teilen die nichtnegative reelle Halbachse in m disjunkte Intervalle J_i auf, die nicht mit den Intervallen I_i zum Log-Rank Test irgendwo übereinstimmen müssen, vgl. nachfolgendes Beispiel,

$$[0, \infty) = J_1 \cup \dots \cup J_m, \quad J_i \cap J_j = \emptyset \text{ für } i \neq j.$$

Bezeichnen wie beim Log-Rank Test r_i die Anzahl der das Intervall J_i lebend erreichen Personen und d_i die Anzahl der im Intervall J_i verstorbenen Personen, dann berechnen wir für jedes Intervall J_i die Größe $s_i = 1 - d_i/r_i$ und durch sukzessive Multiplikation dieser Terme erhalten wir die von Kaplan/Meier (1958) vorgeschlagene Schätzung der Survival Function $S(t)$:

$$\hat{S}(t) = \hat{S}_j = \prod_{i \leq j} s_i, \quad s_i = 1 - \frac{d_i}{r_i}, \text{ für } t \in J_j.$$

Die Varianz von $\hat{S}(t)$ wird geschätzt durch

$$\hat{V}(t) = \hat{V}_j = \hat{S}_j^2 \sum_{i \leq j} f_i, \quad f_i = \frac{d_i}{r_i(r_i - d_i)}, \text{ für } t \in J_j.$$

Approximativ gilt zu einem Zeitpunkt t :

$$\hat{S}(t) \sim \mathcal{N}(S(t), \hat{V}(t)),$$

so daß etwa durch:

$$KI(S(t)) = \hat{S}(t) \pm 1.96 \sqrt{\hat{V}(t)}$$

ein 95%-Konfidenzintervall für die Überlebenswahrscheinlichkeit im Zeitpunkt t gegeben wird. Bilden wir das Konfidenzintervall zu jedem Zeitpunkt t , so ergibt sich ein z.B. 95%-Konfidenzstreifen um $\hat{S}(t)$.

5.3.2. Schätzen der medianen Überlebenszeit und Konfidenzintervall

Sind etwa aus Studiengründen die Intervalle recht groß, s. Beispiel, so bietet sich eine Linearisierung der Treppenfunktion $\hat{S}(t)$ an, wenn man z. B. annimmt, daß die zu einem Zeitpunkt festgestellten Ereignisse etwa gleichverteilt schon vorher aufgetreten sind, aber nicht erkannt werden konnten.

Die Punkte (t_{i-1}, A_{i-1}) und (t_i, A_i) im $t \times A$ -Koordinatenkreuz werden für $t \in [t_{i-1}, t_i]$ verbunden durch die Gerade

$$\phi_i(t | A_{i-1}, A_i) = A_{i-1} + \frac{A_{i-1} - A_i}{t_i - t_{i-1}} t_{i-1} + \frac{A_{i-1} - A_i}{t_i - t_{i-1}} t.$$

Setzt man nun

$$\tilde{S}(t) = \phi_j(t | \hat{S}_{j-1}, \hat{S}_j) \text{ für } t \in J_j,$$

und verstetigen wir entsprechend die Standardabweichungen:

$$\sqrt{\tilde{V}(t)} = \phi_j \left(t \mid \sqrt{\hat{V}_{j-1}}, \sqrt{\hat{V}_j} \right) \text{ für } t \in J_j,$$

so erhält man stetige Modifikationen von $\hat{S}(t)$ und $\sqrt{\hat{V}(t)}$, die in den Eckpunkten mit den ursprünglichen Treppenfunktionen übereinstimmen. Ähnlich verfährt man mitunter auch mit der empirischen Verteilungsfunktion.

Das erleichtert die Bestimmung der stets interessierenden **Median Survival Time** $\hat{T}_{0.5}$ als den Zeitpunkt, an dem gilt:

$$\tilde{S}(\hat{T}_{0.5}) = 0.5.$$

Ein $(1 - \alpha)$ -Konfidenzintervall für die mediane Überlebenszeit $T_{0.5}$, für die bei stetiger Funktion $S(t)$ gilt: $S(T_{0.5}) = 0.5$, ist dann gegeben als:

$$KI(T_{0.5}) = \left\{ t \mid -u_{1-\alpha/2}\sqrt{\tilde{V}(t)} \leq \tilde{S}(t) - 0.5 \leq u_{1-\alpha/2}\sqrt{\tilde{V}(t)} \right\}.$$

Entsprechende Aussagen kann man zu anderen Quantilen als dem Median ableiten. [Vgl. auch Elpelt/Hartung (2004).]

5.3.3. Beispiel Progressionsfreies Überleben PFS: Bestimmung der Kaplan-Meier PFS-Kurven und der medianen Überlebenszeiten

Wir setzen hier das Beispiel aus Abschnitt 5.1.3 fort. Jedoch legen wir hier Intervalle J_i wie folgt fest:

$$\begin{aligned} J_1 &= [t_0, t_1), J_2 = [t_1, t_2) \text{ und } J_3 = [t_3, \infty), \text{ wobei} \\ t_0 &= 0 \text{ Monate, } t_1 = 3 \text{ Monate und } t_3 = 6 \text{ Monate.} \end{aligned}$$

Alle Ereignisse aus dem Intervall I_1 in 5.1.3, s. Tab. 13, werden J_2 zugerechnet und aus I_2 zu J_3 , d. h. die Ereignisse werden alle an den jeweils geplanten Untersuchsterminen 3 Monate und 6 Monate nach Beginn der medikamentösen Behandlung gezählt.

In **Tab. 15** sind die Berechnungen der Kaplan-Meier Treppenfunktionen und ihrer Varianzen für die Gruppen E und C dokumentiert, einschließlich der Konfidenzbereiche.

Die Linearisierung ist hier im Beispiel wirklich angebracht, z. B. ist davon auszugehen, daß die nach 6 Monaten eruierten Progressionen etwa gleichverteilt schon in der Zeit zwischen 3. und 6. Monat stattgefunden haben ohne registriert werden zu können. Wir verzichten jedoch auf eine explizite numerische Angabe der linearisierten Funktionen $\tilde{S}_E(t)$, $\tilde{S}_C(t)$ sowie $\tilde{V}_E(t)^{1/2}$

Tab. 15: Progressionsfreies Überleben, s. Tab. 13, Berechnung der Kaplan-Meier Überlebens-Funktionen und 95%-Konfidenzbereiche für die Gruppen E und C

PFS	J ₁	J ₂	J ₃
Variable	[0, t ₁)	[t ₁ = 3 Mon., t ₂)	[t ₂ = 6 Mon., ∞)
s _{iE}	1	1 - 23/100 = 0.77	1 - 54/70 = 0.22857
Ŝ _{iE}	1	0.77	0.176
f _{iE}	0	0.002987	0.0482143
Σ f _{iE}	0	0.002987	0.0512
Ŷ _{iE}	0	0.001771	0.0015897
0.95 - KI	[1, 1]	[0.6875, 0.8525]	[0.0979, 0.2541]
s _{iC}	1	1 - 77/100 = 0.23	1 - 6/20 = 0.70
Ŝ _{iC}	1	0.23	0.161
f _{iC}	0	0.0334783	0.0214286
Σ f _{iC}	0	0.0334783	0.0549069
Ŷ _{iC}	0	0.001771	0.0014232
0.95 - KI	[1, 1]	[0.1475, 0.3125]	[0.0871, 0.2349]

und $\tilde{V}_C(t)^{1/2}$. Vielmehr haben wir sie in **Abb. 17** graphisch veranschaulicht, indem wir dort zu $\tilde{S}_E(t)$ und $\tilde{S}_C(t)$ auch die 95%-Konfidenzstreifen eingezeichnet haben:

$$\tilde{S}_E(t) \pm 1.96\sqrt{\tilde{V}_E(t)} \text{ und } \tilde{S}_C(t) \pm 1.96\sqrt{\tilde{V}_C(t)}.$$

Ebenfalls skizziert ist in der Abbildung die Ermittlung der Schätzungen für die medianen progressionsfreien Zeiten und der zugehörigen 95%-Konfidenzintervalle in den Gruppen E und C. Der Abbildung entnehmen wir:

$$\hat{T}_{0.5}(C) = 59 \text{ Tage und } \hat{T}_{0.5}(E) = 132 \text{ Tage, sowie}$$

$$KI(T_{0.5}(C)) = [53 \text{ Tage}, 66 \text{ Tage}] \text{ und}$$

$$KI(T_{0.5}(E)) = [120 \text{ Tage}, 145 \text{ Tage}].$$

Die exakten Angaben lassen sich dann leicht numerisch ermitteln. Wir erkennen, wie schon beim Log-Rank Test in 5.1.3, einen beträchtlichen Unterschied zwischen beiden Gruppen, hier bzgl. des medianen PFS. Gerechnet haben wir dabei mit 52 Wochen für ein Jahr.

5.4. Kombination von Überlebenszeiten

Häufig ist etwa eine Medikation in einer Reihe von Time-to-Event Studien eingesetzt, durchaus mit wechselnden Vergleichspräparaten. Es interessiert dann jede Art der zusammenfassenden Beurteilung.

Survival Analysis

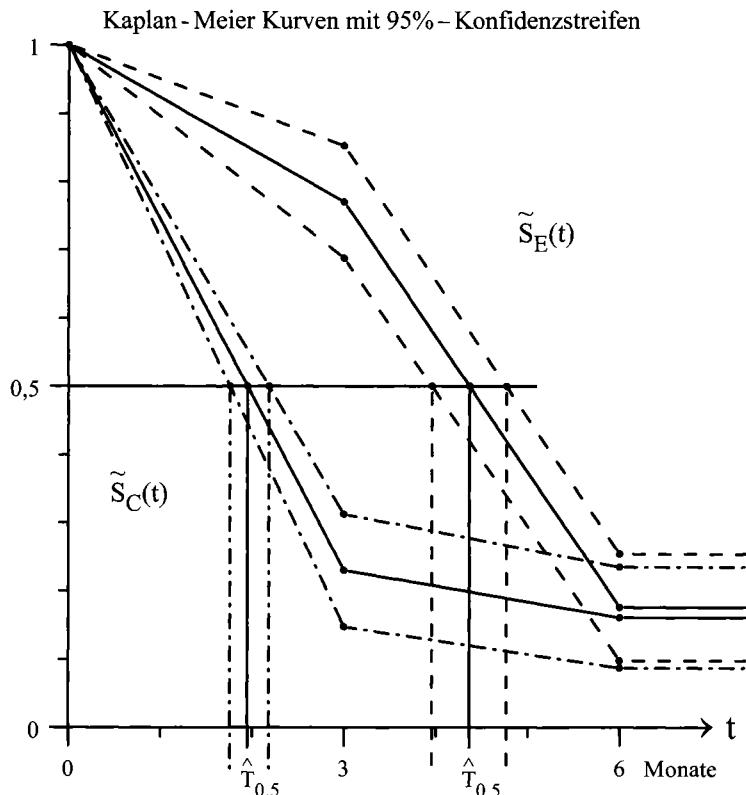


Abb. 17: Tumoren – Progressionsfreies Überleben PFS

5.4.1. Kombination der Kaplan-Meier Überlebenskurven

Von $j = 1, \dots, K$ Time-To-Event Analysen liegen Kaplan-Meier Kurven und zugehörige Varianzfunktionen vor:

$$\hat{S}(t)_1, \hat{V}(t)_1, \dots, \hat{S}(t)_K, \hat{V}(t)_K.$$

Bei gruppierten Daten brauchen die zugehörigen Intervalleinteilungen für die einzelnen Funktionen $\hat{S}(t)_j$ natürlich nicht übereinzustimmen; die Funktionen sind überall für $t \geq 0$ definiert.

Zu jedem festen Zeitpunkt t läßt sich eine Meta-Analyse gemäß eines Fixed Effects oder Random Effects Modells durchführen, vgl. Abschnitt 2 und 3, so daß man dann eine entsprechend kombinierte Überlebenszeitfunktion $\hat{S}(t)_{\text{COMB}}$ mit zugehöriger Varianzfunktion $\hat{V}(t)_{\text{COMB}}$ erhält. Für eine Meta-Analyse im Fixed Effects Modell, zum Beispiel, sieht die Kombination wie

folgt aus. Wir setzen:

$$y_j(t) = \hat{S}(t)_j, \quad \psi(t) = S(t), \quad \hat{\sigma}_j^2(t) = \hat{V}(t)_j$$

und berechnen dann:

$$v_j(t) = 1/\hat{\sigma}_j^2(t), \quad v_{\Sigma}(t) = \sum_{j=1}^K v_j(t) \text{ und } d_j^*(t) = v_j(t)/v_{\Sigma}(t).$$

Es sind dann

$$\hat{S}(t)_{COMB} = \sum_{j=1}^K d_j^*(t) \cdot y_j(t) \text{ und } \hat{V}(t)_{COMB} = \frac{1}{v_{\Sigma}(t)},$$

und der kombinierte $(1 - \alpha)$ -Konfidenzstreifen zu $S(t)$ an jedem Zeitpunkt t ist gegeben als:

$$KI(S(t)) = \hat{S}(t)_{COMB} \pm u_{1-\alpha/2} \sqrt{1/v_{\Sigma}(t)}.$$

Alle übrigen Methoden aus den Abschnitten 2 und 3 lassen sich hier genau so gut anwenden.

5.4.2. Kombination der medianen Überlebenszeiten

Um hier zu einer bequemen Handhabung zu gelangen, wird die kombinierte Kaplan-Meier Kurve $\hat{S}(t)_{COMB}$, s. oben, verstetigt, indem wir die Treppenfunktion stückweise linearisieren, vgl. Abschnitt 5.3.2, d. h. wir bilden die stetige kombinierte Überlebenszeitfunktion:

$$\tilde{S}(t)_{COMB} = \phi_k \left(t \mid \hat{S}(t)_{COMB,k-1}, \hat{S}(t)_{COMB,k} \right)$$

für $t \in [t_{k-1}, t_k]$, wenn $\hat{S}(t)_{COMB}$ in t_{k-1} den Wert $\hat{S}(t)_{COMB,k-1}$ und in t_k den Wert $\hat{S}(t)_{COMB,k}$ hat und sich dazwischen kein weiterer Wert von $\hat{S}(t)_{COMB}$ befindet. Entsprechend verstetigen wir die Standardabweichungen zu $\sqrt{\tilde{V}(t)_{COMB}}$, s. Abschnitt 5.3.2.

Die gemeinsame Schätzung $\hat{T}_{0.5}^{COMB}$ der medianen Überlebenszeit $T_{0.5}$ von $S(t)$ ist dann gegeben als Lösung von:

$$\tilde{S} \left(\hat{T}_{0.5}^{COMB} \right)_{COMB} = 0.5,$$

und das gemeinsame $(1 - \alpha)$ -Konfidenzintervall für $T_{0.5}$ berechnet sich aus, s. oben,

$$KI(T_{0.5})_{COMB} = \left\{ t \mid \left| \tilde{S}(t)_{COMB} - 0.5 \right| \leq u_{1-\alpha/2} \sqrt{\tilde{V}(t)_{COMB}} \right\}.$$

5.4.3. Kombination der Konfidenzintervalle für Überlebenswahrscheinlichkeiten zu markanten Zeitpunkten

In Time-To-Event Analysen werden i. d. R. auch Angaben in Form von Konfidenzintervallen zu Überlebenswahrscheinlichkeiten an markanten Zeitpunkten wie etwa 3, 6 oder 12 Monate nach Beobachtungsbeginn publiziert.

Liegen zu einem identischen Zeitpunkt etwa K 95%-Konfidenzintervalle zur Überlebenswahrscheinlichkeit vor,

$$[L_1, R_1], \dots, [L_K, R_K],$$

so können wir wie folgt rekonstruieren:

$$y_j = L_j + \frac{R_j - L_j}{2} \text{ und } R_j = y_j + 1.96\sqrt{\hat{\sigma}_j^2},$$

woraus sich für die unbekannte Varianz ergibt:

$$\hat{\sigma}_j^2 = \left(\frac{R_j - y_j}{1.96} \right)^2.$$

Eine Kombination z. B. gemäß Fixed Effects Modell mit:

$$v_j = 1/\hat{\sigma}_j^2, \quad v_{\Sigma} = \sum v_j \text{ und } d_j^* = v_j/v_{\Sigma}$$

führt zur gemeinsamen Schätzung $\hat{\psi}$ der Überlebenswahrscheinlichkeit an dem betrachteten Zeitpunkt und dem kombinierten 95%-Konfidenzintervall wie folgt:

$$\hat{\psi} = \sum_{j=1}^K d_j^* y_j \text{ und } KI_{COMB} = \hat{\psi} \pm 1.96\sqrt{1/v_{\Sigma}}.$$

5.5. Beispiel Progressionsfreies Überleben PFS: Kombination der Kaplan-Meier Überlebenskurven, der medianen Überlebenszeiten und der 6-Monats-Konfidenzintervalle

Wir setzen hier das Beispiel aus den Abschnitten 5.1.3 und 5.3.3 fort und werden zu Demonstrationszwecken die beiden Gruppen E und C kombinieren, obwohl der Log-Rank Test in Abschnitt 5.1.3 die Verschiedenheit der beiden Gruppen gezeigt hat.

In Tab. 16 haben wir die Berechnungen der gemäß Fixed Effects Modell kombinierten Kaplan-Meier Überlebenszeitfunktionen der Gruppen E und C zusammengestellt und auch das 95%-Konfidenzband zur kombinierten Überlebenszeitkurve berechnet:

$$\hat{S}(t)_{COMB} = \hat{S}_{i,COMB} \text{ für } t \in J_i, \text{ und}$$

$$KI(S(t))_{COMB} = \hat{S}_{i,COMB} \pm 1.96\sqrt{\hat{V}_{i,COMB}} \text{ für } t \in J_i,$$

 Survival Meta-Analysis

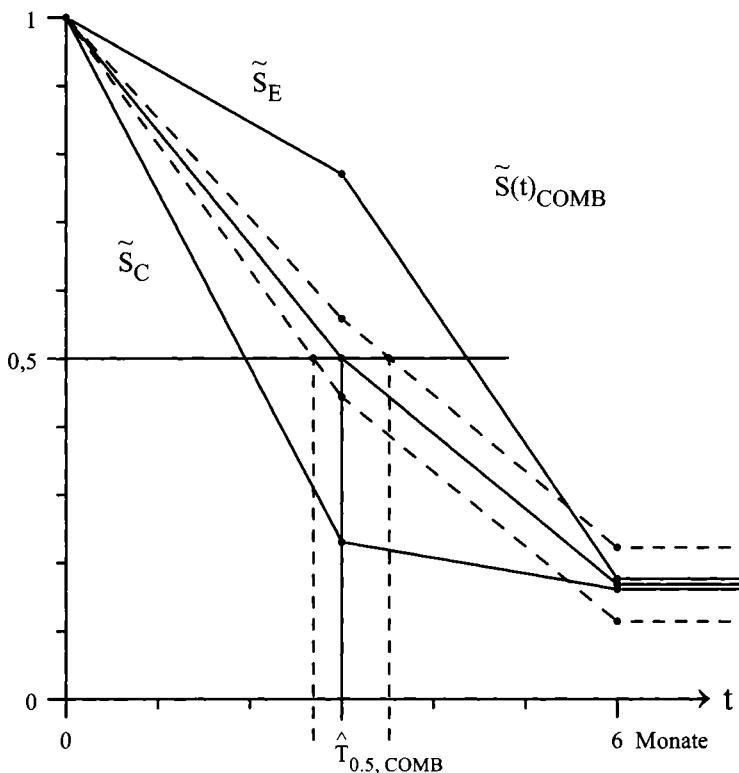


Abb. 18: Combined Kaplan - Meier Survival Function

wobei die Terme $\hat{S}_{i, COMB}$ und $\hat{V}_{i, COMB}$ in Tab. 16 angegeben sind. Man erkennt, daß die kombinierten Konfidenzintervalle erheblich schmäler als die entsprechenden einzelnen Konfidenzintervalle aus Tab. 15 sind.

Die linearisierte kombinierte Überlebenszeitfunktion $\tilde{S}(t)_{COMB}$ und das linearisierte 95%-Konfidenzband sind in **Abb. 18** graphisch veranschaulicht.

Ebenfalls skizziert ist darin die Ermittlung des gemeinsamen Schätzers $\hat{T}_{0.5}^{COMB}$ für die mediane Überlebenszeit sowie zugehörigem 95%-Konfidenzintervall. Der Zeichnung können wir direkt entnehmen:

$$\hat{T}_{0.5}^{COMB} = 91 \text{ Tage, und } KI(T_{0.5})_{COMB} = [82 \text{ Tage}, 107 \text{ Tage}].$$

Die genauen Werte lassen sich leicht rechnerisch ermitteln.

Für die Wahrscheinlichkeit, 6 Monate progressionsfrei zu überleben, werden aus Tab. 15 die 95%-Konfidenzintervalle:

$$[0.098, 0.254] \text{ und } [0.087, 0.235]$$

Tab. 16: Progressionsfreies Überleben, s. Tab. 15, Berechnung der kombinierten Kaplan-Meier Überlebenszeitkurven

PFS	J ₁	J ₂	J ₃
Variable	[0, t ₁]	[t ₁ = 3 Mon., t ₂)	[t ₂ = 6 Mon., ∞)
v _{iE}	0	564.65274	629.04951
v _{iC}	0	564.65274	702.64193
v _{iΣ}	0	1129.30548	1331.69144
d _{iE} *	0	0.5	0.47237
d _{iC} *	0	0.5	0.52763
Ŝ _{iCOMB}	1	0.5	0.168
Ŷ _{iCOMB}	0	0.000855	0.000751
95%-KI	[1,1]	[0.443, 0.558]	[0.114, 0.222]
$v_{iE} = 1/\hat{V}_{iE}$, $v_{iC} = 1/\hat{V}_{iC}$, $v_{iΣ} = v_{iE} + v_{iC}$, $d_{iE}^* = v_{iE}/v_{iΣ}$, $d_{iC}^* = v_{iC}/v_{iΣ}$, $\hat{S}_{iCOMB} = d_{iE}^* \hat{S}_{iE} + d_{iC}^* \hat{S}_{iC}$, $\hat{V}_{iCOMB} = 1/v_{iΣ}$			

der halben Breiten 0.078 und 0.0735 berichtet. Wir rekonstruieren:

$$\begin{aligned} y_1 &= 0.176, \quad \hat{\sigma}_1^2 = [(0.254 - 0.176)/1.96]^2 = 0.00158, \\ y_2 &= 0.161, \quad \hat{\sigma}_2^2 = [(0.235 - 0.161)/1.96]^2 = 0.001425 \text{ und} \\ v_1 &= 1/\hat{\sigma}_1^2 = 632.911, \quad v_2 = 701.534, v_{iΣ} = 1334.445, \\ d_1^* &= v_1/v_{iΣ} = 0.474, \quad d_2^* = 0.526, \end{aligned}$$

so daß der kombinierte Schätzer $\hat{\psi}_{COMB}$ für die Wahrscheinlichkeit ψ , 6 Monate progressionsfrei zu überleben, sich ergibt zu: $\hat{\psi}_{COMB} = \sum d_j^* y_j = 0.16806$. Das zugehörige kombinierte 95%-Konfidenzintervall errechnet sich zu:

$$KI_{COMB} = \hat{\psi}_{COMB} \pm 1.96\sqrt{1/v_{iΣ}} = [0.114, 0.222],$$

dessen halbe Breite mit 0.054 um ca. 30% geringer ist als bei den einzelnen Intervallen. Das rekonstruierte kombinierte Intervall stimmt mit dem auf normalem Weg berechneten Intervall in Tab. 16 überein, was es ja auch muß.

5.6. Cox's Proportional Hazards Regression Model: Vergleich mehrerer Gruppen, Kovariablen und Kombination der Effekte

Bezeichnen $h_0(t)$ eine Basis Hazardrate und β_1, \dots, β_k sowie x_1, \dots, x_k weitere Modellparameter, so lautet das von Cox (1972) eingeführte Proportional Hazards Regression Model für die Hazardrate des i-ten Objekts:

$$h_i(t) = h_0(t) \cdot e^{\sum_{j=1}^k \beta_j x_{ji}}, \quad t > 0.$$

Wie in der Regressionsanalyse, vgl. Kap. X, stehen x_1, \dots, x_k für vorliegende Einflußfaktoren oder Kovariablen mit Werten x_{1i}, \dots, x_{ki} beim i-ten Objekt und die Koeffizienten β_1, \dots, β_k sind die Regressionskoeffizienten.

Wie in der Varianzanalyse, vgl. Kap. XI, lassen sich mittels Indikatorvariablen auch mehr als zwei Behandlungen innerhalb derselben Studie modellieren.

Beispiel: Wir betrachten den Fall, daß 3 Behandlungen zugleich verglichen werden. Als Kovariable berücksichtigen wir das Geschlecht und als metrische Variable die Größe, was einem Modell der Kovarianzanalyse entspricht. Die Behandlung 1 und Geschlecht 1 erhalten keine eigene Variable, um eine **Überparametrisierung** des Modells zu vermeiden. Sie werden dadurch berücksichtigt, daß im Falle ihres Vorliegens die Gegenvariablen, für Behandlung 2 oder 3 und für Geschlecht 2, auf 0 gesetzt sind, d. h. wir setzen in unserem Beispiel für das i-te Objekt:

$$\begin{aligned}x_{1i} &= 1 && \text{falls Behandlung 2 vorliegt, sonst } x_{1i} = 0, \\x_{2i} &= 1 && \text{falls Behandlung 3 vorliegt, sonst } x_{2i} = 0, \\x_{3i} &= 1 && \text{falls Geschlecht 2 vorliegt, sonst } x_{3i} = 0, \\x_{4i} &= 174 && \text{falls das i-te Objekt 174 cm groß ist.}\end{aligned}$$

Durch die Maximierung der partiellen Likelihood werden die Koeffizienten geschätzt, vgl. Cox (1972), Klein/Moeschberger (1997).

Wie in der sonstigen Regressionsanalyse besteht auch hier die Gefahr der **Vermengung** oder des **Confounding**, d. h. dadurch, daß i. d. R. Kovariablen nicht in balancierter Form auftreten, überlagern sich die Einflußfaktoren, wodurch schon mancher Behandlungserfolg unentdeckt blieb, vgl. auch die Ausführungen zur allgemeinen **Versuchsplanung** in Abschnitt 5 vom Kap. V, die natürlich auch hier zu beachten sind.

Andererseits besteht natürlich die Möglichkeit, die Behandlungseffekte rein zu schätzen, d. h. bereinigt von den Einflüssen der Kovariablen.

Taucht in K derartigen mit Placebo C kontrollierten Studien unter den Behandlungen stets eine interessierende Behandlung E^* auf und ist der Behandlungseffekt y_i von E^* im Vergleich zu C in der i-ten Studie sauber geschätzt, so können wir y_i, \dots, y_K **meta-analytisch** kombinieren zu einem Gesamtschätzer für den Effekt von E^* , bereinigt von den Einflüssen der Kovariablen. Zur Verfügung stehen dann auch Random Effects Verfahren, um eine eventuell vorhandene Variabilität zwischen den Studien zu berücksichtigen.

6. Generalized Common Mean (GCM) Problem und Varianzschätzung bei Unkenntnis der Einzelvarianzen

Das Verallgemeinerte Common Mean Problem besteht darin, für einen kombinierten Schätzer oder eine kombinierte Prognose die Varianz zu bestimmen, ohne dabei auf zuverlässige Varianzangaben zu den einzelnen Studien oder Prognosen zurückgreifen zu können.

6.1. Das GCM-Problem

Formal lautet das Generalized Common Mean Problem (GCM) für unabhängige y_1, \dots, y_K wie folgt:

$$E y_i = \psi, \quad \text{Var } y_i = \xi_i, \text{ keine zuverlässige Information über } \xi_i.$$

Sind die Varianzen bekannt oder liegen unverzerrte Schätzungen für ξ_i vor, so spricht man vom gewöhnlichen Common Mean Problem, welches dann allerdings identisch mit dem FE-Modell in Abschnitt 2 ist.

Mit einer irgendwie zustande gekommenen Gewichtung oder Bewertung b_1, \dots, b_K der Einzelergebnisse y_1, \dots, y_K wird das gemeinsame Mittel ψ geschätzt oder prognostiziert gemäß:

$$\hat{\psi}_b = \sum_{i=1}^K b_i y_i, \quad b_i > 0, \quad \sum_{i=1}^K b_i = 1, \quad \text{mit } \widehat{\text{Var}} \hat{\psi}_b = \sum_{i=1}^K b_i^2 \cdot \widehat{\text{Var}} y_i$$

und dem $(1 - \alpha)$ -Konfidenzintervall für ψ :

$$\text{KI}(\psi_b)_0 = \hat{\psi}_b \pm \sqrt{\widehat{\text{Var}} \hat{\psi}_b} \cdot u_{1-\alpha/2},$$

wenn $\widehat{\text{Var}} y_i$ bekannt und unabhängig voneinander sind, wobei $u_{1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil der $\mathcal{N}(0, 1)$ -Verteilung bezeichnet. Sind keine $\widehat{\text{Var}} y_i$ bekannt, so kann mit Hilfe der folgenden Schätzformel $\text{Var}(\hat{\psi}_b)$ direkt geschätzt werden.

6.2. Direkter Varianzschatz von Hartung-Böckenhoff-Knapp

Akzeptiert man bei der freien Wahl der Gewichte b_i die Einschränkung, daß b_i nicht als 0.5 gewählt wird, so kann man die Varianz von $\hat{\psi}_b$ unverzerrt schätzen mittels:

$$\begin{aligned} \widehat{\text{Var}}_{\text{HBK}} \hat{\psi}_b &= \frac{1}{1 + \sum_{j=1}^K \frac{b_j^2}{1-2b_j}} \sum_{i=1}^K \frac{b_i^2}{1-2b_i} \\ &\cdot \left(y_i - \sum_{k=1}^K b_k y_k \right)^2, \quad b_i > 0, \quad b_i \neq 1/2, \quad \sum_{i=1}^K b_i = 1. \end{aligned}$$

Es gilt für den Erwartungswert

$$E \widehat{\text{Var}}_{\text{HBK}} \hat{\psi}_b = \text{Var } \hat{\psi}_b = \sum_{i=1}^K b_i^2 \cdot \widehat{\text{Var}} y_i.$$

Wählt man alle $b_i < 1/2$, so ist die **Varianzschatzung stets positiv**, wenn die y_i nicht gerade alle identisch ausfallen. Das ist natürlich nur für

$K \geq 3$ möglich. Wählt man alle $b_i = 1/2$, so ist $K = 2$ und $\hat{\psi}_b = \bar{\psi}_b$. Allgemein haben wir für $K \geq 2$ und $b_i = 1/K$:

$$\bar{\psi}_b = \sum_{i=1}^K \frac{1}{K} y_i, \quad \widehat{\text{Var}} \bar{\psi}_b = \frac{1}{K(K-1)} \sum_{i=1}^K (y_i - \bar{\psi}_b)^2,$$

was für $K \geq 3$ aus obiger allgemeinen Formel folgt und für $K = 2$ extra gezeigt werden muß, vgl. Hartung/Böckenhoff/Knapp (2003). Es gilt für den Erwartungswert:

$$\begin{aligned} E \widehat{\text{Var}} \bar{\psi}_b &= \frac{1}{K^2} \sum_{i=1}^K \text{Var } y_i \\ &= \frac{1}{K} \sigma^2 \quad \text{falls } \text{Var } y_i = \sigma^2, \end{aligned}$$

was dann wieder die uns wohl bekannte Varianzschätzung für das arithmetische Mittel ergibt, wenn alle y_i die selbe Varianz haben. Eine Anwendung dieser Formel folgt u. a. direkt im nächsten Abschnitt.

Für die Herleitung der Varianzformel wird keine Normalverteilungsannahme benötigt, lediglich bei der Angabe von Konfidenzintervallen mittels $\hat{\psi}_b$, was in der Praxis keine wesentliche Einschränkung bedeuten dürfte, vgl. auch Hotelling (1960), da aufgrund des Zentralen Grenzwertsatzes, vgl. Abschnitt 9 in Kap. II, das durch Mittelung der y_i entstandene $\hat{\psi}_b$ schon bei wenigen Summanden recht gut normalverteilt ist, auch wenn die y_i selbst das nicht sein sollten, vgl. auch den einleitenden Abschnitt 1.10. In bösartigen Fällen steht uns immer noch die Tschebyscheff Ungleichung zur Verfügung, vgl. Abschnitt 8.2 in Kapitel II, mit der das Niveau des Konfidenzintervalls auf jeden Fall abgeschätzt werden kann.

7. Wirtschafts-Prognosen mit Scoringssystem

Wir betrachten hier Situationen, für welche das nachfolgend ausführlich besprochene Beispiel als typisch angesehen werden kann. Eine Anzahl von K Wirtschaftsinstituten gibt jährlich Prognosen für die zukünftige Entwicklung z. B. eines Landes ab mit einem Prognosehorizont von in der Regel einem Jahr.

7.1. Entwicklung eines Scoringssystems

Die Prognosen werden jährlich abgegeben von identischen Instituten. Somit liegen über eine Zahl J von Jahren historische Prognosen vor.

7.1.1. Bekannte Referenzwerte

Zugleich ist historisch auch bekannt, wie sich die Zielgrößen der Prognosen am Ende der jeweiligen Jahre, für welche die Prognosen bestimmt waren, tatsächlich entwickelt haben. Die Abweichung zwischen Prognosewert und realisiertem Wert lässt sich als Prognosefehler bezeichnen. Hierbei ist es in

manchen Fällen ratsam, zwischen positiver und negativer Abweichung zu unterscheiden.

Nehmen wir vereinfachend an, positive und negative Abweichungen seien gleichwertig. Dann haben wir für Institut i im Jahr j die

$$\text{Prognose-Differenz } D_{ij}, \quad i = 1, \dots, K, \quad j = 1, \dots, J,$$

die aber nicht unbedingt die einfache betragliche Differenz sein muß, sondern mit einem geeigneten erscheinenden Maß, vgl. Beispiel, gemessen wird. Bewertet wird das i -te Institut nun mit dem reziproken Fehler D_{ij}^{-1} für das Jahr j . Über die Jahre gemittelt, eventuell derart mit w_{ij} gewichtet, daß weiter zurückliegende Jahre weniger stark gewichtet werden und man dem Institut einen Lernprozess unterstellt, ergibt sich für das i -te Institut der

$$\text{Scoring Point: } SCP(i) = \sum_{j=1}^y w_{ij} D_{ij}^{-1}.$$

7.1.2. Referenzbezug Gruppenmittel

Liegen keine wahren Referenzwerte vor, so können die jährlichen Gruppenmittel der K Prognosen als jeweilige Bezugspunkte herangezogen werden. Dabei unterstellt man implizit, daß im Mittel schon die Wahrheit stets gut getroffen wird, was, wie wir wissen, nicht immer der Fall sein muß.

Die Idee hierzu ist dem technischen Bereich entlehnt. Dort ist als tückisches Problem bekannt, die Präzision von Meßinstrumenten zu bestimmen, wenn die vorzunehmende Prüfung zerstörend ist, vgl. Abschnitt 2.5 in Kapitel X von Hartung/Elpelt (2007) und Hartung/Klössener (1987) sowie Hartung (1999d). Am leichtesten stellt man sich dabei Instrumente zur Geschwindigkeitsmessung von Geschossen vor. Wiederholungsmessungen am selben Objekt sind wie bei Wirtschaftsprognosen nicht möglich, da man das Geschoß nicht ein zweites Mal abschießen kann.

Man behilft sich dann damit, K Instrumente gleichzeitig zu verwenden und diese kollektive Messung öfters durchzuführen, an verschiedenen Objekten wie im Prognosebereich in mehreren Jahren. Die Präzisionsbestimmung findet dann im kollektiven Vergleich statt. Grundlage hierfür sind die Grubbs'schen Meßmodelle, die in Hartung/Elpelt (2007) dargelegt sind und eine Präzisionsbestimmung erlauben, ohne jegliche Normalverteilungsannahme.

Auf unsere Prognosen übetragen haben die Präzisionen hier gerade die Bedeutung von Zuverlässigkeiten, die jeder Prognose als quantitative Größe zugeordnet werden können und zur Varianzschätzung verwandt werden oder als Grundlage zur Festlegung der Scoringpunkte dienen.

7.2. Kombinierte Prognosen mit Scoringpunkten

Haben wir unabhängige Prognosen y_1, \dots, y_K mit entsprechend zugeordneten Scoringpunkten $SCP(1), \dots, SCP(K)$ und wird die gemeinsame Zielgröße der Prognosen mit ψ bezeichnet, so ist die gescored kombinierte Prognose

für ψ gegeben als

$$\hat{\psi}_{SCP} = \sum_{i=1}^K \frac{SCP(i)}{\sum_{k=1}^K SCP(k)} y_i.$$

Setzen wir dann

$$b_i = SCP(i) / \sum_{k=1}^K SCP(k)$$

und sind alle b_i ungleich 0.5, so können wir mit der Varianzformel nach Hartung/Böckenhoff/Knapp (2003), die in Abschnitt 6.2 angegeben ist, die Varianz von $\hat{\psi}_{CSP}$ schätzen als $\widehat{\text{Var}} \hat{\psi}_{CSP}$. Hiermit ist dann ein $(1 - \alpha)$ -Konfidenzintervall für die gemeinsame Zielgröße der Prognosen angebbar als:

$$KI(\psi) = \hat{\psi}_{SCP} \pm \sqrt{\widehat{\text{Var}} \hat{\psi}_{SCP}} \cdot u_{1-\alpha/2},$$

wobei $u_{1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil der $\mathcal{N}(0, 1)$ -Verteilung ist.

7.3. Beispiel: Prognosen des BIP von sieben Wirtschaftsinstituten

Die Prognosen des Brutto-Inlands-Produktes (BIP), d. h. eigentlich dessen Steigerungsrate in %, für die Jahre 1990 bis 1996, aufgestellt von 7 Wirtschaftsinstituten, sind in **Tab 17** zusammengestellt, vgl. Klapper (1999). Anhand dieser mittlerweile historischen Daten leiten Hartung/Argac (2009) ein Scoringssystem her und vergleichen die gescored kombinierten Prognosen, nebst jeweils zugehörigem 95%-Konfidenzintervall, mit den wahren Werten der Jahre 1990 bis 1996. Die Lernphase für das Scoringssystem beginnt allerdings schon im Jahre 1984.

Tab. 17: Prognosen von sieben Wirtschaftsinstituten zum BIP [Steigerungsraten in %] für die Jahre 1990 bis 1996.

Jahr	Institut						
	1	2	3	4	5	6	7
1990	3.50	3.00	3.00	3.00	3.00	2.80	3.20
1991	3.50	3.25	3.50	3.00	3.50	2.70	3.00
1992	1.00	1.50	1.50	2.00	2.50	2.20	1.80
1993	-1.00	-0.50	0.00	0.50	0.00	0.90	0.50
1994	-0.50	1.00	1.00	1.00	0.00	0.50	0.40
1995	2.00	3.00	3.50	3.00	3.00	3.84	2.80
1996	1.00	1.75	1.70	2.50	2.00	2.50	2.40

Das Abstandsmaß wird als Quadrat der Distanz gewählt und die Jahre werden nicht unterschiedlich gewichtet. Bezeichnet ψ_j den wahren Wert des

Tab. 18: Die Gewichte b_i für die BIP Daten.

Jahr	Gewichte						
	b_1	b_2	b_3	b_4	b_5	b_6	b_7
1990	0.042	0.068	0.240	0.029	0.460	0.048	0.112
1991	0.044	0.069	0.240	0.030	0.458	0.049	0.112
1992	0.082	0.067	0.248	0.029	0.432	0.043	0.099
1993	0.061	0.173	0.300	0.028	0.304	0.034	0.101
1994	0.063	0.173	0.299	0.028	0.303	0.034	0.100
1995	0.063	0.173	0.298	0.029	0.302	0.034	0.100
1996	0.167	0.154	0.264	0.026	0.268	0.030	0.090

BIP im Jahr j und y_{ij} die Prognose für das BIP im Jahr j vom i -ten Institut, dann erhalten wir für 1990 als $\text{SCP}(i)$ für das i -te Institut

$$\text{SCP}(i)_{1990} = \sum_{j=1984}^{1989} \frac{1}{(y_{ij} - \psi_j)^2},$$

analog für die Jahre bis 1996. Zur Berechnung, etwa für 1990, der kombinierten Prognose benötigen wir die Gewichte

$$b_i(1990) = \text{SCP}(i)_{1990} / \sum_{k=1}^7 \text{SCP}(k)_{1990}$$

der 7 beteiligten Institute, die in **Tab 18** für die Jahre 1990 bis 1996 zusammengestellt sind. Damit erhalten wir die kombinierte Prognose für 1990

$$\widehat{\psi}_{\text{SCP}}[1990] = \sum_{i=1}^7 b_i(1990) \cdot y_i(1990) = 3.034.$$

Die Varianz der kombinierten Prognose wird mit der Varianzformel aus Abschnitt 6.2 geschätzt zu

$$\widehat{\text{Var}} \widehat{\psi}_{\text{SCP}}[1990] = 0.00110,$$

und wir können damit das kombinierte 95%-Konfidenzintervall für das BIP [1990], aus Sicht des Jahres 1989, angeben als

$$\text{KI}(\psi[1990]) = 3.034 \pm \sqrt{0.00110} \cdot 1.96 = [2.969, 3.099],$$

das mit Abstand engste Konfidenzintervall unter den betrachteten Jahren.

Am Ende von 1990 konnte man dann die Steigerungsrate des BIP für 1990 feststellen als $\psi[1990] = 4.7$, was mancherlei politische Gründe hatte, die

Tab. 19: BIP: Gegenüberstellung kombinierte Prognose und wahrer Wert

Kombinierte			
Jahr	Wahrer Wert	Prognose	95%-Konfidenzintervall
1990	4.7	3.034	[2.969, 3.099]
1991	3.7	3.373	[3.153, 3.593]
1992	1.6	1.966	[1.157, 2.774]
1993	-1.7	-0.053	[-0.274, 1.098]
1994	2.4	0.525	[-0.047, 1.098]
1995	1.9	3.095	[2.753, 3.436]
1996	1.4	1.780	[1.454, 2.106]

in 1989 nicht zu erahnen waren. Die **Tab. 19** enthält alle wahren Werte, kombinierten Prognosen und die zugehörigen 95%-Konfidenzintervalle für die Jahre 1990 bis 1996. In **Abb. 7** im einleitenden Abschnitt 1.5 sind für die Jahre 1990 bis 1996 die wahren BIP-Steigerungen und die kombinierten Prognosen nebst 95%-Konfidenzintervallen graphisch veranschaulicht.

8. Combining Forecasts mit geschätzten Zuverlässigkeit und Korrelationen

Hier gehen wir davon aus, daß zu den K Prognosen y_1, \dots, y_K der gemeinsamen Zielgröße ψ Zuverlässigkeit Z_1, \dots, Z_K geschätzt sind, die als inverse geschätzte Varianzen interpretierbar sind, $Z_i = 1/\hat{\sigma}_i^2$, $i = 1, \dots, K$. Besteht kein Anlaß, eine Korrelation zwischen den y_i zu vermuten, so ist es gebräuchlich, mit $Z_i / \sum Z_j$ gewichtete Mittel der y_i als kombinierte Prognose zu berechnen, d. h. es wird das Fixed Effects Modell (FE) aus Abschnitt 2 angewandt. Wir wollen dies beispielhaft demonstrieren.

8.1. Beispiel: Kombinierte Prognose in der Investitionsberatung

Wir betrachten eine bei Investitionsentscheidungen, vgl. etwa Pflaumer (2009), typische Situation. Vor einer tatsächlich zu tätigenden Investition werden Unternehmensberater hinzugezogen, um die im eigenen Betrieb ermittelte Abschätzung des zu erwartenden **Investitionsgewinns** oder **Cash Flow Profits** mit externen Prognosen zu vergleichen und gegebenenfalls auch zu kombinieren.

Im Beispiel gehen wir vom Normalfall $K = 2$ aus, in dem ein Betrieb B die Steigerung des Cash Flow (SCF) bei geplanter Investition prognostiziert und ein Unternehmensberater U zwecks zweiter Prognose hinzugezogen wird. Wir haben als Datenlage eine extreme Konstellation gewählt. Die uns bekannten Informationen werden jedoch nur so verwandt, wie es das jeweilige Verfahren voraussetzt. Später werden schließlich alle uns vorliegenden Informationen zur Kombination der beiden Prognosen verwandt.

Die jeweils errechneten Prognosen sind [$i = 1$] $y_B = 15$ mit Zuverlässigkeit

$Z_B = 1/\hat{\sigma}_1^2 = 7/8$ und $[i = 2] y_U = 10$ mit $Z_U = 1/\hat{\sigma}_2^2 = 7/2$. In Tab. 20 sind diese Angaben mit weiteren Hilfsgrößen zusammengestellt, wobei $y_1 = 15$ und $y_2 = 10$ gesetzt sind.

Unter Verwendung der Begriffe aus Abschnitt 2 erhalten wir mit den bereits in Tab. 20 berechneten Angaben dann:

$$\hat{\psi}_{FE} = \sum d_i y_i = 11 \text{ sowie } \hat{\sigma}^2(\hat{\psi}_{FE})_C = \frac{1}{v_{\Sigma}} = 8/35 = 0.2285,$$

und damit das 95%-Konfidenzintervall für ψ ,

$$KI_{FE}(\psi)_C = \hat{\psi}_{FE} \pm \sqrt{\hat{\sigma}^2(\hat{\psi}_{FE})_C \cdot 1.96} = [10.063, 11.937].$$

Weiterhin sind wegen

$$Q_C = \sum v_i (y_i - \hat{\psi}_{FE})^2 = 17.5 > \chi^2_{1;0.995} = 7.879$$

die beiden Prognosen hochsignifikant heterogen.

Eine mögliche Heterogenität zwischen den Prognosen explizit zu berücksichtigen, ist der Literatur zur Kombination von Prognosen fremd, vgl. etwa diesbezüglich in der Einleitung 1.6 zu diesem Kapitel zitierte Referenzen. Deshalb zeigen wir eine Kombination gemäß Random Effects Modell (RE) für obiges Beispiel, obwohl der Fall $K = 2$ dafür wenig geeignet ist.

8.2. Beispiel: Investitionsberatung im RE-Modell

Unter Hilfenahme der Berechnungen aus Tab. 20, vgl. auch Abschnitt 3, berechnen wir zunächst etwa den Hedges Schätzer für den Heterogenitätsparameter τ^2 zu

$$\hat{\tau}_{HED}^2 = \frac{1}{K-1} \sum (y_i - \bar{y})^2 - \frac{1}{K} \sum \hat{\sigma}_i^2 = 25/2 - 5/7 = 11.786$$

und damit dann im RE-Modell:

$$\begin{aligned} \hat{\psi}_{RE} &= \sum c_i y_i = 12.415, \quad c_i = w_i / w_{\Sigma}, \quad w_i = (\hat{\tau}_{HED}^2 + \hat{\sigma}_i^2)^{-1}, \\ \hat{\sigma}^2(\hat{\psi}_{RE})_C &= \frac{1}{w_{\Sigma}} = 6.242, \text{ und} \\ \hat{\sigma}^2(\hat{\psi}_{RE})_H &= q_H = \frac{1}{K-1} \sum c_i (y_i - \hat{\psi}_{RE})^2 = 6.243. \end{aligned}$$

Ein 95%-Konfidenzintervall für ψ ist dann etwa

$$KI_{RE}(\psi)_C = \hat{\psi}_{RE} \pm \sqrt{\frac{1}{w_{\Sigma}}} \cdot 1.96 = [7.518, 17.312],$$

also erwartungsgemäß breiter als wenn man Homogenität unterstellt.

Tab. 20: Investitionsberatung im FE- und RE-Modell

i	v_i	d_i	$d_i y_i$	η_i	δ_i	c_i	$c_i y_i$	ϵ_i	w_i
1	7/8	0.2	3	14	25/4	0.483	7.245	3.220	0.07735
2	7/2	0.8	8	3.5	25/4	0.517	5.170	3.023	0.08284
$\sum :$	35/8	1	11	17.5	50/4	1	12.415	6.243	0.1602
	$v_i = 1/\hat{\sigma}_i^2, v_{\Sigma} = \sum v_i, d_i = v_i/v_{\Sigma}, w_i = 1/(\hat{\tau}_{HED}^2 + \hat{\sigma}_i^2), w_{\Sigma} = \sum w_i$ $c_i = w_i/w_{\Sigma}, \eta_i = v_i(y_i - \hat{\psi}_{FE})^2, \delta_i = (y_i - \bar{y})^2, \epsilon_i = c_i(y_i - \psi_{RE})^2$								

8.3. Kombination von Prognosen bei geschätzten Korrelationen aber homogenen Prognosen [EFE-Modell]

Ab hier müssen wir zur Matrix-Schreibweise übergehen. Zur Auffrischung der Kenntnisse zur Vektor- und Matrizenrechnung so wie wir sie anwenden sei auf die nur 17 Seiten umfassende und mit Beispielrechnungen versehene Darstellung in Abschnitt 4 von Kapitel I in Hartung/Elpelt (2007) hingewiesen.

Liegen K Prognosen y_1, \dots, y_K vor, so haben der Vektor y der K Prognosen und die zugehörige geschätzte Kovarianzmatrix S die Gestalt

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_K \end{pmatrix}, \quad S = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{1j} \\ \hat{\sigma}_{ij} & \ddots \\ \hat{\sigma}_{kj} & \hat{\sigma}_K^2 \end{pmatrix}$$

mit $\hat{\sigma}_i^2 = \widehat{\text{Var}} y_i$ und $\hat{\sigma}_{ij} = \widehat{\text{Cov}}(y_i, y_j) = \hat{\rho}_{ij} \hat{\sigma}_i \hat{\sigma}_j$ mit geschätzten Korrelationen $\hat{\rho}_{ij}$, vgl. Kap. IX, für $i \neq j$, vgl. auch Abschnitt 8.4 in Kap. II.

Die y_i haben den Erwartungswert ψ , d. h. $E y_i = \psi$ für alle $i = 1, \dots, K$. Weiterhin sei mit e derjenige K -dimensionale Vektor bezeichnet, dessen Komponenten alle 1 sind, d. h. der zu e transponierte Vektor e^\top hat die Gestalt $e^\top = (1, \dots, 1)_K$. Bezeichnet S^{-1} die inverse Matrix von S , dann wird ψ in diesem **Extended Fixed Effects Model (EFE)** geschätzt gemäß:

$$\hat{\psi}_{EFE} = \frac{1}{e^\top S^{-1} e} e^\top S^{-1} y.$$

Seine Varianz wird geschätzt als:

$$\widehat{\text{Var}}_H \hat{\psi}_{EFE} = \frac{1}{e^\top S^{-1} e}.$$

Daraus ergibt sich mit dem $\mathcal{N}(0, 1)$ -Quantil $u_{1-\alpha/2}$ ein $(1 - \alpha)$ -Konfidenzintervall für ψ , wenn wir y als multivariat normalverteilt annehmen, zu

$$KI_{EFE}(\psi)_H = \hat{\psi}_{EFE} \pm \sqrt{\frac{1}{e^\top S^{-1} e} \cdot u_{1-\alpha/2}}.$$

[Mit Wahrscheinlichkeit 1 ist S invertierbar. Ist dies jedoch nicht der Fall, so ersetzen wir oben S^{-1} durch die pseudo-inverse Matrix S^+ , die ebenso leicht zu berechnen ist, vgl. Hartung/Elpelt (2007).]

8.4. Beispiel: Investitionsberatung im EFE-Modell und „Borrowing of Strength“

Für y , S und e haben wir im Beispiel

$$y = \begin{pmatrix} 15 \\ 10 \end{pmatrix}, \quad S = \frac{1}{7} \begin{pmatrix} 8 & -3 \\ -3 & 2 \end{pmatrix}, \quad e = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

wobei wir bewußt eine starke negative Korrelation gewählt haben,

$$\hat{\rho} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_1^2 \hat{\sigma}_2^2}} = -\frac{3}{4}.$$

Wir berechnen zunächst die Inverse zu S als:

$$S^{-1} = \begin{pmatrix} 2 & 3 \\ 3 & 8 \end{pmatrix},$$

als dann

$$e^\top S^{-1} = (5 \ 11), \quad e^\top S^{-1} e = 16, \quad e^\top S^{-1} y = 5 \times 15 + 11 \times 10 = 185,$$

so daß wir erhalten:

$$\hat{\psi}_{EFE} = \frac{185}{16} = 11.563, \text{ sowie}$$

$$\widehat{\text{Var}}_H \hat{\psi}_{EFE} = \frac{1}{16} = 0.0625,$$

und damit das 95%-Konfidenzintervall:

$$KI_{EFE}(\psi)_H = 11.563 \pm \sqrt{\frac{1}{16} \cdot 1.96} = [11.073, 12.023].$$

Im Vergleich zu Abschnitt 8.1, wo ohne Kenntnis der Korrelation im normalen FE-Modell gerechnet wird, haben wir hier zum einen eine höhere Gesamtprognose und zum anderen eine wesentlich geringere Varianz für den Gesamtschätzer, nämlich 0.0625, als im FE-Modell, wo diese 0.2285 beträgt. Dieser Effekt ist der negativen Korrelation zu verdanken und fällt unter den Begriff „Borrowing of Strength“ von benachbarten Verhältnissen, wie er von **J. W. Tukey** geprägt wurde, vgl. Draper et al. (1992).

8.5. Kombination von Prognosen bei geschätzten Korrelationen und heterogenen Prognosen [ERE-Modell]

Wie schon in Abschnitt 8.2 angesprochen, ist erst recht bei der Verwendung geschätzter Korrelationen in der Kombination von Prognosen die Idee, eine Heterogenität τ^2 zwischen den Prognosen zu berücksichtigen, nirgends zu finden. Als Verallgemeinerung des Hedges Schätzers berechnen wir zunächst:

$$\hat{\tau}_{HH}^2 = \frac{1}{K-1} \sum_{i=1}^K (y_i - \bar{y})^2 - \frac{1}{K-1} \left[\sum_{i=1}^K \hat{\sigma}_i^2 - \frac{1}{K} \sum_{i=1}^K \sum_{j=1}^K \hat{\sigma}_{ij} \right],$$

wobei

$$\bar{y} = \frac{1}{K} \sum_{k=1}^K y_k \text{ und } \hat{\sigma}_{ii} = \hat{\sigma}_i^2, \quad i = 1, \dots, K.$$

Dann bezeichne W die folgende Matrix:

$$W = \begin{pmatrix} \hat{\sigma}_1^2 + \hat{\tau}_{HH}^2 & \hat{\sigma}_{ij} \\ \ddots & \ddots \\ \hat{\sigma}_{ij} & \hat{\sigma}_K^2 + \hat{\tau}_{HH}^2 \end{pmatrix},$$

die also aus S hervorgeht, indem man in S zu den geschätzten Varianzen $\hat{\sigma}_i^2$ die Schätzung $\hat{\tau}_{HH}^2$ addiert. Damit erhalten wir nun im **Extended Random Effects Model (ERE)** die kombinierte Prognose gemäß:

$$\hat{\psi}_{ERE} = \frac{1}{e^\top W^{-1} e} e^\top W^{-1} y.$$

Ihre Varianz wird geschätzt als:

$$\widehat{\text{Var}}_H \hat{\psi}_{ERE} = \frac{1}{e^\top W^{-1} e}.$$

Dies führt zum $(1 - \alpha)$ -Konfidenzintervall für ψ :

$$KI_{ERE}(\psi)_H = \hat{\psi}_{ERE} \pm \sqrt{\frac{1}{e^\top W^{-1} e} \cdot u_{1-\alpha/2}}.$$

8.6. Beispiel: Investitionsberatung im ERE-Modell

Zunächst berechnen wir den Verallgemeinerten Hedges-Schätzer,

$$\hat{\tau}_{HH}^2 = 25/2 - \left[10/7 - \frac{1}{2} 4/7 \right] = 11.4.$$

Damit berechnen sich die Matrix W und ihre Inverse W^{-1} zu:

$$W = \begin{pmatrix} 12.54 & -0.43 \\ -0.43 & 11.68 \end{pmatrix} \text{ und } W^{-1} = \begin{pmatrix} 0.080 & 0.003 \\ 0.003 & 0.086 \end{pmatrix}.$$

Hieraus erhalten wir nun:

$$W^{-1}y = \begin{pmatrix} 1.2272 \\ 0.9013 \end{pmatrix}, \quad W^{-1}e = \begin{pmatrix} 0.083 \\ 0.089 \end{pmatrix},$$

und damit die gewünschte kombinierte Prognose, die auch **Heterogenität zwischen Betrieb und Berater** berücksichtigt, gemäß:

$$\hat{\psi}_{ERE} = \frac{1}{e^T W^{-1} e} e^T W^{-1} y = \frac{2.1285}{0.1714} = 12.418.$$

Ihre Varianz können wir schätzen durch:

$$\widehat{\text{Var}}_H \hat{\psi}_{ERE} = \frac{1}{e^T W^{-1} e} = 5.834,$$

so daß ein 95%-Konfidenzintervall für ψ gegeben ist als:

$$KI_{EFE}(\psi)_H = \hat{\psi}_{EFE} \pm \sqrt{\frac{1}{e^T W^{-1} e}} \cdot 1.96 = [7.684, 17.152],$$

welches erwartungsgemäß durch die Berücksichtigung von Heterogenität zwischen den Prognosen breiter ausfällt als im EFE-Modell in Abschnitt 8.4.

8.7. Ein stabiler Hartung–Varianzschätzer für die kombinierte Prognose im ERE-Modell

Die gemeinsame Zielgröße ψ wird wie eben geschätzt als:

$$\hat{\psi}_{ERE} = \frac{1}{e^T W^{-1} e} e^T W^{-1} y,$$

jedoch als stabile Ergänzung zum Varianzschätzer in 8.5, schlagen wir hier noch – in Verallgemeinerung von q_H aus Abschnitt 3 – den folgenden Varianzschätzer für $\hat{\psi}_{ERE}$ vor:

$$\widehat{\text{Var}}_H^* \hat{\psi}_{ERE} = \frac{1}{K-1} \cdot \frac{1}{e^T W^{-1} e} y^T \left(W^{-1} - \frac{1}{e^T W^{-1} e} W^{-1} e e^T W^{-1} \right) y,$$

der weniger sensitiv ist in Bezug auf Schwankungen des Schätzers für den Heterogenitätsparameter τ^2 . Zugeordnet ist dem Schätzer eine χ^2_{K-1} -Verteilung, so daß ein damit konstruiertes Konfidenzintervall für ψ das t-Quantil $t_{K-1; 1-\alpha/2}$ verwendet.

In unserem **Beispiel** ergibt sich dieser Varianzschätzer zu:

$$\widehat{\text{Var}}_H^* \hat{\psi}_{ERE} = 5.776,$$

etwas kleiner als der entsprechende Schätzer oben, der ist 5.834.

Tab. 21: Ergebnisse der 4 Modelle im Beispiel zur Investitionsberatung

Combined Forecast		
$\hat{\tau}^2$	Modell	95%-Konfidenzintervall
	<u>FE</u>	
0	$\hat{\psi}$	11
	KI(ψ)	[10.36, 11.937]
	<u>EFE</u>	
0	$\hat{\psi}$	11.563
	KI(ψ)	[11.073, 12.023]
	<u>RE</u>	
11.8	$\hat{\psi}$	12.415
	KI(ψ)	[7.518, 17.312]
	<u>ERE</u>	
11.4	$\hat{\psi}$	12.418
	KI(ψ)	[7.684, 17.152]

8.8. Vergleich der vier Prognose-Modelle im Beispiel zur Investitionsberatung

Beim Vergleich der Modelle, s. Tab. 21, gehen wir getrennt vor, vgl. auch die Abb. 8 im einleitenden Abschnitt 1.6, in der die beiden Einzelprognosen nebst 95%-Konfidenzintervallen sowie die vier Combined Forecasts nebst 95%-Konfidenzintervallen graphisch veranschaulicht sind.

8.8.1. Vergleich der kombinierten Prognosen

Beim Vergleich der kombinierten Schätzer für die gemeinsame Zielgröße fällt auf, daß in den jeweils erweiterten Modellen, welche die geschätzte Korrelation in die Gewichtung der einzelnen Prognosen einbeziehen, höhere Werte erzielt werden, wenn auch im ERE-Modell nur geringfügig gegenüber dem RE-Modell. Dieser Effekt ist auf die negative Korrelation zurückzuführen. Die Addition des hohen Wertes für $\hat{\tau}^2$ zu den Varianzen in den RE-Modellen hat zur Folge, daß in diesen Modellen die rechnerische Korrelation in der Matrix W nur noch -0.036 ist, wohingegen sie in der Ausgangsmatrix -0.75 beträgt.

8.8.2. Vergleich der kombinierten Konfidenzintervalle

Entsprechend sind die Unterschiede bei den FE-Modellen wesentlich größer. Hier bewirkt die starke negative Korrelation im EFE-Modell nicht nur die oben erwähnte höhere kombinierte Prognose, sondern auch ein wesentlich schmaleres Konfidenzintervall von der Breite 0.950 im Vergleich zur Breite 1.577 des Konfidenzintervalls im FE-Modell.

Wegen der oben erwähnten wesentlich geringeren Korrelation in W, ist der

Unterschied der Breiten 9.468 im ERE- und 9.794 im RE-Modell für die Konfidenzintervalle ebenfalls wesentlich geringer. Da in den FE-Modellen $\tau^2 = 0$ gesetzt wird, verwundert es nicht, daß in den RE-Modellen mit den hohen Werten für τ^2 die Konfidenzintervalle auch beträchtlich breiter ausfallen als in den FE-Modellen.

9. Kombination unabhängiger p-Werte oder unabhängiger Tests

Nehmen wir an, daß in der i-ten von $i = 1, \dots, K \geq 2$ unabhängigen Studien ein Parameter θ_i von Interesse ist. Mittels einer Teststatistik T_i werde die Nullhypothese

$$H_{0i} : \theta_i = 0 \text{ gegen } H_{1i} : \theta_i > 0$$

verworfen, wenn $T_i >$ kritischem Wert (i) ist. Ist t_i die Realisation von T_i in der i-ten Studie, dann ist der p-Wert P_i der i-ten Studie die Wahrscheinlichkeit dafür, daß T_i bei Gültigkeit von H_{0i} den Wert t_i übertrifft, d. h.

$$P_i = P\{T_i > t_i \mid H_{0i}\}.$$

Ist F_{0i} die Verteilungsfunktion von T_i bei Gültigkeit von H_{0i} , so berechnet sich P_i als $1 - F_{0i}(t_i)$, was auch gemeint ist, wenn man üblicherweise schreibt

$$P_i = 1 - F_{0i}(T_i).$$

Bei der Kombination der p-Werte oder Tests aus den K Studien prüft man dann die globale Nullhypothese

$$H_0 : \theta_1 = \dots = \theta_K \text{ gegen } H_1 : \theta_{i_1} \neq \theta_{i_2}$$

für wenigstens zwei Studien i_1, i_2 . Stehen die θ_i in einem Zusammenhang, etwa der Art, daß sie Behandlungseffekte des selben Medikamentes beschreiben, welche aber in unterschiedlichen klinischen Parametern untersucht werden, so prüft man damit auf einen Gesamteffekt des Medikamentes. Sind die Parameter ohnehin als identisch anzusehen, $\theta_1 = \dots = \theta_K = \theta$, so prüft man mittels Kombination der K Studien die Hypothese $H_0 : \theta = 0$ gegen $H_1 : \theta > 0$, wobei man i. d. R. davon ausgeht, daß die einzelnen Studien keine signifikanten Ergebnisse zeigen oder unterschiedliche, auch widersprüchliche Ergebnisse vorliegen. Exakt gelten die den folgenden Verfahren zu Grunde liegenden Verteilungseigenschaften für stetige Verteilungsfunktionen. Praktisch gelten sie aber eher konservativ ebenso für andere Fälle, vgl. auch Hartung/Knapp/Sinha (2008) und Hartung (2006).

9.1. Inverse Normal- und Stouffer-Methode, Probits

Die Verteilungsfunktion der Standardnormalverteilung wird mit Φ bezeichnet, vgl. Kap. IV, und Φ^{-1} bezeichnet die inverse Funktion zu Φ . Nach Stouffer et al. (1949) wird zum Niveau α die globale Nullhypothese verworfen, wenn:

$$Z_{STOU} = \frac{1}{\sqrt{K}} \sum_{i=1}^K \Phi^{-1}(1 - P_i) > u_{1-\alpha},$$

wobei $u_{1-\alpha} = \Phi^{-1}(1-\alpha)$ das $(1-\alpha)$ -Quantil der $\mathcal{N}(0, 1)$ -Verteilung ist. Die Größen $Z_i = \Phi^{-1}(1 - P_i)$ bezeichnet man auch als **Probits** oder **Normal-Probits**. In Tab. 2 im Anhang ist Φ^{-1} vertafelt.

Mit vorab gewählten Gewichten $\lambda_1 > 0, \dots, \lambda_K > 0$, d. h. in Unkenntnis der Studienergebnisse, wird H_0 zum Niveau α verworfen mittels der **Gewichteten Inversen Normalmethode**, wenn:

$$Z_{GIN} = \frac{1}{\sqrt{\sum_{j=1}^K \lambda_j}} \sum_{i=1}^K \sqrt{\lambda_i} \cdot \Phi^{-1}(1 - P_i) > u_{1-\alpha}.$$

9.2. Inverse χ^2 - und Fisher-Methode, Fisher-Probits

Mit die älteste Kombinationsmethode für p-Werte stammt von R. A. Fisher aus dem Jahre 1932. Verwerfe die globale Nullhypothese H_0 zum Niveau α , wenn:

$$S_{FISH} = \sum_{i=1}^K -2 \ln P_i > \chi_{2K;1-\alpha}^2,$$

mit dem $(1-\alpha)$ -Quantil der χ_{2K}^2 -Verteilung, wobei wir $-2 \ln P_i$ als **Fisher-Probit** bezeichnen.

Mit $F_{\chi_\nu^2}^{-1}$ bezeichnen wir die inverse Funktion zu $F_{\chi_\nu^2}$, die Verteilungsfunktion der χ^2 -Verteilung mit ν Freiheitsgraden. Wie oben werden vorab Gewichte $\lambda_1, \dots, \lambda_K$ festgelegt, die jetzt aber besser als natürliche Zahlen gewählt werden. Zum Niveau α wird dann H_0 mittels **Generalisierter Inverser χ^2 -Methode** verworfen, wenn:

$$S_{GIC} = \sum_{i=1}^K F_{\chi_{\lambda_i}^2}^{-1}(1 - P_i) > \chi_{\Lambda;1-\alpha}^2 \text{ mit } \Lambda = \sum_{j=1}^K \lambda_j.$$

Für $\lambda_i = 2$ ergibt sich wieder die Fisher-Methode. Bevorzugte, asymptotische Testeigenschaften diskutieren z. B. Hedges/Olkin (1985), Hartung/Knapp/Sinha (2008), die auch weitere Verfahren zur Kombination von p-Werten besprechen. Grob besagen diese Eigenschaften, daß in „großen Stichproben“ die Inverse χ^2 -Methode zu den „meisten signifikanten Ergebnissen“ führt. In dieser Hinsicht „schärfert“ als die Fisher-Methode ist die Inverse χ^2 -Methode mit $\lambda_i = 1$, vgl. Hartung (2006), mit:

$$F_{\chi_1^2}^{-1}(1 - P_i) = [\Phi^{-1}(1 - P_i/2)]^2.$$

Allgemein berechnen wir $F_{\chi_{\lambda_i}^2}^{-1}(\gamma) = \chi_{\lambda_i;\gamma}^2$ mittels Tab. 4 im Anhang und der dort umseitig angegebenen Approximation nach Wilson und Hilferty.

9.3. Heterogenität: Hartung-Knapp-Methode

Liegen heterogene Studienverhältnisse vor, etwa vergleichbar zu Situationen, die zum Random Effects Modell führen, so sind die üblichen Methoden der p-Werte Kombinationen nicht in der Lage, darauf zu reagieren. Daher schlagen Hartung/Knapp (2005a) in solchen Fällen ein modifiziertes Verfahren vor. Ausgehend von den Probits und z. B. Stichprobenumfängen n_i , werden die p-Werte P_i zunächst transformiert gemäß

$$z_i = \frac{1}{\sqrt{n_i}} \Phi^{-1}(1 - P_i),$$

wodurch der Effekt der Stichprobenumfänge auf die p-Werte zumindest gedämpft wird. Haben wir Studien mit unbalancierten Vergleichsgruppen E_i und C_i , so empfiehlt sich das harmonische Mittel der Stichprobenumfänge n_{E_i} und n_{C_i} als n_i zu nehmen. Wir erreichen damit, daß bei Gültigkeit von H_0 die z_i , wenn auch behaftet mit zufälligen Störungen a_i , den selben Erwartungswert ζ besitzen. Mit sinnvoll gewählten Gewichten

$$b_i > 0, \dots, b_K > 0 \text{ und } \sum_{i=1}^K b_i = 1,$$

erhalten wir dann den kombinierten Schätzer für ζ als:

$$\hat{\zeta}_b = \sum_{i=1}^K b_i z_i.$$

Mit dem allgemeinen Varianzschätzer aus Abschnitt 6.2 wird nun die Varianz von ζ_b als $\widehat{\text{Var}} \zeta_b$ geschätzt und:

$$\Delta = \max \left\{ \widehat{\text{Var}} \zeta_b, \sum_{i=1}^K b_i^2 / n_i \right\}$$

bestimmt, wobei der zweite Term hier die Varianz von $\hat{\zeta}_b$ bei Gültigkeit von H_0 beschreibt, die ohne Heterogenitäten zwischen den Studien vorliegen würde. Die globale Nullhypothese H_0 wird dann zum Niveau α verworfen, wenn:

$$T_b = \frac{\hat{\zeta}_b}{\sqrt{\Delta}} \geq u_{1-\alpha}.$$

Als Gewichtung schlagen wir $b_i = n_i / \sum n_j$ vor.

9.4. Beispiel Verhaltensforschung: Konformität bei Männern und Frauen

Untersucht wurden Geschlechtsunterschiede im Grad der Konformität, vgl. Hedges/Olkin (1985, p. 35). In **Tab. 22** sind, neben einigen Hilfsgrößen, die

Tab. 22: Verhaltensforschung: Geschlechtsunterschiede im Grad der Konformität, Nullhypothese: Konformität bei Frauen \leq Konformität bei Männern

Studie	Umfang	p-Wert	Transformierter	
			Probit	Probit
i	n _i	P _i	Z _i = $\Phi^{-1}(1 - P_i)$	z _i = Z _i / $\sqrt{n_i}$
1	126	0.0029	2.758	0.246
2	40	0.0510	1.635	0.259
3	62	0.6310	-0.335	-0.043
4	92	0.9783	-2.020	-0.211
5	32	0.0034	2.706	0.478
6	45	0.0305	1.873	0.279
7	30	0.0341	1.824	0.333
8	10	0.0367	1.790	0.566
9	70	0.9740	-1.942	-0.232
10	59	0.3517	0.381	0.050

p-Werte von 10 Studien angegeben. Die Vergleichstests waren in den Studien so angelegt, daß kleine p-Werte P_i einen höheren Grad an Konformität für Frauen anzeigen. Die Stichprobenumfänge n_i sind harmonische Mittel der Gruppenumfänge.

Wir wenden die Inverse Normalmethode an mit Gewichten λ_i = 1 [Stouffer-Methode] und λ_i = n_i. Für die Inverse χ²-Methode nehmen wir die Gewichte λ_i = 2 [Fisher-Methode] und λ_i = n_i.

Das Verfahren von Hartung-Knapp wird angewandt mit b_i = n_i / $\sum_{j=1}^{10} n_j$. Wir erhalten mit den errechneten Werten z_i aus Tab. 22 dann: $\hat{\zeta}_b = \sum_{i=1}^{10} b_i z_i = 0.0875$. Die Varianz von $\hat{\zeta}_b$ unter homogenen Studienverhältnissen ist: $\sum_{i=1}^{10} b_i^2 / n_i = 0.0018$, und mittels der Varianzformel aus Abschnitt 6.2 wird die Varianz von $\hat{\zeta}_b$ geschätzt zu: $\widehat{\text{Var}} \hat{\zeta}_b = 0.0078$, so daß also: Δ = 0.0078, und somit:

$$T_b = \frac{\hat{\zeta}_b}{\sqrt{\Delta}} = 1.009 \leq u_{1-\alpha}$$

für α mit Werten von 0 bis 0.1564, wobei der letzte Wert gerade der p-Wert von T_b ist, da: P(T_b) = 1 - Φ(1.009) = 0.1564. Anzumerken ist, daß wir hier für Δ = $\sum b_i^2 / n_i$ wieder die Inverse Normalmethode mit λ_i = n_i erhalten würden. In **Tab. 23** sind zu allen aufgeführten Tests die beobachteten Testwerte, die kritischen Werte zum Niveau α = 0.025 und zur besseren Vergleichsmöglichkeit auch die „globalen“ p-Werte zu den jeweils realisierten Testwerten angegeben. In **Abb. 4**, die im Rahmen des einleitenden Abschnitts 1.3 plaziert ist, sind die extremen Studienergebnisse sowie die drei

kombinierten p-Werte nach Fisher, Stouffer und Hartung-Knapp graphisch veranschaulicht.

Die extrem unterschiedlichen, ja sogar gegensätzlichen Studienergebnisse, vgl. Tab. 22 und Abb. 4, erklären das vorsichtige Verhalten der Teststatistik T_b . Die anderen Methoden zur Kombination der p-Werte liefern hier ein Bild, das zu Fehlinterpretationen verleitet. Wir verweisen deshalb auf die Ausführungen zu Beginn dieses Abschnitts, die auch bei der Interpretation ähnlicher Testsituationen in einer konkreten Anwendung beobachtet werden sollten.

Tab. 23: Verhaltensforschung: Resultate der p-Werte Kombinationen zu den Studienergebnissen aus Tab. 22 bzgl. H_0 : Konformität von Frauen \leq Konformität von Männern

Teststatistik	Testwert	Kritischer Wert	Globaler p-Wert
<u>Invers Normal</u>			
		<u>$u_{0.975}$</u>	
Stouffer	2.74	1.96	0.003
$\lambda_i = n_i$	2.08	1.96	0.019
<u>Invers χ^2</u>			
		<u>$\chi^2_{\sum \lambda_i; 0.975}$</u>	
Fisher	52.46	34.17	0.0001
$\lambda_i = n_i$	652.81	633.82	0.007
<u>Hartung-Knapp</u>			
		<u>$u_{0.975}$</u>	
T_b	1.009	1.96	0.156

10. Combining Forecasts bei unbekannten Korrelationen und Kombination abhängiger p-Werte oder Tests

Liegen Prognosen von $i = 1, \dots, K \geq 2$ Einrichtungen, Beratern, Experten [Delphi Methode] oder verschiedenen Modellspezifikationen vor, so ist mitunter unklar, wie weit diese K Prognosen miteinander korrelieren. Die Unabhängigkeit der Prognosen ist also in Zweifel gestellt.

Daß eine Kombination der Prognosen erhebliche Vorteile verschafft gegenüber einer individuellen Prognose, wurde bereits 1992 vom US National Research Council ausführlich diskutiert und der methodische Ausbau empfohlen, vgl. Draper et al. (1992). Dort befinden sich schon umfangreiche Literaturangaben, die insbesondere die Bereiche Business, Economy, Risk Assessment sowie militärische Feinderkennung betreffen.

10.1. Strategie zu Combined Forecasting

Wie in Draper et al. (1992) dargelegt, geht es zunächst einmal darum, die K verschiedenen Prognosen „auf einen gemeinsamen Nenner“ zu bringen, wenn zudem die eventuell jeweils beigefügten eigenen Angaben zu Zuverlässigkeit oder zum Streubereich der gestellten Prognose nicht verwertbar sind, d. h. wenn wir die Verfahren des Abschnittes 8 nicht anwenden können.

Eine Möglichkeit, eine vergleichbare Situation herbeizuführen, besteht nach Draper et al. (1992) darin, daß der Entscheidungsträger, der die K Prognosen kombinieren will, einen gemeinsamen **Threshold Value** vergibt, der für den Entscheidungsträger von entscheidender Bedeutung ist, etwa für eine geplante Aktion. Die Prognosegeber sind dann gefragt, mit ihrem jeweiligen Instrumentarium eine Aussage über diese allen gemeinsame kritische Schranke zu treffen, etwa: Wird sie überschritten oder nicht, oder mit welcher Wahrscheinlichkeit würde der Schwellenwert überschritten, wenn die geplante Aktion getätigter wäre. Derartige Aussagen sind dann direkt vergleichbar.

Wir erhalten Ergebnisse, die wir interpretieren können als p-Werte P_1, \dots, P_K , vgl. Abschnitt 9. **Combining Information** oder **Model Mixing** wird dann erreicht durch eine geeignete Kombination der eventuell nicht unabhängigen p-Werte zu einem kombinierten p-Wert p_{COMB} .

10.2. Hartung-Methode zur Kombination von abhängigen p-Werten oder Tests

Liegen möglicherweise abhängige p-Werte P_1, \dots, P_K vor, wie sie formal zu Beginn des Abschnitts 9 beschrieben sind, so vergeben wir zunächst zugehörige Bewertungen:

$$\lambda_1, \dots, \lambda_K \text{ mit } \sum_{i=1}^K \lambda_i \neq 0.$$

Statt der originalen p-Werte betrachten wir die Probits:

$$Z_1 = \Phi^{-1}(1 - P_1), \dots, Z_K = \Phi^{-1}(1 - P_K),$$

und damit die Quadratform:

$$q_{Ha} = \frac{1}{K-1} \sum_{i=1}^K \left(Z_i - \frac{1}{K} \sum_{j=1}^K Z_j \right)^2.$$

Nach Hartung (1999b) ist $1 - q_{Ha}$ ein unverzerrter Schätzer der zu Grunde liegenden (durchschnittlichen) Korrelationen zwischen den Probits [unter H_0 , s. Abschnitt 9], so daß wir die durchschnittliche Korrelation schätzen mittels:

$$\hat{\rho} = \max \left\{ -\frac{1}{K-1}, 1 - q_{Ha} \right\}.$$

Ferner sei $\kappa > 0$ ein gewählter Regularitätsparameter mit dem wir die Hilfsgröße:

$$h(\kappa) = \hat{\rho} + \kappa \sqrt{\frac{2}{K+1}} (1 - \hat{\rho})$$

bilden, die wir für folgende Teststatistik benötigen, vgl. Hartung (1999b),

$$T_H = \frac{\sum_{i=1}^K \lambda_i Z_i}{\sqrt{\sum_{i=1}^K \lambda_i^2 + h(\kappa) \left[\left(\sum_{i=1}^K \lambda_i \right)^2 - \sum_{i=1}^K \lambda_i^2 \right]}} \stackrel{(H_0)}{\approx} \mathcal{N}(0, 1),$$

die approximativ $\mathcal{N}(0, 1)$ -verteilt ist, unter der Annahme, daß die den p-Werten zugrunde liegenden Nullhypothesen gültig sind. Der kombinierte p-Wert ist folglich

$$p_{COMB} = 1 - \Phi(T_H).$$

Zur Wahl des Regularitätsparameters κ schlägt Hartung (1999b) generell $\kappa = \kappa_1 = 0.2$ oder $\kappa = \kappa_2 = (1 + \frac{1}{K-1} - \hat{\rho}) 0.1$ vor, κ_2 bevorzugt, wenn für $K \geq 3$ mit negativen Korrelationen zu rechnen ist.

Simulationen zeigen, daß T_H in allen denkbaren Konstellationen von Korrelation zwischen den P_1, \dots, P_K das Testniveau ziemlich genau einhält, lediglich bei negativen Korrelationen kann T_H sehr konservativ werden, vgl. Hartung (1999b).

Demetrescu/Hassler/Tarcolea (2006) führen weitere theoretische Untersuchungen der Teststatistik T_H durch und zeigen weitere Simulationsergebnisse, auch für kreuz-korrelierte Zeitreihen Panels, vgl. auch Abschnitt 11.

10.3. Hartung-Methode für Fisher-Probits

Da die in Abschnitt 9.2 vorgestellte Fisher-Methode bevorzugte asymptotische Testeigenschaften hat, Näheres s. Hedges/Olkin (1985), Hartung/Knapp/Sinha (2008) und Abschnitt 9.2, ist es sinnvoll, diese auch für den Fall abhängiger p-Werte auszunutzen.

Liegen p-Werte P_1, \dots, P_K vor, so berechnen wir zunächst die Fisher-Probits: $-2 \ln P_i$, und transformieren sie dann gemäß der inversen Wilson/Hilferty (1931) Transformation:

$$Z_i^F = 3 \left(\sqrt[3]{- \ln P_i} - \frac{8}{9} \right), \text{ für } i = 1, \dots, K.$$

Die Z_i^F sind dann praktisch normalverteilt [unter H_0] und wir können mit diesen Z_i^F anstelle der Z_i die Teststatistik T_H aus vorigem Abschnitt berechnen, die wir dann mit T_H^F bezeichnen. Folglich ist hier der kombinierte p-Wert dann gegeben als:

$$p_{COMB}^F = 1 - \Phi(T_H^F).$$

10.4. Beispiel: Investitionsprognosen mit unbekannter Korrelation zwischen Betrieb und Berater

Wir setzen hier das Beispiel zur Investitionsberatung aus Abschnitt 8 fort. Wir gehen von jeweils unverzerrten normalverteilten Prognosen y_B und y_U

aus, mit $y_B = 15\%$ und $y_U = 10\%$ und recht unterschiedlichen Varianzen, $\text{Var } y_B = 8/7$ und $\text{Var } y_U = 2/7$. Desweiteren sind die Prognosen stark negativ korreliert, $\text{Corr}(y_B, y_U) = -3/4$, was wir hier aber nicht verwenden wollen. Eine negative Korrelation ist in der Praxis auch durch längere Erfahrung bekannt. Der **Threshold Value** bezüglich der Steigerung des Cash Flows [SCF] ist vom Entscheidungsträger festgesetzt auf: $\text{THR}V = 11.5\%$. Sowohl B als auch U prüfen nun jeweils die Hypothese

$$H_0 : \text{SCF} \leq 11.5 \text{ gegen } H_1 : \text{SCF} > 11.5$$

mittels der Teststatistiken

$$t_B(\mu) = \frac{15 - \mu}{\sqrt{8/7}} \text{ und } t_U(\mu) = \frac{10 - \mu}{\sqrt{2/7}},$$

die unter H_0 , d. h. wenn $\mu = 11.5$ gültig ist, nach $\mathcal{N}(0, 1)$ verteilt sind. Wir erhalten

$$t_B(11.5) = 3.270 \text{ und } t_U(11.5) = -2.806$$

und damit die p-Werte

$$P_B = 1 - \Phi(3.27) = 0.00123 \text{ und } P_U = 1 - \Phi(-2.806) = 0.99745,$$

die recht extrem gegensätzlich ausfallen. Mit den Bezeichnungen aus vorstehendem Abschnitt 10.2 erhalten wir $q_{Ha} = 18.46$, also $\hat{\rho} = -1$ und für

$$\kappa = \kappa_1 : T_H = 0.574, \quad \kappa = \kappa_2 : T_H = 0.469,$$

so daß wir die kombinierten p-Werte erhalten als:

$$p_{COMB} = 0.28 \text{ bzw. } p_{COMB} = 0.32,$$

ein doch recht erfreuliches Kombinationsergebnis, welches eine Entscheidung zu Gunsten der geplanten Investition nachhaltig unterstützt.

Wenden wir jetzt die inverse Wilson-Hilferty Transformation auf die Fisher-Probits an, vgl. Abschnitt 10.3, so ergeben sich

$$Z_B^F = 3.2129 \text{ und } Z_U^F = -2.2589$$

sowie $q_{Ha} = 14.97$, $\hat{\rho} = -1$ und daraus dann für

$$\kappa = \kappa_1 : T_H^F = 1.180, \quad \kappa = \kappa_2 : T_H^F = 0.964,$$

bzw. die kombinierten p-Werte:

$$p_{COMB}^F = 0.12 \text{ bzw. } p_{COMB}^F = 0.17.$$

Durch die Schätzung $\hat{\rho} = -1$ in beiden Methoden ist eine negative Korrelation erkennbar, so daß dann für $\kappa = \kappa_1$ die Teststatistiken hier im Falle $K = 2$ weniger konservativ sind, s. oben. Ausschlaggebend sollte also der kombinierte p-Wert $p_{COMB}^F = 12\%$ sein, der die Investitionsentscheidung eindeutig befürwortet.

11. Kombination von kreuz-korrelierten Zeitreihen zu einem Panel

Hat man eine Zeitreihe (Y_t^*) vorliegen, von der man vermutet, daß sie instationär ist, vgl. die im Rahmen des einleitenden Abschnitts 1.7 plazierte **Abb. 9**, so wird man mittels einfacher [z. B. Betrieb 3 in Abb. 9] oder auch mehrfacher [z. B. 2-facher für Betrieb 1 in Abb. 9] Differenzenbildung eine Zeitreihe (Y_t) erhalten, die dann, hoffentlich, stationär ist [z. B. Betriebe 2 und 4 in Abb. 9], vgl. Abschnitt 2.4 in Kap. XII.

11.1. Dickey-Fuller Unit Root Test auf Stationarität einer Zeitreihe

Mittelwertbereinigt folge die Zeitreihe (Y_t) gemäß Abschnitt 2.1 in Kap. XII einem Autoregressiven Prozeß 1. Ordnung, d. h. AR(1)-Prozeß, gemäß:

$$Y_t = \phi Y_{t-1} + \epsilon_t,$$

wobei ϵ_t ein weißes Rauschen ist mit: $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

Wir haben dort für AR(p)-Prozesse eine Bedingung für Stationarität formuliert, die für $p = 1$ lautet: Die Gleichung

$$1 - \phi z = 0$$

muß eine Lösung $z = 1/\phi$ größer 1 haben, d. h. aber: $\phi < 1$. Der Unit Root Test von Dickey/Fuller (1979, 1981) prüft nun zu einem Niveau α die Hypothese

$$H_0 : \phi = 1 \text{ gegen } H_1 : \phi < 1,$$

also auf signifikantes Vorliegen einer stationären Zeitreihe (Y_t).

Wir greifen uns einen Ausschnitt

$$Y_t, Y_{t-1}, \dots, Y_{t-n}$$

der Länge $n + 1$ aus der Zeitreihe (Y_t) und berechnen gemäß Abschnitt 1.6 in Kap. X als Schätzung für ϕ :

$$\hat{\phi} = \frac{\sum_{i=0}^{n-1} Y_{t-i} Y_{t-1-i}}{\sum_{j=0}^{n-1} Y_{t-1-j}^2},$$

sowie als Schätzung für die Varianz von ϵ_t :

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-1} \sum_{i=0}^{n-1} (Y_{t-i} - \hat{\phi} Y_{t-1-i})^2.$$

Aus Abschnitt 1.1 in Kap. X erhalten wir als Schätzung für die Varianz von $\hat{\phi}$:

$$\hat{\sigma}(\hat{\phi})^2 = \hat{\sigma}_\epsilon^2 \left/ \sum_{i=0}^{n-1} \left(Y_{t-1-i} - \frac{1}{n} \sum_{j=0}^{n-1} Y_{t-1-j} \right)^2 \right..$$

Die Dickey-Fuller Teststatistik ist formal die des üblichen t-Tests,

$$T_{DF} = \frac{\hat{\phi} - 1}{\sqrt{\hat{\sigma}(\hat{\phi})^2}},$$

die jedoch unter H_0 für $n \rightarrow \infty$ gegen eine von der Standardnormalverteilung verschiedene Grenzverteilung konvergiert, die unabhängig von σ_ϵ^2 und dem Ausgangswert ist.

Bezeichnet $cv(DF)_\alpha$ das α -Quantil dieser Grenzverteilung, so wird zum Niveau α :

$$H_0 \text{ verworfen, falls } T_{DF} < cv(DF)_\alpha.$$

Bezeichnet F_{DF} die Verteilungsfunktion dieser Grenzverteilung, so wird der p-Wert P von T_{DF} hier berechnet als:

$$P = F_{DF}(T_{DF}).$$

Kritische Werte $cv(DF)$ sind vertafelt oder können simuliert werden, s. Dickey/Fuller (1979, 1981) oder z. B. Hamilton (1994).

Den p-Wert kann man sich aber für ein festes n auch günstig über eine Simulation berechnen: Im j -ten Lauf werden die ϵ_t alle erzeugt aus einer $\mathcal{N}(0, a^2)$ -Verteilung, dann setzt man $Y_{t-n} = Y_{t-n-1} + \epsilon_{t-n}$ oder nimmt Y_{t-n-1} ebenfalls aus einer $\mathcal{N}(0, a^2)$ -Verteilung, dann bildet man $Y_{t-n+1} = Y_{t-n} + \epsilon_{t-n+1}$, usw. Damit wird der simulierte Wert $S_j(T_{DF})$ der Teststatistik im j -ten Lauf berechnet und dann mit dem tatsächlich berechneten Wert T_{DF} der vorliegenden Zeitreihe verglichen, ob kleiner oder nicht. Bei $j = 1$ bis M Läufen ist dann

$$P = \hat{P} = \#\{j \mid S_j(T_{DF}) < T_{DF}\}/M.$$

Die Simulationsvarianz kann man als $a^2 = \hat{\sigma}_\epsilon^2$ wählen, da die „Grenzverteilung“ unabhängig davon ist, oder aber z. B. a aus $\mathcal{N}(\bar{\sigma}_\epsilon, \hat{\sigma}_\epsilon^2/16)$ wählen.

Bei mehreren Zeitreihen kann man in jedem Simulationslauf auch gleich alle Testwerte vergleichen und erhält so simultan alle p-Werte.

Die p-Werte $P = F_{DF}(T_{DF})$ lassen sich jedoch nach MacKinnon (1994) hinreichend genau wie folgt berechnen. Mit den Polynomen

$$b(x) = 0.6344 + 1.2378 \cdot x + 0.032496 \cdot x^2 \text{ sowie}$$

$$B(x) = 0.4797 + 0.93557 \cdot x - 0.06999 \cdot x^2 + 0.033066 \cdot x^3$$

und der Verteilungsfunktion $\Phi(x)$ der $\mathcal{N}(0, 1)$ -Verteilung, vgl. Tab. 1 im Anhang, errechnen sich die Werte der Verteilungsfunktion $F_{DF}(T_{DF})$ gemäß der Fallunterscheidung:

$$\begin{aligned} T_{DF} < -19.04 & : F_{DF}(T_{DF}) \approx 0, \\ -19.04 \leq T_{DF} < -1.04 & : F_{DF}(T_{DF}) = \Phi[b(T_{DF})], \text{ und} \\ -1.04 \leq T_{DF} & : F_{DF}(T_{DF}) = \Phi[B(T_{DF})]. \end{aligned}$$

Zum Beispiel für $T_{DF} = -1.1$ ergibt sich mit $b(-1.1) = -0.69$ der p-Wert $P = 0.25$, und für $T_{DF} = -1$ mit $B(-1) = -0.56$ der p-Wert $P = 0.28$. Für $T_{DF} = 0$ erhält man $P = 0.68$ und nicht 0.5 wie bei der $\mathcal{N}(0, 1)$ -Verteilung oder der [zentralen] t-Verteilung.

11.2. Prüfung eines Panels kreuz-korrelierter Zeitreihen auf Stationarität

Haben wir ein Panel von K kreuz-korrelierten Zeitreihen vorliegen, vgl. Abschnitt 4 in Kap. XII, in dem das Kreuzkorrelogramm von jeweils zwei Zeitreihen behandelt wird,

$$(Y_t)_1, \dots, (Y_t)_K,$$

die jeweils bereits die Gestalt der Zeitreihe aus vorigem Abschnitt haben, so berechnen wir mit dem zuletzt beschriebenen Verfahren für jede Zeitreihe den p -Wert der jeweiligen Dickey-Fuller Teststatistik T_{DF} , so daß wir erhalten: P_1, \dots, P_K .

11.2.1. Kombination der Dickey-Fuller p -Werte nach Hartung

Mit der Inversen Φ^{-1} der $\mathcal{N}(0, 1)$ -Verteilungsfunktion Φ werden die Normal-Probits berechnet gemäß:

$$Z_1 = \Phi^{-1}(1 - P_1), \dots, Z_K = \Phi^{-1}(1 - P_K).$$

Hiermit berechnen wir dann die Teststatistik T_H aus Abschnitt 10.2 und daraus den kombinierten p -Wert:

$$p_{COMB} = 1 - \Phi(T_H).$$

Ist p_{COMB} nicht allzu groß, so können wir das Panel als stationär betrachten.

Demetrescu/Hassler/Tarcolea (2006) haben umfangreiche Simulationen für dieses Verfahren durchgeführt. Die untersuchten Zeitreihen haben Längen bis $n = 250$ und die Panelbreite geht dabei bis zu $K = 25$. Das Verfahren bricht nicht zusammen und hält in etwa das Testniveau. Je höher die Korrelationen der Zeitreihen sind, um so genauer trifft der Test mit T_H das Testniveau.

Clo (2001) führte die Anwendung der üblichen Inversen Normalmethode für unabhängige Zeitreihen in diesen Bereich ein.

11.2.2. Kombination der Dickey-Fuller p -Werte nach Hartung mit den Fisher-Probits

Ebenfalls wie Clo (2001) kombinieren Maddala/Wu (1999) die p -Werte unabhängiger Zeitreihen und zwar mit dem von ihnen bevorzugten Fisher Test. Die Erweiterung auf korrelierte Zeitreihen geschieht nun mit Hilfe der inversen Wilson-Hilferty Transformation.

Wir berechnen zunächst die Fisher-Probits:

$$-2 \ln P_1, \dots, -2 \ln P_K,$$

und wenden nun das Verfahren aus Abschnitt 10.3 an. Wir erhalten dann den kombinierten p -Wert zur entsprechenden Teststatistik T_H^F als:

$$p_{COMB}^F = 1 - \Phi(T_H^F).$$

11.3. Interpretation des kombinierten Tests auf Stationarität

Sind die Originalreihen stationär, dann liegen die zu beobachtenden Schwankungen um ein langfristiges Mittel und die Varianz ist endlich und nicht zeitabhängig.

Hingegen instationäre Zeitreihen zeigen keine Tendenz, auf einen langfristig deterministischen Pfad zurückzukehren und die Varianz ist zeitabhängig. Nicht-stationäre Zeitreihen erleiden permanent Einwirkungen von zufälligen Schocks und folgen somit einem Random Walk.

Etwa Hassler/Tarcolea (2005) geben eine Interpretation in Hinsicht auf **langfristige Zinsraten**, wenn in einem kombinierten Länder-Panel der wie oben kombinierte Dickey-Fuller Test eine Evidenz für Instationarität aufweist.

In einem **Aktien-Portfolio**, zum Beispiel, bilden die Zeitreihen der Aktienkurse sicherlich ein Panel kreuz-korrelierter Zeitreihen. In „ruhigen Zeiten“ sind die Originalreihen stationär. Der kombinierte p-Wert p_{COMB} der Dickey-Fuller Tests eignet sich nun gut zu einer **automatischen Überwachung** des Portfolios. Zeigt er in „ruhigen Zeiten“ einen recht kleinen Wert, so deutet ein sukzessiver Anstieg auf Veränderungen mindestens eines Kurses in Richtung Instationarität.

12. Delphi Methode für Vorhersagen und zur Gewinnung Subjektiver Wahrscheinlichkeiten

In den Abschnitten 1.9 und 1.10 von Kap. XIII wird im Rahmen der Zuverlässigkeitssanalyse von komplexen Systemen und mehrphasigen Missionen die **Fehlerbaumanalyse** besprochen. Die dort ausführlich vorgestellten realen Beispiele sind Sicherheitsanalysen eines Herzkreislaufentlastungssystems und des Feuerbekämpfungsmanagements beim Brand einer chemischen Fabrik. Für die im Fehlerbaum erfaßten Ereignisse lassen sich die Wahrscheinlichkeiten für ihr Eintreten nicht experimentell ermitteln, sondern nur durch eine Befragung von Experten subjektiv erfassen.

12.1. Anwendungsbereiche der Delphi Methode

Wie in der Risiko Analyse lassen sich in vielen Bereichen Informationen nur durch individuelle oder kollektive Beurteilungen gewinnen. Ein Verfahren zur kollektiven Informationsbildung ist die Delphi Methode, die schon Pill (1971) hinsichtlich Substanz, Kontext und Kritik durchleuchtet.

Delphi Methode und Entscheidungsfindung bezüglich Studien zu Quality of Life werden von Dalkey (1972) vorgestellt. Etwa Chakravarti et al. (1998) prognostizieren den technischen Fortschritt mittels Delphi Methode. Z. B. Häder/Häder (2000) setzen die Delphi Technik in den Sozialwissenschaften ein.

Im Bereich Wirtschaft, sagt etwa Woll (2008), habe die ergänzend eingesetzte Delphi Methode gegenüber reinen quantitativen Prognoseverfahren, wie z. B. der Regressionsanalyse, vgl. Kap. X, den Vorteil, daß zusätzliche Informationen berücksichtigt werden können.

12.2. Kombination der Einzelprognosen oder subjektiven Risikobeurteilungen mittels der Delphi Methode

Wir gehen von der Situation aus, daß von K Experten:

$$y_1, \dots, y_K \text{ Prognosen oder Risikoeinschätzungen} \\ [\text{zwischen } 0 \text{ und } 1] \text{ für } \psi_{\text{DEL}}$$

vorliegen, wobei ψ_{DEL} die interessierende Größe sei, und wir nach einem Scoringssystem oder Bonitätsschema normierte

$$\lambda_1, \dots, \lambda_k \text{ Bonituren}$$

vergeben, die wir den entsprechenden Prognosen oder Risikoeinschätzungen zuordnen. Hierbei mögen Erfahrungen, eingeschätzte Qualität oder Risikofreude, oder auch extern eingeholte Glaubwürdigkeitsbeurteilungen für die im Delphi Prozeß stehenden Experten zur Festlegung der Bonitätspunkte herangezogen werden.

Die kombinierte Delphi Prognose oder Subjektive Wahrscheinlichkeit ist dann gegeben als:

$$\hat{\psi}_{\text{DEL}} = \sum_{i=1}^K \lambda_i y_i, \quad \sum_{i=1}^K \lambda_i = 1.$$

Können die K Experten als unabhängig angesehen werden und ist nichts bekannt zur Varianz der y_i , so schätzen wir die Varianz etwa der kombinierten Prognose mit Hilfe der Formel aus Abschnitt 6.2 als $\widehat{\text{Var}}_{\text{HBK}} \hat{\psi}_{\text{DEL}}$ und können ein $(1 - \kappa)$ -Konfidenzintervall für $\hat{\psi}_{\text{DEL}}$ wie folgt angeben:

$$\text{KI}(\hat{\psi}_{\text{DEL}})_{\text{HBK}} = \hat{\psi}_{\text{DEL}} \pm \sqrt{\widehat{\text{Var}}_{\text{HBK}} \hat{\psi}_{\text{DEL}}} \cdot D_\kappa,$$

wobei $D_\kappa = u_{1-\kappa/2}$ ist, mit dem $(1 - \kappa/2)$ -Quantil $u_{1-\kappa/2}$ der $\mathcal{N}(0, 1)$ -Verteilung, wenn zumindest $\hat{\psi}_{\text{DEL}}$ als genähert normalverteilt angesehen werden kann.

Andernfalls setzen wir $D_{0.11} = 3$, so daß ausgehend von der Tschebyscheff-Ungleichung, vgl. Abschnitt 8.2A in Kap. II, das Niveau des Konfidenzintervalls für $\hat{\psi}_{\text{DEL}}$ mindestens ca. 89% beträgt, bzw. bei $D_{0.25} = 2$ ca. 75%.

Müssen die Experten als abhängig betrachtet werden, so sollte die Fragestellung derart aufbereitet werden, daß die Antworten als p-Werte interpretiert werden können, vgl. Abschnitt 10.1. Dies ermöglicht dann eine Vorgehensweise gemäß Abschnitt 10.2 für abhängige p-Werte.

Sind den einzelnen Prognosen oder Risikoeinschätzungen Streubereiche in numerischer Form beigelegt, so können Kombinationsverfahren des Abschnitts 8 herangezogen werden.

12.3. Beispiel Investitionsberatung: Cash Flow Vorhersagen ohne Zuverlässigkeitssangaben

Für die in Abschnitt 8.1 dargelegte Situation der Investitionsberatung nehmen wir jetzt die Vorhersage des Betriebs $y_B = 15\%$ und des Unternehmensberaters $y_U = 10\%$ zur Delphi-Kombination für die Vorhersage der Cash Flow Steigerung. Die Bonituren vergeben wir in Unkenntnis der dort angegebenen Varianzen gleichwertig als:

$$\lambda_1 = 0.5 \text{ und } \lambda_2 = 0.5.$$

Folglich lautet die kombinierte Prognose

$$\hat{\psi}_{\text{DEL}} = 12.5\%,$$

und mit der Varianzformel aus Abschnitt 6.2 ergibt sich die Schätzung der Varianz von $\hat{\psi}_{\text{DEL}}$ zu:

$$\widehat{\text{Var}}_{\text{HBK}}(\hat{\psi}_{\text{DEL}}) = 6.25,$$

bzw. die geschätzte Standardabweichung zu 2.5%. Mittels Tschebyscheff Ungleichung erhalten wir das 75%-Konfidenzintervall für ψ_{DEL} als [7.5%, 17.5%], was bei nur zwei Prognosen und ohne jegliche Angabe zur jeweiligen Zuverlässigkeit der Prognosen, wobei zusätzlich auch keine Verteilungsannahme getroffen wurde, ein recht ordentliches Resultat ist. Nehmen wir approximativ Normalverteilung für die kombinierte Prognose an, so ist $u_{0.75} = 0.675$ und damit dann das 75%-Konfidenzintervall für ψ_{DEL} gegeben als:

$$\text{KI}(\psi_{\text{DEL}}) = [10.82\%, 14.19\%].$$

Der Fall abhängiger Prognosen wird für dieses Beispiel in Abschnitt 10.4 ausführlich dargestellt. Im Abschnitt 8 werden Streuungsangaben zu den Prognosen mitverarbeitet, wie auch Korrelationen, wenn bekannt.

12.4. Iterative Delphi Methode

Etwa bei Gruppenentscheidungen mit einem verzerrnden Einfluß durch einen dominierenden Meinungsführer (risk-shift Phänomen) werden die Gruppenbefragungen wiederholt durchgeführt, um einen Ausgleich dieser Verzerrungen zu erreichen, vgl. z. B. Bamberg/Baur/Krapp (2008), Woll (2008).

Liegt das kombinierte Resultat einer ersten Delphi Befragung vor, so wird dieses mit oder ohne Angabe des Streubereichs wieder an alle Experten zurückgegeben zwecks Überdenkens etwa ihrer zuerst abgegebenen Prognose. Liegt diese weit entfernt von der kombinierten Prognose, mag das Anlaß dafür geben zu überprüfen, ob man vielleicht wesentliche Punkte beim ersten Mal übersehen hat, um nun daraufhin die zweite Prognose zu verbessern.

So entsteht eine zweite Delphi kombinierte Gesamteinschätzung für ψ_{DEL} . Diesen Vorgang kann man solange wiederholen bis praktisch kein Unterschied mehr zwischen den Prognosen besteht.

13. Heterogenitätsmaß für Meta-Analysen: Verallgemeinerter Intraklass-Korrelationskoeffizient

Mit bekannten Varianzen ist im Random Effects Modell $y_i = \psi + a_i + e_i$ mit $\text{Var}(a_i) = \tau^2$ und $\text{Var}(e_i) = \sigma_i^2$, vgl. Abschnitt 3, der gemeinsame unverzerrte Schätzer kleinster Varianz, der Gauß-Markov Schätzer, für ψ gegeben als [vgl. Abschnitt 17.5]:

$$\hat{\psi}_{GM} = \sum_{i=1}^K \gamma_i y_i, \quad \gamma_i = \omega_i / \omega_\Sigma, \quad \omega_i = (\tau^2 + \sigma_i^2)^{-1}, \quad \omega_\Sigma = \sum_{j=1}^K \omega_j,$$

dessen Varianz zerfällt in:

$$\text{Var } \hat{\psi}_{GM} = \tau^2 \sum_{i=1}^K \gamma_i^2 + \sum_{i=1}^K \gamma_i^2 \sigma_i^2.$$

Mit $\tau^2 + \xi_i^2 = \text{Var}(a_i + e_{ij})$ bezeichnen wir die totale Varianz einer, wenn auch mitunter nur gedachten, Versuchseinheit in der i -ten Studie: Ist in einer einarmigen Studie $y_i = \sum_{j=1}^{n_i} y_{ij}/n_i$ bzw. $e_i = \sum_{j=1}^{n_i} e_{ij}/n_i$ mit $y_{ij} = \psi + a_i + e_{ij}$, dann ist $\xi_i^2 = \text{Var } e_{ij} = n_i \sigma_i^2$, in einer zweiarmigen Studie mit den Beobachtungen $y_{E_{ij_1}}$ und $y_{C_{ij_2}}$ ist $\xi_i^2 \approx \text{Var}(e_{E_{ij_1}} - e_{C_{ij_2}})$, so daß wir allgemein, auch z. B. für Effektmaße binärer Daten, setzen:

$$\xi_i^2 = \bar{n}_{ih} \sigma_i^2 \text{ mit } \bar{n}_{ih} = \frac{2}{1/n_{E_i} + 1/n_{C_i}},$$

wenn die Gruppen E und C der Umfänge n_{E_i} und n_{C_i} verglichen werden, sonst $\bar{n}_{ih} = n_i$ bei einarmiger Studie des Umfangs n_i . Wird die Rest-Varianz ξ_i^2 einer Versuchseinheit VE_i der i -ten Studie mit Δ_i gewichtet, so erhalten wir für eine durchschnittliche Versuchseinheit \bar{VE} mit $\text{Var } \bar{VE} = \xi_{\text{Total}}^2$ die **Varianzzerlegung**:

$$\xi_{\text{Total}}^2 = \tau^2 + \sum_{i=1}^K \Delta_i \xi_i^2 \text{ bzw. } \tau^2 = I_{H\Delta} \times \xi_{\text{Total}}^2,$$

wobei $I_{H\Delta}$ den Anteil von τ^2 an der Totalvarianz angibt und somit als **Maß für die Inhomogenität** des RE-Modells gegeben ist als

$$I_{H\Delta} = \frac{\tau^2}{\tau^2 + \sum_{i=1}^K \Delta_i \xi_i^2}.$$

Für $\sum \Delta_i = 1$ ist $I_{H\Delta}$ eine direkte Erweiterung des **Intraklass-Korrelationskoeffizienten** auf den Fall heteroskedastischer Fehlervarianzen, der in der Genetik auch **Heritabilitätskoeffizient** genannt wird, vgl. Abschnitt 4.1 in Kap. XI. Vernünftige Festlegungen sind:

$$\Delta_i = 1/K \text{ oder } \Delta_i = \lambda_i = \gamma_i^2 / \left(\sum_{j=1}^K \gamma_j^2 \right).$$

Für $\Delta_i = \lambda_i / \bar{n}_{ih}$ erhalten wir das Higgins/Thompson (2002) zurechenbare Maß I^2 , welches außer der Heterogenität simultan auch die Stichprobenumfänge der beteiligten Studien bewertet: Steigen die Umfänge an, so konvergiert I^2 gegen 1, auch wenn $\tau^2 > 0$ dabei unverändert bleibt.

13.1. Hartung-Konfidenzintervalle für die Heterogenitätsmaße

Bezeichnen $S_\Delta = \sum_{i=1}^K \Delta_i \xi_i^2$ und $\widehat{S}_\Delta = \sum_{i=1}^K \Delta_i \widehat{\xi}_i^2$, mit $\widehat{\xi}_i^2 = \bar{n}_{ih} \widehat{\sigma}_i^2$, so setzen wir

$$W_{H\Delta} = R_H / \widehat{S}_\Delta, \text{ mit}$$

$$R_H = \frac{1}{K-1} \sum_{i=1}^K \Gamma_i (y_i - \widehat{\psi}_{GM})^2, \text{ wobei } \Gamma_i = \gamma_i / \sum_{j=1}^K \gamma_j^2.$$

Wir erhalten dann mit den Quantilen der F-Verteilung, vgl. Tab. 5 im Anhang, die genäherte Aussage: Mit Wahrscheinlichkeit $1 - \alpha$ gilt

$$F_{K-1, \nu(\Delta); \alpha/2} \leq \frac{W_{H\Delta} S_\Delta}{\tau^2 + \sum_{i=1}^K \lambda_i \sigma_i^2} \leq F_{K-1, \nu(\Delta); 1-\alpha/2}.$$

Hierbei wird durch Gleichsetzen der ersten zwei Momente die Verteilung von \widehat{S}_Δ durch eine skalierte χ^2 -Verteilung approximiert gemäß Patnaik (1949):

$$\nu(\Delta) \widehat{S}_\Delta / E \widehat{S}_\Delta \stackrel{\text{approx}}{\sim} \chi_{\nu(\Delta)}^2 \text{ mit } \nu(\Delta) = 2 \frac{(E \widehat{S}_\Delta)^2}{\text{Var } \widehat{S}_\Delta},$$

wobei $\text{Var } \widehat{S}_\Delta$ gegebenenfalls mit dem Gaußschen Fehlerfortpflanzungsgesetz, vgl. Abschnitt 4.2 in Kap. V, bestimmt wird, s. Hartung (1999a) für eine Anwendung u. a. im Fall des binären Effektmaßes Risiko Differenz. Wenn die Studienumfänge nicht ausgesprochen klein sind, können wir auch $\nu(\Delta) = \infty$ verwenden. Weiterhin setzen wir:

$$\theta = \frac{\tau^2}{S_\Delta} \text{ und } \Lambda = \frac{\sum_{i=1}^K \lambda_i \sigma_i^2}{S_\Delta}$$

und erhalten damit das $(1 - \alpha)$ -Konfidenzintervall für θ :

$$\begin{aligned} L = \frac{W_{H\Delta}}{F_{K-1, \nu(\Delta); 1-\alpha/2}} - \Lambda &\leq \theta \leq \frac{W_{H\Delta}}{F_{K-1, \nu(\Delta); \alpha/2}} - \Lambda = U \\ &\leq U^* = \frac{W_{H\Delta}}{F_{K-1, \nu(\Delta); \alpha/2}}. \end{aligned}$$

Für $L \geq 0$ und $U \geq 0$ wenden wir jetzt die monotone Transformation $g(x) = x/(1+x)$ sowohl auf beide Grenzen an als auch auf θ :

$$g(\theta) = \frac{\tau^2}{S_\Delta} / \left(1 + \frac{\tau^2}{S_\Delta} \right) = \frac{\tau^2}{\tau^2 + S_\Delta} = I_{H\Delta}.$$

Falls $L < 0$ oder $U < 0$, so werden L bzw. U auf 0 gesetzt. So erhalten wir das $(1 - \alpha)$ -Konfidenzintervall für das Heterogenitätsmaß $I_{H\Delta}$ als:

$$\frac{L}{1+L} \leq I_{H\Delta} \leq \frac{U}{1+U} \leq \frac{U^*}{1+U^*}.$$

Für eine Realisation werden die verbliebenen Varianzen in den Grenzen geschätzt, wie auch Λ , außer für $\Delta_i = \lambda_i/\bar{n}_{ih}$, d. h. für I^2 , da in dem Fall $\Lambda = 1$ ist. Dabei sei darauf hingewiesen, daß Λ in U^* nicht mehr enthalten ist.

Das $(1 - \alpha)$ -Konfidenzintervall für $\theta = \tau^2/S_\Delta$ mit $\Delta_i = \lambda_i/\bar{n}_{ih}$, d. h. $S_\Delta = \sum_{i=1}^K \lambda_i \sigma_i^2$, eignet sich für eine **Homogenitätsanalyse** der beteiligten Studien oder Prognosen: Ist $L > 0$, so ist zum Niveau $\alpha/2$ signifikant: $\tau^2 > 0$, und somit sind die Studien oder Prognosen **signifikant inhomogen**. Ist hingegen $U \leq 0$, so ist zum Niveau $\alpha/2$ signifikant: $\tau^2 = 0$, und somit sind die Studien oder Prognosen dann als **signifikant homogen** nachgewiesen.

13.2. Beispiel Landwirtschaft–Genetik: Erblichkeit der Besamungsfähigkeit

Wir setzen hier das Beispiel aus Abschnitt 3.4 fort, da das Inhomogenitätsmaß $I_{H\Delta}$ hier positiv verstanden wird als Heritabilitätskoeffizient, der für den Grad der Ähnlichkeit oder Erblichkeit bezüglich der Besamungsfähigkeit der Bullen des Zuchttamms steht. Ist $I_{H\Delta} = 0$, so bedeutet das hier, daß züchterisch dieses Merkmal nicht beeinflußt werden kann.

Hier ist $\hat{\xi}_i^2 = s_i^2$ die Stichprobenvarianz aus Tab. 5 in Abschnitt 3.4, woher wir auch die DSL-Schätzung 66.8 für τ^2 übernehmen und damit die geschätzten Gewichte $\hat{\gamma}_i = c_i$ berechnen. Wir betrachten die folgenden Funktionen der Umweltvarianzen s_i^2 :

$$S_1 = \hat{S}_\Delta(1) = \frac{1}{K} \sum_{i=1}^K s_i^2,$$

$$S_2 = \hat{S}_\Delta(2) = \sum_{i=1}^K c_i^2 s_i^2 \Big/ \sum_{j=1}^K c_j^2 \quad \text{und}$$

$$S_3 = \hat{S}_\Delta(3) = \sum_{i=1}^K (c_i^2/n_i) s_i^2 \Big/ \sum_{j=1}^K c_j^2, \quad \text{sowie}$$

$$\widehat{W}_{Hk} = \widehat{R}_H / S_k \quad \text{für } k = 1, 2, 3, \text{ hierbei ist}$$

$$\widehat{R}_H = q_H \Big/ \sum_{i=1}^K c_i^2 = 117.85088,$$

s. Abschnitt 3.4 und **Tab. 24**, in der die hier benötigten Berechnungen zusammengestellt sind, so daß wir sofort erhalten:

$$S_1 = 232.42898, \quad S_2 = 193.85161 \quad \text{und} \quad S_3 = 33.60096.$$

Die beste invariante unverzerrte Schätzung für deren Varianzen ermitteln wir nach Hartung/Voet (1986) gemäß:

$$\widehat{\text{Var}} \widehat{S}_{\Delta} = 2 \sum_{i=1}^K \widehat{\Delta}_i^2 s_i^4 / (n_i + 1),$$

und erhalten:

$$\widehat{\text{Var}} S_1 = 3693.31016, \widehat{\text{Var}} S_2 = 1857.76174 \text{ und } \widehat{\text{Var}} S_3 = 65.18058.$$

Mit Hilfe des Steinerschen Verschiebungssatzes, s. Abschnitt 8.2.A in Kap. II, schätzen wir die Freiheitsgrade der jeweils angepaßten χ^2 -Verteilung gemäß:

$$\widehat{\nu}_i = \widehat{\nu}(S_i) = 2 \frac{s_i^2}{\widehat{\text{Var}} S_i} - 2$$

zu: $\widehat{\nu}_1 = 27$, $\widehat{\nu}_2 = 38$ und $\widehat{\nu}_3 = 32$.

Das Heterogenitätsmaß $I_{H\Delta}$ bezeichnen wir entsprechend mit I_{H1} und als Konfidenzniveau wählen wir $1 - \alpha = 0.95$. Für S_1 berechnen wir: $\widehat{W}_{H1} = 0.50704$, $\widehat{\Lambda} = 33.60/232.429 = 0.1446$, $F_{5,27;0.975} = 3.063$, $F_{5,27;0.025} = 0.1602$, $L = [0.50704/3.063] - \widehat{\Lambda} = 0.021$, $U = [0.50704/0.1602] - \widehat{\Lambda} = 3.0205$ und $U^* = 3.165$, so daß wir schließlich das 95%-Konfidenzintervall für das Inhomogenitätsmaß oder den verallgemeinerten Heritabilitätskoeffizienten I_{H1} erhalten als:

$$\frac{L}{1+L} = 0.021 \leq I_{H1} \leq 0.751 = \frac{U}{1+U} \leq 0.76 = \frac{U^*}{1+U^*}.$$

Für S_2 erhalten wir entsprechend: $\widehat{W}_{H2} = 0.60794$, $\widehat{\Lambda} = 33.60/193.85 = 0.173$, $F_{5,38;0.975} = 3$, $F_{5,38;0.025} = 0.1617$, $L = [0.60794/3] - 0.173 = 0.0293$ und $U = [0.60794/0.1616] - 0.173 = 3.6121$, sowie $U^* = 3.7854$, so daß wir für das zweite Inhomogenitätsmaß oder den zweiten verallgemeinerten Heritabilitätskoeffizienten I_{H2} das 95%-Konfidenzintervall erhalten gemäß:

$$\frac{L}{1+L} = 0.028 \leq I_{H2} \leq 0.783 = \frac{U}{1+U} \leq 0.791 = \frac{U^*}{1+U^*}.$$

Für S_3 erhalten wir analog: $\widehat{W}_{H3} = 3.50737$, $\widehat{\Lambda} = 1$, $F_{5,32;0.975} = 3.026$, $F_{5,32;0.025} = 1/F_{32,5;0.975} = 0.1606$, $L = [3.50737/3.026] - 1 = 0.1591$ und $U = [3.50737/0.1606] - 1 = 20.8387$, so daß das 95%-Konfidenzintervall für das Heterogenitätsmaß $I_{H3} = I^2$ [Higgins/Thompson] gegeben ist als:

$$\frac{L}{1+L} = 0.14 \leq I_{H3} = I^2 \leq 0.95 = \frac{U}{1+U},$$

welches wesentlich höhere Grenzen aufweist als die beiden anderen Konfidenzintervalle. Mit der DSL-Schätzung 66.8 für τ^2 erhalten wir z. B. folgende

Punktschätzungen:

$$\begin{aligned}\hat{I}_{H1} &= \frac{66.8}{66.8 + 232.43} = 0.223, \\ \hat{I}_{H2} &= \frac{66.8}{66.8 + 193.85} = 0.256 \text{ und} \\ \hat{I}_{H3} &= \hat{I}^2 = \frac{66.8}{66.8 + 33.6} = 0.665.\end{aligned}$$

In diesem Beispiel liegt der durchschnittliche Stichprobenumfang der beteiligten Studien unter 6, so daß die Unterschiede in den Punktschätzungen für $I_{H3} = I^2$ und I_{H1} bzw. I_{H2} nicht so extrem ausfallen.

13.3. Ein weiterer expliziter Hartung-Schätzer für den Heterogenitätsparameter

Als Nebenprodukt erhalten wir hier einen neuen Schätzer für τ^2 . Für die in Abschnitt 13.1 oben eingeführte Statistik R_H gilt im Erwartungswert:

$$E R_H = \tau^2 + \sum_{i=1}^K \lambda_i \sigma_i^2, \text{ mit } \lambda_i = \gamma_i^2 / \sum_{j=1}^K \gamma_j^2.$$

Folglich ist ein unverzerrter Schätzer für τ^2 gegeben durch: $R_H - \sum_{i=1}^K \lambda_i \hat{\sigma}_i^2$, so daß wir entsprechend einen neuen Schätzer für τ^2 erhalten als:

$$\hat{\tau}_{HART}^2 = \hat{R}_H - \sum_{i=1}^K c_i^2 \hat{\sigma}_i^2 / \sum_{j=1}^K c_j^2.$$

Beispiel: Im eben behandelten **Bullen**-Beispiel haben wir berechnet:

$$\hat{R}_H = 117.85 \text{ und } S_3 = 33.60,$$

so daß wir erhalten:

$$\hat{\tau}_{HART}^2 = 117.85 - 33.60 = 84.25.$$

Damit liegt der Schätzer in der Region des Hedges Schätzers $\hat{\tau}_{HED}^2 = 89.68$, vgl. Abschnitt 3.4.

Da der neue Schätzer ebenfalls vom Prinzip der Unverzerrtheit ausgeht, kann er natürlich auch negativ ausfallen: Im **Beispiel** der 4 Studien zur **Multiplen Sklerose (MS)** erhalten wir aus Abschnitt 3.3 für q_H den Wert 1.18303, so daß mit den Berechnungen aus **Tab. 25** folgt:

$$\hat{\tau}_{HART}^2 = \frac{1.18303}{0.35785} - \frac{1.29188}{0.35785} = -0.304.$$

Bereits negativ in diesem Beispiel sind der Hedges Schätzer: -3.11 und der DerSimonian-Laird Schätzer: -1.28 in Abschnitt 3.1 berechnet, während der positive Schätzer $\hat{\tau}_{HAR}^2$ den Wert 1.58 annimmt.

Tab. 24: Landwirtschaft, vgl. Tab. 5, Heterogenitätsmaß bzw. Erblichkeit der Besamungsfähigkeit des Bullen-Zuchtstammes, Berechnungen mit $\hat{\tau}_{DSL}^2 = 66.8$.

Bulle					
i	n _i	s _i ²	c _i ²	c _i ² s _i ²	c _i ² s _i ² /n _i
1	5	175.8276	0.02764	4.86064	0.97213
2	2	60.5284	0.03050	1.84614	0.92307
3	7	132.9409	0.03903	5.18897	0.74128
4	5	505.3504	0.01020	5.15222	1.03044
5	7	270.6025	0.02584	6.99225	0.99889
6	9	249.3241	0.03216	8.01702	0.89078
$\Sigma:$	35	1394.5739	0.16537	32.05724	5.55659

Tab. 25: Multiple Sklerose - Arbeitstabelle zur Berechnung des neuen Schätzers $\hat{\tau}_{HART}^2$ für den Heterogenitätsparameter τ^2 ; vgl. Tab. 2 und Tab. 4 in den Abschnitten 2 und 3

MS-Studien				
i	c _i	σ_i^2	c _i ²	c _i ² · σ_i^2
1	0.5305	1.924	0.28143	0.54147
2	0.1573	10.2	0.092474	0.25238
3	0.19541	7.94	0.03819	0.030319
4	0.11617	14.437	0.01350	0.19483
$\Sigma:$	-	-	0.35785	1.29188

13.4. Median Unverzerrte Hartung-Schätzer für die Heterogenitätsmaße

Da die Schätzungen für τ^2 recht unterschiedlich ausfallen, sind die damit berechneten Punktschätzungen für $I_{H\Delta}$ ebenfalls recht verschieden. Außerdem hat eine derart konstruierte Schätzung keine schätz-theoretischen Eigenschaften, sondern stellt nur eine Hilfslösung dar. Daher stellt der folgende neue Schätzer, der $I_{H\Delta}$ direkt schätzt, eine echte Lösung dieses Schätzproblems dar.

Wir setzen, s. obigen Abschnitt 13.1,

$$H_\Delta = \frac{W_{H\Delta}}{F_{K-1, \nu(\Delta); 0.5}} - \Lambda, \text{ wobei genähert:}$$

$$F_{K-1, \nu(\Delta); 0.5} \approx \exp \left\{ \frac{2}{3} \left(\frac{1}{K-2} + \frac{1}{\nu(\Delta)-1} - 1 \right) \cdot \left(\frac{1}{K-2} - \frac{1}{\nu(\Delta)-1} \right) \right\},$$

dann wird der Median Unverzerrte Schätzer für $I_{H\Delta}$ definiert als:

$$\hat{I}_{H\Delta, HAR} = \frac{H_\Delta}{1 + H_\Delta},$$

d. h. der Schätzer $\hat{I}_{H\Delta, HAR}$ fällt mit gleicher Wahrscheinlichkeit sowohl kleiner als auch größer als das wahre Maß $I_{H\Delta} = \tau^2 / (\tau^2 + S_\Delta)$ aus. Darüberhinaus liegt er in allen obigen $(1 - \alpha)$ -Konfidenzintervallen für $I_{H\Delta}$, wenn α zwischen 0 und 1 variiert. Für eine Realisation werden dann die verbliebenen Varianzen in H_Δ durch ihre Schätzungen ersetzt.

Beispiel: Für das in Abschnitt 13.2 behandelte Bullen-Beispiel erhalten wir mit den dort bereits berechneten Größen und eingeführten Bezeichnungen:

$$\begin{aligned} H_\Delta(1) &= \frac{\widehat{W}_{H1}}{F_{5,27;0.5}} - \widehat{\Lambda} = \frac{0.50704}{0.89244} - 0.1446 = 0.42355, \\ H_\Delta(2) &= \frac{0.60794}{0.88595} - 0.173 = 0.51320 \text{ und} \\ H_\Delta(3) &= \frac{3.50737}{0.88893} - 1 = 2.94560, \end{aligned}$$

so daß wir die Median Unverzerrten Schätzungen gemäß $H_\Delta(k) / (1 + H_\Delta(k))$ für I_{Hk} erhalten als:

$$\hat{I}_{H1, HAR} = 0.30, \quad \hat{I}_{H2, HAR} = 0.34 \text{ und } \hat{I}_{H3, HAR} = \hat{I}_{HAR}^2 = 0.75.$$

Obige Näherungsformel ergibt etwa für $F_{5,27;0.5} = 0.89244$ in $H_\Delta(1)$ den guten Approximationswert $e^{-0.10035} = 0.90452$.

In der im Rahmen des einleitenden Abschnitts 1.4 plazierten Abb. 6 sind diese Schätzungen mit den zugehörigen 95%-Konfidenzintervallen aus 13.2 auch graphisch veranschaulicht.

13.5. Zusammenfassendes Beispiel: Zweiarmige Studien zu Multipler Sklerose MS

Um die Heterogenitätsmaße auch für zweiarmige Studien zu demonstrieren, betrachten wir hierfür die in Abschnitt 2.5 eingeführten 4 Studien zur Progression der Behinderung bei Patienten mit Multipler Sklerose (MS), deren Versuchsdaten in der dort plazierten Tab. 1 aufgeführt sind. Daraus ergeben sich die Varianzschätzungen $\hat{\sigma}_i^2 = s_{E_i}^2/n_{E_i} + s_{C_i}^2/n_{C_i}$ und $\hat{\xi}_i^2 = \bar{n}_{ih} \cdot \hat{\sigma}_i^2$ mit $\bar{n}_{ih} = 2/(1/n_{E_i} + 1/n_{C_i})$. Mit den Gewichten $\Delta_i(I) = 1/K$, $\Delta_i(II) =$

$\gamma_i^2 / \sum \gamma_j^2 = \lambda_i$ und $\Delta_i(\text{III}) = \lambda_i / \bar{n}_{ih}$ erhalten wir entsprechend aus Tab. 26 für $\widehat{S}_\Delta = \sum \Delta_i \cdot \widehat{\xi}_i^2$ hier die Schätzungen:

$$S_I = 279.813, S_{II} = 307.052 \text{ und } S_{III} = 3.610.$$

Aus Abschnitt 3.3 und Tab. 25 entnehmen wir weiter:

$$\widehat{R}_H = q_H / \sum_{i=1}^4 c_i^2 = \frac{1.18303}{0.35785} = 3.30594,$$

so daß sich für $W_{H\Delta} = R_H / \widehat{S}_\Delta$ die entsprechenden Schätzungen ergeben zu:

$$W_I = 0.01181, W_{II} = 0.01077 \text{ und } W_{III} = 0.91577.$$

Die zweiten Freiheitsgrade $\nu(\Delta)$ in der obigen F-Statistik können wir hier auf Grund der nicht ausgesprochen geringen Fallzahlen in der Summe der 4 Studien auf Unendlich setzen: $\nu(\Delta) = \infty$. So bietet es sich dann an, die F-Quantile $F_{K-1,\infty;\kappa}$ mittels der ausführlicher vertafelten χ^2_{K-1} -Quantile zu ersetzen durch $\chi^2_{K-1;\kappa} / (K - 1)$. Damit ergibt sich direkt:

$$H_\Delta = \frac{(K - 1) \cdot W_{H\Delta}}{\chi^2_{K-1;0.50}} - \Lambda, \text{ mit } \Lambda = \frac{S_{\Delta(\text{III})}}{S_\Delta},$$

und mit $\chi^2_{3;0.50} = 2.366$ ergeben sich daraus gemäß $H_\Delta(i) = 3 \cdot W_i / 2.366 - \widehat{\Lambda}_i$, für $i = I, II, III$, die Schätzungen:

$$H_\Delta(I) = 0.0020746, H_\Delta(II) = 0.0018960 \text{ und } H_\Delta(III) = 0.1611623.$$

Mittels $\widehat{I}_{H\Delta}(i) = H_\Delta(i) / [1 + H_\Delta(i)]$ erhalten wir schließlich die **Median Unverzerrten Schätzungen** für die Inhomogenitätsmaße $I_{H\Delta}(i)$ als:

$$\begin{aligned}\widehat{I}_{H\Delta}(I)_{\text{HAR}} &= 0.00207, \widehat{I}_{H\Delta}(II)_{\text{HAR}} = 0.00189 \text{ und} \\ \widehat{I}_{H\Delta}(III)_{\text{HAR}} &= 0.13879 = \widehat{I}_{\text{HAR}}^2.\end{aligned}$$

Bedingt durch die höheren Stichprobenumfänge im Vergleich zu vorigem Beispiel ist hier die letzte Schätzung [für I^2] rund **70** mal größer als die beiden anderen Schätzungen.

Zur Berechnung der $(1 - \alpha)$ -Konfidenzintervalle benötigen wir die Hilfsgrößen:

$$L = \frac{(K - 1) \cdot W_{H\Delta}}{\chi^2_{K-1;1-\alpha/2}} - \Lambda \text{ und } U = \frac{(K - 1) \cdot W_{H\Delta}}{\chi^2_{K-1;\alpha/2}} - \Lambda, \quad U^* = U + \Lambda.$$

Mit $\chi^2_{3;0.975} = 9.348$ und $\chi^2_{3;0.025} = 0.216$ für $\alpha = 0.05$ ergeben sich diese Größen zu:

$$L_I = -0.00911, L_{II} = -0.00830 \text{ und } L_{III} = -0.70611, \text{ sowie}$$

$$U_I = 0.13138, U_I^* = 0.14428, U_{II} = 0.13782, U_{II}^* = 0.14958 \text{ und}$$

$$U_{III} = 11.71902,$$

so daß wir für die Heterogenitätsmaße $I_{H\Delta(i)}$ in den 4 zweiarmigen MS-Studien die folgenden 95%-Konfidenzintervalle erhalten:

$$\begin{aligned} 0 \leq I_{H\Delta(I)} &\leq \frac{U_I}{1+U_I} = 0.116 &< \frac{U_I^*}{1+U_I^*} = 0.126, \\ 0 \leq I_{H\Delta(II)} &\leq \frac{U_{II}}{1+U_{II}} = 0.121 &< \frac{U_{II}^*}{1+U_{II}^*} = 0.130 \text{ und} \\ 0 \leq I_{H\Delta(III)} &= I^2 \leq \frac{U_{III}}{1+U_{III}} = 0.921, \end{aligned}$$

wobei hier die obere Grenze für I^2 rund das **7-fache** der anderen oberen Konfidenzgrenzen beträgt. In **Abb. 19** sind all die interessanten Gegebenheiten auch graphisch veranschaulicht.

Inhomogenitätsmaße
Vier Zweiarmige Studien zu Multipler Sklerose MS

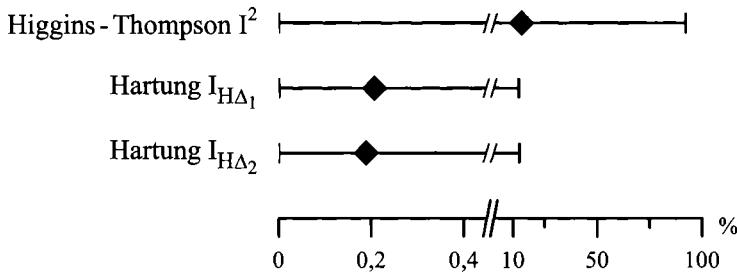


Abb. 19: Median Unverzerrte Schätzer und
95% – Konfidenzintervalle für die Inhom. – Maße

14. Normalverteilte Daten: Exakte Verfahren und spezielle Effektmaße

In diesem Abschnitt gehen wir davon aus, daß in der i-ten Studie, von K unabhängigen Studien, unabhängige Beobachtungen aus realisierten Normalverteilungen vorliegen:

$$\begin{aligned} x_{E_{ij}} &\sim \mathcal{N}(\mu_E, \xi_E^2), \quad \text{für } j = 1, \dots, n_{E_i}, \text{ und} \\ x_{C_{ij}} &\sim \mathcal{N}(\mu_C, \xi_C^2), \quad \text{für } j = 1, \dots, n_{C_i}, \end{aligned}$$

mit den Stichprobenmitteln

$$\bar{x}_{E_i} = \frac{1}{n_{E_i}} \sum_{j=1}^{n_{E_i}} x_{E_{ij}} \quad \text{und} \quad \bar{x}_{C_i} = \frac{1}{n_{C_i}} \sum_{j=1}^{n_{C_i}} x_{C_{ij}},$$

Tab. 26: Multiple Sklerose - Arbeitstabelle für Heterogenitätsmaße, vgl. Versuchsdaten in Tab. 1, Abschnitt 2.5, und Tab. 25 in Abschnitt 13.3

MS-Studien					
i	n _{E_i}	n _{C_i}	\bar{n}_{ih}	$\hat{\xi}_i^2 = \bar{n}_{ih}\hat{\sigma}_i^2$	$c_i^2\hat{\xi}_i^2$
1	162	175	168.25	323.713	91.10255
2	15	20	17.14	174.828	4.32524
3	30	32	30.97	245.902	9.39100
4	27	25	25.96	374.807	5.05989
$\sum:$	-	-	-	1119.250	109.87868
$\bar{n}_{ih} = 2/(1/n_{E_i} + 1/n_{C_i})$, $\hat{\sigma}_i^2 = s_{E_i}^2/n_{E_i} + s_{C_i}^2/n_{C_i}$, $c_i = w_i/w_\Sigma$					
$w_i = 1/(\hat{\tau}_{HAR}^2 + \hat{\sigma}_i^2)$, $\hat{\tau}_{HAR}^2 = 1.58418$, $w_\Sigma = \sum w_i$					
$S_I = \sum \hat{\xi}_i^2 / K = 279.813$, $S_{II} = \sum c_i^2 \hat{\xi}_i^2 / \sum c_j^2 = 307.052$,					
$S_{III} = \sum c_i^2 \hat{\sigma}_i^2 / \sum c_j^2 = 3.610$					
$\hat{\Lambda}_I = S_{III}/S_I = 0.01290$, $\hat{\Lambda}_{II} = S_{III}/S_{II} = 0.01176$, $\Lambda_{III} = 1$					

den Stichprobenvarianzen, für $n_{E_i} \geq 2$ und $n_{C_i} \geq 2$,

$$s_{E_i}^2 = \frac{1}{n_{E_i} - 1} \sum_{j=1}^{n_{E_i}} (x_{E_{ij}} - \bar{x}_{E_i})^2 \text{ und } s_{C_i}^2 = \frac{1}{n_{C_i} - 1} \sum_{j=1}^{n_{C_i}} (x_{C_{ij}} - \bar{x}_{C_i})^2,$$

sowie der gepoolten Varianz, vgl. Abschnitt 1.1.2 in Kap. VIII,

$$s_i^2 = \frac{(n_{E_i} - 1)s_{E_i}^2 + (n_{C_i} - 1)s_{C_i}^2}{n_{E_i} + n_{C_i} - 2}.$$

Die mit einem Parameter λ gebildete Differenz der zu Grunde liegenden Responsevariablen X_E und X_C , d. h.

$$D(\lambda) = X_E - \lambda X_C, \text{ wobei } X_E \sim \mathcal{N}(\mu_E, \xi_E^2), X_C \sim \mathcal{N}(\mu_C, \xi_C^2),$$

wird in der i-ten Studie geschätzt durch $\hat{D}(\lambda)_i = \bar{x}_{E_i} - \lambda \bar{x}_{C_i}$. Die zugehörige Varianz wird mit $\nu_i = \nu_i(\lambda)$ zugeordneten Freiheitsgraden geschätzt mittels:

$$\hat{\sigma}_i^2(\lambda) = \left(\frac{1}{n_{E_i}} + \frac{\lambda^2}{n_{C_i}} \right) s_i^2,$$

mit $\nu_i = n_{E_i} + n_{C_i} - 2 = \nu_i(\lambda)$, falls $\xi_E^2 = \xi_C^2$, andernfalls:

$$\hat{\sigma}_i^2(\lambda) = \frac{s_{E_i}^2}{n_{E_i}} + \lambda^2 \frac{s_{C_i}^2}{n_{C_i}}, \text{ und gemäß Satterthwaite (1946):}$$

$$\nu_i = \nu_i(\lambda) = \frac{(s_{E_i}^2/n_{E_i} + \lambda^2 \cdot s_{C_i}^2/n_{C_i})^2}{(s_{E_i}^2/n_{E_i})^2 / (n_{E_i} - 1) + \lambda^4 (s_{C_i}^2/n_{C_i})^2 / (n_{C_i} - 1)},$$

oder: $\nu_i = n_{E_i} + n_{C_i} - 2 = \nu_i(\lambda)$, falls $n_{E_i} \approx n_{C_i}$, vgl. Müller (1980).

Des Weiteren bezeichne $F_{t_{\nu_i}}$ die Verteilungsfunktion der t_{ν_i} -Verteilung und Φ^{-1} die Inverse der $\mathcal{N}(0, 1)$ -Verteilungsfunktion Φ , s. Abschnitte 1.1 und 1.2.2.B in Kap. IV. Φ^{-1} ist in Tab. 2 im Anhang vertafelt. $F_{t_{\nu_i}}$ ist in praktisch jedem statistischen Programm paket numerisch verfügbar, was auch für die Verteilungsfunktion der nicht-zentralen t-Verteilung aus kommendem Abschnitt 14.3.1 gilt. Wir geben jedoch für alle Fälle, in denen diese Funktionen auftauchen, gute Approximationen an, die elementar zu berechnen sind.

14.1. Effektmaß – Differenz der Mittelwerte

Die bisherigen Verfahren in den Fixed und Random Effects Modellen werden z. B. anhand der MS-Studien in den Abschnitten 2.5 und 3.3 für normalverteilte Daten dargelegt. Im Fixed Effects Modell lassen sich hier jedoch auch sogenannte exakte Methoden anwenden.

14.1.1. Exakte Punkt- und Intervallschätzer im Fixed Effects Modell [Hartung-Knapp-Verfahren]

Mit den obigen Bezeichnungen setzen wir hier:

$$\begin{aligned}\psi &= D(1) = \mu_E - \mu_C, \quad y_i = \bar{x}_{E_i} - \bar{x}_{C_i}, \quad \hat{\sigma}_i^2 = \widehat{\sigma}_i^2(1), \text{ und} \\ T_i^D(\psi) &= \frac{y_i - \psi}{\hat{\sigma}_i} \sim t_{\nu_i}, \quad \nu_i = \nu_i(1),\end{aligned}$$

und erhalten, vgl. Hartung/Knapp (2005a),

$$d(\psi) = \frac{1}{\sqrt{K}} \sum_{i=1}^K \Phi^{-1} \left(F_{t_{\nu_i}} [T_i^D(\psi)] \right) \sim \mathcal{N}(0, 1),$$

wobei diese Verteilungen im engeren Sinne exakt sind, wenn wir $\xi_E^2 = \xi_C^2$ annehmen. Ferner ist $d(\psi)$ eine streng monoton fallende Funktion von ψ , so daß ein exaktes $(1 - \alpha)$ -Konfidenzintervall für die Differenz der Erwartungswerte gegeben ist als:

$$KI(\mu_E - \mu_C) = [D_L, D_R], \text{ mit } d(D_L) = u_{1-\alpha/2}, \quad d(D_R) = -u_{1-\alpha/2}.$$

Hierbei sind die Lösungen D_L und D_R der beiden Gleichungen eindeutig und $u_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ ist das $(1 - \alpha/2)$ -Quantil der $\mathcal{N}(0, 1)$ -Verteilung.

Ferner ergibt sich für die eindeutige Lösung $\hat{\psi}_D$ von

$$\begin{aligned}d(\hat{\psi}_D) &= 0 : \quad \hat{\psi}_D = \text{Median Unverzerrter Pseudo Maximum} \\ &\quad \cdot \text{Likelihood [MUPML] Gesamt-Schätzer} \\ &\quad \text{für } \mu_E - \mu_C,\end{aligned}$$

d. h. $\hat{\psi}_D$ maximiert die Verteilungsdichte von $d(\psi)$ und liegt mit gleicher Wahrscheinlichkeit sowohl unterhalb als auch oberhalb der wahren Behandlungsdifferenz $\mu_E - \mu_C$.

Zur t_{ν_i} -Verteilung gehört die Varianz $\nu_i/(\nu_i - 2)$, so daß wir $d(\psi)$ gut approximieren können durch die praktisch normalverteilte Statistik:

$$d^*(\psi) = \frac{1}{\sqrt{K}} \sum_{i=1}^K W_i^D (\bar{x}_{E_i} - \bar{x}_{C_i} - \psi), \text{ mit } W_i^D = \sqrt{\frac{\nu_i - 2}{\nu_i \cdot \hat{\sigma}_i^2}}.$$

Damit erhalten wir das $(1 - \alpha)$ -Konfidenzintervall für die Differenz der Mittelwerte als:

$$KI(\mu_E - \mu_C)^* = \sum_{i=1}^K \frac{W_i^D (\bar{x}_{E_i} - \bar{x}_{C_i})}{\sum_{j=1}^K W_j^D} \pm \frac{u_{1-\alpha/2}}{\sum_{j=1}^K W_j^D} \cdot \sqrt{K}.$$

Der Mittelpunkt dieses Intervalls ist der **approximierende Median Unverzerrte Pseudo Maximum Likelihood [MUPML] Gesamt-Schätzer \widehat{D}_{ML}** für $\mu_E - \mu_C$, d. h.

$$\widehat{D}_{ML} = \sum_{i=1}^K \frac{W_i^D (\bar{x}_{E_i} - \bar{x}_{C_i})}{\sum_{j=1}^K W_j^D}, \text{ mit } \widehat{\text{Var}} \widehat{D}_{ML} = \frac{\sum_{i=1}^K (1 - 2/v_i)}{\left(\sum_{j=1}^K W_j^D \right)^2}.$$

Da die Mittelwertschätzer \bar{x}_{E_i} und \bar{x}_{C_i} (stochastisch) unabhängig von den Varianzschätzern $\hat{\sigma}_i^2$ sind, ist \widehat{D}_{ML} auch ein unverzerrter bzw. erwartungstreuer Schätzer für $\mu_E - \mu_C$:

$$\begin{aligned} E \widehat{D}_{ML} &= \sum_{i=1}^K E \left(\frac{W_i^D}{\sum_{j=1}^K W_j^D} \right) \cdot E(\bar{x}_{E_i} - \bar{x}_{C_i}) \\ &= E \left(\sum_{i=1}^K \frac{W_i^D}{\sum_{j=1}^K W_j^D} \right) (\mu_E - \mu_C) = 1 \cdot (\mu_E - \mu_C). \end{aligned}$$

Auffallend ist hier, daß invers zur geschätzten Standardabweichung gewichtet wird und nicht wie bisher mit der inversen geschätzten Varianz [vgl. Abschnitt 17.5].

14.1.2. Beispiel: Zusammenfassung zweier Noninferiority Studien zu Asthma Bronchiale [Differenz der Mittelwerte]

Bei $K = 2$ Nichtunterlegenheitsstudien zur Behandlung von Asthma Bronchiale, vgl. Tab. 27, ist die Vergleichsvariable der Lungenfunktionsparameter FEV₁: forced expiratory volume in 1 sec., gemessen in Litern (l). Die zugelassene Verschlechterungsgrenze liegt bei $\Delta_D = 0.15l$, d. h. es ist zu zeigen, daß mit erlaubter Irrtumswahrscheinlichkeit von $\alpha/2 = 0.025$ gilt: $\mu_E > \mu_C - 0.15l$, was gleichbedeutend für das 95%-Konfidenzintervall der Differenz $\mu_E - \mu_C$ beinhaltetet, daß dieses oberhalb von -0.15 liegt.

Mit $W_1^D = 6.928$, $W_2^D = 4.220$, $\bar{x}_{E_1} - \bar{x}_{C_1} = 0.12$ und $\bar{x}_{E_2} - \bar{x}_{C_2} = 0.14$ ergeben sich das 95%-Konfidenzintervall und der Med. Univ. Ps. Max. Lik. Schätzer für die Mittelwertdifferenz zu:

$$KI(\mu_E - \mu_C)^* = [-0.117, 0.373] \text{ und } \widehat{\mu_E - \mu_C} = 0.128.$$

Tab. 27: Asthma Bronchiale: Noninferiority Studien bzgl. eines Lungenfunktionsparameters [FEV₁]: Daten [l] für μ_E , μ_C und $\xi_E = \xi_C$

Studie					
i	$n_{E_i} = n_{C_i}$	\bar{x}_{E_i}	\bar{x}_{C_i}	s_i	ν_i
1	64	2.67	2.55	0.81	126
2	28	2.70	2.56	0.87	54

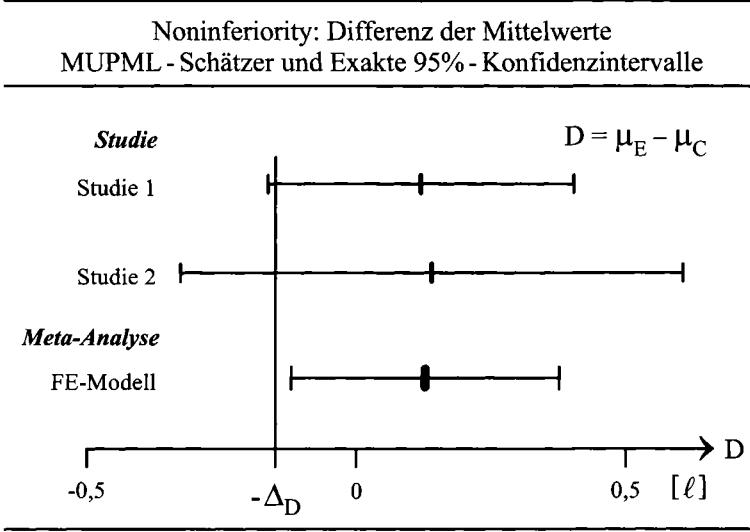
Zusammen mit den 95%-Konfidenzintervallen der beiden Studien:

$$\text{KI(Studie 1)} = 0.12 \pm 1.98 \cdot 0.81/\sqrt{32} = [-0.164, 0.404] \text{ und}$$

$$\text{KI(Studie 2)} = 0.14 \pm 2.005 \cdot 0.87/\sqrt{14} = [-0.326, 0.606]$$

sind alle exakten Konfidenzintervalle in Bezug auf die Nichtunterlegenheitsschranke $-\Delta_D = -0.15$ l in **Abb. 20** graphisch veranschaulicht. Die Varianz der Gesamtschätzung für die mittlere Differenz wird geschätzt zu:

$$\widehat{\text{Var}} \mu_E - \mu_C = (2 - 2/\nu_1 - 2/\nu_2) / (W_1^D + W_2^D)^2 \\ = 1.94709/11.148^2 = 0.01567.$$

**Abb. 20:** Asthma Bronchiale – FEV₁ mit $\Delta_D = 0,15 \ell$

Für weitere exakte und mittels sogenannter generalisierter p-Werte berechnete generalisierte Konfidenzintervalle für das gewöhnliche **Common Mean Problem** oder FE-Modell für normalverteilte Daten verweisen wir auf Hartung/Knapp/Sinha (2008) und Hartung/Knapp (2008).

14.2. Effektmaß – Quotient der Mittelwerte [Hartung-Verfahren]

Der Quotient von Mittelwerten, oder auch proportionale Veränderungen, wird beispielsweise häufig verwandt im Bereich der **Ökologie**, vgl. Hedges/Gurevitch/Curtis (1999). Ebenfalls häufig ergibt sich der Quotient als Effektmaß bei der Zusammenfassung von Klinischen Prüfungen zum Nachweis der Nichtunterlegenheit eines neuen Präparates im Vergleich zu einem Standardpräparat, wenn die Nichtunterlegenheit relativ formuliert ist, wie etwa „höchstens 10% schlechter“.

Das interessierende Effektmaß ist hier der Quotient μ_E/μ_C , wofür wir $\mu_E \geq 0$, $\mu_C > 0$ und entsprechend $\bar{x}_{E_i} \geq 0$, $\bar{x}_{C_i} > 0$ für $i = 1, \dots, K$ annehmen.

14.2.1. Exakte Methoden im Fixed Effects Model für Quotienten

Mit den anfangs eingeführten Bezeichnungen definieren wir hier, für $\lambda = \psi$ gesetzt,

$$\psi = \frac{\mu_E}{\mu_C}, \quad 0 \leq \psi < \infty, \text{ und } y_i(\psi) = \bar{x}_{E_i} - \psi \cdot \bar{x}_{C_i}, \text{ so daß}$$

$$E y_i(\psi) = 0 \text{ und } T_i(\psi) = y_i(\psi)/\hat{\sigma}_i(\psi) \sim t_{\nu_i},$$

wobei $\nu_i = \nu_i(\tilde{\psi}_0)$ mit $\tilde{\psi}_0 = (1/K) \sum_{i=1}^K \bar{x}_{E_i}/\bar{x}_{C_i}$ falls $\xi_E^2 \neq \xi_C^2$ und $n_{E_i} \neq n_{C_i}$. Kann $\xi_E^2 = \xi_C^2$ angenommen werden, so ist mit $\nu_i = n_{E_i} + n_{C_i} - 2$ diese t_{ν_i} -Verteilung exakt. Wir haben:

$$T_i(0) = \sqrt{n_{E_i}} \cdot \bar{x}_{E_i}/s_{E_i} \text{ und setzen } T_i(\infty) = -\sqrt{n_{C_i}} \cdot \bar{x}_{C_i}/s_{C_i}.$$

Falls $\xi_E^2 = \xi_C^2$, so ersetzen wir hier s_{E_i} und s_{C_i} durch s_i . Es gilt nun für die Statistik

$$a(\psi) = \frac{1}{\sqrt{K}} \sum_{i=1}^K \Phi^{-1}(F_{t_{\nu_i}}[T_i(\psi)]) \sim \mathcal{N}(0, 1)$$

und $a(\psi)$ ist streng monoton fallend in $\psi \geq 0$. Folglich ist ein $(1 - \alpha)$ -Konfidenzintervall für den Quotienten der Mittelwerte gegeben als:

$$KI(\mu_E/\mu_C)_a = [\psi_L^a, \psi_R^a], \text{ wobei:}$$

ψ_L^a löst $a(\psi_L^a) = u_{1-\alpha/2}$ falls $a(0) > u_{1-\alpha/2}$, sonst: $\psi_L^a = 0$, und

ψ_R^a löst $a(\psi_R^a) = -u_{1-\alpha/2}$ falls $a(\infty) < -u_{1-\alpha/2}$, sonst: $\psi_R^a = \infty$.

Ein **Median Unverzerrter Pseudo Maximum Likelihood [MUPML]**

Gesamt-Schätzer $\hat{\psi}_a$ für μ_E/μ_C ist implizit gegeben als:

$$\hat{\psi}_a \text{ löst: } a(\hat{\psi}_a) = 0.$$

Die Funktion $a(\psi)$ wird gut approximiert durch die praktisch normalverteilte Statistik

$$A(\psi) = \frac{1}{\sqrt{K}} \sum_{i=1}^K \sqrt{\frac{\nu_i - 2}{\nu_i}} \cdot T_i(\psi) \sim \mathcal{N}(0, 1),$$

so daß wir damit entsprechende Punkt- und Intervallschätzungen erhalten, s. nachfolgendes Beispiel. $A(\psi)$ ist ebenfalls streng monoton fallend in $\psi \geq 0$.

14.2.2. Beispiel Asthma Bronchiale: Zusammenfassung zweier Noninferiority Studien [Quotient der Mittelwerte]

Wir setzen das Beispiel aus Abschnitt 14.1.2 fort, nehmen hier jedoch an, daß die zugelassene Verschlechterungsgrenze **relativ** als $\Delta = 5\%$ formuliert ist, d. h. zum Niveau $\alpha/2 = 0.025$ ist $\mu_E \geq 0.95\mu_C$ zu zeigen bzw. das 95%-Konfidenzintervall für μ_E/μ_C soll oberhalb von $1 - \Delta = 0.95$ liegen, damit Behandlung E als der Kontrollbehandlung C nicht unterlegen angesehen werden kann. Mit den Daten aus der dortigen Tab. 27 erhält die Funktion $A(\psi)$ folgende Gestalt:

$$A(\psi) = \frac{1}{\sqrt{2}} \left\{ \sqrt{\frac{124 \cdot 64}{126}} \cdot \frac{2.67 - \psi \cdot 2.55}{0.81\sqrt{1 + \psi^2}} + \sqrt{\frac{52 \cdot 28}{54}} \cdot \frac{2.70 - \psi \cdot 2.56}{0.87\sqrt{1 + \psi^2}} \right\}.$$

Wir haben zu lösen, z. B. mittels „*regula falsi*“, s. Abschnitt 1.2.2 in Kap. XII,

$$A(\psi_L^A) = 1.96, \quad A(\psi_R^A) = -1.96 \text{ und } A(\hat{\psi}_A) = 0,$$

so daß wir das 95%-Konfidenzintervall für μ_E/μ_C und den **approximierenden Median Unverzerrten Pseudo Maximum Likelihood [MUPML] Gesamt-Schätzer** $\hat{\psi}_A$ für μ_E/μ_C erhalten als:

$$KI(\mu_E/\mu_C)_A = [0.95478, 1.15114] \text{ und } \hat{\psi}_A = 1.050.$$

Bezeichnen $A_1(\psi)$ und $A_2(\psi)$ die Summanden in der Klammer von $A(\psi)$, so sind dies die entsprechenden Statistiken zur Ermittlung der Konfidenzintervalle in den einzelnen Studien und ihre obige Kombination ergibt $A(\psi)_{COMB} = \{A_1(\psi) + A_2(\psi)\}/\sqrt{2} = A(\psi)$. Die eindeutigen Lösungen x von $A_i(x) = 0$ und $A_i(x) = \pm 1.96$ ergeben die Schätzer und 95%-Konfidenzintervalle:

$$\text{Studie 1: } KI(\psi) = [0.939, 1.168], \quad \hat{\psi}_{A_1} = 1.047$$

$$\text{Studie 2: } KI(\psi) = [0.884, 1.261], \quad \hat{\psi}_{A_2} = 1.055.$$

In Abb. 21 wird die **graphische Ermittlung** der Konfidenzgrenzen und Punktschätzer demonstriert. Abb. 22 zeigt dann die Ergebnisse in Bezug auf die relative Noninferiority Schranke $1 - \Delta = 0.95$.

Im noch folgenden Abschnitt 14.2.5 wird eine **alternative Methode** mit **expliziten** Lösungen vorgestellt.

14.2.3. Random Effects Model für „Quotient der Mittelwerte“

Wir betrachten zunächst die E-Gruppen aller Studien getrennt und führen für diese eine Meta-Analyse gemäß Abschnitt 3.2 durch, so daß wir für μ_E den Schätzer $\hat{\mu}_E(\text{RE})$ erhalten, etwa mit der geschätzten Varianz $q_H(E)$.

Analog erhalten wir getrennt für die C-Gruppen aller K Studien gemäß Abschnitt 3.2 für μ_C den RE-Schätzer $\hat{\mu}_C(\text{RE})$ mit geschätzter Varianz $q_H(C)$. Hier handelt es sich natürlich insgesamt nicht mehr um ein exaktes Verfahren in obigem Sinne.

Quotient der Mittelwerte: Ermittlung der
MUPML - Schätzer und 95%-Konfidenzgrenzen

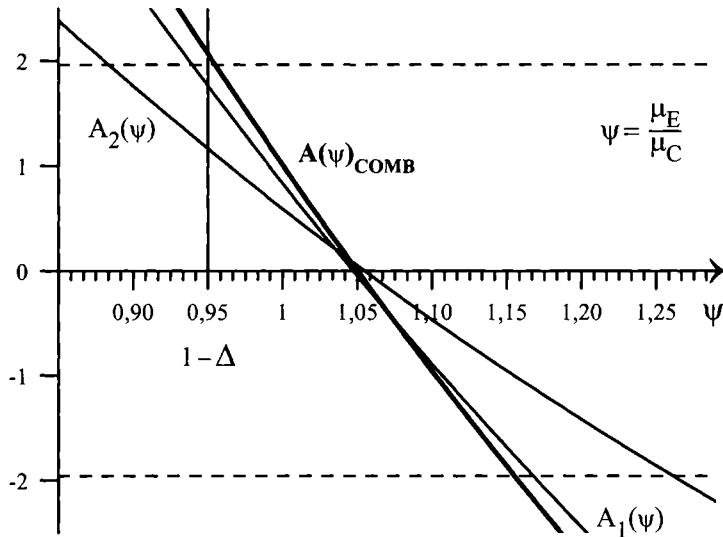


Abb. 21: Asthma Bronchiale – $A_1(\psi)$, $A_2(\psi)$ und $A(\psi)_{\text{COMB}}$

Die den Varianzschätzern jeweils zugeordneten Quadratformen $Q_H(E)$ und $Q_H(C)$, vgl. Abschnitt 3.2, entstammen beide einer χ^2_{K-1} -Verteilung, so daß wir die Freiheitsgrade der folgenden Statistik als $2(K-1)$ annehmen können, vgl. die Ausführungen zum Behrens-Fisher-Problem in Müller (1980). So erhalten wir die genäherte Aussage:

$$H(\psi) = \frac{\hat{\mu}_E(\text{RE}) - \psi \cdot \hat{\mu}_C(\text{RE})}{\sqrt{q_H(E) + \psi^2 q_H(C)}} \sim t_{2(K-1)}.$$

Ferner ist $H(\psi)$ streng monoton fallend bzgl. $\psi \geq 0$. Analog zu oben ist ein $(1-\alpha)$ -Konfidenzintervall für $\psi = \mu_E/\mu_C$ nun gegeben als:

$$\text{KI}_{\text{RE}}(\mu_E/\mu_C)_H = \left\{ \psi \geq 0 \mid |H(\psi)| \leq t_{2(K-1);1-\alpha/2} \right\},$$

das entsprechend zu oben durch Gleichsetzen an den Rändern berechnet wird. Durch Quadrieren gelangt man zu einer quadratischen Gleichung und deren Lösungen führen auf ein explizites Konfidenzintervall von der Gestalt, wie es Fieller (1940) im Prinzip für einen Quotienten hergeleitet hat, siehe nächsten Abschnitt.

Nach der Momentenmethode, vgl. Abschnitt 1.A in Kap. III, erhält man für μ_E/μ_C hier einen Gesamt-Schätzer, der zugleich ein **Median Unverzerrter**

Noninferiority: Quotient der Mittelwerte
MUPML - Schätzer und Exakte 95%-Konfidenzintervalle

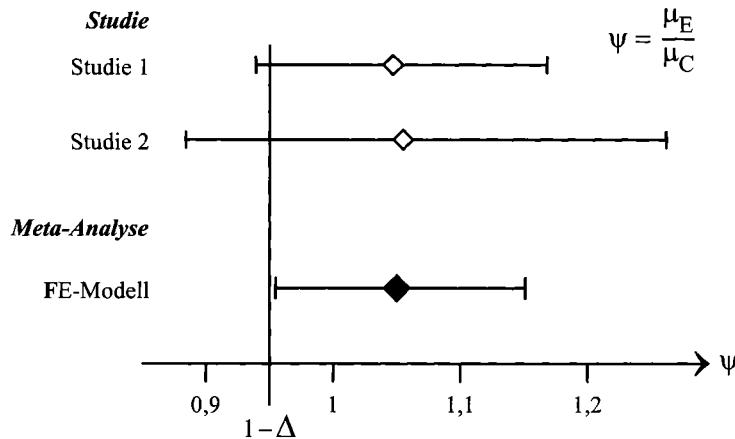


Abb. 22: Asthma Bronchiale – $H_1: \psi = \mu_E / \mu_C > 1 - \Delta = 0,95$

Pseudo Maximum Likelihood [MUPML] Schätzer im RE-Modell ist:

$$\widehat{\psi}_H = \widehat{\mu}_E(\text{RE}) / \widehat{\mu}_C(\text{RE}).$$

Nach dem Gaußschen Fehlerfortpflanzungsgesetz, vgl. Abschnitt 4.2(δ) in Kap. V, läßt sich seine Varianz schätzen als:

$$\widehat{\text{Var}} \widehat{\psi}_H = \left\{ q_H(E) + \widehat{\psi}_H^2 \cdot q_H(C) \right\} / \mu_C(\text{RE})^2.$$

Sind X_E und X_C eigentlich **lognormalverteilt**, so erhält man durch Logarithmieren normalverteilte Größen und aus dem Quotienten wird eine Differenz, so daß wir das Verfahren aus vorigem Abschnitt anwenden können, sowie die Methoden aus den Abschnitten 2 und 3.

14.2.4. Beispiel MS-Studien: RE-Model für „Quotient der Mittelwerte“

Wir setzen hier das Beispiel der 4 Studien zur Progression der Behinderung bei Patienten mit Multipler Sklerose aus Abschnitt 2.5 fort. Für die Versuchsdaten in der dort plazierten Tab. 1 legen wir jetzt das Effektmaß Quotient der Mittelwerte zu Grunde und für die Meta-Analyse wählen wir das Random Effects Model, wie schon in Abschnitt 3.3, in dem wir allerdings die 4 Studien für das Effektmaß Differenz der Mittelwerte auswerten.

Zunächst greifen wir uns nur die experimentellen E-Gruppen aus den 4 Studien heraus und werten sie gemäß eines Fixed Effect Model aus, vgl. Tab. 28, um zu einer Schätzung für den Heterogenitätsparameter τ^2 zu gelangen. Da die Schätzer von Hedges (-3.99) und DerSimonian-Laird (-2.83) negativ

sind, arbeiten wir mit $\hat{\tau}_{\text{HAR}}^2 = 0.05$ weiter, vgl. **Tab. 29**. Daraus erhalten wir als Schätzer für μ_E und zugehörige Varianz [vgl. Abschnitt 3.2]:

$$\hat{\mu}_E(\text{RE}) = 2.567 \text{ und } q_{\text{Hadj}}(E) = 0.627 \text{ [für } q_H(E)].$$

Völlig analog verfahren wir mit den Placebo-Kontrollgruppen C, vgl. **Tab. 30**, aus der wir $\hat{\tau}_{\text{HAR}}^2 = 1.10582$ entnehmen, um das Kollektiv der 4 C-Gruppen im Random Effects Model auszuwerten, vgl. **Tab. 31**. Wir erhalten daraus als Schätzer für μ_C und zugehörige Varianz:

$$\hat{\mu}_C(\text{RE}) = 4.974 \text{ und } q_H(C) = 0.780.$$

Uns interessieren Konfidenzintervalle der Form:

$$\text{KI}(\psi) = \{\psi \geq 0 \mid -t \leq H(\psi) \leq t\} = [L, R],$$

wobei $H(\psi)$ von folgender Struktur ist:

$$H(\psi) = \frac{a - \psi \cdot b}{\sqrt{c + \psi^2 \cdot d}}, \text{ mit } a \geq 0, b > 0, c \geq 0, d > 0.$$

Die Intervallgrenzen lassen sich explizit angeben, wobei wir einige **Fallunterscheidungen** treffen müssen:

$$(i) \text{ Ist } a^2 \leq t^2 c \quad \text{so ist:}$$

$$L = 0.$$

$$(ii) \text{ Ist } b^2 \leq t^2 d \quad \text{so ist:}$$

$$R = \infty.$$

$$(iii) \text{ Sind } b^2 = t^2 d, a^2 \geq t^2 c \quad \text{so sind:}$$

$$L = (a^2 - t^2 c) / 2ab \text{ und } R = \infty.$$

$$(iv) \text{ Sind } b^2 < t^2 d, a^2 \geq t^2 c \quad \text{so sind:}$$

$$R = \infty \text{ und}$$

$$L = \frac{ab}{b^2 - t^2 d} - \frac{1}{b^2 - t^2 d} \sqrt{a^2 b^2 - (a^2 - t^2 c)(b^2 - t^2 d)}.$$

$$(v) \text{ Sind } a^2 > t^2 c \text{ und } b^2 > t^2 d \text{ so sind:}$$

$$L = \frac{ab}{b^2 - t^2 d} - \frac{1}{b^2 - t^2 d} \sqrt{a^2 b^2 - (a^2 - t^2 c)(b^2 - t^2 d)} \text{ und}$$

$$R = \frac{ab}{b^2 - t^2 d} + \frac{1}{b^2 - t^2 d} \sqrt{a^2 b^2 - (a^2 - t^2 c)(b^2 - t^2 d)}.$$

Damit können wir zunächst die exakten Konfidenzintervalle für μ_E/μ_C in jeder der 4 Studien berechnen, vgl. **Tab. 32**, und dann das 95%-Konfidenzintervall für μ_E/μ_C in der RE-Meta-Analyse, vgl. **Tab. 33**. Schließlich haben wir in **Abb. 23** die gesamten Verhältnisse graphisch veranschaulicht.

Multiple Sklerose – Progression der Behinderung
Effektmaß Quotient der Mittelwerte: Random Effects Modell

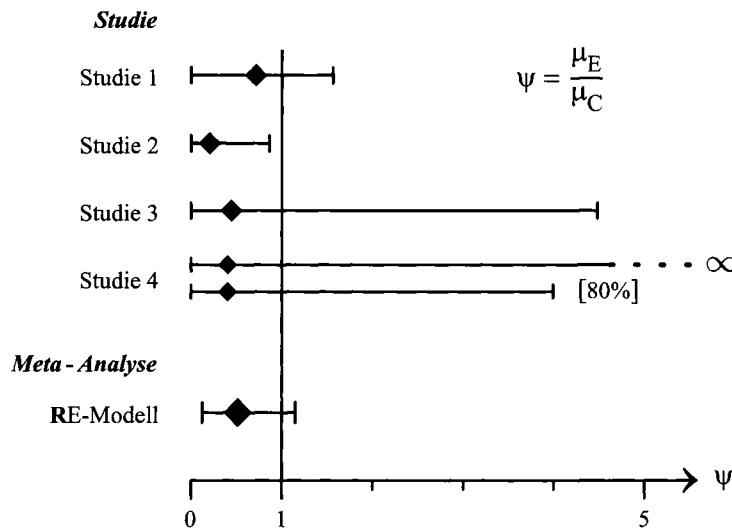


Abb. 23: MUPML – Schätzer und Exakte Konfidenzint. für ψ zum Konfidenz - Niveau 95% bzw. 80% (Studie 4)

14.2.5. Alternatives FE-Model für Quotienten mit Beispiel Asthma Bronchiale

Im obigen Abschnitt 14.2.3 setzen wir in dem dort betrachteten Random Effects Modell den Heterogenitätsparameter $\tau^2 = 0$ und erhalten so ein zu Abschnitt 14.2.1 alternatives Konzept im Fixed Effects Model.

Zunächst fassen wir die Versuchsdaten der experimentellen E-Gruppen aller K Studien: \bar{x}_{E_i} , $s_{E_i}^2$, n_{E_i} gemäß des exakten Vorgehens in Abschnitt 14.1.1 zusammen, d. h. wir berechnen den gemeinsamen Schätzer aller E-Gruppen für μ_E wie folgt:

$$\hat{\mu}_E(\text{FE}) = \sum_{i=1}^K W_i^E \bar{x}_{E_i} \Bigg/ \sum_{j=1}^K W_j^E, \text{ mit } W_i^E = \sqrt{\frac{(\nu_i - 2) \cdot n_{E_i}}{\nu_i \cdot s_{E_i}^2}},$$

wobei $\nu_i = n_{E_i} - 1$ oder $\nu_i = n_{E_i} + n_{C_i} - 2$, wenn $s_{E_i}^2 = s_i^2$, s_i^2 gepoolte Schätzung. Die Varianz von $\hat{\mu}_E(\text{FE})$ schätzen wir ebenfalls wie dort:

$$\widehat{\text{Var}} \hat{\mu}_E(\text{FE}) = \sum_{i=1}^K (1 - 2/\nu_i) \Bigg/ \left(\sum_{j=1}^K W_j^E \right)^2.$$

Völlig analog wenden wir uns ausschließlich den C-Kontrollgruppen der K Studien zu und erhalten aus deren Kollektiv die entsprechenden Schätzungen zu eben, wenn wir da E gegen C vertauschen.

Die alternative Meta-Analyse im Fixed Effects Modell für Quotienten liefert nun das gemeinsame $(1 - \alpha)$ -Konfidenzintervall für $\psi = \mu_E/\mu_C$ als:

$$KI_{FE}(\psi)_H = \{\psi \geq 0 \mid -u_{1-\alpha/2} \leq H(\psi)_{FE} \leq u_{1-\alpha/2}\}, \text{ mit}$$

$$H(\psi)_{FE} = \frac{\hat{\mu}_E(FE) - \psi \cdot \hat{\mu}_C(FE)}{\sqrt{\widehat{\text{Var}} \hat{\mu}_E(FE) + \psi^2 \cdot \widehat{\text{Var}} \hat{\mu}_C(FE)}},$$

und den gemeinsamen Schätzer $\hat{\psi}_H(FE) = \frac{\hat{\mu}_E(FE)}{\hat{\mu}_C(FE)}$. Im letzten Abschnitt oben ist die explizite Darstellung eines solchen Intervalls angegeben. Die Varianz des Schätzers lässt sich analog zu oben schätzen als:

$$\widehat{\text{Var}} \hat{\psi}_H(FE) = \left\{ \widehat{\text{Var}} \hat{\mu}_E + \hat{\psi}_H(FE)^2 \cdot \widehat{\text{Var}} \hat{\mu}_C(FE) \right\} / \hat{\mu}_C(FE)^2.$$

Beispiel: Wir betrachten hier die 2 Studien zu Asthma Bronchiale aus Tab. 27 in Abschnitt 14.1.2 mit dem Effektmaß Quotient der Mittelwerte. Zunächst berechnen wir mit obigem expliziten Verfahren exakte 95%-Konfidenzintervalle für die einzelnen Studien, vgl. **Tab. 34**; die mit den in Abschnitt 14.2.2 ermittelten Intervallen praktisch identisch sind, und dann für das alternative Meta-Analyse Konzept im FE-Modell:

Die dortige exakte Meta-Analyse ergibt für μ_E/μ_C den Schätzer 1.050 und das 95%-Konfidenzintervall [0.95478, 1.15114], während die alternative Meta-Analyse, vgl. **Tab. 35** und **Abb. 22**, die Schätzung 1.04942 für μ_E/μ_C und das 95%-Konfidenzintervall

$$KI_{FE}(\psi)_H = [0.95548, 1.15310]$$

liefert, was mit den bisherigen Resultaten gut übereinstimmt.

Tab. 28: Multiple Sklerose - Experimentelle Gruppen E. Arbeitstabelle im Fixed Effects Modell, Versuchsdaten in Tab. 1, Abschnitt 2

Experimentelle Gruppen E [MS]					
Studie					
i	$y_i = \bar{x}_{E_i}$	$\hat{\sigma}_i^2 = s_{E_i}^2/n_{E_i}$	$v_i = 1/\hat{\sigma}_i^2$	$v_i y_i$	$v_i y_i^2$
1	3.0	0.98148	1.01887	3.05661	9.16983
2	1.7	5.4000	0.18519	0.31481	0.53519
3	2.0	4.03333	0.24793	0.49587	0.99174
4	1.7	7.03700	0.14211	0.24158	0.41069
$\sum :$	8.4	17.15941	1.59410	4.10887	11.12446

$\hat{\mu}_{FE} = \sum v_i y_i / v_\Sigma = 2.578$, $Q_C = \sum v_i y_i^2 - v_\Sigma \hat{\mu}_{FE}^2 = 0.534$
 $v_\Sigma = \sum v_i$, $\bar{y} = \sum y_i / 4 = 2.1$, $\sum y_i^2 = 18.75$, $\sum v_i^2 = 1.154056$
 $\hat{\tau}_{HED}^2 = -3.99$, $\hat{\tau}_{DSL}^2 = -2.83$, $\hat{\tau}_{HAR}^2 = 0.05$

Tab. 29: Multiple Sklerose - Experimentelle Gruppen E. Arbeitstabelle im Random Effects Modell, vgl. Tab. 28, mit $\hat{\tau}_{HAR}^2 = 0.05$

Experimentelle Gruppen E [MS]					
Studie					
i	$\hat{\tau}_{HAR}^2 + \hat{\sigma}_i^2$	$w_i = (\hat{\tau}_{HAR}^2 + \hat{\sigma}_i^2)^{-1}$	$w_i y_i$	$w_i y_i^2$	
1	1.03148	0.96948	2.90844	8.72533	
2	5.45000	0.18349	0.31193	0.53028	
3	4.0833	0.24490	0.48980	0.97960	
4	7.0870	0.14110	0.23988	0.40779	
$\sum :$	-	1.53897	3.95005	10.64300	

$\hat{\mu}_{RE} = \sum w_i y_i / w_\Sigma = 2.567$, $q_H = \frac{1}{3} (\sum w_i y_i^2 / w_\Sigma - \hat{\mu}_{RE}^2) = 0.109$,
 $q_{Hadj} = \max\{q_H, 1/v_\Sigma\} = 0.627$, vgl. Tab. 28, $w_\Sigma = \sum w_i$

Tab. 30: Multiple Sklerose - Kontrollgruppen C. Arbeitstabelle im Fixed Effects Modell, Versuchsdaten in Tab. 1, Abschnitt 2

Kontrollgruppen C [MS]					
Studie					
i	$y_i = \bar{x}_{C_i}$	$\hat{\sigma}_i^2 = s_{C_i}^2/n_{C_i}$	$v_i = 1/\hat{\sigma}_i^2$	$v_i y_i$	$v_i y_i^2$
1	4.2	0.94286	1.06061	4.45455	18.70909
2	8.3	4.80000	0.20833	1.72917	14.35208
3	4.5	3.90625	0.25600	1.15200	5.18400
4	4.2	7.40000	0.13514	0.56757	2.38378
$\sum :$	21.2	17.04911	1.66008	7.90329	40.62895

$\hat{\mu}_{FE} = \sum v_i y_i / v_\Sigma = 4.761$, $Q_C = \sum v_i y_i^2 - v_\Sigma \hat{\mu}_{FE}^2 = 3.003$
 $v_\Sigma = \sum v_i$, $\bar{y} = \sum y_i / 4 = 5.3$, $\sum y_i^2 = 124.42$, $\sum v_i^2 = 1.25209$
 $\hat{\tau}_{HED}^2 = -0.242$, $\hat{\tau}_{DSL}^2 = -0.004$, $\hat{\tau}_{HAR}^2 = 1.10582$

Tab. 31: Multiple Sklerose - Kontrollgruppen C. Arbeitstabelle im Random Effects Modell, vgl. Tab. 30, mit $\hat{\tau}_{HAR}^2 = 1.10582$

Kontrollgruppen C [MS]				
Studie				
i	$\hat{\tau}_{HAR}^2 + \hat{\sigma}_i^2$	$w_i = (\hat{\tau}_{HAR}^2 + \hat{\sigma}_i^2)^{-1}$	$w_i y_i$	$w_i y_i^2$
1	2.04868	0.48812	2.05010	8.61042
2	5.90582	0.16932	1.40539	11.66476
3	5.01207	0.19952	0.89783	4.04025
4	8.50582	0.11757	0.49378	2.07387
$\sum :$	-	0.97453	4.84710	26.38930

$\hat{\mu}_{RE} = \sum w_i y_i / w_\Sigma = 4.974$, $q_H = \frac{1}{3} (\sum w_i y_i^2 / w_\Sigma - \hat{\mu}_{RE}^2) = 0.780$

Tab. 32: Multiple Sklerose - Exakte Konfidenzintervalle für μ_E/μ_C der 4 Studien aus Tab. 1 in Abschnitt 2.5

Multiple Sklerose $\psi = \mu_E/\mu_C$				
Parameter	Studie 1	Studie 2	Studie 3	Studie 4
$a = \bar{x}_{E_i}$	3.0	1.7	2.0	1.7
$b = \bar{x}_{C_i}$	4.2	8.3	4.5	4.2
$c = s_{E_i}^2/n_{E_i}$	0.9815	5.4	4.0333	7.037
a/\sqrt{c}	3.028	0.732	0.996	0.64
$d = s_{C_i}^2/n_{C_i}$	0.9429	4.8	3.9063	7.4
b/\sqrt{d}	4.325	3.789	2.277	1.544
$\nu_i = n_{E_i} + n_{C_i} - 2$	335	33	60	50
$t = t_{\nu_i; 0.975}$	1.968	2.036	2.000	2.009
$\hat{\psi}_i = a/b$	0.714	0.205	0.444	0.405
95%-KI	[0.238, 1.564]	[0, 0.863]	[0, 4.478]	[0, ∞)
$t = t_{50; 0.90}$	-	-	-	1.299
80%-KI	-	-	-	[0, 3.299]

Tab. 33: Multiple Sklerose - Effektmaß Quotient der Mittelwerte μ_E/μ_C . Meta-Analyse im Random Effects Modell der 4 Studien aus Tab. 1 in Abschnitt 2.5 mit Hilfsgrößen aus Tab. 29 und Tab. 31

Multiple Sklerose $\psi = \mu_E/\mu_C$	
Parameter	Meta-Analyse im RE-Modell mit $\hat{\tau}_{HAR}^2$
$a = \hat{\mu}_E(\text{RE})$	2.567
$\hat{\tau}_{HAR}^2(E)$	0.05
$b = \hat{\mu}_C(\text{RE})$	4.974
$\hat{\tau}_{HAR}^2(C)$	1.10582
$c = q_{Hadj}(E)$	0.627
$d = q_H(C)$	0.780
$t = t_{6; 0.975}$	2.447
$[t = t_{2(K-1); 1-\alpha/2}]$	
a/\sqrt{c}	3.248
b/\sqrt{d}	5.632
$\hat{\psi}_H = \hat{\mu}_E(\text{RE})/\hat{\mu}_C(\text{RE})$	0.516
$\widehat{\text{Var}} \hat{\psi}_H$	0.03374
$\bar{\psi} = \sum \hat{\psi}_i/4$	0.442
95% - KI _{RE} (ψ) _H	[0.12289, 1.14947]

Tab. 34: Asthma Bronchiale - Effektmaß Quotient der Mittelwerte μ_E/μ_C . Exakte Konfidenzintervalle für μ_E/μ_C in den 2 Studien aus Tab. 27 in Abschnitt 14.1.2

Asthma Bronchiale $\psi = \mu_E/\mu_C$		
Parameter	Studie 1	Studie 2
$a = \bar{x}_{E_i}$	2.67	2.70
a/\sqrt{c}	26.4	16.4
$b = \bar{x}_{C_i}$	2.55	2.56
b/\sqrt{d}	25.2	15.6
$c = d = 2s_i^2/n_i$	0.010252	0.027032
$\nu_i = n_i - 2$	126	54
$[n_i = n_{E_i} + n_{C_i}]$		
$t = t_{\nu_i; 0.975}$	1.98	2.005
95%-KI	[0.940, 1.168]	[0.883, 1.262]
$\hat{\psi}_i = a/b$	1.047	1.055

14.3. Effektmaß – Standardisierte Mittlere Differenzen

Das Effektmaß Standardisierte Differenz der Mittelwerte SDD ist definiert als:

$$SDD = \frac{X_E - X_C}{\xi}, \text{ mit } \xi = (\xi_E + \xi_C)/2.$$

Während die gebräuchliche Differenz D der Mittelwerte nicht skaleninvariant ist und der Quotient der Mittelwerte nicht translationsinvariant ist, ist die Standardisierte Differenz SDD sowohl **skalen-** als auch **translationsinvariant**, d. h. transformiert man X_E und X_C gemäß $\beta \cdot X_E + \gamma$ und $\beta \cdot X_C + \gamma$ mit $\beta \neq 1$ und $\gamma \neq 0$, was inhaltlich nichts verändert, so bleibt von den drei hier betrachteten Effektmaßen lediglich SDD unverändert. Diese Eigenschaft lässt die SDD bevorzugt dann anwenden, wenn die Meßskalen der beteiligten Studien nicht absolut vergleichbar sind.

14.3.1. Exakte Methoden im FE-Model für Standardisierte Mittlere Differenzen [Hartung-Verfahren]

Mit den anfangs eingeführten Bezeichnungen setzen wir

$$\psi = SDD \text{ und } y_i = \frac{\bar{x}_{E_i} - \bar{x}_{C_i}}{s_i}, s_i^2 \text{ gepoolter Varianzschatzer.}$$

Ferner bezeichne $t_{\nu_i}[\delta]$ die **nichtzentrale t-Verteilung** mit ν_i Freiheitsgraden und Nichtzentralitätsparameter δ , vgl. Abschnitt 1.2.2.B in Kap. IV, und $F_{t_{\nu_i}[\delta]}(\bullet)$ die zugehörige Verteilungsfunktion. Wir erhalten dann folgende

Tab. 35: Asthma Bronchiale - Effektmaß Quotient der Mittelwerte μ_E/μ_C . Alternative Meta-Analyse im Fixed Effects Modell der 2 Studien aus Tab. 27 in Abschnitt 14.1.2

Parameter	Asthma Bronchiale $\psi = \mu_E/\mu_C$	Alternative Meta-Analyse
	im FE-Modell	
$a = \hat{\mu}_E(\text{RE})$	2.68	
a/\sqrt{c}	31.0	
$b = \hat{\mu}_C(\text{RE})$	2.55379	
b/\sqrt{d}	29.5	
$c = d = \widehat{\text{Var}} \hat{\mu}$	0.007833	
$t = t_{\infty; 0.975} = u_{0.975}$	1.96	
$\hat{\psi}_H = \hat{\mu}_E(\text{RE})/\hat{\mu}_C(\text{RE})$	1.04942	
$\bar{\psi} = (\hat{\psi}_1 + \hat{\psi}_2)/2$	1.051	
$\widehat{\text{Var}} \hat{\psi}_H$	0.002524	
95% - KI _{FE} (μ_E/μ_C) _H	[0.95548, 1.15310]	
$\hat{\mu}_E(\text{FE}) = \sum W_i \hat{x}_{E_i} / \sum W_j$, $\hat{\mu}_C(\text{FE}) = \sum W_i \hat{x}_{C_i} / \sum W_j$,		
$\widehat{\text{Var}} \hat{\mu} = \sum (1 - 2/\nu_i) / (\sum W_j)^2$, $W_1^2 = (124 \cdot 64) / (126 \cdot 0.81^2) = 9.79784^2$		
$W_2^2 = (52 \cdot 28) / (54 \cdot 0.87^2) = 5.968492^2$, vgl. Tab 27		

Eigenschaften:

$$b_i \cdot y_i \sim t_{\nu_i}[b_i \cdot \psi], \text{ mit } b_i^2 = \frac{n_{E_i} n_{C_i}}{n_{E_i} + n_{C_i}}, \nu_i = n_{E_i} + n_{C_i} - 2, \text{ sowie}$$

$$g(\psi) = \frac{1}{\sqrt{K}} \sum_{i=1}^K \Phi^{-1}[F_{t_{\nu_i}[b_i \cdot \psi]}(b_i \cdot y_i)] \sim \mathcal{N}(0, 1),$$

und $g(\psi)$ ist eine streng monoton fallende Funktion von ψ . Somit ergibt sich das exakte $(1 - \alpha)$ -Konfidenzintervall für $\psi = \text{SDD}$ als:

$$\text{KI}(\text{SDD})_g = [\text{SDD}_L, \text{SDD}_R], \text{ wobei}$$

$$g(\text{SDD}_L) = u_{1-\alpha/2} \text{ und } g(\text{SDD}_R) = -u_{1-\alpha/2}.$$

Der **Median Unverzerrte Pseudo Maximum Likelihood [MUPML] Gesamt-Schätzer** $\hat{\psi}_g = \widehat{\text{SDD}}$ für die Standardisierte Differenz ist gegeben als:

$$\widehat{\text{SDD}} \text{ löst eindeutig: } g(\widehat{\text{SDD}}) = 0.$$

Obige Verteilungen sind im engeren Sinne exakt, wenn wir $\xi_E^2 = \xi_C^2$ annehmen. Andernfalls sind sie jedoch gut approximierend.

Mit der praktisch normalverteilten Statistik

$$G(\psi) = \frac{1}{\sqrt{K}} \sum_{i=1}^K \frac{y_i^* - \psi}{\sqrt{V_i}} \sim \mathcal{N}(0, 1), \text{ wobei } V_i = \hat{\sigma}^2(y_i^*) \text{ und}$$

$$y_i^* = \left(1 - \frac{3}{4n_i - 9}\right) y_i, \quad n_i = n_{E_i} + n_{C_i} \text{ sowie } V_i = \frac{1}{b_i^2} + \frac{y_i^2}{2\nu_i},$$

läßt sich die Funktion $g(\psi)$ gut annähern, so daß wir das $(1 - \alpha)$ -Konfidenzintervall für $\psi = \text{SDD}$ erhalten gemäß:

$$\text{KI}(\text{SDD})_G = \sum_{i=1}^K \frac{y_i^*/\sqrt{V_i}}{\sum_{j=1}^K 1/\sqrt{V_j}} \pm \frac{u_{1-\alpha/2} \cdot \sqrt{K}}{\sum_{j=1}^K 1/\sqrt{V_j}}.$$

Der Mittelpunkt dieses Konfidenzintervall ist der **approximierende Median Unverzerrte Pseudo Maximum Likelihood [MUPML] Gesamtschätzer $\hat{\psi}_G$** für SDD,

$$\hat{\psi}_G = \sum_{i=1}^K \frac{y_i^*/\sqrt{V_i}}{\sum_{j=1}^K 1/\sqrt{V_j}}, \quad \text{mit } \widehat{\text{Var}} \hat{\psi}_G = \frac{K}{\left(\sum_{j=1}^K 1/\sqrt{V_j}\right)^2}.$$

Der Schätzer ist praktisch auch unverzerrt bzgl. des Erwartungswertes, d. h. erwartungstreue [vgl. Abschnitt 17.5]. Hier fällt wieder auf, daß y_i^* invers zur geschätzten Standardabweichung gewichtet wird und nicht wie sonst üblich invers zur geschätzten Varianz.

14.3.2. Beispiel: Zusammenfassung zweier kleiner Superiority Studien zu Acne Papulopustulosa [Standardisierte Mittlere Differenzen] – Klinische Relevanz

In $K = 2$ Studien mit den Behandlungsgruppen E und Kontrollgruppen C ist die Response-Variablen für Patienten mit Acne Papulopustulosa die Bakterienreduktion nach sechs Behandlungswochen, welche auf Agarplatten als logCFU/cm² gemessen werden, mit CFU: colony forming units, vgl. Lehmann/Wassmer (1999). Mit den Daten:

$$n_{E_1} = n_{C_1} = 12, \quad y_1 = 1.177 \text{ und } n_{E_2} = n_{C_2} = 6, \quad y_2 = 1.023$$

erhalten wir die Hilfsgrößen: $y_1^* = 1.13$, $V_1 = 0.198$, $y_2^* = 0.99$, $V_2 = 0.391$, $1/\sqrt{V_1} + 1/\sqrt{V_2} = 3.847$, $y_1^*/\sqrt{V_1} + y_2^*/\sqrt{V_2} = 4.136$, so daß das 95%-Konfidenzintervall und der Med. Unv. Ps. Max. Lik. Schätzer $\hat{\psi}_G$ für die Standardisierte Differenz SDD sich ergeben zu:

$$\text{KI}(\text{SDD})_G = \frac{4.136}{3.847} \pm \frac{1.96 \cdot \sqrt{2}}{3.847} = [0.355, 1.796] \text{ und } \hat{\psi}_G = 1.075.$$

Die Behandlungsdifferenz $\mu_E - \mu_C$ ist folglich zum Niveau $\alpha/2 = 0.025$ signifikant größer als oder gleich $0.355 \cdot \xi$, also deutlich **klinisch relevant**,

SDD - Superiority: Klinische Relevanz
MUPML - Schätzer und Exakte 95%-Konfidenzintervalle

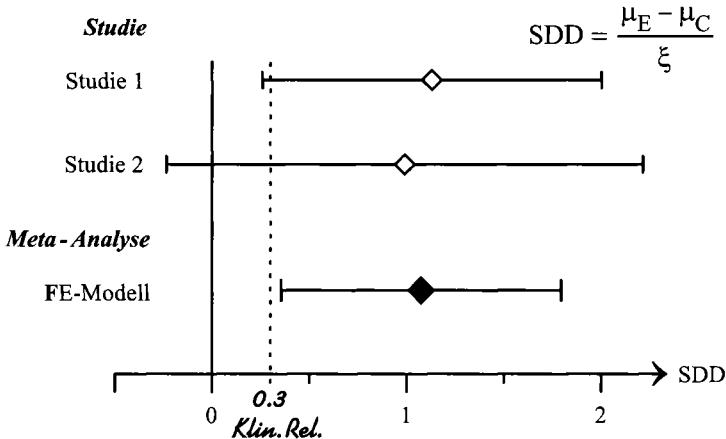


Abb. 24: Acne Papulopustulosa – Klinische Relevanz
für $\mu_E - \mu_C \geq 0.3 \cdot \xi$ [$\xi^2 = \text{Var } X_E = \text{Var } X_C$]

wobei $\xi^2 = \xi_E^2 = \xi_C^2$ die Varianz in den beiden den Gruppen E und C zugrunde liegenden Patienten-Grundgesamtheiten bezeichnet. In Abb. 24 sind die Studienergebnisse samt Meta-Analyse für die Standardisierte Mittlere Differenz SDD veranschaulicht, wobei die Grenze für die **Klinische Relevanz** der Behandlungsdifferenz hier für die Bakterienreduktion als $0.3 \cdot \xi$ gesetzt ist, was durchaus auch diskutiert wird.

Bezüglich einer allgemeinen Diskussion der Klinischen Relevanz verweisen wir auf Victor (1987).

14.3.3. Bisherige Methoden im FE- und RE-Model für Standardisierte Mittlere Differenzen

Mit obigen Bezeichnungen setzen wir hier

$$\psi = \text{SDD} \text{ und } y_i = \left(1 - \frac{3}{4n_i - 9}\right) \frac{\bar{x}_{E_i} - \bar{x}_{C_i}}{s_i} = \widehat{\text{SDD}}_i, \text{ sowie}$$

$$\widehat{\sigma}_i^2 = \frac{n_{E_i} + n_{C_i}}{n_{E_i} \cdot n_{C_i}} + \frac{y_i^2}{2\nu_i}, \text{ mit } \nu_i = n_i - 2 \text{ und } n_i = n_{E_i} + n_{C_i}.$$

Damit sind die benötigten Größen ψ , y_i und $\widehat{\sigma}_i^2$ wohl definiert und wir können alle Methoden aus den Abschnitten 2 und 3 hier anwenden, vgl. auch Hartung/Knapp/Sinha (2008).

Tab. 36: Arbeitsbelastung bei Angina Pectoris – Daten

Protocol	Amlodipine 10mg (E)			Placebo (C)		
	n _E	\bar{x}_E	s _E ²	n _C	\bar{x}_C	s _C ²
154	46	0.2316	0.2254	48	-0.0027	0.0007
156	30	0.2811	0.1441	26	0.0270	0.1139
157	75	0.1894	0.1981	72	0.0443	0.4972
162A	12	0.0930	0.1389	12	0.2277	0.0488
163	32	0.1622	0.0961	34	0.0056	0.0955
166	31	0.1837	0.1246	31	0.0943	0.1734
303A	27	0.6612	0.7060	27	-0.0057	0.9891
306	46	0.1366	0.1211	47	-0.0057	0.1291

14.3.4. Beispiel: Verbesserung der Belastbarkeit bei Angina Pectoris [Standardisierte Mittlere Differenzen] – Klinische Relevanz

Nach einem standardisierten Schema wurden in K = 8 Studien der Effekt von Amlodipine 10 mg (Gruppe E) im Vergleich zu Placebo (Gruppe C) auf die Arbeitsbelastbarkeit von Personen mit Angina Pectoris untersucht, vgl. Li/Shi/Roth (1994). In **Tab. 36** sind die wesentlichen Originaldaten aus den Studien zusammengestellt. Daraus ergeben sich die Standardisierten Differenzen $y_i = \widehat{SDD}$ mit zugehörigen Standardfehlern $\widehat{\sigma}_i$, die in **Tab. 37** aufgeführt sind. Dabei haben wir zur Berechnung der Standardabweichung die **gepoolte Schätzung** verwandt, welche recht **robust** bezüglich unterschiedlichen zugrunde liegenden Varianzen ist.

Wir haben Auswertungen sowohl im Fixed Effects Modell als auch im Random Effects Modell durchgeführt. Die Ergebnisse sind in **Tab. 38** zusammengestellt und in **Abb. 25** graphisch veranschaulicht, indem wir die SDD-Schätzungen der acht Studien und der Meta-Analyse in beiden Modellen zusammen mit den jeweils zugehörigen 95%-Konfidenzintervallen eingezeichnet haben.

Bis zu einem α -Niveau von 20% wird die Homogenitätshypothese nicht verworfen, was nicht zwingend zum Verbleib im FE-Modell führt. Im RE-Modell haben wir zum einen die DerSimonian-Laird Schätzung $\widehat{\tau}_{DSL}^2 = 0.0227$ für τ^2 zu Grunde gelegt und zum anderen, s. den nächsten Abschnitt 15, die wesentlich kleiner ausfallende Restrinierte Maximum Likelihood Schätzung $\widehat{\tau}_{REML}^2 = 0.0076$. Entsprechend reagieren die klassischen Konfidenzintervalle recht stark, während die Konfidenzintervalle nach Hartung nur wenig differieren.

Auffallend ist zudem der große Unterschied zwischen den Konfidenzintervallen im FE-Modell und im RE-Modell beim klassischen Vorgehen, wenn im RE-Modell die größere DSL-Schätzung für τ^2 zu Grunde liegt.

Die entsprechenden Konfidenzintervalle nach Hartung differieren geringfügig.

Tab. 37: Arbeitsbelastung bei Angina Pectoris: Standardisierte Mittlere Differenzen \widehat{SDD}_i und zugehörige Standardfehler $\widehat{\sigma}_i$ zu den Daten aus Tab. 36

Standardisierte		
Protocol	Mittlere Differenzen	Standardfehler
154	0.6987	0.2125
156	0.6946	0.2759
157	0.2459	0.1656
162A	-0.4245	0.4128
163	0.5000	0.2501
166	0.2287	0.2548
303A	0.7139	0.2807
306	0.3989	0.2095

gig. Eine Fehlspezifikation in der Modellwahl Fixed Effects Modell oder Random Effects Modell hat beim Konfidenzintervall nach Hartung deutlich geringere Auswirkungen als beim klassischen Vorgehen.

Liegt die Grenze für die **Klinische Relevanz** der Behandlungsdifferenz hier, wie für diesen Bereich mitunter gefordert, bei 20% der Standardabweichung, so wird diese mit den klassischen Konfidenzintervallen erreicht und mit den Konfidenzintervallen nach Hartung knapp verpaßt, vgl. Abb. 25.

15. Konfidenzintervalle und implizite Schätzer für den Heterogenitätsparameter

Für die Ergebnisse y_1, \dots, y_K von K unabhängigen Studien, Experimenten oder Prognosen liege das Random Effects Modell aus Abschnitt 3 zu Grunde:

$$y_i = \psi + a_i + e_i, \quad a_i \sim \mathcal{N}(0, \tau^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2), \quad \widehat{\sigma}_i^2 \text{ gegeben.}$$

Explizite Schätzer für den Heterogenitätsparameter τ^2 sind in den Abschnitten 3.1 und 13.3 bereits angegeben.

15.1. Hartung-Knapp-Konfidenzintervall für den Heterogenitätsparameter

In Abschnitt 3.2 wird die Quadratform Q_H eingeführt. Wird darin τ^2 nicht geschätzt, sondern als variabel angesehen, so ergibt sich als Funktion von τ^2 die Quadratform

$$Q_H(\tau^2) = \sum_{i=1}^K \frac{1}{\tau^2 + \widehat{\sigma}_i^2} \left(y_i - \frac{1}{\sum_{k=1}^K \frac{1}{\tau^2 + \widehat{\sigma}_k^2}} \sum_{j=1}^K \frac{1}{\tau^2 + \widehat{\sigma}_j^2} y_j \right)^2.$$

Mit einer Schätzung $\widehat{\tau}^2$ für τ^2 ist also $Q_H(\widehat{\tau}^2) = Q_H$ unsere bisherige Quadratform. Hartung/Knapp (2005b) zeigen, daß die approximativ χ_{K-1}^2 -verteilte

Tab. 38: Arbeitsbelastung bei Angina Pectoris: Ergebnisse zu den Daten aus Tab. 36 für die Standardisierten Mittleren Differenzen

Modell	Variable	Ergebnisse
FE	$\hat{\psi}$	0.4204
	$KI_{FE}(\psi)_C$	[0.2570, 0.5835]
	$KI_{FE}(\psi)_H$	[0.1880, 0.6529]
	Q_C	9.782
RE [DSL]	$\hat{\psi}$	0.4238
	$\hat{\tau}_{DSL}^2$	0.0227
	$KI_{RE}(\psi)_C$	[0.2259, 0.6218]
	$KI_{RE}(\psi)_H$	[0.1757, 0.6720]
RE [REML]	$\hat{\psi}$	0.4229
	$\hat{\tau}_{REML}^2$	0.0076
	$KI_{RE}(\psi)_C$	[0.2469, 0.5987]
	$KI_{RE}(\psi)_H$	[0.1841, 0.6616]

Statistik $Q_H(\tau^2)$ eine streng monoton fallende Funktion von τ^2 ist. Damit ist ein $(1 - \alpha)$ -Konfidenzintervall für τ^2 gegeben als:

$$KI(\tau^2)_{HK} = [\tau_L^2, \tau_R^2], \text{ mit}$$

$$Q_H(\tau_L^2) = \chi_{K-1; 1-\alpha/2}^2 \text{ und } Q_H(\tau_R^2) = \chi_{K-1; \alpha/2}^2.$$

Da die Lösungen jeweils eindeutig sind, bietet sich die „regula falsi“ zum Lösen an, vgl. Abschnitt 1.2.2 in Kap. XII, wo diese Methode ausführlich angewandt wird. Falls $\tau_L^2 < 0$, so kann τ_L^2 durch 0 ersetzt werden. Ist $\tau_R^2 < 0$, so setzen wir beide Grenzen auf 0.

Ist $\tau_L^2 > 0$, so sind die Studien etc. signifikant inhomogen zum Niveau $\alpha/2$. Zum selben Niveau sind sie signifikant homogen, wenn $\tau_R^2 \leq 0$ ausfällt.

15.2. Beispiel: Fortsetzung MS-Studien

Für die $K = 4$ Studien zu Multipler Sklerose aus Abschnitt 2.5 entnehmen wir die Werte für y_i und $\hat{\sigma}_i^2$ der dortigen Tab. 2 und bilden damit die Quadratfarm $Q_H(\tau^2)$, deren Verlauf in **Abb. 26** graphisch dargestellt ist.

Der Wert 9.348 des $\chi_{3, 0.975}^2$ -Quantils wird gerade für $\tau_L^2 = -3.2661825$ angenommen. Der Wert des unteren Quantils $\chi_{3; 0.025}^2 = 0.216$ wird von $Q_H(\tau^2)$ erst „weit draußen“ für $\tau_R^2 = 68.448295$ angenommen. Wir erhalten also für τ^2 das 95%-Konfidenzintervall:

$$KI(\tau^2)_{HK} = [0, 68.448].$$

 Angina Pectoris: Standardisierte Differenzen SDD

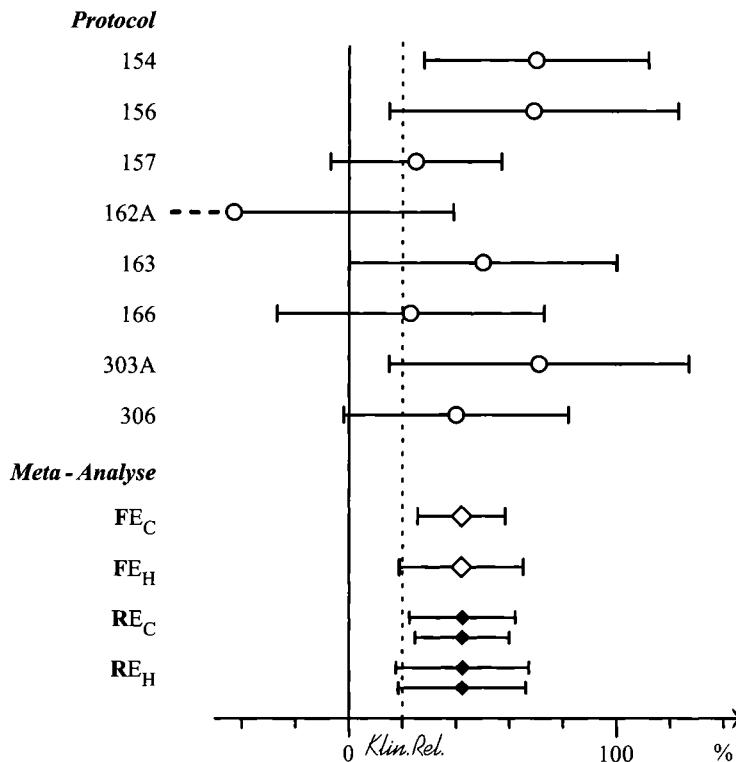


Abb. 25: Amlodipine – Steigerung der Belastbarkeit

15.3. Weitere Konfidenzintervalle und Tests für den Heterogenitätsparameter

Weitere Konfidenzintervalle für τ^2 , auch solche in expliziter Form und eine leichte Modifikation von $Q_H(\tau^2)$ für normalverteilte Daten, werden in Hartung/Knapp (2005b) und Knapp/Biggerstaff/Hartung (2006) diskutiert. Dabei zeigen sich die universelle Überlegenheit und Eignung des obigen Konfidenzintervalls, auch für Effektmaße in Studien mit binären Auskommern, vgl. auch Viechtbauer (2007).

Ein iterativ definiertes Konfidenzintervall für τ^2 geben Hartung/Argac (2002a) an. Einen verallgemeinerten Welch (1951) Test zum Testen von Hypothesen der Art: $H_0 : \tau^2 > \tau_0^2$ versus $H_1 : \tau^2 < \tau_0^2$ bzw. $H_0^* : \tau^2 \leq \tau_0^2$ versus $H_1^* : \tau^2 > \tau_0^2$, mit vorgegebener Schranke $\tau_0^2 > 0$, leiten Hartung/Argac (2002b) her.

In Hartung/Knapp/Sinha (2008) werden zudem noch Likelihood Methoden

 Vier Studien zur Multiplen Sklerose: Quadratform Q_H

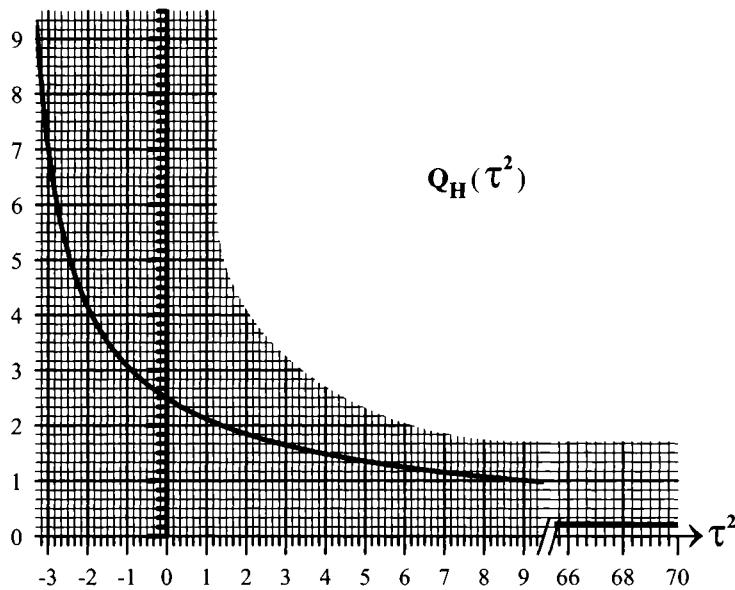


Abb. 26: Zur Ermittlung von Konfidenzintervall
und impliziten Schätzern für τ^2

und Konfidenzintervalle vom Waldschen Typ diskutiert.

15.4. Implizit definierte Schätzer für den Heterogenitätsparameter

Außer den eingangs erwähnten expliziten Schätzern für τ^2 sind in den unterschiedlichen Anwendungsbereichen auch implizite bzw. iterative Schätzer gebräuchlich, die im Folgenden aufgeführt sind, neben zwei neuen derartigen Schätzern.

15.4.1. Mandel-Paule Schätzer

Schon Mandel/Paule (1970) schlagen vor, den Schätzer $\hat{\tau}_{MP}^2$ für τ^2 zu bestimmen als Lösung der Gleichung

$$Q_H(\hat{\tau}_{MP}^2) = K - 1,$$

wobei $K - 1$ gerade der Erwartungswert der χ_{K-1}^2 -Verteilung ist.

15.4.2. Median Unverzerrte Schätzer nach Hartung

Der Median Unverzerrte Schätzer $\hat{\tau}_{\text{MU(H)}}^2$ für τ^2 löst:

$$Q_H(\hat{\tau}_{\text{MU(H)}}^2) = \chi_{K-1;0.5}^2 \approx (K-1) \left[1 - \frac{2}{9(K-1)} \right]^3.$$

Er fällt mit gleicher Wahrscheinlichkeit sowohl kleiner als auch größer als der wahre Wert von τ^2 aus und hat den Vorzug, in allen Konfidenzintervallen für τ^2 aus obigem Abschnitt 15.1 enthalten zu sein, wenn das Konfidenzniveau $1-\alpha$ zwischen 1 und 0 variiert.

15.4.3. Pseudo Maximum Likelihood Schätzer nach Hartung

Da die Dichte der χ_{K-1}^2 -Verteilung, vgl. Abschnitt 1.2.1.A in Kap. IV, maximal wird an der Stelle $K-3$ für $K \geq 4$, definieren wir für $K \geq 4$ den Pseudo Maximum Likelihood Schätzer $\hat{\tau}_{\text{PML(H)}}^2$ für τ^2 als Lösung der Gleichung

$$Q_H(\hat{\tau}_{\text{PML(H)}}^2) = K-3, \text{ für } K \geq 4.$$

Für die χ_{K-1}^2 -Verteilung gilt die Beziehung:

$$K-1 > \chi_{K-1;0.50}^2 > K-3.$$

$Q_H(\tau^2)$ ist streng monoton fallend in τ^2 , so daß die Schätzer in folgender Beziehung zu einander stehen:

$$\hat{\tau}_{\text{MP}}^2 < \hat{\tau}_{\text{MU(H)}}^2 < \hat{\tau}_{\text{PML(H)}}^2.$$

15.4.4. Beispiel: Fortsetzung MS-Studien

Mit der Quadratform $Q_H(\tau^2)$ aus Abschnitt 15.2 für die $K=4$ MS-Studien aus Abschnitt 2.5 erhalten wir, vgl. auch die dort bereits eingeführte Abb. 26:

$$Q_H(x) = 3 : \quad x = -0.999213,$$

$$Q_H(x) = 2.366 = \chi_{3;0.50}^2 : \quad x = 0.159616,$$

$$Q_H(x) = 1 : \quad x = 8.900602,$$

Also nehmen die drei hier eingeführten Schätzer für die vier MS-Studien die folgenden Werte an:

$$\hat{\tau}_{\text{MP}}^2 = -0.999, \quad \hat{\tau}_{\text{MU(H)}}^2 = 0.160 \text{ und } \hat{\tau}_{\text{PML(H)}}^2 = 8.901.$$

15.4.5. Maximum Likelihood Schätzer

Beginnend für $h=0$ mit einem Wert x_0 , erhalten wir im $(h+1)$ -ten Schritt für den Maximum Likelihood Schätzer $\hat{\tau}_{\text{ML}}^2$ den Wert x_{h+1} , der gegeben ist durch

$$x_{h+1} = \frac{\sum_{i=1}^K w_i(x_h)^2 \left\{ [y_i - \psi(x_h)]^2 - \hat{\sigma}_i^2 \right\}}{\sum_{j=1}^K w_j(x_h)^2},$$

für $h = 0, 1, \dots$. Hierbei sind:

$$w_i(x) = \frac{1}{x + \hat{\sigma}_i^2}, \quad \psi(x) = \frac{1}{w_{\Sigma}(x)} \sum_{i=1}^K w_i(x) y_i, \quad w_{\Sigma}(x) = \sum_{j=1}^K w_j(x).$$

15.4.6. Restringierte Maximum Likelihood Schätzer

Für $h = 0$ beginnen wir mit einem Wert x_0 und erhalten dann im $(h+1)$ -ten Schritt für den Restringierten Maximum Likelihood Schätzer $\hat{\tau}_{REML}^2$ den Wert x_{h+1} gemäß

$$x_{h+1} = \frac{\sum_{i=1}^K w_i(x_h)^2 \left\{ [y_i - \psi(x_h)]^2 - \hat{\sigma}_i^2 \right\}}{\sum_{j=1}^K w_j(x_h)^2} + \frac{1}{\sum_{k=1}^K w_k(x_h)},$$

für $h = 0, 1, \dots$. Dabei sind $w_i(x)$ und $\psi(x)$ wie oben definiert.

15.4.7. Beispiel: Fortsetzung MS-Studien

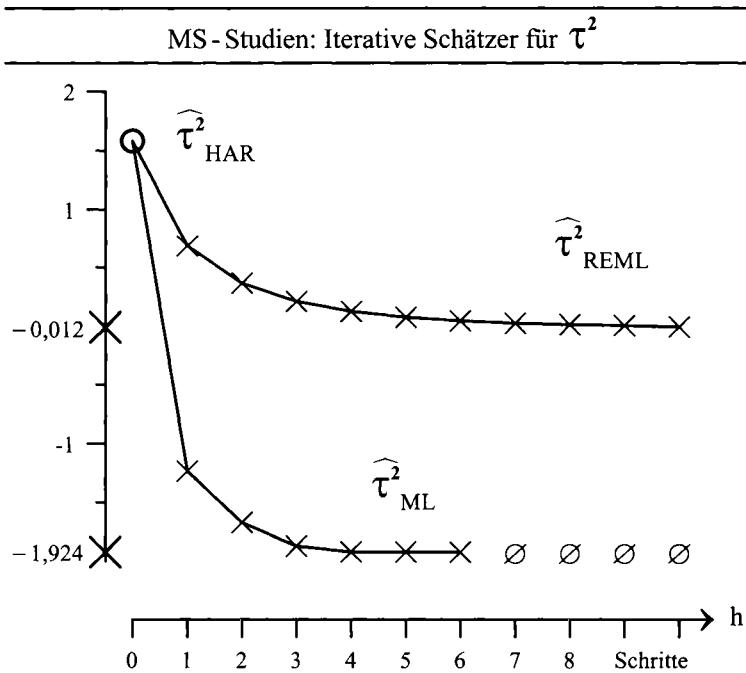


Abb. 27: Verläufe von Maximum Likelihood Schätzer $\hat{\tau}_{ML}^2$ und Restringiertem Maximum Likelihood Schätzer $\hat{\tau}_{REML}^2$

Mit den Daten y_i und $\hat{\sigma}_i^2$ aus Tab. 2 in Abschnitt 2.5 für die $K = 4$ dort eingeführten Studien zur Progression der Behinderung bei Patienten mit Multipler Sklerose (MS) und dem Startpunkt $x_0 = \hat{\tau}_{\text{HAR}}^2 = 1.58$ aus Abschnitt 3.1 für $h = 0$ ergeben die ersten Iterationen für $h = 1$ bis 10 jeweils Werte für $\hat{\tau}_{\text{ML}}^2$ und $\hat{\tau}_{\text{REML}}^2$ wie wir sie in **Abb. 27** eingezeichnet haben. Dabei werden die Werte für $\hat{\tau}_{\text{ML}}^2$ ab $h = 1$ und für $\hat{\tau}_{\text{REML}}^2$ ab $h = 10$ negativ. Ab $h = 5$ bzw. $h = 18$ erreichen die Iterationen konstant die Werte:

$$\hat{\tau}_{\text{ML}}^2 = -1.924 \text{ und } \hat{\tau}_{\text{REML}}^2 = -0.012.$$

Hierbei brechen für $\hat{\tau}_{\text{ML}}^2$ ab $h = 7$ die Iterationen zusammen und wir schreiben den zuletzt berechneten Wert -1.924 fort. Zum Vergleich aller bisher für dieses Beispiel berechneten 9 Schätzer für τ^2 aus den Abschnitten 3.1, 13.3 und 15.4.4 ergibt eine geordnete Zusammenstellung die 6 negativen Schätzungen:

$$\begin{aligned}\hat{\tau}_{\text{HED}}^2 &= -3.11, \quad \hat{\tau}_{\text{ML}}^2 = -1.92, \quad \hat{\tau}_{\text{DSL}}^2 = -1.28, \\ \hat{\tau}_{\text{MP}}^2 &= -1.00, \quad \hat{\tau}_{\text{HART}}^2 = -0.30, \quad \hat{\tau}_{\text{REML}}^2 = -0.012,\end{aligned}$$

sowie die 3 positiven Schätzungen:

$$\hat{\tau}_{\text{MU(H)}}^2 = 0.16, \quad \hat{\tau}_{\text{HAR}}^2 = 1.58, \quad \hat{\tau}_{\text{PML(H)}}^2 = 8.90.$$

Im folgenden Beispiel fallen alle Schätzungen der positiven Varianzkomponente τ^2 auch positiv aus.

15.5. Zusammenfassendes Beispiel: Fortsetzung Landwirtschaft – Genetik

Wir betrachten hier wieder die Erhebung zur Besamungssicherheit der $K = 6$ Bullen aus Abschnitt 3.4. Die Erhebungsdaten sind in der dortigen Tab. 5 eingetragen und ergeben die hier benötigten Statistiken $y_i = \bar{x}_i$ und $\hat{\sigma}_i^2 = s_i^2/n_i$. Damit bilden wir zunächst die Quadratform $Q_H(\tau^2)$, so wie wir sie auch in **Abb. 28** gezeichnet haben, und die nachstehenden nichtlinearen Gleichungen haben die gesuchten eindeutigen Lösungen wie folgt:

$$\begin{aligned}Q_H(x) &= 12.83 = \chi_{5,0.975}^2 & : x &= 5.7482 \\ Q_H(x) &= 5 = K - 1 & : x &= 81.4035 \\ Q_H(x) &= 4.351 = \chi_{5,0.50}^2 & : x &= 100.5575 \\ Q_H(x) &= 3 = K - 3 & : x &= 167.5034 \\ Q_H(x) &= 0.831 = \chi_{5,0.025}^2 & : x &= 738.1540\end{aligned}$$

Hieraus ergeben sich das 95%-Konfidenzintervall für τ^2 :

$$\text{KI}(\tau^2)_{\text{HK}} = [5.748, 738.154]$$

und die impliziten Schätzungen

$$\hat{\tau}_{\text{MP}}^2 = 81.40, \quad \hat{\tau}_{\text{MU(H)}}^2 = 100.56 \text{ und } \hat{\tau}_{\text{PML(H)}}^2 = 167.50.$$

 Besamungssicherheit der sechs Bullen: Quadratform Q_H

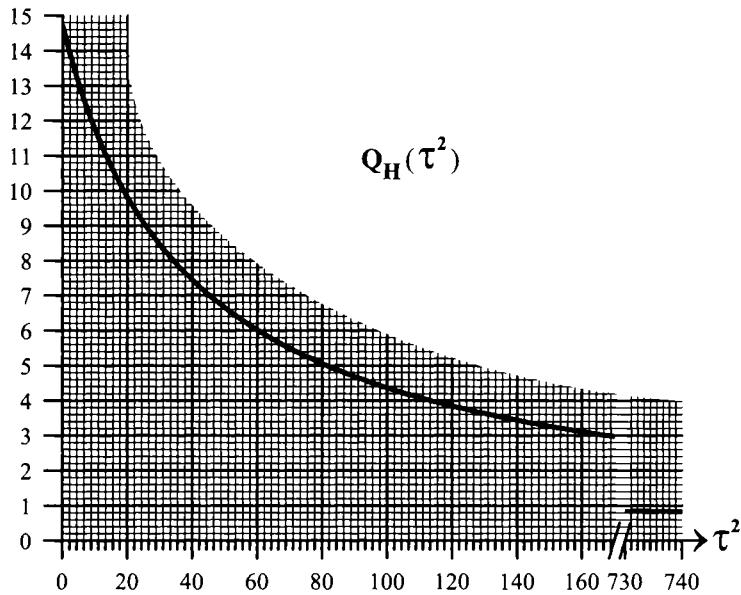


Abb. 28: Zur Ermittlung des Konfidenzintervalls
und der impliziten Schätzer für τ^2

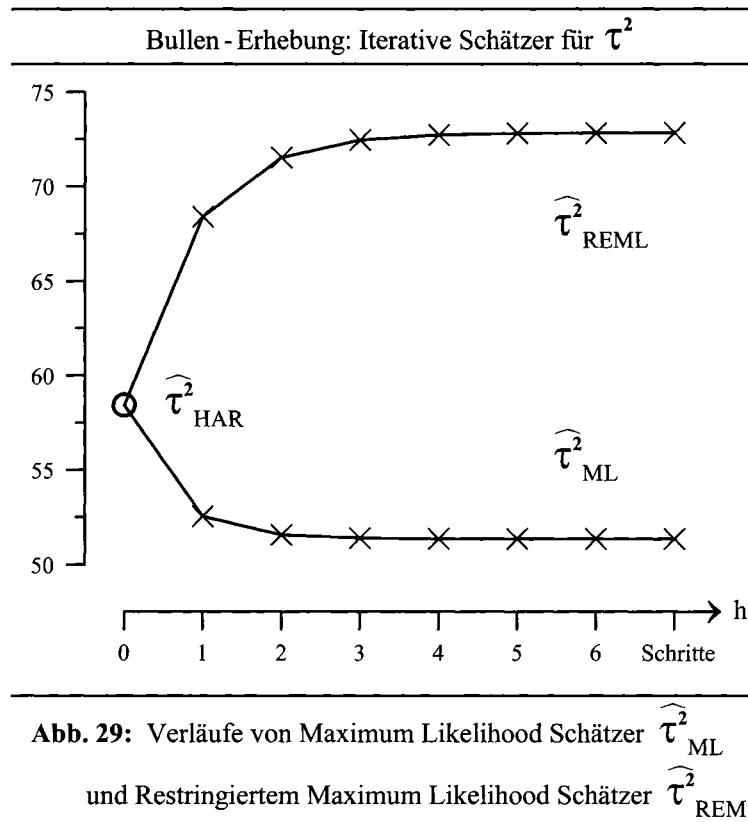
Beginnend mit $\hat{\tau}_{\text{HAR}}^2 = 58.44$, vgl. Abschnitt 3.4, für $h = 0$ ergeben für $h = 1$ bis 7 die Iterationen für $\hat{\tau}_{\text{ML}}^2$ und $\hat{\tau}_{\text{REML}}^2$ die in **Abb. 29** eingezeichneten Werte, die sich da schon stabilisiert haben auf:

$$\hat{\tau}_{\text{ML}}^2 = 51.35 \text{ und } \hat{\tau}_{\text{REML}}^2 = 72.83.$$

Mit den in den Abschnitten 3.4 und 13.3 berechneten expliziten Schätzern erhalten wir zum Vergleich die folgende geordnete Zusammenstellung der 9 für dieses Beispiel berechneten Schätzungen für den Heterogenitätsparameter τ^2 :

$$\begin{aligned} \hat{\tau}_{\text{ML}}^2 &= 51.35, \quad \hat{\tau}_{\text{HAR}}^2 = 58.44, \quad \hat{\tau}_{\text{DSL}}^2 = 66.80, \\ \hat{\tau}_{\text{REML}}^2 &= 72.83, \quad \hat{\tau}_{\text{MP}}^2 = 81.40, \quad \hat{\tau}_{\text{HART}}^2 = 84.25, \\ \hat{\tau}_{\text{HED}}^2 &= 89.68, \quad \hat{\tau}_{\text{MU(H)}}^2 = 100.56, \quad \hat{\tau}_{\text{PML(H)}}^2 = 167.50. \end{aligned}$$

Es bleibt daran zu erinnern, daß $\hat{\tau}_{\text{MU(H)}}$ in allen $(1 - \alpha)$ -Konfidenzintervallen für τ^2 liegt, wenn man das Konfidenzniveau $1 - \alpha$ variiert, und für $1 - \alpha < 9.2\%$ ist $\hat{\tau}_{\text{MU(H)}}$ die einzige von obigen Schätzungen, welche im $(1 - \alpha)$ -Konfidenzintervall verbleibt, vgl. auch Abb. 28.



Weitere Schätzer für τ^2 , auch solche, die stets positive Werte liefern wenn nicht alle y_i gerade identisch ausfallen, ergeben sich aus den allgemeinen diesbezüglichen Schätzformeln in Elpelt (1982, 1983, 1984, 1989), Elpelt/Hartung (1982), Heine (1993), Hartung/Elpelt/Voet (1997), Hartung/Makambi (2002a), Hartung (1981; 1999c, d; 2003), Hartung/Böckenhoff (2003) und Hartung/Knapp/Sinha (2008).

16. Portfolioselektion im Wertpapiermanagement

Die Zusammenstellung eines Wertpapier-Portfolios wird bereits im Unterabschnitt 1.9 des einführenden Abschnitts 1 zu diesem Kapitel angesprochen, auf den wir auch bzgl. allgemeiner Literaturangaben verweisen. Die dort bereits verwandten Begriffe Rendite und Volatilität werden zunächst präzisiert.

16.1. Rendite und Volatilität, mit Beispiel

Steht die Zeitabschnittsbezeichnung **Periode** etwa für 1 Tag, 1 Börsentag, 1 Woche, 1 Monat oder 1 Quartal, so greifen wir uns zum Zeitpunkt t bei-

spielsweise die Aktienkurse der letzten $n + 1$ Perioden heraus:

$$A_t, A_{t-1}, \dots, A_{t-n}.$$

Damit definieren wir nun die folgenden Größen für $j = 1, \dots, n$:

$$r_j = \ln \frac{A_{t+1-j}}{A_{t-j}} \approx \frac{A_{t+1-j} - A_{t-j}}{A_{t-j}} : \text{Rendite der } j\text{-ten Periode,}$$

$$\bar{r} = \frac{1}{n} \sum_{j=1}^n r_j : \text{mittlere Rendite,}$$

$$s_{\text{Per}}^2 = \frac{1}{n-1} \sum_{j=1}^n (r_j - \bar{r})^2 : \text{Varianz der Renditen pro Periode, und}$$

$$s_{\text{vol}} = \sqrt{T} \cdot s_{\text{Per}} : \text{jährliche Volatilität, wobei}$$

T = Anzahl der Perioden pro Jahr,

z. B. $T = 365, 250, 52, 12$ oder 4 wenn die Perioden gemäß obiger Aufzählung gewählt sind.

Als **Beispiel** betrachten wir folgende vierteljährlichen Kurse der letzten $n + 1 = 5$ Quartale: 100, 116.2, 135.5, 110.5, 100, mit den $n = 4$ Renditen $r_1 = -0.15$, $r_2 = -0.15$, $r_3 = 0.2$ und $r_4 = 0.1$. Es ergeben sich die mittlere Rendite $\bar{r} = 0$ und $s_{\text{Per}}^2 = 0.03167$ bzw. $s_{\text{Per}} = 0.178$. Da bei Quartalswerten $T = 4$ ist, erhält man für die jährliche Volatilität:

$$s_{\text{vol}} = \sqrt{4} \cdot 0.178 = 0.356.$$

Mit den Methoden der Regressions- und Zeitreihenanalyse, vgl. Kap. X und XII, lassen sich ausgehend von den so ermittelten historischen Werten entsprechende Prognosen für Rendite und Volatilität im **Planungshorizont** gewinnen, mit denen dann optional eine Portfolioauswahl vorgenommen werden kann. Analog können reale oder erwartete **Dividendenzahlungen** den Renditen hinzugerechnet werden.

16.2. Pareto-Optimale Portfolioauswahl

Für $i = 1, \dots, K$ Wertpapiere bezeichne:

$$y_i = \text{Rendite und } \sigma_i = \text{Volatilität [z. B. } s_{\text{vol}}]$$

des i -ten Wertpapiers. Im Portfolio aus den K Wertpapieren, deren Renditen y_i wir zunächst als unabhängig annehmen, definiere dann:

x_i : Anteil des i -ten Wertpapiers, und

$$x = (x_1, \dots, x_K) : \text{Anlage[strategie], mit } x_i \geq 0 \text{ und } \sum_{i=1}^K x_i = 1,$$

$$\psi(x) = \sum_{i=1}^K x_i \cdot y_i : \text{Rendite des Portfolios, und}$$

$$\sigma(x) = \sqrt{\sum_{i=1}^K x_i^2 \cdot \sigma_i^2} : \text{Volatilität des Portfolios.}$$

Ziel ist es nun, $\psi(x)$ zu maximieren und zugleich $\sigma(x)$ zu minimieren bzw. $-\sigma(x)$ ebenfalls zu maximieren. Es liegt also das Problem der Mehrziel- oder Vektormaximierung [v-max] vor:

$$\text{v-max} \left\{ F(x) = \begin{pmatrix} \psi(x) \\ -\sigma(x) \end{pmatrix} \middle| \sum_{i=1}^K x_i = 1, x_1 \geq 0, \dots, x_K \geq 0 \right\}.$$

Eine Anlagestrategie $x^{\text{Par}} = (x_1^{\text{Par}}, \dots, x_K^{\text{Par}})$ heißt **effiziente** oder **Pareto-optimale** Lösung dieser Optimierungsaufgabe, wenn es **keine** Anlage $x = (x_1, \dots, x_K)$, wobei $\sum_{i=1}^K x_i = 1, x_i \geq 0$, gibt mit der Eigenschaft:

$$\psi(x) \geq \psi(x^{\text{Par}}) \text{ und } \sigma(x) \leq \sigma(x^{\text{Par}}) \text{ sowie } F(x) \neq F(x^{\text{Par}}).$$

Faßt man y_i als Zufallsvariable mit $E y_i = \psi$ auf, so liefert der Gauß-Markov Schätzer für ψ , vgl. Abschnitte 2 und 17.5, als ausgezeichnete Pareto-optimale Lösung diejenige Anlagestrategie, welche die **geringste Volatilität** aufzuweisen hat:

$$\hat{\psi}_{V_{\min}} = \sum_{i=1}^K d_i \cdot y_i, \quad d_i = v_i/v_{\Sigma}, \quad v_i = 1/\sigma_i^2, \quad v_{\Sigma} = \sum_{i=1}^K v_i, \text{ mit}$$

$$\sigma(x)_{\min} = \sigma(d) = \sqrt{1/v_{\Sigma}}, \text{ wobei } x^{\text{Par}} = d.$$

Bezeichne $u_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ das $(1-\alpha/2)$ -Quantil der $\mathcal{N}(0, 1)$ -Verteilung, so erhalten wir durch die Wahl von $\alpha \in (0, 1)$ eine Familie von Ersatz-optimierungsaufgaben: Bestimme

$$R_{\alpha} = R_{\alpha}(x^{\alpha}) = \max \left\{ R_{\alpha}(x) = \psi(x) - u_{1-\alpha/2} \cdot \sigma(x) \middle| \sum_{i=1}^K x_i = 1, \right.$$

$$\left. x_1 \geq 0, \dots, x_K \geq 0 \right\}.$$

Die Lösungen x^{α} für $\alpha \in (0, 1)$ sind jeweils Pareto-optimal, und bezieht man deren Grenzwerte für $\alpha \rightarrow 0$ und $\alpha \rightarrow 1$ mit ein, so werden hierdurch sämtliche Pareto-optimalen Anlagen beschrieben, z. B. Hartung (1972a, 1974a). Speziell erhalten wir für $\alpha \rightarrow 0$: $x^{\alpha} \rightarrow d$ bzw. $\hat{\psi}(x^{\alpha}) \rightarrow \hat{\psi}_{V_{\min}}$ und $\sigma(x^{\alpha}) \rightarrow \sigma(x)_{\min} = 1/\sqrt{v_{\Sigma}}$, z. B. Hartung (1976e, 1985).

Haben wir uns für eine Anlage x entschieden, so ist ein $(1-\alpha)$ -Konfidenzintervall oder $(1-\alpha)$ -Risikobereich für die erwartete Rendite $\psi = E \psi(x) = \sum_{i=1}^K x_i \cdot E y_i$ gegeben als:

$$\psi(x) - u_{1-\alpha/2} \cdot \sigma(x) \leq \psi \leq \psi(x) + u_{1-\alpha/2} \cdot \sigma(x),$$

wenn wir von nun an die y_i zumindest genähert als normalverteilt annehmen. Da die untere Grenze gerade $R_{\alpha}(x)$ ist, können wir $R_{\alpha}(x)$ als **Risikofunktion** zum Niveau α und speziell für das übliche $1-\alpha = 0.95$ lediglich als Risikofunktion bezeichnen, so daß z. B. die Maximierung von $R_{0.05}(x)$ zur Anlage mit größter unterer 95%-Konfidenzgrenze, d. h. zur Anlage mit **geringstem Risiko** $\hat{\psi}_{RI_{\min}}$ führt.

16.3. Die Rendite-Volatilitäts-Projektionsebene, mit Beispiel

Wir beginnen mit einem **Beispiel**, dessen Daten aus Pflaumer/Heine/Hartung (2009) entnommen sind. Das Portfolio bestehe aus den $K = 2$ unabhängigen Aktien WIR-AG mit [in %] $y_1 = 20$, $\sigma_1 = 30$ und ICH-AG mit $y_2 = 10$, $\sigma_2 = 20$. Da im Portfolio $x_1 + x_2 = 1$ ist, können wir

$$\begin{aligned}\sigma(x) &= \sqrt{x_1^2 \cdot 30^2 + (1 - x_1)^2 \cdot 20^2} = \sigma(x_1) \quad \text{und} \\ \psi(x) &= x_1 \cdot 20 + (1 - x_1) \cdot 10 = \psi(x_1)\end{aligned}$$

in einer $x_1 - \sigma(x_1) - \psi(x_1)$ -Darstellung veranschaulichen mit $0 \leq x_1 \leq 1$ vgl.

Portfolioauswahl: Entscheidungsraum

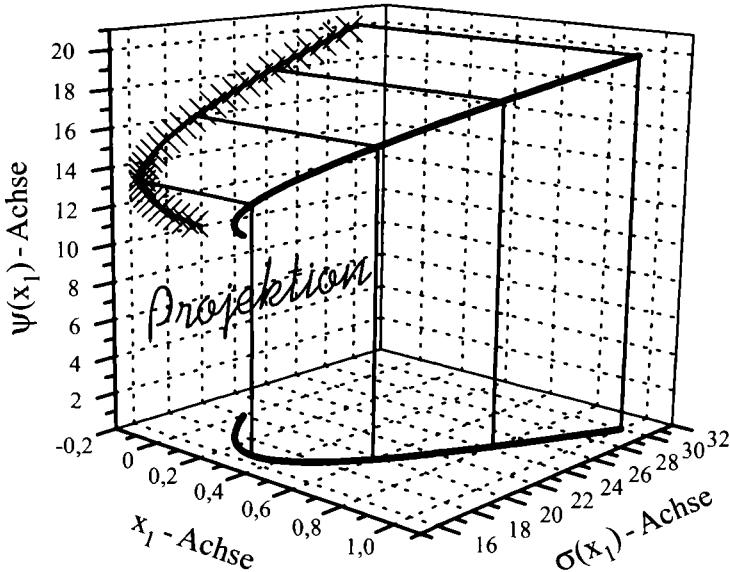


Abb. 30: $Rendite \quad \psi(x_1) = 20x_1 + 10(1 - x_1),$
 $Volatilität \quad \sigma(x_1) = \sqrt{900x_1^2 + 400(1 - x_1)^2},$
 $Anlage \quad x_1 \cdot \text{WIR-AG} + (1 - x_1) \cdot \text{ICH-AG}, \quad 0 \leq x_1 \leq 1$

Abb. 30, die so den **Entscheidungsraum** widerspiegelt. Dessen **Projektion** auf die $\psi(x)$ - $\sigma(x)$ -Ebene, mit $x = (x_1, x_2)$, $x_1 + x_2 = 1$, $x_1 \geq 0$, $x_2 \geq 0$, ist in **Abb. 31** getrennt dargestellt. Die Kurve zwischen den darin eingezeichneten Punkten A und B charakterisiert alle Pareto-optimalen Anlagen, wobei A für eine Anlage mit minimaler Volatilität $\hat{\psi}_{V_{\min}}$ steht. Im ebenfalls eingezeichneten Punkt D wird die Risikofunktion $\psi(x) - 1 \cdot \sigma(x) = R_{0,3}(x)$

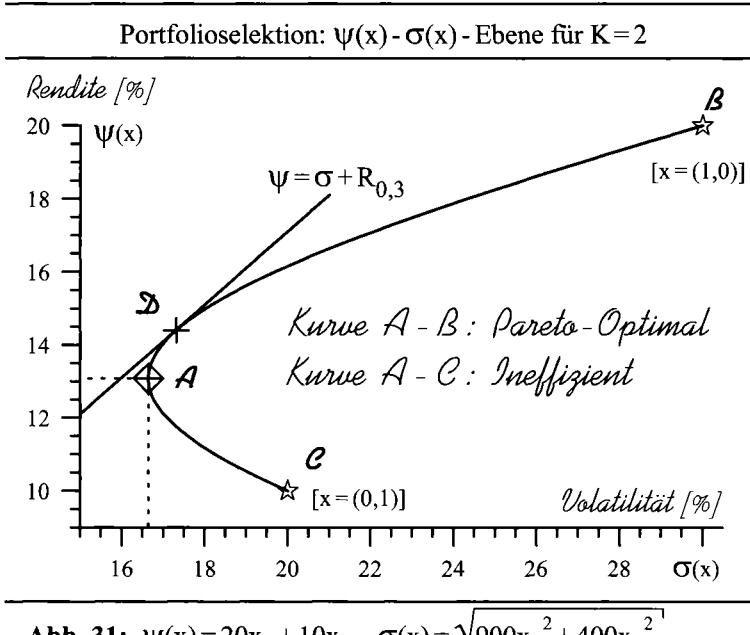


Abb. 31: $\psi(x) = 20x_1 + 10x_2$, $\sigma(x) = \sqrt{900x_1^2 + 400x_2^2}$,

$$\mathcal{A} \triangleq \widehat{\psi}_{V\min}, \quad \mathcal{D} \triangleq R_{0,3}; \quad 0 \leq x_1, 0 \leq x_2, \quad x_1 + x_2 = 1$$

[da $u_{0.85} = 1$] maximal [= $R_{0,3}$], d. h. D ist Pareto-optimal und liegt zudem auf der Tangente $\psi = \sigma + R_{0,3}$.

Allgemein wird für jedes $x^0 = (x_1^0, \dots, x_K^0)$ zunächst $\sigma(x^0) = \sigma(x_1^0, \dots, x_K^0)$ auf der $\sigma(x)$ -Achse eingezeichnet und dann darüber die zum selben $\sigma(x^0)$ führenden Werte von $\psi(x) = \psi(x_1, \dots, x_K)$, wobei x den gesamten zulässigen Bereich $\{x \mid \sum_{i=1}^K x_i = 1, x_1 \geq 0, \dots, x_K \geq 0\}$ durchläuft. Im nächsten Abschnitt veranschaulichen wir den Fall $K = 3$.

16.4. Rendite-Schwelle und Volatilitäts-Schranke, mit Beispiel

Eine weitere Möglichkeit, die Pareto-optimalen Anlagen sinnvoll einzuschränken, besteht in der Vorgabe einer Rendite-Schwelle **RES**, welche vom Portfolio nicht unterschritten werden sollte, und/oder einer Volatilitäts-Schranke **VS**, die es im Portfolio einzuhalten gilt. Als **Beispiel** betrachten wir ein Portfolio bestehend aus den $K = 3$ unabhängigen Aktien SIE-AG mit [in %] $y_1 = 30$, $\sigma_1 = 40$, WIR-AG mit $y_2 = 20$, $\sigma_2 = 30$ und ICH-AG mit $y_3 = 10$, $\sigma_3 = 20$. Die zugehörige $\psi(x) - \sigma(x)$ -Ebene ist in **Abb. 32** graphisch dargestellt. Unterhalb der Volatilitäts-Schranke $VS = 25\%$ und oberhalb der Rendite-Schwelle $RES = 20\%$ entsteht das **Entscheidungs- oder Anlagefenster** zwischen den eingezeichneten Punkten E und F. Die darin enthaltene, als Punkt G markierte mittlere Anlagestrategie im Anlagefenster maximiert zugleich die Risikofunktion $R_{0.5}(x) = \psi(x) - 0.69 \cdot \sigma(x)$ [da $0.69 \approx u_{0.75}$].

Zum 95%-Risikobereich gehört die untere Grenze $R_{0.05}(x) = \psi(x) - 1.96 \cdot$

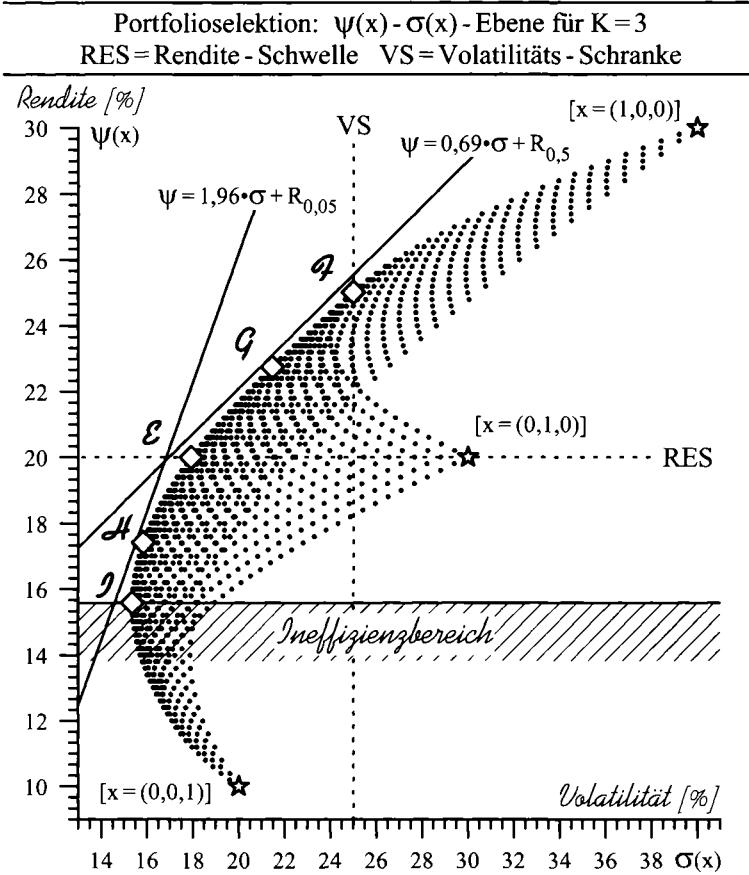


Abb. 32: $\psi(x) = 30x_1 + 20x_2 + 10x_3$, $\sigma(x) = \sqrt{1600x_1^2 + 900x_2^2 + 400x_3^2}$
 $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$, $x_1 + x_2 + x_3 = 1$

\mathcal{E} - \mathcal{Q} : Anlagefenster, $\mathcal{Q} \triangleq R_{0,5}$ mittlere Anlage im Fenster,

$\mathcal{H} \triangleq \hat{\psi}_{RI_{min}}$ minimales Risiko [$R_{0,05}$], $\mathcal{I} \triangleq \hat{\psi}_{V_{min}}$ minimale Volatilität

$\sigma(x)$, die im markierten Punkt H maximal wird, der somit für eine Anlage mit dem **geringsten Risiko** $\hat{\psi}_{RI_{min}}$ steht. Ebenfalls in Abb. 32 markiert der Punkt I diejenige Anlage mit der **geringsten Volatilität** $\hat{\psi}_{V_{min}}$, die zugleich für alle $\alpha \in (0, 1)$ den jeweils kürzesten $(1 - \alpha)$ -Risikobereich bedingt und somit die **stabilste Anlage** darstellt. Die letzten beiden Anlagestrategien $\hat{\psi}_{RI_{min}}$ und $\hat{\psi}_{V_{min}}$ sind mit zugehörigen 95%-Risikobereichen in Abb. 12 im Rahmen des einführenden Abschnitts 1.9 auch graphisch veranschaulicht.

Allgemein führt die Forderung einer Renditeschwelle RES unter Einhaltung einer Volatilitätschranke VS auf das folgende nichtlineare restriktierte

Maximierungsproblem:

$$\max \psi(x) - \lambda \cdot \sigma(x), \quad \lambda > 0,$$

unter den Nebenbedingungen:

$$\psi(x) \geq \text{RES}, \quad \sigma(x) \leq \text{VS}, \quad \sum_{i=1}^K x_i = 1, \quad x_1 \geq 0, \dots, x_K \geq 0.$$

Prädestinierte Lösungsmethoden sind die sequentiellen **Penalty-Verfahren**: Die stabile Innere Penalty-Methode von Hartung (1978c), deren Zwischenlösungen alle im Inneren des zulässigen Bereiches liegen, die stabilen äußeren Penalty-Methoden von Hartung (1980, 1985), die sich dem Maximum von außerhalb des zulässigen Bereichs beginnend nähern, oder die gemischte äußere-innere Penalty-Methode gemäß Elpelt/Hartung (1983b). Am Ende des nächsten Abschnitts 16.5 zeigen wir für den allgemeinen Fall, wenn auch Korrelationen in $\sigma(x)$ berücksichtigt werden, daß die zu maximierende Zielfunktion **konkav** ist und die Nebenbedingungen einen **konvexen** Bereich beschreiben.

16.5. Korrelation der Wertpapierkurse, mit Beispiel

Haben die Renditen y_i und y_j die Korrelation ρ_{ij} , so ändert sich die Volatilität des Portfolios, vgl. Abschnitt 8.2.A(3) in Kap. II, in:

$$\sigma(x) = \sigma_\rho(x) = \sqrt{\sum_{i=1}^K x_i^2 \cdot \sigma_i^2 + \sum_{i=1}^K \sum_{j=1, j \neq i}^K \rho_{ij} \cdot x_i \cdot x_j \cdot \sigma_i \cdot \sigma_j}.$$

Die ausgezeichnete Anlagestrategie minimaler Volatilität $\hat{\psi}_{V_{\min}}$ bestimmen wir analog zu $\hat{\psi}_{EFE}$ in Abschnitt 8.3. Die übrigen bisherigen Überlegungen bleiben unverändert gültig. Im **Beispiel** wenden wir uns wieder dem Portfolio aus Abschnitt 16.3 zu, bestehend aus WIR-AG mit $y_1 = 20$, $\sigma_1 = 30$ und ICH-AG mit $y_2 = 10$, $\sigma_2 = 20$. Dabei betrachten wir die folgenden 3 Konstellationen der Korrelation $\rho_{12} = \rho$:

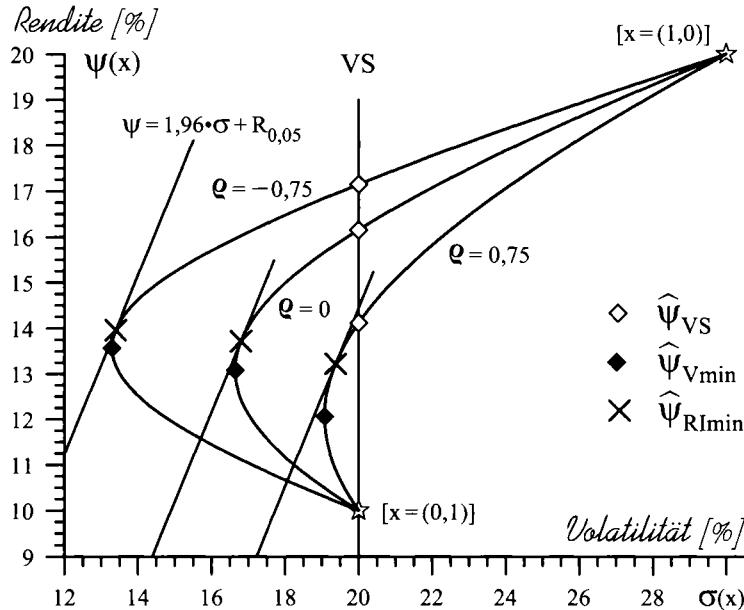
$$\rho = 0, \quad \rho = 3/4 \text{ und } \rho = -3/4.$$

In **Abb. 33** sind die zugehörigen $\psi(x) - \sigma(x)$ -Darstellungen wiedergegeben. Eingezeichnet sind zudem die Geraden $\psi = 1.96 \cdot \sigma + R_{0.05}$, wobei jeweils für die verschiedenen Korrelationen ρ :

$$R_{0.05} = \max \{ \psi(x) - 1.96 \cdot \sigma_\rho(x) \mid x_1 + x_2 = 1, \quad x_1 \geq 0, \quad x_2 \geq 0 \}.$$

Deren tangentiale Berührungspunkte mit dem jeweiligen $\psi(x) - \sigma_\rho(x)$ -Graphen liefern die Anlagen minimalen Risikos $\hat{\psi}_{RI_{\min}}$. Die Anlagen mit höchster Rendite $\hat{\psi}_{VS}$ bei einer Volatilitäts-Schranke $VS = 20\%$ sind ebenfalls in Abb. 33 markiert, wie auch die stabilsten Anlagen $\hat{\psi}_{V_{\min}}$ mit minimaler Volatilität. Schließlich haben wir in **Abb. 34** diese 3 Anlagestrategien zusammen mit

Wertpapiemanagement

 $\Psi(x) - \sigma(x)$ -Ebene für $K=2$ bei Korrelation Q **Abb. 33:** $\Psi(x) = 20 \cdot x_1 + 10 \cdot x_2, x_1 \geq 0, x_2 \geq 0, x_1 + x_2 = 1,$

$$\sigma(x) = \sqrt{30^2 \cdot x_1^2 + 20^2 \cdot x_2^2 + 2 \cdot Q \cdot x_1 \cdot x_2 \cdot 20 \cdot 30},$$

$$Q = 0; Q = 0,75; Q = -0,75$$

den zugehörigen 95%-Risikobereichen, bzw. deren untere Grenzen, für alle 3 Konstellationen von hier betrachteten Korrelationen dargestellt.

Eine bisher nicht übliche **Erweiterung** der Vorgehensweise erhalten wir durch eine Einbettung der Portfolioauswahl in das **Random Effects Modell**, vgl. Abschnitte 3 [RE] und 8.5 [ERE].

Sind mögliche Korrelationen nicht bekannt, so können wir durch eine RE-Modellierung in der praktischen Wirkung einen großen Teil der tatsächlich vorhandenen Korrelationen „einfangen“, vgl. auch die Diskussion in Abschnitt 8.8.

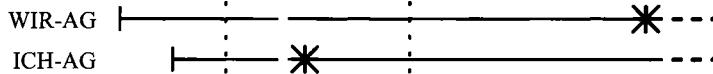
Mit den Bezeichnungen aus Abschnitt 8.3 entspricht mit der Kovarianzmatrix S :

$$\sigma(x)^2 \doteq \text{Var}(x^T y) = x S x^T,$$

und S ist positiv semidefinit, d. h. $b^T S b \geq 0$ für alle $b = (b_1, \dots, b_K)^T$. Zugleich ist $2 \cdot S$ die Matrix der zweifachen partiellen Ableitungen von $\sigma(x)^2$

 Wertpapiermanagement – Einfluß der Korrelation

Renditen und 95% – Risikobereiche

Aktien


Meta-Analyse: Portfolio aus WIR-AG und ICH-AG mit Renditekorrelation ρ

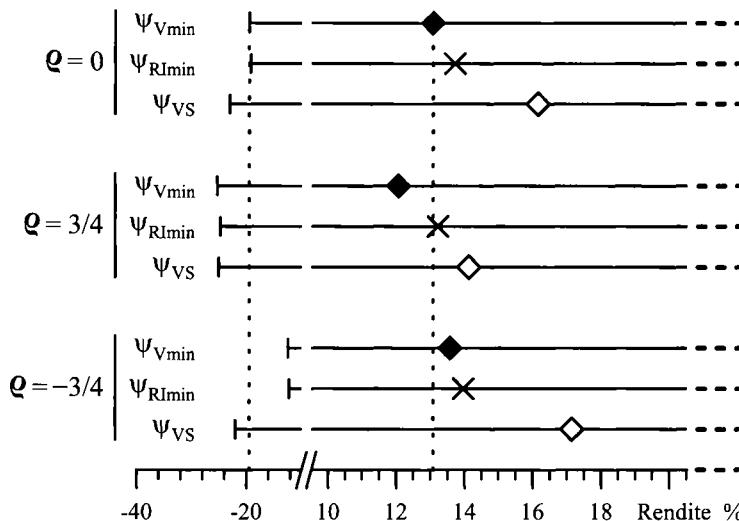


Abb. 34: Portfolioselektion $\hat{\psi}_{V\min}$ mit minimaler Volatilität, $\hat{\psi}_{R\min}$ mit minimalem Risiko sowie $\hat{\psi}_{VS}$ bei Volatilitätsschranke: VS = 20% und 95% – Risikobereiche [untere Grenzen] für $\psi_{V\min}$, $\psi_{R\min}$ sowie ψ_{VS}

nach x_i und x_j , so daß folglich $\sigma(x)^2$ eine konvexe Funktion und, da $\sigma(x)^2 \geq 0$ ist, $\sqrt{\sigma(x)^2} = \sigma(x)$ eine quasi-konvexe Funktion in $x = (x_1, \dots, x_K)$ ist, vgl. Abschnitt 17.5. Weiterhin ist $\sigma(x)$ positiv homogen, d. h. $\sigma(\lambda \cdot x) = \lambda \cdot \sigma(x)$ für $\lambda > 0$, so daß dann mit Hartung (1975a, 1976a) folgt, dass $\sigma(x)$ sogar eine konvexe Funktion bzw. $-\sigma(x)$ eine konkave Funktion in $x = (x_1, \dots, x_K)$ ist. Damit besitzt $-\sigma(x)$ nur ein einziges Maximum über $\{x | \sum_{i=1}^K x_i = 1, x_i \geq 0\}$, welches zudem auch nur an einer einzigen Stelle angenommen wird, wenn die Kovarianzmatrix S invertierbar bzw. positiv definit ist, d. h. wenn $b^T S b > 0$ für alle $b \neq 0$ gilt. Das Selbe gilt dann insbesondere für die oben betrachteten zusammengesetzten Risiko-Funktionen der Gestalt: $\psi(x) - \lambda \cdot \sigma(x)$, mit $\lambda > 0$, für deren Maximumsbetrachtung $\sigma(x)$ nicht durch $\sigma(x)^2$ ersetzt werden kann.

17. Weitere Aspekte der Meta-Analyse

Obwohl in den bisherigen Abschnitten ein weites Spektrum der Meta-Analyse dargestellt wird, bleiben doch noch einige erwähnenswerte Aspekte.

17.1. Ergänzende Tests auf Homogenität

Welch (1951) verbessert den Cochran'schen Homogenitätstest aus Abschnitt 2.4 für den Fall normalverteilter Versuchsauskommen. Ebenfalls für diese Anwendung werden eine Reihe weiterer Tests auf Homogenität analysiert in Hartung/Argac/Makambi (2002, 2003) und speziell hier für zweiarmige Studien in Hartung/Argac (2001). Anstatt die Gleichheit der Mittelwerte im Fixed Effects Modell zu testen, kann man im Random Effects Modell den Heterogenitätsparameter τ^2 auf 0 testen. Man vergleiche hierzu Cochran (1954), Hartung/Argac/Makambi (2001) sowie Hartung/Argac (2002a, b) und Hartung/Knapp (2000b, 2005b). Speziell Studien mit binären Versuchsauskommen werden diesbezüglich in Hartung/Knapp (2001f, 2004) betrachtet. Des Weiteren sei hier auch auf Hartung/Böckenhoff (2000, 2003) und Hartung/Knapp/Sinha (2008) hingewiesen.

17.2. „Partial Weighting“ nach Yates-Cochran bei Dominanz einer Studie oder Prognose und Verallgemeinertes RE-Modell

Ist etwa für eine der in eine Meta-Analyse einbezogenen Studien oder Prognosen ein im Vergleich zu den übrigen Studien oder Prognosen extrem kleiner Streubereich angegeben, der nicht unbedingt als Fehlanalyse eingestuft werden muß, so führt eine Gewichtung der Studien oder Prognosen invers zur selbst angegebenen Varianz zu einer Dominanz der betroffenen Studie oder Prognose in der Kombination aller Beteiligten zur Gewinnung eines gemeinsamen Schätzers oder einer gemeinsamen Prognose, wodurch mitunter eine Verfälschung des Gesamtbildes entstehen kann. Deshalb schlagen bereits **Yates/Cochran (1938)** zur Stützung des dominierenden Gewichtes ein „Partial Weighting“ vor. Dieses Verfahren wird in Hartung/Böckenhoff/Knapp (2003) und Hartung/Knapp (2005a) in formalisierter Darstellung geschildert und an Hand von realen Beispielen diskutiert.

In diesem Zusammenhang wird von Hartung/Böckenhoff/Knapp (2003) und Hartung/Knapp (2005a) auch ein Verallgemeinertes Random Effects Modell in dem Sinn diskutiert, daß, wie bereits von **Wald (1940)** vorgeschlagen, die Wechselwirkungen a_i unterschiedliche Varianzen τ_i^2 besitzen, was in der Varianzanalyse zu **Erweiterten ANOVA Modellen** führt, wie sie auch in Hartung/Knapp (2003b) betrachtet werden, wo speziell Konfidenzbereiche für $(\tau_1^2, \dots, \tau_K^2)$ und ein Homogenitätstest für $H_{0,\text{hom}} : \tau_1^2 = \dots = \tau_K^2$ entwickelt werden.

17.3. Meta-Regression, mit Beispiel „Tuberkulose Studien“

Werden externe Einflußfaktoren x_1, \dots, x_k , welche in den K vorliegenden Studien sich verändernde Werte annehmen, vermutet, so erhalten wir das Modell, vgl. Abschnitt 3,

$$y_i = \psi_i + a_i + e_i, \quad a_i \sim \mathcal{N}(0, \tau^2), \quad e_i \sim \mathcal{N}(0, \sigma_i^2), \quad \hat{\sigma}_i^2 \text{ gegeben},$$

wobei jetzt der Erwartungswert ψ_i in einer Regressionsbeziehung, vgl. Kap. X, steht:

$$\psi_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ji}.$$

Hierbei bezeichne x_{ji} denjenigen Wert, den der Einflußfaktor j in der i -ten Studie annimmt, β_0 das allgemeine Mittel und β_j den Regressionskoeffizienten des j -ten Einflußfaktors.

In Kapitel 10 von Hartung/Knapp/Sinha (2008) wird dieses **Random Effects Modell der Meta-Regression** ausführlich besprochen, vgl. auch Knapp/Hartung (2003a), zu deren vorgeschlagener Varianzschätzung in diesem Modell Sidik/Jonkman (2005) in einer vergleichenden Simulationsstudie feststellen: „The improved variance estimator of Knapp and Hartung yields the best performance“.

Ein bekanntes **Beispiel** sind die in verschiedenen Ländern durchgeführten vergleichenden Untersuchungen zur Verhinderung der **Tuberkulose (TB)** durch die Bacillus-Calmette-Guérin (BCG) Impfung. Da die Länderergebnisse recht unterschiedlich ausfielen, wurde nach Erklärungen gesucht. Als signifikant erwies sich der Breitengrad des Landes, in dem die Studie stattfand, vgl. Hartung/Knapp/Sinha (2008). Dabei stellt der Breitengrad eine **Surrogat-Variable** für das Vorhandensein von Mykrobakterien in der Umwelt dar, welche einen gewissen natürlichen Schutz gegen TB bieten.

Bezeichnet hierbei ein Studienergebnis y_i das Relative Risiko, vgl. Abschnitt 4.3, an TB zu erkranken, und zwar von nicht geimpften Studienteilnehmern im Vergleich zu geimpften Teilnehmern, so sind die Studienergebnisse der $K = 13$ Studien in Abb. 35 veranschaulicht. Die y_i -Werte sind dort in Abhängigkeit vom jeweils zugeordneten absoluten Breitengrad x des Studienortes eingetragen. Ebenfalls dargestellt sind die zugehörigen 95%-Konfidenzintervalle, deren Längen die recht unterschiedlichen Studienumfänge widerspiegeln. Eingezeichnet ist zudem das zur von Hartung/Knapp (2005a) im Random Effects Meta-Regression Model berechneten Regressionsgeraden

$$\ln \hat{y}(x) = -0.1894 + 0.0273 \cdot x \text{ gehörige } \hat{y}(x).$$

Man erkennt an Hand der Abbildung aber auch, daß die Studien grob in 2 **Cluster** aufteilbar sind: Zum einen bestehend aus den 7 Studien, welche in Ländern südlich des 35-sten Breitengrades durchgeführt wurden, und zum anderen aus den 6 Studien in Ländern nördlich des 35-sten Breitengrades, auf dessen Höhe etwa Kreta und Memphis, Tennessee liegen. Zur Anwendung kommen dann die Methoden für Studien-Cluster aus Abschnitt 3.5, die sich als Spezialfall in die Random Effects Meta-Regressionsanalyse einordnen lassen.

17.4. Prospektive Meta-Analyse

Wird etwa mittels einer Meta-Analyse von mehreren klinischen Prüfungen die behördliche Marktzulassung eines Medikamentes angestrebt, so ist es für

Epidemiologie: Tuberkulose (TB)

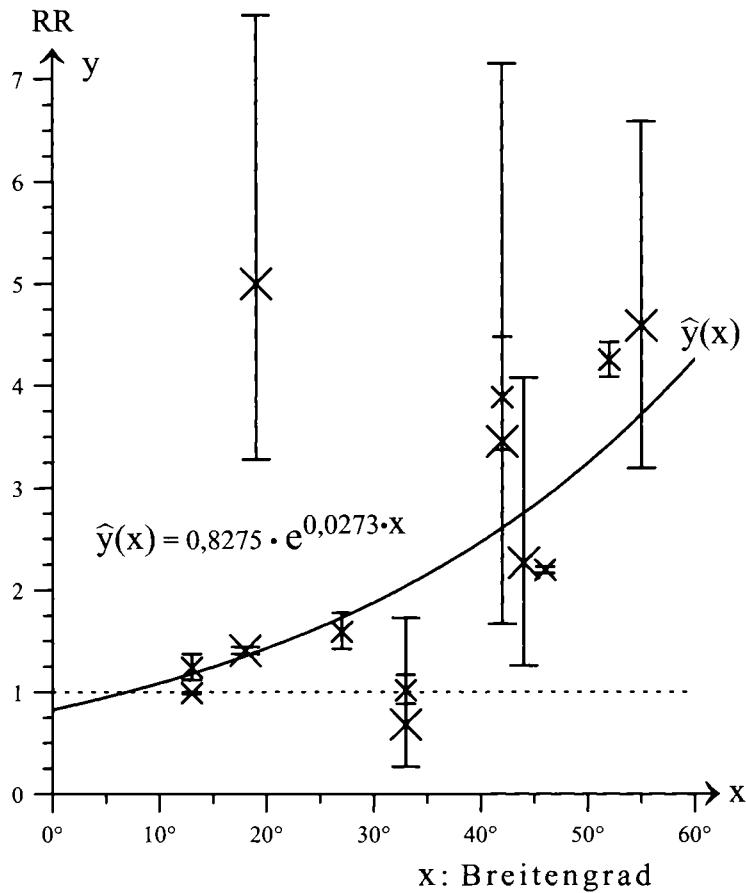


Abb. 35: Wirksamkeit der BCG - Impfung gegen TB

das Zulassungsverfahren günstig, wenn die Meta-Analyse prospektiv angelegt war: Vor Beginn der Studien ist nachweisbar ein Plan für deren mögliche Anzahl, Anlage und Durchführung festgelegt. Die diesbezügliche Zulassungs-Richtlinie EMEA (2001) formuliert zum Beispiel: „When the need for a meta-analysis is prospectively identified, the protocol for the meta-analysis should be prospectively specified during the planning of the clinical development program.“ Hierfür dienliche Überlegungen werden diskutiert in Whitehead (1997, 2002) und Hartung/Knapp (2005a).

17.4.1. Adaptive und „Self-designing“ Versuchsanlagen zur Prospektiven Meta-Analyse

Legt man die Anzahl der definitiv durchzuführenden Studien als K fest und bleiben die Studien unabhängig voneinander, so können die K Studien geplant werden wie eine multizentrische Studie mit K Zentren, vgl. Bock (1998).

Wird lediglich eine Maximalzahl K festgelegt, so ist die Gesamtheit der Studien zu behandeln wie ein einziger gruppen-sequentieller Versuch, in dem die Stufen des Versuchs von den Studien gebildet werden, vgl. Pocock (1977), O'Brien/Fleming (1979), Jennison/Turnbull (2000).

Werden dabei die Studien sukzessive derart **adaptiv** durchgeführt, daß an Hand der Ergebnisse bereits durchgeföhrter Studien die Fallzahl der nächstfolgenden Studie bestimmt wird, so ist die damit verbundene sukzessive Meta-Analyse so durchzuführen wie die Auswertung eines adaptiven gruppen-sequentiellen Versuches, vgl. Hartung (2006) und dortige Referenzen, sowie Hartung/Knapp (2007, 2009a, b).

Zusätzlich zur adaptiven Berechnung der Fallzahlen kann hierbei auch die Anzahl k der insgesamt durchzuführenden Studien adaptiv in Abhängigkeit von den Ergebnissen der bereits durchgeföhrten Studien bestimmt werden. Die Studienanzahl k ist dann eine Zufallsvariable, die von den Studienergebnissen selbst bestimmt wird. Die ebenfalls sukzessive Meta-Analyse der Studien erfolgt dann gemäß der Auswertung eines „**Self-designing**“ gruppen-sequentiellen Versuchs, vgl. Hartung (2000, 2001, 2006), Hartung/Knapp (2003c, 2005a, 2006, 2009c) und in den Publikationen zitierte Referenzen. Im eher wirtschaftlich geprägten Umfeld werden solche Studienanlagen auch als „**Self-organizing**“ gefördert, vgl. Hartung/Knapp (2005c).

Eine Spezialität sind Noninferiority Zulassungsstudien, die ein neues Testpräparat als einem Standardpräparat nicht unterlegen nachweisen sollen. Dabei ist seitens der Zulassungsbehörden die gleichzeitige Mitführung einer Placebogruppe zwecks Gesamtkontrolle gewünscht. Hier erbringt ein adaptiver Verlaufsplan eine erhöhte **Kosteneinsparung**, vgl. Hartung/Knapp (2009d) und das folgende Beispiel.

Ist die Zielsetzung nicht ein Vergleich zweier oder mehrerer Gruppen, sondern die Genauigkeit [accuracy], vgl. Abschnitt 4.3.4 in Kap. V, der Bestimmung eines Parameterschätzers, so entsteht die bereits in Abschnitt 1.4 von Kap. IV diskutierte Aufgabe, $(1-\alpha)$ -Konfidenzintervalle von einer gewünschten **Breite** zu konstruieren. Schon Tukey (1953) forderte, daß ein diesbezügliches Konstruktionsverfahren die geforderte Breite bzw. Genauigkeit auch mit einer gewissen Wahrscheinlichkeit oder **Power** erreichen müsse. Ein solches, das Vorgehen von Stein (1945) und Seelbinder (1953), vgl. Abschnitt 1.4.1 in Kap. IV, adaptiv erweiterndes Verfahren wird von Hartung/Knapp (2009e) entwickelt.

17.4.2. Beispiel: Adaptive Placebo-kontrollierte prospektive Nicht-unterlegenheits-Studien zu Asthma Bronchiale

Zur Illustration einer prospektiven und adaptiven Meta-Analyse betrachten wir in einer auf bis zu $K = 3$ Stufen angelegten gruppen-sequentiellen Versuchsanlage gemäß Pocock (1977), allerdings mit **flexiblen Gruppenumfängen** gemäß adaptiver Fallzahlenplanung wie in Hartung (2006), mit-

Prospektive Meta - Analyse : Adaptiv

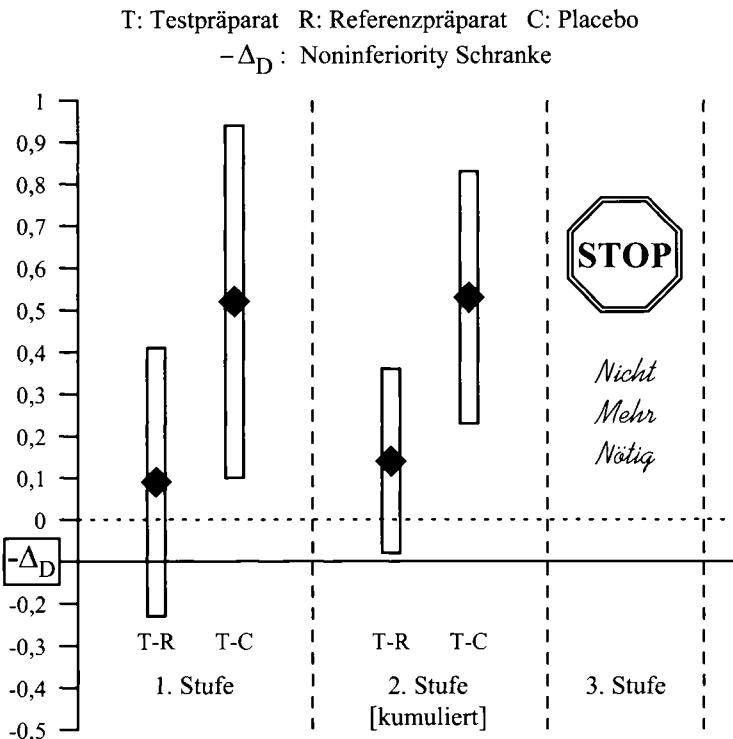


Abb. 36: Asthma Bronchiale – Controlled Noninferiority Trials.
Bis K = 3 Stufen adaptiv geplant mit 4:2:1 Randomisation.

tels Placebo (C) kontrollierte Nichtunterlegenheits-Studien einer neuen Testmedikation (T) bzgl. einer Standard-Referenzmedikation (R) für Patienten mit Asthma Bronchiale. Der klinische Vergleichsparameter ist FEV₁: forced expiratory volume in 1 sec [in l] und das Effektmaß ist die gewöhnliche Differenz der Mittelwerte, so daß wir wie in Abschnitt 14.1.1 auswerten können, und zwar sukzessive, indem wir dort zunächst K = 1 setzen und die 1. Studie auswerten, dann K = 2 setzen und kumuliert die 1. Studie und die 2. Studie gemeinsam auswerten, usw. Dabei ist jedoch der **kritische Wert** $u_{1-\alpha/2} = 1.96$ für $\alpha = 0.05$ zu ersetzen durch **2.289** hier im Fall von maximal 3 Stufen, z. B. Hartung (2006). Dank des höheren kritischen Wertes besitzen alle auf den Stufen $i = 1, 2$ und gegebenenfalls 3 berechneten Konfidenzintervalle $KI(T - C)_i$ für die mittlere Differenz der Gruppen T und C sowie $KI(T - R)_i$ für die mittlere Differenz der Gruppen T und R jeweils ein Konfidenzniveau von mindestens 95%. Die Fallzahlen der Gruppen T, R und C in den verschiedenen Stufen sind nun adaptiv, d. h. aufbauend auf bereits erzielte Resultate, derart zu bestimmen, daß in einer Stufe

Tab. 39: Placebo C kontrollierte Nichtunterlegenheitsstudien eines neuen Testpräparats T im Vergleich zu einem Standard-Referenzpräparat R [bzgl. FEV₁ mit $\Delta_D = 0.11$] für Patienten mit Asthma Bronchiale – Daten von 2 realisierten Studien, adaptiv angelegt in 7-er Blöcken mit der Randomisation 4(T) : 2(R) : 1(C); Bis K = 3 Studien-Stufen geplant, vorzeitiger Stopp wegen signifikanter Ergebnisse: T – C in 1. und 2. Stufe, T – R in 2. Stufe

Stufe bzw. Studie	Adaptive Fallzahlen			Versuchsergebnisse [in l]			Freiheits- grade		
	i	n _{T_i}	n _{R_i}	n _{C_i}	\bar{x}_{T_i}	\bar{x}_{R_i}	\bar{x}_{C_i}	s _i	ν_i
1	116	58	29	2.65	2.56	2.13	0.87	200	
2	96	48	24	2.69	2.51	2.15	0.81	165	
3					STOPP				
					Wegen bereits gezeigter signifikanter Resultate				

$j \in \{1, 2, 3\} : KI(T - C)_j > 0$ und $KI(T - R)_j > -\Delta_D$ erreicht wird, wobei hier $\Delta_D = 0.11$ der vorgegebene Nichtunterlegenheitsparameter ist, vgl. Hartung (2006), Hartung/Knapp (2009d).

In **Tab. 39** sind die Daten zu unserer Versuchsanlage zusammengestellt. Dabei wurde für die Gruppen T, R und C die Gleichheit der Varianzen angenommen: $\xi_T^2 = \xi_R^2 = \xi_C^2$, so daß s_i^2 die gepoolte Varianz dieser 3 Gruppen in der i-ten Stufe bzw. Studie ist, der folglich die Freiheitsgrade $\nu_i = n_{T_i} + n_{R_i} + n_{C_i} - 3$ zugeordnet sind. Mit obigen Modifikationen zur Auswertung der Versuche gemäß Abschnitt 14.1.1 erhalten wir die Konfidenzintervalle

$$\begin{aligned} KI(T - R)_1 &= [-0.23, 0.41] & \text{und } KI(T - R)_2 &= [-0.09, 0.36], \\ KI(T - C)_1 &= [0.10, 0.94] & \text{und } KI(T - C)_2 &= [0.23, 0.83], \end{aligned}$$

so wie wir sie in **Abb. 36** graphisch veranschaulicht haben. Das Konfidenzniveau eines jeden Konfidenzintervalls beträgt mindestens 0.95. Da $KI(T - C)_2$ oberhalb von 0 und $KI(T - R)_2$ oberhalb von $-\Delta_D = -0.11$ liegen, ist die **Aufgabe der Prospektiven Meta-Analyse** somit hier bereits **erfüllt** und wir können die optionale 3. Stufe des Versuchsanlage **einsparen**. Da bereits in der 1. Stufe $KI(T - C)_1 > 0$, hätten wir schon die Placebogruppe bereits in der 2. Stufe einsparen können. Hier wurde sie jedoch weiter mitgeführt, um z. B. mögliche Nebenwirkungen von T und R auch weiterhin neutral kontrollieren zu können.

17.5. Eigenschaften des Gauß-Markov-Schätzers und der Varianzschätzer sowie verwandter Schätzer

Für $i = 1, \dots, K$ mögen die [stochastisch] unabhängigen Zufallsvariablen y_i den Erwartungswert $Ey_i = \psi$ und die Varianz $Var y_i = V_i > 0$ besitzen. Ein mit den Gewichten b_i gebildeter Schätzer $\hat{\psi}_b = \sum_{i=1}^K b_i \cdot y_i$ für ψ ist erwartungstreu, wenn $E \hat{\psi}_b = \sum_{i=1}^K b_i E y_i = \sum_{i=1}^K b_i \psi = \psi$ und somit $\sum_{i=1}^K b_i = 1$ gilt. Soll die Varianz von $\hat{\psi}_b$: $Var \hat{\psi}_b = \sum_{i=1}^K b_i^2 V_i$ als positive und konvexe Funktion der b_i **minimal** werden unter allen erwartungstreuen

Schätzern der Form $\hat{\psi}_b$, so bilden wir die Lagrange-Funktion:

$$L(b_1, \dots, b_K; \lambda) = \frac{1}{2} \sum_{i=1}^K b_i^2 V_i + \lambda \left(1 - \sum_{i=1}^K b_i \right)$$

und setzen die partiellen Ableitungen gleich Null:

$$\frac{\partial L}{\partial b_i} = b_i \cdot V_i - \lambda = 0, \quad \frac{\partial L}{\partial \lambda} = 1 - \sum_{i=1}^K b_i = 0,$$

woraus folgt:

$$\begin{aligned} b_i &= \lambda / V_i, \quad \sum_{i=1}^K b_i = \lambda \cdot \sum_{i=1}^K 1/V_i = 1, \text{ d. h. } \lambda = 1 / \sum_{i=1}^K 1/V_i \text{ und} \\ b_i &= \frac{1/V_i}{\sum_{j=1}^K 1/V_j}, \quad \text{mit} \\ \text{Var } \hat{\psi}_b &= \frac{1}{\left(\sum_{j=1}^K 1/V_j \right)^2} \cdot \sum_{i=1}^K \frac{1}{V_i^2} \cdot V_i = \frac{1}{\sum_{j=1}^K 1/V_j}. \end{aligned}$$

Dieser **optimale** Schätzer ist folglich der **Gauß-Markov Schätzer**:

$$\hat{\psi}_{GM} = \sum_{i=1}^K \frac{y_i / V_i}{\sum_{j=1}^K 1/V_j} \quad \text{mit} \quad \sigma_{GM}^2 = \frac{1}{\sum_{j=1}^K 1/V_j}.$$

Setzt man zusätzlich **Normalverteilung** voraus, so besitzt der Gauß-Markov Schätzer die kleinste Varianz unter allen erwartungstreuen Schätzern für ψ , also auch unter solchen, die nicht die lineare Form von $\hat{\psi}_b$ haben, vgl. Lehmann/Casella (1998).

Wir haben nun auch **Schätzer** betrachtet, welche **invers zur Standardabweichung gewichten**:

$$\hat{\psi}_H = \sum_{i=1}^K W_i \cdot y_i / \sum_{j=1}^K W_j \quad \text{mit} \quad W_i = 1/\sqrt{V_i},$$

so daß dann:

$$E \hat{\psi}_H = \psi \quad \text{und} \quad \text{Var } \hat{\psi}_H = \frac{K}{\left(\sum_{i=1}^K 1/\sqrt{V_i} \right)^2} = \sigma_H^2.$$

Natürlich muß $\sigma_H^2 \geq \sigma_{GM}^2$ sein, wie obige Ausführungen implizieren, aber läßt sich dies auch einsehen?

Dazu definieren wir die Zufallsvariable X mit $P\{X = W_i\} = 1/K$, so daß:

$$E X = \sum_{i=1}^K \frac{1}{K} W_i \text{ und } E X^2 = \sum_{i=1}^K \frac{1}{K} W_i^2.$$

Die Funktion $h(X) = X^2$ ist eine konvexe Funktion und die **Jensensche Ungleichung**, vgl. Abschnitt 8.1.A in Kap. II, besagt:

$$\begin{aligned} E h(X) &\geq h(E X), \text{ d. h.} \\ E X^2 &= \sum_{i=1}^K \frac{1}{K} W_i^2 \geq \left(\sum_{i=1}^K \frac{1}{K} W_i \right)^2 = (E X)^2, \text{ bzw.} \\ \frac{1}{K} \sum_{i=1}^K W_i^2 &\geq \frac{1}{K^2} \left(\sum_{i=1}^K W_i \right)^2, \text{ oder} \\ \sigma_H^2 &= \frac{K}{\left(\sum_{i=1}^K 1/\sqrt{V_i} \right)^2} = \frac{K}{\left(\sum_{i=1}^K W_i \right)^2} \geq \frac{1}{\sum_{i=1}^K W_i^2} = \frac{1}{\sum_{i=1}^K 1/V_i} = \sigma_{GM}^2. \end{aligned}$$

Diese Überlegungen erleichtern schließlich unsere Einsicht.

Darüber hinaus, sind $\widehat{V}_i > 0$ Schätzer für V_i , die **nicht unverzerrt** sein müssen, jedoch **unabhängig** von den y_i sein sollen, dann bleiben die damit an Stelle der V_i gewichteten Schätzer unverzerrt:

$$\begin{aligned} \text{Für } \widehat{\psi}_{GM} &= \sum_{i=1}^K \frac{y_i/\widehat{V}_i}{\sum_{j=1}^K 1/\widehat{V}_j} \text{ gilt: } E \widehat{\psi}_{GM} = \psi, \text{ und} \\ \text{für } \widehat{\psi}_H &= \sum_{i=1}^K \frac{y_i/\sqrt{\widehat{V}_i}}{\sum_{j=1}^K 1/\sqrt{\widehat{V}_j}} \text{ gilt: } E \widehat{\psi}_H = \psi. \end{aligned}$$

Wir zeigen diese **Erwartungstreue** für den „geschätzten“ Gauß-Markov Schätzer $\widehat{\psi}_{GM}$ und für $\widehat{\psi}_H$ allgemein wie folgt:

Nach Abschnitt 8.1.A in Kap. II gilt für K nicht notwendigerweise unabhängige Paare von unabhängigen Zufallsvariablen X_i und Y_i , mit $E Y_i = \psi$ für alle $i = 1, \dots, K$:

$$\begin{aligned} E \left(\sum_{i=1}^K X_i \cdot Y_i \right) &= \sum_{i=1}^K E X_i \cdot Y_i = \sum_{i=1}^K E X_i \cdot E Y_i \\ &= \sum_{i=1}^K E X_i \cdot \psi = \psi \cdot \sum_{i=1}^K E X_i = \psi \cdot E \left(\sum_{i=1}^K X_i \right) \\ &= \psi \cdot 1 = \psi \text{ falls } \sum_{i=1}^K X_i = 1. \end{aligned}$$

Setzen wir nun:

$$Y_i = y_i \text{ und } X_i = \frac{1/\hat{V}_i}{\sum_{j=1}^K 1/\hat{V}_j} \text{ oder } X_i = \frac{1/\sqrt{\hat{V}_i}}{\left(\sum_{j=1}^K 1/\sqrt{\hat{V}_j}\right)^2},$$

so ist damit die Unverzerrtheit von $\widehat{\psi}_{GM}$ und $\widehat{\psi}_H$ gezeigt. Leider überträgt sich diese Eigenschaft nicht auf die Varianzschätzer:

$$\widehat{\sigma}_{GM}^2 = \frac{1}{\sum_{i=1}^K 1/\hat{V}_i} \text{ und } \widehat{\sigma}_H^2 = \frac{K}{\left(\sum_{i=1}^K 1/\sqrt{\hat{V}_i}\right)^2}.$$

Wir entnehmen Hartung (1972a, 1975a, 1976a) einige Eigenschaften zusammengesetzter Funktionen: Mit $0 < \gamma \leq 1$ sind für positive x_1, \dots, x_K die Funktionen x_i^γ konkav und somit $1/x_i^\gamma$ konvexe Funktionen. Dann ist $f = \sum_{i=1}^K 1/x_i^\gamma$ eine konvexe Funktion in K Variablen und, da f positiv ist, ist $f^{1/\gamma}$, mit $0 < \gamma \leq 1$, ebenfalls konvex. Weiterhin ist $g = \delta/f^{1/\gamma}$, mit $\delta > 0$, eine **quasi-konkave** Funktion, wobei die Funktion $g = g(x_1, \dots, x_K)$ quasi-konkav ist, wenn die oberen Niveau-Mengen:

$$NM_g(\epsilon) = \{(x_1, \dots, x_K) \mid g(x_1, \dots, x_K) \geq \epsilon, x_1 > 0, \dots, x_K > 0\}$$

für alle $\epsilon \geq 0$ konvexe Mengen sind, d. h. mit zwei Punkten (x_1, \dots, x_K) und (u_1, \dots, u_K) liegen auch die Zwischenpunkte $(\beta x_1 + (1-\beta)u_1, \dots, \beta x_K + (1-\beta)u_K)$, mit $0 \leq \beta \leq 1$, in der konvexen Menge. Ferner ist g eine **positiv homogene** Funktion, d. h. für $\lambda > 0$ gilt:

$$g(\lambda x_1, \dots, \lambda x_K) = \frac{\delta}{\left(\sum_{i=1}^K 1/\{\lambda x_i\}^\gamma\right)^{1/\gamma}} = \lambda \cdot g(x_1, \dots, x_K).$$

Für eine quasi-konkave Funktion, die zudem positiv homogen ist, lässt sich nun aber zeigen, vgl. Hartung (1975a, 1976a), daß sie sogar konkav ist. Folglich ist g eine konkav Funktion, d. h. der untere Epigraph von g :

$$EG(g) = \{(z, x_1, \dots, x_K) \mid z \leq g(x_1, \dots, x_K), x_1 > 0, \dots, x_K > 0\}$$

ist eine konvexe Menge [f (quasi-)konvex, wenn $-f$ (quasi-)konkav]. Sind x_1, \dots, x_K Zufallsvariablen, so besagt die **Jensensche Ungleichung** für mehrdimensionale Zufallsvariablen dann:

$$Eg(x_1, \dots, x_K) \leq g(E x_1, \dots, E x_K).$$

Wir setzen $x_i = \hat{V}_i$, und mit $\gamma = \delta = 1$ ergibt sich: $g = \widehat{\sigma}_{GM}^2$, bzw. mit $\gamma = 1/2$ und $\delta = K$: $g = \widehat{\sigma}_H^2$. Folglich erhalten wir die Resultate:

$$\begin{aligned} E \widehat{\sigma}_{GM}^2 &= E \frac{1}{\sum_{i=1}^K 1/\hat{V}_i} \leq \frac{1}{\sum_{i=1}^K 1/E \hat{V}_i} = \sigma_{GM}^2, \text{ falls } E \hat{V}_i = V_i, \text{ und} \\ E \widehat{\sigma}_H^2 &= E \frac{K}{\left(\sum_{i=1}^K 1/\sqrt{\hat{V}_i}\right)^2} \leq \frac{K}{\left(\sum_{i=1}^K 1/\sqrt{E \hat{V}_i}\right)^2} = \sigma_H^2, \text{ falls } E \hat{V}_i = V_i. \end{aligned}$$

Sind also die \widehat{V}_i unverzerrte Schätzer für die V_i , wie von jetzt an auch angenommen, so werden mit den Varianzschätzern $\widehat{\sigma}_{GM}^2$ und $\widehat{\sigma}_H^2$ die wahren Varianzen σ_{GM}^2 und σ_H^2 jeweils **systematisch unterschätzt**.

Darüber hinaus übersteigen natürlich die Varianzen der „geschätzten“ Schätzer $\widehat{\psi}_{GM}$ und $\widehat{\psi}_H$ die Varianzen der mit den wahren Varianzen V_i gewichteten Schätzer:

$$\text{Var } \widehat{\psi}_{GM} \geq \sigma_{GM}^2 \quad \text{und} \quad \text{Var } \widehat{\psi}_H \geq \sigma_H^2.$$

Sind die y_i normalverteilt und werden den \widehat{V}_i jetzt χ^2 -Verteilungen mit ν_i Freiheitsgraden, die proportional zu den involvierten Stichprobenumfängen stehen, zugeordnet:

$$y_i \sim \mathcal{N}(\psi, V_i) \quad \text{und} \quad \nu_i \cdot \widehat{V}_i / V_i \sim \chi_{\nu_i}^2, \quad \text{für } i = 1, \dots, K \geq 2,$$

so leitet etwa **Meier (1953)** für die Varianz des „geschätzten“ Gauß-Markov-Schätzers die folgende Entwicklung in eine **Potenzreihe** bzgl. $1/\nu_i$ her:

$$\text{Var } \widehat{\psi}_{GM} = \sigma_{GM}^2 \cdot \left\{ 1 + 2 \cdot \sum_{i=1}^K \left[\frac{1}{\nu_i} d_i (1 - d_i) + \frac{1}{\nu_i^2} A_{\nu_i} \right] \right\},$$

wobei $d_i = (1/V_i) / \sum_{j=1}^K 1/V_j$ und $A_{\nu_i} \rightarrow A$ für $\nu_i \rightarrow \infty$. Betrachten wir nun unsere Quadratform:

$$q_H = \frac{1}{K-1} \sum_{i=1}^K \widehat{d}_i \left(y_i - \widehat{\psi}_{GM} \right)^2, \quad \text{mit} \quad \widehat{d}_i = \frac{1/\widehat{V}_i}{\sum_{j=1}^K 1/\widehat{V}_j},$$

so gilt, vgl. Hartung/Knapp (2005a), für den **[unbedingten] Erwartungswert von q_H** die folgende **Potenzreihen-Entwicklung**:

$$\begin{aligned} E q_H &= \sigma_{GM}^2 \cdot \left\{ 1 + 2 \cdot \sum_{i=1}^K \left[\frac{1}{\nu_i} \left(\frac{K}{K-1} d_i (1 - d_i) \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{1}{K-1} (1 - d_i)^2 \right) + \frac{1}{\nu_i^2} B_{\nu_i} \right] \right\}, \end{aligned}$$

wobei jetzt $B_{\nu_i} \rightarrow B$ für $\nu_i \rightarrow \infty$. Dabei sind A und B [endliche] Konstanten.

Die entscheidenden ersten Terme von $E q_H$ dominieren die entsprechenden ersten Terme von $\text{Var } \widehat{\psi}_{GM}$, wobei mit wachsendem K die Differenz zum Verschwinden kommt.

Damit liegt zugleich eine der theoretischen Begründungen dafür vor, daß q_H in Simulationen so überzeugende Ergebnisse zeigt.

Anhang

1. Tabellenverzeichnis

1.1. Kritische Werte, Quantile

Inhalt der Tabelle	Kapitel	Tabelle	Seite
Quantile u_γ der Standardnormalverteilung	Anhang	2	1027
Quantile $t_{n; \gamma}$ der t-Verteilung	Anhang	3	1028
Quantile $\chi^2_{n; \gamma}$ der χ^2 -Verteilung	Anhang	4	1029
Quantile $F_{n_1, n_2; \gamma}$ der F-Verteilung	Anhang	5	1031
Quantile $q_{n, v; \gamma}$ der Studentisierten Spannweite	Anhang	6	1038
Kritische Werte $d_{n; 1-\alpha}$ für den Kolmogoroff-Smirnov-Anpassungstest	IV	14	184
Kritische Werte $l_{n; 1-\alpha}^{\text{norm}}$ zum Kolmogoroff-Smirnov-Test auf nicht spezifizierte Normalverteilung	IV	15	185
Kritische Werte $l_{n; 1-\alpha}^{\text{exp}}$ zum Kolmogoroff-Smirnov-Test auf nicht spezifizierte Exponentialverteilung	IV	34	227
Kritische Werte $w_{n; \gamma}$ des Vorzeichenrangtests von Wilcoxon	IV	40	245
Kritische Werte $K_{n; 1-\alpha}$ der Kendallschen K-Statistik	IV	44	249
Kritische Werte $R_{0,05}$ (I, J) und wahre Grenzen des Testniveaus für Youden's extremen Rangsummentest	V	9	339
Quantile $Q_{n; \gamma}$ des David-Hartley-Pearson-Tests auf Ausreißer	V	13	344
Kritische Werte $T_{n; \gamma}$ des Ausreißertests von Grubbs	V	14	345
Kritische Werte der Dixon-Tests auf Ausreißer	V	16	347
Kritische Werte $s_{n; \alpha}$ für die Tests auf ein Ausreißerpaaar	V	17	348
Kritische Werte $W_{n_1, n_2; \gamma}$ des Wilcoxon-Rangsummen-Tests	VIII	4	518
Kritische Werte $z_{2, n; 1-\alpha}$ des Kolmogoroff-Smirnov-Tests zum Vergleich zweier empirischer Verteilungsfunktionen bei gleichen Stichprobenumfängen	VIII	6	521
Approximative kritische Werte $k_{n_1, n_2; 1-\alpha}$ des Kolmogoroff-Smirnov-Tests zum Vergleich zweier empirischer Verteilungsfunktionen bei ungleichen Reihenlängen	VIII	7	522
Kritische Werte $h_{n; \alpha}$ der Hotelling-Pabst-Statistik	IX	5	557
Kritische Werte $h_{3, (n_1, n_2, n_3); 1-\gamma}$ zum Kruskal-Wallis-Test	XI	6	615
Kritische Werte $v_{p, r; 1-\gamma}$ des Friedman-Tests auf signifikante Behandlungsunterschiede im Blockexperiment	XI	10	623
Kritische Werte $k_{n; \gamma}$ der cttot-Teststatistik K_n	XIII	7	778
Kritische Werte $j_{n; \gamma}$ des Hollander-Prochan-Tests	XIII	9	781

1.2. Weitere allgemeine Tabellen

Inhalt der Tabelle	Kapitel	Tabelle	Seite
Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung $N(0,1)$	Anhang	1	1026
Gammafunktion $\Gamma(x)$	Anhang	7	1040
Gleichverteilte Zufallszahlen	Anhang	8	1041
Fehlerarten beim Testen (Fehler 1. und 2. Art)	III	2	133
Korrekturfaktoren γ_n zum Schätzen der Standardabweichung σ mittels der empirischen Standardabweichung bei normalverteilter Grundgesamtheit	IV	1	158
Korrekturfaktoren α_n zum Schätzen der Standardabweichung σ mittels Spannweite bei normalverteilter Grundgesamtheit	IV	2	158
Korrekturfaktoren δ_k zur Berechnung von Konfidenzintervallen für die Standardabweichung σ mittels Spannweite bei normalverteilter Grundgesamtheit	IV	5	163
Toleranzfaktoren λ zur Berechnung von Toleranzintervallen bei normalverteilter Grundgesamtheit	IV	6	165
r-Statistiken von Dixon für Tests auf Ausreißer	V	15	346
Erforderliche exakte Stichprobenumfänge $n_1 = n_2$ für einseitige Tests in der 2×2 -Tafel (Vierfeldertafel)	VII	15	421
Hypothesen für die rxs-Kontingenztafel	VII	26	436
Maximum-Likelihood-Schätzwerte für hierarchische loglineare Modelle 3dimensionaler Kontingenztafeln	VII	65	485
Maximum-Likelihood-Schätzwerte für einige hierarchische loglineare Modelle	VII	66	486
4dimensionaler Kontingenztafeln	VIII	9	527
Vergabe der Ränge zum Ansari-Bradley-Freund-Test	IX	2	548
Die Fishersche Z-Transformation	XI	3	611
Tafel der einfachen Varianzanalyse	XI	7	621
Varianzanalysetafel für das einfache Blockexperiment	XI	11	626
Varianzanalysetafel der zweifachen Varianzanalyse mit Wechselwirkungen bei n Beobachtungen pro Zelle	XI	14	628
Varianzanalysetafel der zweifachen Varianzanalyse ohne Wechselwirkungen bei n Beobachtungen pro Zelle	XI	15	630
Varianzanalysetafel der einfach hierarchischen Klassifikation mit zufälligen Effekten (Modell II)	XIII	16	805
Wertungsfaktoren für im Labor ermittelte Ausfallraten von elektronischen Bauelementen	XIII	17	805
Ausfallraten (Vollauffälle) bei 50°C Umgebungstemperatur für Transistoren, Widerstände und Kondensatoren	XIII		

2. Tabellenanhang

Tab. 1: Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung $N(0, 1)$

Tab. 2: Quantile u_γ der Standardnormalverteilung $N(0, 1)$

Tab. 3: Quantile $t_{n; \gamma}$ der t-Verteilung

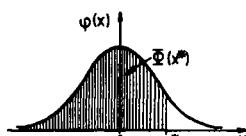
Tab. 4: Quantile $\chi^2_{n; \gamma}$ der χ^2 -Verteilung

Tab. 5: Quantile $F_{n_1, n_2; \gamma}$ der F-Verteilung

Tab. 6: Quantile $q_{n, v; \gamma}$ der Studentisierten Spannweite

Tab. 7: Gammafunktion $\Gamma(x)$

Tab. 8: Gleichverteilte Zufallszahlen



Tab. 1: Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung $N(0, 1)$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

zu Tab. 1:

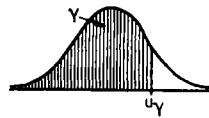
Ablesebeispiel: $\Phi(1,56) = 0,9406$

Erweiterung der Tafel: $\Phi(-x) = 1 - \Phi(x)$

Approximation nach Hastings für $x > 0$:

$$\Phi(x) \approx 1 - \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \cdot (a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5) \quad \text{mit} \quad t = \frac{1}{1 + bx},$$

$$b = 0,2316419, \quad a_1 = 0,31938153, \quad a_2 = -0,356563782, \quad a_3 = 1,781477937, \\ a_4 = -1,821255978, \quad a_5 = 1,330274429.$$

Tab. 2: Quantile u_γ der Standardnormalverteilung $N(0, 1)$

γ	u_γ	γ	u_γ	γ	u_γ	γ	u_γ
0,9999	3,7190	0,9975	2,8070	0,965	1,8119	0,83	0,9542
0,9998	3,5401	0,9970	2,7478	0,960	1,7507	0,82	0,9154
0,9997	3,4316	0,9965	2,6968	0,955	1,6954	0,81	0,8779
0,9996	3,3528	0,9960	2,6521	0,950	1,6449	0,80	0,8416
0,9995	3,2905	0,9955	2,6121	0,945	1,5982	0,79	0,8064
0,9994	3,2389	0,9950	2,5758	0,940	1,5548	0,78	0,7722
0,9993	3,1947	0,9945	2,5427	0,935	1,5141	0,76	0,7063
0,9992	3,1559	0,9940	2,5121	0,930	1,4758	0,74	0,6433
0,9991	3,1214	0,9935	2,4838	0,925	1,4395	0,72	0,5828
0,9990	3,0902	0,9930	2,4573	0,920	1,4051	0,70	0,5244
0,9989	3,0618	0,9925	2,4324	0,915	1,3722	0,68	0,4677
0,9988	3,0357	0,9920	2,4089	0,910	1,3408	0,66	0,4125
0,9987	3,0115	0,9915	2,3867	0,905	1,3106	0,64	0,3585
0,9986	2,9889	0,9910	2,3656	0,900	1,2816	0,62	0,3055
0,9985	2,9677	0,9905	2,3455	0,890	1,2265	0,60	0,2533
0,9984	2,9478	0,9900	2,3263	0,880	1,1750	0,58	0,2019
0,9983	2,9290	0,9850	2,1701	0,870.	1,1264	0,56	0,1510
0,9982	2,9112	0,9800	2,0537	0,860	1,0803	0,54	0,1004
0,9981	2,8943	0,9750	1,9600	0,850	1,0364	0,52	0,0502
0,9980	2,8782	0,9700	1,8808	0,840	0,9945	0,50	0,0000

zu Tab. 2:Ablesebeispiel: $u_{0,95} = 1,6449$ Erweiterung der Tafel: $u_{1-\gamma} = -u_\gamma$ Approximation nach Hastings für $0,5 < \gamma < 1$:

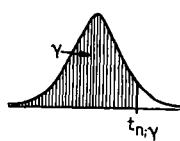
$$u_\gamma \approx t - \frac{a_0 + a_1 t + a_2 t^2}{1 + b_1 t + b_2 t^2 + b_3 t^3} \quad \text{mit} \quad t = \sqrt{-2 \ln(1 - \gamma)},$$

$$a_0 = 2,515517, \quad a_1 = 0,802853, \quad a_2 = 0,010328,$$

$$b_1 = 1,432788, \quad b_2 = 0,189269, \quad b_3 = 0,001308.$$

Tab. 3: Quantile $t_{n;\gamma}$ der t-Verteilung

n \ \gamma	0,990	0,975	0,950	0,900
1	31,821	12,706	6,314	3,078
2	6,965	4,303	2,920	1,886
3	4,541	3,182	2,353	1,638
4	3,747	2,776	2,132	1,533
5	3,365	2,571	2,015	1,476
6	3,143	2,447	1,943	1,440
7	2,998	2,365	1,895	1,415
8	2,896	2,306	1,860	1,397
9	2,821	2,262	1,833	1,383
10	2,764	2,228	1,812	1,372
11	2,718	2,201	1,796	1,363
12	2,681	2,179	1,782	1,356
13	2,650	2,160	1,771	1,350
14	2,624	2,145	1,761	1,345
15	2,602	2,131	1,753	1,341
16	2,583	2,120	1,746	1,337
17	2,567	2,110	1,740	1,333
18	2,552	2,101	1,734	1,330
19	2,539	2,093	1,729	1,328
20	2,528	2,086	1,725	1,325
21	2,518	2,080	1,721	1,323
22	2,508	2,074	1,717	1,321
23	2,500	2,069	1,714	1,319
24	2,492	2,064	1,711	1,318
25	2,485	2,060	1,708	1,316
26	2,479	2,056	1,706	1,315
27	2,473	2,052	1,703	1,314
28	2,467	2,048	1,701	1,313
29	2,462	2,045	1,699	1,311
30	2,457	2,042	1,697	1,310
40	2,423	2,021	1,684	1,303
50	2,403	2,009	1,676	1,299
60	2,390	2,000	1,671	1,296
70	2,381	1,994	1,667	1,294
80	2,374	1,990	1,664	1,292
90	2,369	1,987	1,662	1,291
100	2,364	1,984	1,660	1,290
150	2,352	1,976	1,655	1,287
200	2,345	1,972	1,653	1,286
300	2,339	1,968	1,650	1,284
400	2,336	1,966	1,649	1,284
600	2,333	1,964	1,647	1,283
800	2,331	1,963	1,647	1,283
1000	2,330	1,962	1,646	1,282
∞	2,326	1,960	1,645	1,282

**zu Tab. 3:**Ablesebeispiel: $t_{15,0,95} = 1,753$

Erweiterung der Tafel:

$$t_{n;1-\gamma} = -t_{n;\gamma}$$

und speziell

$$t_{1;\gamma} = \tan(\pi \cdot \{\gamma - \frac{1}{2}\}),$$

$$t_{2;\gamma} = \frac{\sqrt{2} \cdot (2 \cdot \gamma - 1)}{\sqrt{1 - (2 \cdot \gamma - 1)^2}},$$

$$t_{\infty;\gamma} = u_{\gamma}.$$

Approximation für $0,5 < \gamma < 1$:

$$t_{n;\gamma} \approx \frac{c_9 u^9 + c_7 u^7 + c_5 u^5 + c_3 u^3 + c_1 u}{92160 n^4}$$

$$\text{mit } u = u_{\gamma}, \quad c_9 = 79, \quad c_7 = 720n + 776,$$

$$c_5 = 4800n^2 + 4560n + 1482,$$

$$c_3 = 23040n^3 + 15360n^2 + 4080n - 1920,$$

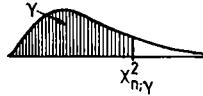
$$c_1 = 92160n^4 + 23040n^3 + 2880n^2 \\ - 3600n - 945;$$

für $n \geq 10$ kann man auch die **Formel von Peizer und Pratt** verwenden:

$$t_{n;\gamma} \approx \sqrt{n \cdot e^{c \cdot u^2} - n} \quad \text{mit } u = u_{\gamma} \quad \text{und}$$

$$c = \frac{n - \frac{5}{6}}{\left(n - \frac{2}{3} + \frac{1}{10n}\right)^2}.$$

(Anmerkung: Die Peizer-Pratt-Approximation liefert bereits für $n=3$ und $0,5 < \gamma < 0,99$ eine passable Anpassung, wobei die absolute Abweichung zum wahren Wert höchstens 0,08 wird.)

Tab. 4: Quantile $x^2_{n;\gamma}$ der χ^2 -Verteilung

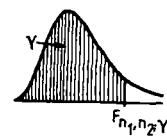
$n \backslash \gamma$	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
1	7,879	6,635	5,024	3,841	2,706	1,323	0,455	0,102	-2,158	-3,93	-4,9,82	-4,1,57	-5,3,93
2	10,60	9,210	7,378	5,991	4,605	2,773	1,386	0,575	0,211	0,103	-2,5,06	-2,2,01	-2,1,00
3	12,84	11,34	9,348	7,815	6,251	4,108	2,366	1,213	0,584	0,352	0,216	0,115	-2,7,17
4	14,86	13,28	11,14	9,488	7,779	5,385	3,357	1,923	1,064	0,711	0,484	0,297	0,207
5	16,75	15,09	12,83	11,07	9,236	6,626	4,351	2,675	1,610	1,145	0,831	0,554	0,412
6	18,55	16,81	14,45	12,59	10,64	7,841	5,348	3,455	2,204	1,635	1,237	0,872	0,676
7	20,28	18,48	16,01	14,07	12,02	9,037	6,346	4,255	2,833	2,167	1,690	1,239	0,989
8	21,96	20,09	17,53	15,51	13,36	10,22	7,344	5,071	3,490	2,733	2,180	1,647	1,344
9	23,59	21,67	19,02	16,92	14,68	11,39	8,343	5,899	4,168	3,325	2,700	2,088	1,735
10	25,19	23,21	20,48	18,31	15,99	12,55	9,342	6,737	4,865	3,940	3,247	2,558	2,156
11	26,76	24,73	21,92	19,68	17,28	13,70	10,34	7,584	5,578	4,575	3,816	3,053	2,603
12	28,30	26,22	23,34	21,03	18,55	14,85	11,34	8,438	6,304	5,226	4,404	3,571	3,074
13	29,82	27,69	24,74	22,36	19,81	15,98	12,34	9,299	7,042	5,892	5,009	4,107	3,565
14	31,32	29,14	26,12	23,68	21,06	17,12	13,34	10,17	7,790	6,571	5,629	4,660	4,075
15	32,80	30,58	27,49	25,00	22,31	18,25	14,34	11,04	8,547	7,261	6,262	5,229	4,601
16	34,27	32,00	28,85	26,30	23,54	19,37	15,34	11,91	9,312	7,962	6,908	5,812	5,142
17	35,72	33,41	30,19	27,59	24,77	20,49	16,34	12,79	10,09	8,672	7,564	6,408	5,697
18	37,16	34,81	31,53	28,87	25,99	21,60	17,34	13,68	10,86	9,390	8,231	7,015	6,265
19	38,58	36,19	32,85	30,14	27,20	22,72	18,34	14,56	11,65	10,12	8,907	7,633	6,844
20	40,00	37,57	34,17	31,41	28,41	23,83	19,34	15,45	12,44	10,85	9,591	8,260	7,434
21	41,40	38,93	35,48	32,67	29,62	24,93	20,34	16,34	13,24	11,59	10,28	8,897	8,034
22	42,80	40,29	36,78	33,92	30,81	26,04	21,34	17,24	14,04	12,34	10,98	9,542	8,643
23	44,18	41,64	38,08	35,17	32,01	27,14	22,34	18,14	14,85	13,09	11,69	10,20	9,260
24	45,56	42,98	39,36	36,42	33,20	28,24	23,34	19,04	15,66	13,85	12,40	10,86	9,886
25	46,93	44,31	40,65	37,65	34,38	29,34	24,34	19,94	16,47	14,61	13,12	11,52	10,52
26	48,29	45,64	41,92	38,89	35,56	30,43	25,34	20,84	17,29	15,38	13,84	12,20	11,16
27	49,64	46,96	43,19	40,11	36,74	31,53	26,34	21,75	18,11	16,15	14,57	12,88	11,81
28	50,99	48,28	44,46	41,34	37,92	32,62	27,34	22,66	19,94	16,93	15,31	13,56	12,46
29	52,34	49,59	45,72	42,56	39,09	33,71	28,34	23,57	19,77	17,71	16,05	14,26	13,12
30	53,67	50,89	46,98	43,77	40,26	34,80	29,34	24,48	20,60	18,49	16,79	14,95	13,79
40	66,77	63,69	59,34	55,76	51,81	45,62	39,34	33,66	29,05	26,51	24,43	22,16	20,71
50	79,49	76,15	71,42	67,50	63,17	56,33	49,33	42,94	37,69	34,76	32,36	29,71	27,99
60	91,95	88,38	83,30	79,08	74,40	66,98	59,33	52,29	46,46	43,19	40,48	37,48	35,53
70	104,2	100,4	95,02	90,53	85,53	77,58	69,33	61,70	55,33	51,74	48,76	45,44	43,28
80	116,3	112,3	106,6	101,9	96,58	88,13	79,33	71,14	64,28	60,39	57,15	53,54	51,17
90	128,3	124,1	118,1	113,1	107,6	98,65	89,33	80,62	73,29	69,13	65,65	61,75	59,20
100	140,2	135,8	129,6	124,3	118,5	109,1	99,33	90,13	82,36	77,93	74,22	70,06	67,33
150	198,4	193,2	185,8	179,6	172,6	161,3	149,3	138,0	128,3	122,7	118,0	112,7	109,1
200	255,3	249,4	241,1	234,0	226,0	213,1	199,3	186,2	174,8	168,3	162,7	156,4	152,2
250	311,3	304,9	295,7	287,9	279,1	264,7	249,3	234,6	221,8	214,4	208,1	200,9	196,2
300	366,8	359,9	349,9	341,4	331,8	316,1	299,3	283,1	269,1	260,9	253,9	246,0	240,7
400	476,6	468,7	457,3	447,6	436,6	418,7	399,3	380,6	364,2	354,6	346,5	337,2	330,9
600	693,0	683,5	669,8	658,1	644,8	623,0	599,3	576,3	556,1	544,2	534,0	522,4	514,5
800	906,8	896,0	880,3	866,9	851,7	826,6	799,3	772,7	749,2	735,4	723,5	709,9	700,7
1000	1119,	1107,	1090,	1075,	1058,	1030,	999,3	969,5	943,1	927,6	914,3	898,9	888,6

zu Tab. 4:

Ablesebeispiel: $\chi^2_{1; 0,05} = -3,93 = 3,93 \cdot 10^{-3} = 0,00393$

Approximation nach Wilson und Hilferty für $0 < \gamma < 1$:

$$\chi^2_{n; \gamma} \approx n \cdot \left[1 - \frac{2}{9n} + u_\gamma \cdot \sqrt{\frac{2}{9n}} \right]^3.$$

**Tab.5:** Quantile $F_{n_1, n_2; \gamma}$ der F-Verteilung

n_2	γ	1	2	3	4	5	6	7	8	9	10	11
1	0,990	4052,	4999,	5403,	5625,	5764,	5859,	5928,	5981,	6022,	6056,	6083,
	0,975	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3	968,6	973,0
	0,950	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0
	0,900	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,20	60,47
2	0,990	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,41
	0,975	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41
	0,950	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40
	0,900	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,401
3	0,990	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,13
	0,975	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,37
	0,950	10,13	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,763
	0,900	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,222
4	0,990	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45
	0,975	12,22	10,65	9,979	9,605	9,364	9,197	9,074	8,980	8,905	8,844	8,793
	0,950	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936
	0,900	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,907
5	0,990	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,962
	0,975	10,01	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681	6,619	6,568
	0,950	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,704
	0,900	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,282
6	0,990	13,75	10,92	9,780	9,148	8,746	8,466	8,260	8,102	7,976	7,874	7,789
	0,975	8,813	7,260	6,599	6,227	5,988	5,820	5,695	5,600	5,523	5,461	5,409
	0,950	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,027
	0,900	3,776	3,463	3,289	3,181	3,108	3,055	3,015	2,983	2,958	2,937	2,919
7	0,990	12,25	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719	6,620	6,538
	0,975	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823	4,761	4,709
	0,950	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,603
	0,900	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725	2,703	2,684
8	0,990	11,26	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911	5,814	5,734
	0,975	7,571	6,059	5,416	5,053	4,817	4,652	4,529	4,433	4,357	4,295	4,243
	0,950	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,313
	0,900	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538	2,518
9	0,990	10,56	8,022	6,992	6,422	2,057	5,802	5,613	5,467	5,351	5,257	5,177
	0,975	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026	3,964	3,912
	0,950	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,102
	0,900	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416	2,396
10	0,990	10,04	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942	4,849	4,771
	0,975	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779	3,717	3,665
	0,950	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,943
	0,900	3,285	2,924	2,728	2,605	2,522	2,461	2,414	2,377	2,347	2,323	2,302

Tab. 5: Fortsetzung

$n_2 \backslash \begin{matrix} n_1 \\ \gamma \end{matrix}$	12	13	14	15	20	24	30	40	60	120	∞	
1	0,990 0,975 0,950 0,900	6106, 976,7 243,9 60,71	6126, 979,8 244,7 60,90	6143, 982,5 245,4 61,07	6157, 984,9 245,9 61,22	6209, 993,1 248,0 61,74	6235, 997,2 249,1 62,00	6261, 1001 250,1 62,26	6287, 1006 251,1 62,53	6313, 1010 252,2 62,79	6339, 1014 253,3 63,06	6366, 1018 254,3 63,33
2	0,990 0,975 0,950 0,900	99,42 39,41 19,41 9,408	99,42 39,42 19,42 9,415	99,43 39,43 19,43 9,420	99,45 39,45 19,45 9,425	99,46 39,46 19,45 9,441	99,47 39,47 19,46 9,450	99,47 39,47 19,47 9,458	99,48 39,48 19,48 9,466	99,49 39,49 19,49 9,475	99,50 39,50 19,50 9,483	
3	0,990 0,975 0,950 0,900	27,05 14,34 8,745 5,216	26,98 14,30 8,729 5,210	26,92 14,28 8,715 5,205	26,87 14,25 8,694 5,200	26,69 14,17 8,657 5,184	26,60 14,12 8,560 5,176	26,50 14,08 8,539 5,168	26,41 14,04 8,511 5,160	26,32 13,99 8,572 5,151	26,22 13,95 8,549 5,143	26,13 13,90 8,526 5,134
4	0,990 0,975 0,950 0,900	14,37 8,751 5,912 3,896	14,31 8,715 5,891 3,885	14,25 8,684 5,873 3,877	14,20 8,657 5,858 3,869	14,02 8,560 5,803 3,844	13,93 8,511 5,774 3,831	13,84 8,461 5,746 3,817	13,75 8,411 5,717 3,804	13,65 8,360 5,688 3,790	13,56 8,309 5,658 3,775	13,46 8,257 5,628 3,761
5	0,990 0,975 0,950 0,900	9,888 6,525 4,678 3,268	9,824 6,487 4,655 3,257	9,770 6,455 4,636 3,247	9,722 6,428 4,619 3,238	9,553 6,329 4,558 3,207	9,466 6,278 4,527 3,191	9,379 6,227 4,496 3,174	9,291 6,175 4,464 3,157	9,202 6,123 4,431 3,140	9,112 6,069 4,398 3,123	9,020 6,015 4,365 3,105
6	0,990 0,975 0,950 0,900	7,718 5,366 4,000 2,905	7,657 5,329 3,976 2,892	7,605 5,297 3,956 2,881	7,559 5,269 3,938 2,871	7,396 5,168 3,874 2,836	7,313 5,117 3,841 2,818	7,229 5,065 3,808 2,800	7,143 5,012 3,774 2,781	7,057 4,959 3,740 2,762	6,969 4,904 3,705 2,742	6,880 4,849 3,669 2,722
7	0,990 0,975 0,950 0,900	6,469 4,666 3,575 2,668	6,410 4,628 3,550 2,654	6,359 4,596 3,529 2,643	6,314 4,568 3,511 2,632	6,155 4,467 3,445 2,595	6,074 4,415 3,410 2,575	5,992 4,362 3,376 2,555	5,908 4,309 3,340 2,535	5,824 4,254 3,304 2,514	5,737 4,199 3,267 2,493	5,650 4,142 3,230 2,471
8	0,990 0,975 0,950 0,900	5,667 4,200 3,284 2,502	5,609 4,162 3,259 2,488	5,558 4,129 3,237 2,475	5,515 4,101 3,218 2,464	5,359 3,999 3,150 2,425	5,279 3,947 3,115 2,404	5,198 3,894 3,079 2,383	5,116 3,840 3,043 2,361	5,032 3,784 3,005 2,339	4,946 3,728 2,967 2,316	4,859 3,670 2,928 2,293
9	0,990 0,975 0,950 0,900	5,111 3,868 3,073 2,379	5,054 3,830 3,047 2,364	5,005 3,798 3,025 2,351	4,962 3,769 3,006 2,340	4,808 3,667 2,936 2,298	4,729 3,614 2,900 2,277	4,649 3,560 2,864 2,255	4,567 3,505 2,826 2,232	4,483 3,449 2,787 2,208	4,398 3,392 2,748 2,184	4,311 3,333 2,707 2,159
10	0,990 0,975 0,950 0,900	4,706 3,621 2,913 2,284	4,649 3,583 2,887 2,269	4,600 3,550 2,864 2,255	4,558 3,522 2,845 2,244	4,405 3,419 2,774 2,201	4,327 3,365 2,737 2,178	4,247 3,311 2,700 2,155	4,165 3,255 2,661 2,132	4,082 3,198 2,621 2,107	3,996 3,140 2,580 2,082	3,909 3,080 2,538 2,055

Tab. 5: Fortsetzung

n ₂	n ₁ γ	1	2	3	4	5	6	7	8	9	10	11
11	0,990	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,744	4,632	4,539	4,462
	0,975	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588	3,526	3,473
	0,950	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,818
	0,900	3,225	2,860	2,660	2,536	2,451	2,389	2,342	2,304	2,273	2,248	2,227
12	0,990	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,388	4,296	4,219
	0,975	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,436	3,374	3,321
	0,950	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,717
	0,900	3,177	2,807	2,605	2,480	2,394	2,331	2,283	2,245	2,214	2,188	2,166
13	0,990	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191	4,100	4,024
	0,975	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312	3,250	3,197
	0,950	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,634
	0,900	3,136	2,763	2,560	2,434	2,347	2,283	2,234	2,195	2,164	2,138	2,115
14	0,990	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030	3,939	3,863
	0,975	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209	3,147	3,094
	0,950	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,565
	0,900	3,102	2,726	2,522	2,395	2,307	2,243	2,193	2,154	2,122	2,095	2,073
15	0,990	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,004	3,895	3,805	3,730
	0,975	6,199	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123	3,060	3,007
	0,950	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,506
	0,900	3,073	2,695	2,490	2,361	2,273	2,208	2,158	2,119	2,086	2,059	2,036
16	0,990	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,780	3,691	3,616
	0,975	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049	2,986	2,933
	0,950	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,456
	0,900	3,048	2,668	2,462	2,333	2,244	2,178	2,128	2,088	2,055	2,028	2,005
17	0,990	8,400	6,112	5,185	4,669	4,336	4,101	3,927	3,791	3,682	3,593	3,518
	0,975	6,042	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985	2,922	2,869
	0,950	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,412
	0,900	3,026	2,645	2,437	2,308	2,218	2,152	2,102	2,061	2,028	2,001	1,977
18	0,990	8,285	6,013	5,092	4,579	4,248	4,015	3,841	3,705	3,597	3,508	3,433
	0,975	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929	2,866	2,813
	0,950	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,374
	0,900	3,007	2,624	2,416	2,286	2,196	2,130	2,079	2,038	2,005	1,977	1,953
19	0,990	8,185	5,926	5,010	4,500	4,171	3,939	3,765	3,631	3,523	3,434	3,359
	0,975	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880	2,817	2,764
	0,950	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,340
	0,900	2,990	2,606	2,397	2,266	2,176	2,109	2,058	2,017	1,984	1,956	1,932
20	0,990	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457	3,368	3,293
	0,975	5,871	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837	2,774	2,720
	0,950	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,310
	0,900	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965	1,937	1,913

Tab. 5: Fortsetzung

$n_2 \backslash \gamma$	n_1	12	13	14	15	20	24	30	40	60	120	∞
11	0,990	4,397	4,341	4,293	4,251	4,099	4,021	3,941	3,860	3,776	3,690	3,602
	0,975	3,430	3,391	3,358	3,330	3,226	3,173	3,118	3,061	3,004	2,944	2,883
	0,950	2,788	2,761	2,738	2,719	2,646	2,609	2,570	2,531	2,490	2,448	2,404
	0,900	2,209	2,193	2,179	2,167	2,123	2,100	2,076	2,052	2,026	2,000	1,972
12	0,990	4,155	4,099	4,051	4,010	3,858	3,780	3,701	3,619	3,535	3,449	3,361
	0,975	3,277	3,239	3,206	3,177	3,073	3,019	2,963	2,906	2,848	2,787	2,725
	0,950	2,687	2,660	2,637	2,617	2,544	2,505	2,466	2,426	2,384	2,341	2,296
	0,900	2,147	2,131	2,117	2,105	2,060	2,036	2,011	1,986	1,960	1,932	1,904
13	0,990	3,960	3,905	3,857	3,815	3,665	3,587	3,507	3,425	3,341	3,255	3,165
	0,975	3,153	3,115	3,081	3,053	2,948	2,893	2,837	2,780	2,720	2,659	2,595
	0,950	2,604	2,577	2,553	2,533	2,459	2,420	2,380	2,339	2,297	2,252	2,206
	0,900	2,097	2,080	2,066	2,053	2,007	1,983	1,958	1,931	1,904	1,876	1,846
14	0,990	3,800	3,745	3,697	3,656	3,505	3,427	3,348	3,266	3,181	3,094	3,004
	0,975	3,050	3,011	2,978	2,949	2,844	2,789	2,732	2,674	2,614	2,552	2,487
	0,950	2,534	2,507	2,483	2,463	2,388	2,349	2,308	2,266	2,223	2,178	2,131
	0,900	2,054	2,037	2,022	2,010	1,962	1,938	1,912	1,885	1,857	1,828	1,797
15	0,990	3,666	3,611	3,563	3,522	3,372	3,294	3,214	3,132	3,047	2,959	2,868
	0,975	2,963	2,924	2,891	2,862	2,756	2,701	2,644	2,585	2,524	2,461	2,395
	0,950	2,475	2,448	2,424	2,403	2,328	2,288	2,247	2,204	2,160	2,114	2,066
	0,900	2,017	2,000	1,985	1,972	1,924	1,899	1,873	1,845	1,817	1,787	1,755
16	0,990	3,553	3,497	3,450	3,409	3,259	3,181	3,101	3,018	2,933	2,845	2,753
	0,975	2,889	2,850	2,817	2,788	2,681	2,625	2,568	2,509	2,447	2,383	2,316
	0,950	2,425	2,397	2,373	2,352	2,276	2,235	2,194	2,151	2,106	2,059	2,010
	0,900	1,985	1,968	1,953	1,940	1,891	1,866	1,839	1,811	1,782	1,751	1,718
17	0,990	3,455	3,400	3,353	3,312	3,162	3,084	3,003	2,920	2,835	2,746	2,653
	0,975	2,825	2,786	2,752	2,723	2,616	2,560	2,502	2,442	2,380	2,315	2,247
	0,950	2,381	2,353	2,329	2,308	2,230	2,190	2,148	2,104	2,058	2,011	1,960
	0,900	1,958	1,940	1,925	1,912	1,862	1,836	1,809	1,781	1,751	1,719	1,686
18	0,990	3,371	3,316	3,268	3,227	3,077	2,999	2,919	2,835	2,749	2,660	2,566
	0,975	2,769	2,730	2,696	2,667	2,559	2,503	2,444	2,384	2,321	2,256	2,187
	0,950	2,342	2,314	2,290	2,269	2,191	2,150	2,107	2,063	2,017	1,968	1,917
	0,900	1,933	1,915	1,900	1,887	1,837	1,810	1,783	1,754	1,723	1,691	1,657
19	0,990	3,297	3,241	3,194	3,153	3,003	2,925	2,844	2,761	2,674	2,584	2,489
	0,975	2,720	2,680	2,646	2,617	2,509	2,452	2,394	2,333	2,270	2,203	2,133
	0,950	2,308	2,280	2,255	2,234	2,155	2,114	2,071	2,026	1,980	1,930	1,878
	0,900	1,912	1,894	1,878	1,865	1,814	1,787	1,759	1,730	1,699	1,666	1,631
20	0,990	3,231	3,176	3,129	3,088	2,938	2,859	2,778	2,695	2,608	2,517	2,421
	0,975	2,676	2,636	2,602	2,573	2,464	2,408	2,349	2,287	2,223	2,156	2,085
	0,950	2,278	2,249	2,225	2,203	2,124	2,082	2,039	1,994	1,946	1,896	1,843
	0,900	1,892	1,874	1,859	1,845	1,794	1,767	1,738	1,708	1,677	1,643	1,607

Tab. 5: Fortsetzung

$n_2 \backslash \begin{matrix} n_1 \\ \gamma \end{matrix}$	1	2	3	4	5	6	7	8	9	10	11	
22	0,990	7,945	5,719	4,817	4,313	3,988	3,758	3,587	3,453	3,346	3,258	3,183
	0,975	5,786	4,383	3,783	3,440	3,215	3,055	2,934	2,839	2,763	2,700	2,646
	0,950	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342	2,297	2,258
	0,900	2,949	2,561	2,351	2,219	2,128	2,060	2,008	1,967	1,933	1,904	1,880
24	0,990	7,823	5,614	4,718	4,218	3,895	3,667	3,496	3,363	3,256	3,168	3,094
	0,975	5,717	4,319	3,721	3,379	3,155	2,995	2,874	2,779	2,703	2,640	2,586
	0,950	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255	2,216
	0,900	2,927	2,538	2,327	2,195	2,103	2,035	1,983	1,941	1,906	1,877	1,853
26	0,990	7,721	5,526	4,637	4,140	3,818	3,591	3,421	3,288	3,182	3,094	3,020
	0,975	5,659	4,265	3,670	3,329	3,105	2,945	2,824	2,729	2,653	2,590	2,536
	0,950	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265	2,220	2,181
	0,900	2,909	2,519	2,307	2,174	2,082	2,014	1,961	1,919	1,884	1,855	1,830
28	0,990	7,636	5,453	4,568	4,074	3,754	3,528	3,358	3,226	3,120	3,032	2,958
	0,975	5,610	4,221	3,626	3,286	3,063	2,903	2,782	2,687	2,611	2,547	2,493
	0,950	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190	2,151
	0,900	2,894	2,503	2,291	2,157	2,064	1,996	1,943	1,900	1,865	1,836	1,811
30	0,990	7,562	5,390	4,510	4,018	3,699	3,473	3,304	3,173	3,067	2,979	2,905
	0,975	5,568	4,182	3,589	3,250	3,026	2,867	2,746	2,651	2,575	2,511	2,457
	0,950	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165	2,125
	0,900	2,881	2,489	2,276	2,142	2,049	1,980	1,927	1,884	1,849	1,819	1,794
40	0,990	7,314	5,179	4,313	3,828	3,514	3,291	3,124	2,993	2,888	2,801	2,727
	0,975	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,452	2,388	2,334
	0,950	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077	2,037
	0,900	2,835	2,440	2,226	2,091	1,997	1,927	1,873	1,829	1,793	1,763	1,737
60	0,990	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,718	2,632	2,558
	0,975	5,286	3,925	3,343	3,008	2,786	2,627	2,507	2,412	2,334	2,270	2,215
	0,950	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993	1,952
	0,900	2,791	2,393	2,177	2,041	1,946	1,875	1,819	1,775	1,738	1,707	1,680
80	0,990	6,964	4,882	4,036	3,564	3,256	3,037	2,872	2,743	2,639	2,552	2,478
	0,975	5,219	3,865	3,285	2,951	2,730	2,571	2,451	2,356	2,278	2,214	2,158
	0,950	3,961	3,111	2,719	2,486	2,329	2,214	2,127	2,057	1,999	1,952	1,910
	0,900	2,770	2,370	2,154	2,017	1,921	1,849	1,793	1,748	1,711	1,680	1,652
120	0,990	6,851	4,787	3,949	3,480	3,174	2,956	2,792	2,663	2,559	2,472	2,398
	0,975	5,152	3,805	3,227	2,894	2,674	2,515	2,395	2,299	2,222	2,157	2,101
	0,950	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959	1,910	1,869
	0,900	2,748	2,347	2,130	1,992	1,896	1,824	1,767	1,722	1,684	1,652	1,625
∞	0,990	6,635	4,605	3,782	3,319	3,017	2,802	2,639	2,511	2,407	2,321	2,247
	0,975	5,024	3,689	3,116	2,786	2,567	2,408	2,288	2,192	2,114	2,048	1,992
	0,950	3,841	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880	1,831	1,788
	0,900	2,706	2,303	2,084	1,945	1,847	1,774	1,717	1,670	1,632	1,599	1,570

Tab. 5: Fortsetzung

$n_2 \backslash \gamma$	n_1	12	13	14	15	20	24	30	40	60	120	∞
22	0,990	3,121	3,066	3,019	2,978	2,827	2,749	2,667	2,583	2,495	2,403	2,305
	0,975	2,602	2,562	2,528	2,498	2,389	2,332	2,272	2,210	2,145	2,076	2,003
	0,950	2,226	2,197	2,172	2,151	2,071	2,028	1,984	1,938	1,889	1,838	1,783
	0,900	1,859	1,841	1,825	1,811	1,759	1,731	1,702	1,671	1,639	1,604	1,567
24	0,990	3,032	2,977	2,930	2,889	2,738	2,659	2,577	2,492	2,403	2,310	2,211
	0,975	2,541	2,501	2,467	2,437	2,327	2,269	2,209	2,146	2,080	2,010	1,935
	0,950	2,183	2,154	2,129	2,108	2,027	1,984	1,939	1,892	1,842	1,790	1,733
	0,900	1,832	1,813	1,797	1,783	1,730	1,702	1,672	1,641	1,607	1,571	1,533
26	0,990	2,958	2,903	2,856	2,815	2,664	2,585	2,503	2,417	2,327	2,233	2,131
	0,975	2,491	2,451	2,417	2,387	2,276	2,217	2,157	2,093	2,026	1,954	1,878
	0,950	2,148	2,119	2,093	2,072	1,990	1,946	1,901	1,853	1,803	1,749	1,691
	0,900	1,809	1,790	1,774	1,760	1,706	1,677	1,647	1,615	1,581	1,544	1,504
28	0,990	2,896	2,841	2,794	2,753	2,602	2,522	2,440	2,353	2,263	2,167	2,064
	0,975	2,448	2,408	2,374	2,344	2,232	2,174	2,112	2,048	1,980	1,907	1,829
	0,950	2,118	2,088	2,063	2,041	1,959	1,915	1,869	1,820	1,769	1,714	1,654
	0,900	1,790	1,770	1,754	1,740	1,685	1,656	1,625	1,592	1,558	1,520	1,478
30	0,990	2,843	2,788	2,741	2,700	2,549	2,469	2,386	2,299	2,208	2,111	2,006
	0,975	2,412	2,372	2,337	2,307	2,195	2,136	2,074	2,009	1,940	1,866	1,787
	0,950	2,092	2,062	2,037	2,015	1,932	1,887	1,841	1,792	1,740	1,684	1,622
	0,900	1,773	1,753	1,737	1,722	1,667	1,638	1,606	1,573	1,538	1,499	1,456
40	0,990	2,665	2,610	2,563	2,522	2,369	2,288	2,203	2,114	2,019	1,917	1,805
	0,975	2,288	2,247	2,212	2,182	2,068	2,007	1,943	1,875	1,803	1,724	1,637
	0,950	2,003	1,973	1,947	1,924	1,839	1,793	1,744	1,693	1,637	1,577	1,509
	0,900	1,715	1,695	1,677	1,662	1,605	1,574	1,541	1,506	1,467	1,425	1,377
60	0,990	2,496	2,441	2,393	2,352	2,198	2,115	2,028	1,936	1,836	1,726	1,601
	0,975	2,169	2,128	2,092	2,061	1,944	1,882	1,815	1,744	1,667	1,581	1,482
	0,950	1,917	1,886	1,860	1,836	1,748	1,700	1,649	1,594	1,534	1,467	1,389
	0,900	1,657	1,637	1,619	1,603	1,543	1,511	1,476	1,437	1,395	1,348	1,291
80	0,990	2,416	2,361	2,313	2,272	2,116	2,033	1,944	1,849	1,746	1,630	1,491
	0,975	2,112	2,070	2,034	2,003	1,885	1,821	1,753	1,679	1,598	1,507	1,396
	0,950	1,876	1,844	1,817	1,793	1,703	1,654	1,602	1,545	1,482	1,410	1,322
	0,900	1,629	1,608	1,590	1,574	1,513	1,479	1,443	1,403	1,358	1,306	1,242
120	0,990	2,336	2,281	2,233	2,192	2,035	1,950	1,860	1,763	1,656	1,533	1,381
	0,975	2,055	2,013	1,976	1,945	1,825	1,760	1,690	1,614	1,530	1,433	1,310
	0,950	1,834	1,802	1,774	1,750	1,659	1,608	1,554	1,495	1,429	1,352	1,254
	0,900	1,601	1,580	1,561	1,545	1,482	1,447	1,409	1,368	1,320	1,265	1,193
∞	0,990	2,185	2,129	2,080	2,039	1,878	1,791	1,696	1,592	1,473	1,325	1,000
	0,975	1,945	1,902	1,865	1,833	1,708	1,640	1,566	1,484	1,388	1,268	1,000
	0,950	1,752	1,719	1,691	1,666	1,571	1,517	1,459	1,394	1,318	1,221	1,000
	0,900	1,546	1,523	1,504	1,487	1,421	1,383	1,342	1,295	1,240	1,169	1,000

zu Tab. 5:

Ablesebeispiel: $F_{7,20;0,99} = 3,699$

$$\text{Erweiterung der Tafel: } F_{n_1, n_2; 1-\gamma} = \frac{1}{F_{n_2, n_1; \gamma}},$$

$$F_{1, n_2; \gamma} = (t_{n_2; (1+\gamma)/2})^2, \quad F_{1, \infty; \gamma} = (u_{(1+\gamma)/2})^2$$

$$F_{n_1, \infty; \gamma} = \frac{1}{n_1} \cdot \chi_{n_1; \gamma}^2, \quad F_{\infty, \infty; \gamma} = 1.$$

Interpolation nach Laubscher: Gesucht ist $F_{n_1, n_2; \gamma}$. Gibt es dann natürliche Zahlen $n_3 \leq n_1 < n_5$ sowie $n_4 \leq n_2 < n_6$ derart, daß die Quantile $F_{n_3, n_4; \gamma}$, $F_{n_3, n_6; \gamma}$, $F_{n_5, n_4; \gamma}$ und $F_{n_5, n_6; \gamma}$ vertafelt sind, so gilt:

$$F_{n_1, n_2; \gamma} = (1 - c_1) \cdot (1 - c_2) \cdot F_{n_3, n_4; \gamma} + (1 - c_1) \cdot c_2 \cdot F_{n_3, n_6; \gamma} \\ + c_1 \cdot (1 - c_2) \cdot F_{n_5, n_4; \gamma} + c_1 \cdot c_2 \cdot F_{n_5, n_6; \gamma}$$

$$\text{für } c_1 = \frac{n_5(n_1 - n_3)}{n_1(n_5 - n_3)} \text{ und } c_2 = \frac{n_6(n_2 - n_4)}{n_2(n_6 - n_4)}.$$

Läßt sich $n_3 = n_1$ wählen, so wird offensichtlich $c_1 = 0$, wie für $n_4 = n_2$ auch $c_2 = 0$ ist. In diesen Fällen vereinfacht sich die Interpolationsformel entsprechend.

Approximation für $0,5 < \gamma < 1$: $F_{n_1, n_2; \gamma} \approx e^{u \cdot a - b}$ mit $u = u_\gamma$,

$$a = \sqrt{2 \cdot d + c \cdot d^2}, \quad b = 2 \cdot \left(\frac{1}{n_1 - 1} - \frac{1}{n_2 - 1} \right) \cdot \left(c + \frac{5}{6} - \frac{d}{3} \right),$$

$$c = \frac{(u_\gamma)^2 - 3}{6} \quad \text{und} \quad d = \frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}.$$

Tab. 6: Quantile $q_{n,v;\gamma}$ der Studentisierten Spannweite

$v \backslash \gamma$	n	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0,99	90,0	135,	164,	186,	202,	216,	227,	237,	246,	253,	260,	266,	272,	277,
	0,95	18,0	27,0	32,8	37,1	40,4	43,1	45,4	47,4	49,1	50,6	52,0	53,2	54,3	55,4
2	0,99	14,0	19,0	22,3	24,7	26,6	28,2	29,5	30,7	31,7	32,6	33,4	34,1	34,8	35,4
	0,95	6,08	8,33	9,80	10,9	11,7	12,4	13,0	13,5	14,0	14,4	14,8	15,1	15,4	15,7
3	0,99	8,26	10,6	12,2	13,3	14,2	15,0	15,6	16,2	16,7	17,1	17,5	17,9	18,2	18,5
	0,95	4,50	5,91	6,83	7,50	8,04	8,48	8,85	9,18	9,46	9,72	9,95	10,1	10,3	10,5
4	0,99	6,51	8,12	9,17	9,96	10,6	11,1	11,6	11,9	12,3	12,6	12,8	13,1	13,3	13,5
	0,95	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,03	8,21	8,37	8,52	8,66
5	0,99	5,70	6,98	7,80	8,42	8,91	9,32	9,67	9,97	10,2	10,5	10,7	10,9	11,1	11,2
	0,95	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	7,00	7,17	7,32	7,47	7,60	7,72
6	0,99	5,24	6,33	7,03	7,56	7,97	8,32	8,61	8,87	9,10	9,30	9,49	9,65	9,81	9,95
	0,95	3,46	4,34	4,90	5,30	5,63	5,90	6,12	6,32	6,49	6,65	6,79	6,92	7,03	7,14
7	0,99	4,95	5,92	6,54	7,01	7,37	7,68	7,94	8,17	8,37	8,55	8,71	8,86	9,00	9,12
	0,95	3,34	4,17	4,68	5,06	5,36	5,61	5,81	6,00	6,16	6,30	6,43	6,55	6,66	6,76
8	0,99	4,75	5,64	6,20	6,63	6,96	7,24	7,47	7,68	7,86	8,03	8,18	8,31	8,44	8,55
	0,95	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05	6,17	6,29	6,39	6,48
9	0,99	4,60	5,43	5,96	6,35	6,66	6,92	7,13	7,33	7,50	7,65	7,78	7,91	8,02	8,13
	0,95	3,20	3,95	4,41	4,76	5,02	5,24	5,43	5,60	5,74	5,87	5,98	6,09	6,19	6,28
10	0,99	4,48	5,27	5,77	6,14	6,43	6,67	6,87	7,06	7,21	7,36	7,48	7,60	7,71	7,81
	0,95	3,15	3,88	4,33	4,65	4,91	5,12	5,31	5,46	5,60	5,72	5,83	5,94	6,03	6,11
12	0,99	4,32	5,05	5,50	5,84	6,10	6,32	6,51	6,67	6,81	6,94	7,06	7,17	7,26	7,36
	0,95	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,39	5,51	5,62	5,71	5,80	5,88
15	0,99	4,17	4,84	5,25	5,56	5,80	5,99	6,16	6,31	6,44	6,56	6,66	6,76	6,84	6,93
	0,95	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,31	5,40	5,49	5,57	5,65
18	0,99	4,07	4,70	5,09	5,38	5,60	5,79	5,94	6,08	6,20	6,31	6,41	6,50	6,60	6,66
	0,95	2,97	3,61	4,00	4,28	4,50	4,67	4,82	4,96	5,07	5,17	5,27	5,35	5,43	5,50
20	0,99	4,02	4,64	5,02	5,29	5,51	5,69	5,84	5,97	6,09	6,19	6,29	6,37	6,45	6,52
	0,95	2,95	3,58	3,96	4,23	4,44	4,62	4,77	4,90	5,01	5,11	5,20	5,28	5,36	5,43
24	0,99	3,96	4,55	4,91	5,17	5,37	5,54	5,69	5,81	5,92	6,02	6,11	6,19	6,26	6,33
	0,95	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,91	5,01	5,10	5,18	5,25	5,32
30	0,99	3,89	4,46	4,80	5,05	5,24	5,40	5,54	5,65	5,76	5,85	5,93	6,01	6,08	6,14
	0,95	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	4,92	5,00	5,08	5,15	5,21
40	0,99	3,82	4,37	4,70	4,93	5,11	5,27	5,39	5,50	5,60	5,69	5,76	5,84	5,90	5,96
	0,95	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,74	4,82	4,90	4,98	5,04	5,11
60	0,99	3,76	4,28	4,60	4,82	4,99	5,13	5,25	5,36	5,45	5,53	5,60	5,67	5,73	5,78
	0,95	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81	4,88	4,94	5,00
120	0,99	3,70	4,20	4,50	4,71	4,87	5,01	5,12	5,21	5,30	5,38	5,44	5,50	5,56	5,61
	0,95	2,80	3,36	3,69	3,92	4,10	4,24	4,36	4,47	4,56	4,64	4,71	4,78	4,84	4,90
∞	0,99	3,64	4,12	4,40	4,60	4,76	4,88	4,99	5,08	5,16	5,23	5,29	5,35	5,40	5,45
	0,95	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62	4,69	4,74	4,80

Tab. 6: Fortsetzung

$\nu \backslash \gamma$	n	16	18	20	22	24	26	28	30	32	36	40	60	80	100
1	0,99	282,	290,	298,	305,	311,	316,	321,	326,	330,	338,	345,	370,	387,	400,
	0,95	56,3	58,0	59,6	60,9	62,1	63,2	64,2	65,1	66,0	67,6	68,9	74,0	77,4	80,0
2	0,99	36,0	37,0	37,8	38,8	39,5	40,2	40,8	41,3	41,8	42,8	43,6	46,7	48,8	50,4
	0,95	15,9	16,4	16,8	17,1	17,4	17,7	18,0	18,3	18,5	18,9	19,3	20,7	21,6	22,3
3	0,99	18,8	19,3	19,8	20,2	20,5	20,9	21,2	21,4	21,7	22,2	22,6	24,1	25,2	26,0
	0,95	10,7	11,0	11,2	11,5	11,7	11,9	12,1	12,2	12,4	12,6	12,9	13,8	14,4	14,8
4	0,99	13,7	14,1	14,4	14,7	14,9	15,2	15,4	15,6	15,8	16,1	16,4	17,5	18,2	18,8
	0,95	8,79	9,03	9,23	9,42	9,58	9,74	9,87	10,0	10,1	10,3	10,5	11,2	11,7	12,1
5	0,99	11,4	11,8	11,9	12,2	12,4	12,5	12,7	12,9	13,0	13,3	13,5	14,4	15,0	15,5
	0,95	7,83	8,03	8,21	8,37	8,51	8,64	8,76	8,88	8,98	9,17	9,33	9,95	10,4	10,7
6	0,99	10,1	10,3	10,5	10,7	10,9	11,1	11,2	11,3	11,5	11,7	11,9	12,7	13,2	13,5
	0,95	7,24	7,43	7,59	7,73	7,86	7,98	8,09	8,19	8,28	8,45	8,60	9,16	9,55	9,84
7	0,99	9,24	9,46	9,65	9,81	9,97	10,1	10,2	10,4	10,5	10,7	10,9	11,5	12,0	12,3
	0,95	6,85	7,02	7,17	7,30	7,42	7,53	7,63	7,73	7,81	7,97	8,11	8,63	8,99	9,26
8	0,99	8,66	8,85	9,03	9,18	9,32	9,45	9,57	9,68	9,78	9,96	10,1	10,8	11,2	11,5
	0,95	6,57	6,73	6,87	7,00	7,11	7,21	7,31	7,40	7,48	7,62	7,76	8,25	8,59	8,84
9	0,99	8,23	8,41	8,57	8,72	8,85	8,97	9,07	9,18	9,27	9,44	9,59	10,2	10,6	10,9
	0,95	6,36	6,51	6,64	6,76	6,87	6,97	7,06	7,15	7,22	7,36	7,49	7,96	8,28	8,53
10	0,99	7,91	8,08	8,23	8,36	8,48	8,60	8,70	8,79	8,88	9,04	9,19	9,73	10,1	10,4
	0,95	6,19	6,34	6,47	6,58	6,69	6,78	6,86	6,95	7,02	7,16	7,28	7,73	8,04	8,28
12	0,99	7,44	7,59	7,73	7,85	7,96	8,07	8,16	8,25	8,33	8,47	8,88	9,09	9,43	9,69
	0,95	5,95	6,09	6,21	6,32	6,41	6,50	6,58	6,66	6,73	6,86	6,97	7,39	7,69	7,91
15	0,99	7,00	7,14	7,26	7,37	7,47	7,57	7,65	7,73	7,80	7,93	8,05	8,49	8,80	9,04
	0,95	5,72	5,85	5,96	6,06	6,15	6,23	6,31	6,38	6,44	6,56	6,67	7,07	7,34	7,55
18	0,99	6,73	6,85	6,97	7,07	7,16	7,25	7,33	7,40	7,46	7,59	7,70	8,11	8,39	8,61
	0,95	5,57	5,69	5,79	5,89	5,98	6,06	6,13	6,19	6,26	6,37	6,47	6,85	7,11	7,31
20	0,99	6,59	6,71	6,82	6,92	7,01	7,09	7,17	7,24	7,30	7,42	7,52	7,92	8,07	8,40
	0,95	5,49	5,61	5,71	5,81	5,89	5,97	6,04	6,10	6,17	6,28	6,37	6,74	6,99	7,19
24	0,99	6,39	6,51	6,61	6,71	6,79	6,87	6,94	7,00	7,06	7,17	7,27	7,64	7,90	8,10
	0,95	5,38	5,49	5,59	5,68	5,76	5,84	5,91	5,97	6,03	6,13	6,23	6,58	6,82	7,01
30	0,99	6,20	6,31	6,41	6,49	6,57	6,64	6,71	6,77	6,83	6,93	7,02	7,37	7,61	7,80
	0,95	5,27	5,38	5,48	5,56	5,64	5,71	5,77	5,83	5,89	5,99	6,08	6,42	6,65	6,83
40	0,99	6,02	6,12	6,21	6,29	6,36	6,43	6,49	6,55	6,60	6,70	6,78	7,10	7,33	7,50
	0,95	5,16	5,27	5,36	5,44	5,51	5,58	5,62	5,70	5,75	5,85	5,93	6,26	6,48	6,64
60	0,99	5,84	5,93	6,01	6,09	6,16	6,22	6,28	6,33	6,38	6,47	6,55	6,84	7,05	7,21
	0,95	5,06	5,15	5,24	5,32	5,39	5,45	5,51	5,57	5,62	5,71	5,79	6,09	6,30	6,46
120	0,99	5,66	5,75	5,83	5,90	5,96	6,02	6,07	6,12	6,16	6,24	6,32	6,59	6,78	6,92
	0,95	4,95	5,04	5,13	5,20	5,27	5,33	5,38	5,43	5,48	5,57	5,64	5,93	6,13	6,28
∞	0,99	5,49	5,57	5,65	5,71	5,77	5,82	5,87	5,91	5,95	6,03	6,09	6,34	6,51	6,64
	0,95	4,84	4,93	5,01	5,08	5,14	5,20	5,25	5,30	5,35	5,43	5,50	5,76	5,95	6,09

Tab. 7: Gammafunktion $\Gamma(x)$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,0	1,0000	0,9943	0,9888	0,9836	0,9784	0,9735	0,9687	0,9642	0,9597	0,9555
1,1	0,9514	0,9474	0,9436	0,9399	0,9364	0,9330	0,9298	0,9267	0,9237	0,9209
1,2	0,9182	0,9156	0,9131	0,9108	0,9085	0,9064	0,9044	0,9025	0,9007	0,8990
1,3	0,8975	0,8960	0,8946	0,8934	0,8922	0,8912	0,8902	0,8893	0,8885	0,8879
1,4	0,8873	0,8868	0,8864	0,8860	0,8858	0,8857	0,8856	0,8856	0,8857	0,8859
1,5	0,8862	0,8866	0,8870	0,8876	0,8882	0,8889	0,8896	0,8905	0,8914	0,8924
1,6	0,8935	0,8947	0,8959	0,8972	0,8986	0,9001	0,9017	0,9033	0,9050	0,9068
1,7	0,9086	0,9106	0,9126	0,9147	0,9168	0,9191	0,9214	0,9238	0,9262	0,9288
1,8	0,9314	0,9341	0,9368	0,9397	0,9426	0,9456	0,9487	0,9518	0,9551	0,9584
1,9	0,9618	0,9652	0,9688	0,9724	0,9761	0,9799	0,9837	0,9877	0,9917	0,9958

Ablesebeispiel: $\Gamma(1,33) = 0,8934$

Erweiterung der Tafel: $\Gamma(x) = (x - 1) \cdot \Gamma(x - 1)$ für $x > 1$;

$$\Gamma(1-x) = \frac{\pi}{\Gamma(x) \cdot \sin \pi x} \quad \text{für } 0 < x < 1; \quad \Gamma(n+1) = n! \quad \text{für } n = 1, 2, 3, \dots$$

Gaußsche Multiplikationsformel für $x > 0$ und $n = 1, 2, 3, \dots$

$$\Gamma(nx) = (2\pi)^{(1-n)/2} \cdot n^{nx-1/2} \cdot \Gamma(x) \cdot \Gamma\left(x + \frac{1}{n}\right) \cdot \dots \cdot \Gamma\left(x + \frac{n-1}{n}\right),$$

$$\text{speziell: } \Gamma(2x) = \frac{2^{2x-1}}{\sqrt{\pi}} \cdot \Gamma(x) \cdot \Gamma\left(x + \frac{1}{2}\right) \quad \text{„Verdoppelungsformel“}$$

$$\text{und } \Gamma(3x) = \frac{3^{3x-0.5}}{2\pi} \cdot \Gamma(x) \cdot \Gamma\left(x + \frac{1}{3}\right) \cdot \Gamma\left(x + \frac{2}{3}\right) \quad \text{„Verdreifachungsformel“}$$

(Man beachte, daß für $n \leq y < n+1$ und $x = y/n$ sich $\Gamma(y)$ durch die Gaußsche Multiplikationsformel auf ein Produkt von Gammafunktionen zurückführen lässt, deren Argumente der Form $x + i/n$ mit $i = 0, 1, \dots, n-1$ alle aus dem Intervall $[1; 2]$ stammen.)

Approximation nach Hastings für $0 < x < 1$:

$$\begin{aligned} \Gamma(x+1) &\simeq 1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + a_6 x^6 + a_7 x^7 + a_8 x^8 \quad \text{mit} \\ a_1 &= -0,577191652, \quad a_2 = 0,988205891, \quad a_3 = -0,897056937, \quad a_4 = 0,918206857, \\ a_5 &= -0,756704078, \quad a_6 = 0,482199394, \quad a_7 = -0,193527818, \quad a_8 = 0,035868343. \end{aligned}$$

Stirlingsche Formel für $x > 1$:

$$\Gamma(x) \simeq e^{-x} \cdot x^{x-0.5} \cdot \sqrt{2\pi} \cdot \left(1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51840x^3} - \frac{571}{2488320x^4}\right).$$

Tab. 8: Gleichverteilte Zufallszahlen

16306	21417	11021	78499	17466	49767	05661	40786	57832	85454	27504	59472	40029
74442	41284	63927	25310	85664	61316	43388	26151	21941	54740	23158	20997	90044
61695	06715	55820	61639	71878	61647	95685	84224	51113	55591	26313	42061	17906
42173	05639	67119	70012	96940	65378	00010	60999	03300	62587	00924	26804	88462
81229	10571	62302	09370	83731	61114	47866	88642	12551	37111	25930	00508	41738
86448	27548	06746	78496	81279	48964	42620	14746	28963	78167	52819	28686	33002
71457	77393	36360	03979	80008	87689	93212	20493	26428	72294	45656	40098	31325
72966	32633	60242	89866	70574	37177	29202	98341	21038	87976	07023	30728	41263
10935	76454	65249	45078	30737	03889	60941	29174	22410	50300	84209	09039	12883
32176	85399	04899	32554	32464	25351	77653	16848	61501	51967	14813	57784	67955
26773	66437	05533	99762	01477	31072	19987	15062	44540	91686	72312	43138	17731
01612	96519	95353	02691	20479	91023	26089	77497	00153	70205	43681	43227	20674
65977	71357	00516	76193	79339	58177	76124	10911	81950	39965	99495	05014	67683
50255	29322	10612	54191	81218	30109	37677	34467	94609	98732	89979	78413	82700
41004	51025	84392	79217	98234	19996	80801	20279	23794	97684	77626	55335	09676
22429	68008	08102	63060	56059	79805	87501	32874	19824	77155	48598	86356	80336
12037	03536	35615	85508	25833	82478	14199	50485	00895	44541	03653	93590	62101
38734	10080	22254	28424	14898	90710	68960	14487	84639	30456	77175	82134	32270
68480	37291	41987	75747	74753	72559	05008	77603	76621	75033	84067	05893	35366
91500	45496	45122	72765	59422	10660	57193	47658	92872	07514	92335	82508	18473
72169	50047	39963	61332	00452	90031	49419	80981	49467	93133	83582	57315	89164
72279	85032	26525	03893	33411	42784	67040	43975	86429	18851	35463	27257	06124
24808	52507	90679	43116	48351	38936	91836	84794	29863	01497	60538	66252	88973
68864	00377	31225	01123	77571	40283	43031	24127	07951	36095	45503	74735	70073
22114	73931	55194	39681	12674	15995	29812	44128	53699	22708	51520	98318	24268
69348	38436	90640	88176	66401	00424	18909	95497	20492	15659	76298	34149	45330
57190	87281	31453	22638	69500	93568	05695	13095	95073	90044	67693	17276	50950
10702	72356	86580	55296	64587	20247	84179	03964	21730	09145	95948	97052	49083
35610	91403	15470	96463	46542	34643	47500	32483	40748	45283	65361	30494	20498
09392	51770	04522	93095	40434	76504	95200	22431	90202	22961	05714	30553	09756
64728	87734	49227	60361	92139	77356	21495	74427	88837	81035	53244	66440	58339
11526	14761	93907	00599	73820	92305	68289	38567	01082	89660	15939	94755	40589
79704	89041	17796	89665	05969	26546	59010	81623	34204	65110	98966	21697	53936
00031	24042	72983	30035	96189	52989	92378	04275	48980	13495	16514	46897	95256
68745	97695	58171	74260	93984	89839	19958	38567	03081	84411	97276	15019	25091
96364	89856	15662	30815	00073	33712	98686	08366	15277	52697	79359	94731	43009
60070	90971	42331	51545	10055	95761	55585	12328	99408	45866	62473	85918	18974
94249	42099	57920	55254	55644	01251	24927	41966	17487	00966	30455	61605	00859
33090	50883	97011	71419	46845	31388	32927	00347	25387	71881	07561	75273	12368
62951	14153	61224	88503	71795	52532	11555	98130	78160	30685	28888	25773	60279
16313	76053	22701	30811	38397	37019	83496	22773	38759	27838	72614	25482	73628
57577	00944	70492	65195	39261	52375	67532	03186	50804	60355	90438	96415	45951
34457	13100	74351	15222	36937	95391	38680	96829	09301	19109	64868	37636	48865
73913	53722	04084	35318	93261	32953	43131	07285	40347	08740	93193	88888	44039
67290	45074	66311	84817	24603	52913	14359	30710	37032	95114	73025	23230	34848

Tab. 8: Fortsetzung

85284	74047	13318	41800	33178	17478	47273	17680	55626	98239	02693	61685	37477
82690	74413	63976	42538	35188	10443	07245	74504	94651	97901	18277	87904	01119
98957	69510	49031	72317	36766	30597	46887	30221	31408	69974	54113	77616	85822
06416	37932	24658	21748	10604	21084	65441	65661	56684	92765	04557	84710	23833
69562	33657	72657	47820	06394	58017	85811	22565	57445	73096	28967	51710	87106
91104	91346	50171	30911	17560	32528	95711	06472	77856	18393	30492	74421	93818
07057	12646	49343	10275	84811	17835	60717	71610	52055	90811	63851	48335	59240
46980	00822	13643	92802	23967	11410	69977	11102	83173	94040	25592	30360	61344
16715	21114	60598	72129	77956	06104	91572	57644	17741	65623	28714	90033	78437
96792	79064	23033	15746	35354	87213	93536	59632	32238	21240	72150	25268	80122
08643	69761	73969	89994	2287	79099	24128	17714	16332	86298	12158	40579	12175
17457	03426	83884	34387	41206	53192	97483	96651	20266	17545	82355	37699	13013
14343	67163	07248	25039	34149	34819	01940	10716	01166	99045	51278	24757	84181
30937	53562	23779	57496	35525	61832	17868	11869	78338	19800	72138	15450	61965
47787	58293	24719	52620	78667	56194	45425	56699	43798	05417	43520	47350	08961
06819	06101	41028	60301	74097	52343	24064	40259	35354	94439	29222	33515	84658
48537	62266	02920	83608	95679	78519	75440	12893	92736	09429	78085	73858	33346
39128	23856	40064	58852	30996	47674	58852	19242	04224	92325	02903	84082	60376
44391	71804	03455	59952	20191	43608	18900	48950	01718	76398	24349	35427	22577
59849	77064	16065	03922	16788	58574	52730	40391	54762	85613	99038	36049	67595
71058	76914	90319	86954	34867	03155	33925	83909	65013	65997	14828	20499	28426
51671	34161	51290	31928	20969	32170	74941	33173	35808	24147	35765	01579	35788
86373	75292	29271	64758	92053	33831	92018	47379	97251	02032	60166	15683	76832
22453	59581	85638	10397	41113	82584	83710	18219	09065	48886	23428	47185	33951
16531	31939	06728	81570	42829	26581	54770	27435	95296	41437	33656	58340	12590
03901	56091	23920	15450	61239	41080	38214	67630	57155	99865	31230	86842	58349
71860	50503	05690	31083	10592	12660	82607	67741	19635	04984	62793	63393	38552
36972	89156	42356	78827	41903	63244	34760	09807	28396	56185	03106	94638	81976
69627	18830	76201	16633	46670	88869	13053	76387	32810	32217	65359	86850	86869
06493	30836	56788	41940	84502	30685	29685	16319	80796	45534	85302	74447	28238
01368	91244	37842	04564	05474	04519	46041	08945	32375	23102	67537	88550	59766
87497	60714	15110	54455	32301	84787	11478	08299	87765	60738	28192	24141	31355
82256	70039	51746	89319	76669	72617	73742	88244	15888	28895	41557	56624	77341
65780	61354	76418	54910	70675	29969	84615	19871	68044	20850	24795	31829	45976
19995	48694	99657	32743	11886	67763	92433	23366	15431	54733	00378	56177	69899
95177	45793	39381	70857	89073	43995	27617	56108	02363	14259	48524	41630	68521
40750	79301	07511	40113	83464	73269	33890	90358	39164	36436	86890	64690	52321
54058	54560	82327	76044	67200	32807	87825	83025	99302	61341	52646	26473	32737
09605	26718	51052	35379	12717	39998	53819	37161	57408	54037	95160	56638	17955
63537	59730	77554	42122	47919	67387	75396	87099	72882	34134	93967	05435	41970
83053	71939	80978	98126	05243	19421	14814	75265	70759	44521	61201	99827	89548
29905	13509	42347	22387	59365	40157	19012	41822	94376	70551	43229	43101	00662
31018	27179	89231	05048	48788	78405	45815	18783	80825	20613	44729	68466	07566
65909	25184	65144	76402	84931	43137	04287	46000	19465	47858	51725	37871	93372
43611	72340	15275	18950	88417	29021	50683	22519	75947	46190	15679	17748	91075

Bezüglich des Umganges mit Zufallszahlen sei auf Abschnitt 8 in Kapitel III verwiesen.

3. Griechisches Alphabet

<i>A</i>	α	Alpha	<i>N</i>	ν	Ny
<i>B</i>	β	Beta	<i>\Xi</i>	ξ	Xi
<i>\Gamma</i>	γ	Gamma	<i>O</i>	o	Omikron
<i>\Delta</i>	δ	Delta	<i>\Pi</i>	π	Pi
<i>E</i>	ε	Epsilon	<i>P</i>	ϱ	Rho
<i>Z</i>	ζ	Zeta	<i>\Sigma</i>	σ	Sigma
<i>H</i>	η	Eta	<i>T</i>	τ	Tau
<i>\Theta</i>	θ	Theta	<i>Y</i>	υ	Ypsilon
<i>I</i>	ι	Jota	<i>\Phi</i>	ϕ, φ	Phi
<i>K</i>	κ	Kappa	<i>X</i>	χ	Chi
<i>\Lambda</i>	λ	Lambda	<i>\Psi</i>	ψ	Psi
<i>M</i>	μ	My	<i>\Omega</i>	ω	Omega

4. Symbolverzeichnis

Symbol	Bedeutung
$\Phi(x)$	Verteilungsfunktion der Standardnormalverteilung $N(0, 1)$.
$\varphi(x)$	Dichte der Standardnormalverteilung $N(0, 1)$, Ausnahmen in Kap. XII und XIII.
\sim	„verteilt nach“, z. B. bedeutet $X \sim N(0, 1)$, daß X einer Standardnormalverteilung $N(0, 1)$ genügt.
$N(\mu, \sigma^2)$	Normalverteilung mit Mittelwert μ und Varianz σ^2 .
$N(0, 1)$	Standardnormalverteilung.
u_γ	γ -Quantil der Standardnormalverteilung $N(0, 1)$.
$B(n, p)$	Binomialverteilung mit den Parametern n und p.
t_n	(Studentsche) t-Verteilung mit n Freiheitsgraden.
$t_{n, \gamma}$	γ -Quantil der t-Verteilung mit n Freiheitsgraden.
χ^2_n	χ^2 -Verteilung mit n Freiheitsgraden.
$\chi^2_{n, \gamma}$	γ -Quantil der χ^2 -Verteilung mit n Freiheitsgraden.
F_{n_1, n_2}	(Fishersche) F-Verteilung mit den Freiheitsgraden n_1 und n_2 .
$F_{n_1, n_2; \gamma}$	γ -Quantil der F-Verteilung mit n_1 und n_2 Freiheitsgraden.
$U(a, b)$	(Stetige) Gleichverteilung (Rechteckverteilung) über dem Intervall $a \leq x \leq b$.
$Ex(\lambda)$	Exponentialverteilung mit Parameter λ .
$Po(\lambda)$	Poisson-Verteilung mit Parameter λ .
$M(n, p_1, \dots, p_k)$	Multinomialverteilung mit den Parametern n, p_1, \dots, p_k .
$H(N, M, n)$	Hypergeometrische Verteilung mit den Parametern N, M, n.
$q_{n, v; \gamma}$	γ -Quantil der Studentisierten Spannweite mit n und v Freiheitsgraden.
$P(A)$	Wahrscheinlichkeit des Ereignisses A.
EX	Erwartungswert der Zufallsvariablen X.
$Var X$	Varianz der Zufallsvariablen X.
$Cov(X, Y)$	Kovarianz der Zufallsvariablen X und Y.
$\rho_{X, Y}$	(Theoretische) Korrelation der Zufallsvariablen X und Y (auch $\rho(X, Y)$).
$Corr(X, Y)$	$Corr(X, Y) = \rho(X, Y)$.
\bar{x}	arithmetisches Mittel.
\tilde{x}_α	empirisches α -Quantil ($\tilde{x}_{0,5}$ – manchmal auch mit m bezeichnet – ist der Median).
$s^2 = s_x^2$	empirische Varianz (von X).
s_{XY}	empirische Kovarianz (von X und Y).
r_{XY}	empirische Korrelation (von X und Y).
ML	Maximum Likelihood.
OC	Operationscharakteristik, die als Funktion mit L bezeichnet wird.
MSE	Mittlerer quadratischer Fehler (mean squared error).
$S_n(x)$	Empirische Verteilungsfunktion einer Meßreihe x_1, \dots, x_n .
$S_{n_i, i}(x)$	Empirische Verteilungsfunktion der i-ten Meßreihe $x_{i_1}, \dots, x_{i_{n_i}}$.
$\cos x$	Cosinusfunktion.
$\sin x$	Sinusfunktion.
$\tan x$	Tangensfunktion.
$\arcsin x$	Arcus-Sinus-Funktion; Umkehrfunktion der Sinusfunktion: $y = \arcsin x$, wenn $x = \sin y$.

Symbol	Bedeutung
$\arctan x$	Arcus-Tangens-Funktion; Umkehrfunktion der Tangensfunktion: $y = \arctan x$, wenn $x = \tan y$.
$\operatorname{arctanh} x$	Arcus-Tangens-Hyperbolicus-Funktion (vgl. Tab. 2 in Kap. IX); $\operatorname{arctanh} x = \frac{1}{2} \ln \frac{1+x}{1-x}.$
$\Gamma(x)$	Gammafunktion.
$\ln x$	Natürlicher Logarithmus (d. h. zur Basis $e = 2,718281828459\dots$) von x : $x = e^y$, wenn $y = \ln x$.
$e^x, \exp\{x\}$	Exponentialfunktion.
$\log x$	Dekadischer Logarithmus (d. h. zur Basis 10) von x : $x = 10^y$, wenn $y = \log x$.
$n!$	„n Fakultät“: $n! = \Gamma(n+1) = 1 \cdot 2 \cdot \dots \cdot (n-1) \cdot n$ und $0! = 1$.
$\binom{n}{k}$	Binomialkoeffizient („n über k“): $\binom{n}{k} = \frac{n!}{k!(n-k)!}.$
\lim	Limes, Grenzwert.
$a_n \xrightarrow{n \rightarrow \infty} a$	„ a_n konvergiert gegen a , wenn n nach unendlich läuft“
\min	Minimum.
\max	Maximum.
$ x $	Absolutbetrag der Zahl x ; z. B. $ 3 = 3$, $ -3 = 3$.
$ A $	Mächtigkeit der Menge A (Anzahl der Elemente von A); z. B. $ \{1, 7, 6\} = 3$.
x^n	n-te Potenz von x („ x hoch n “): $x^n = x \cdot x \cdot \dots \cdot x$ (n-mal); z. B. $x^3 = x \cdot x \cdot x$.
$\sqrt{x} = x^{1/2}$	Quadratwurzel aus x ; z. B. $\sqrt{9} = 3$.
$\sqrt[n]{x} = x^{1/n}$	n-te Wurzel aus x : $y = \sqrt[n]{x}$, wenn $x = y^n$; z. B. $\sqrt[3]{64} = 4$.
$x^{n/m}$	„ x hoch n/m “, m-te Wurzel aus x^n , n-te Potenz der m-ten Wurzel aus x : $x^{n/m} = (x^n)^{1/m} = (\sqrt[n]{x})^m$.
x^y	„ x hoch y “: $x^y = e^{y \cdot \ln x} = 10^{y \cdot \log x}$ für $x > 0$; $x^{-y} = 1/x^y$, z. B.: $x^{-1} = 1/x$.
\sum_i	„Summe über i“: $\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n.$
$\sum_{i,j}$	„Doppelsumme über i und j“: $\sum_{i=1}^n \sum_{j=1}^m x_{ij} = x_{11} + x_{12} + \dots + x_{1m} + x_{21} + x_{22} + \dots + x_{2m} + \dots$ $\dots + x_{n1} + x_{n2} + \dots + x_{nm}.$
\prod_i	„Produkt über i“: $\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n.$
$\int_a^b g(x) dx$	Integralzeichen: $\int_a^b g(x) dx =$ Summe der Flächeninhalte der Flächen zwischen der x-Achse und der Funktion $g(x)$ im Intervall $a \leq x \leq b$, wobei Flächen oberhalb der x-Achse positiv und Flächen unterhalb der x-Achse negativ ge- wertet werden.
$=$	„gleich“.
\approx	„ungefähr gleich“, „rund“.
$\hat{=}$	„entspricht“.
$>$	„größer als“.
$<$	„kleiner als“.
\geq	„größer oder gleich“.

Symbol	Bedeutung
\leq	„kleiner oder gleich“.
\in	„Element von“: $3 \in \{1, 2, 3, 7\}$.
\subset	Teilmenge von: $\{1, 3\} \subset \{1, 2, 3, 7\}$
\emptyset	„unmögliches Ereignis“, „leere Menge“: $\{1, 3\} \cap \{2, 4\} = \emptyset$
\cup	„vereinigt mit“: $\{3, 4\} \cup \{1, 2, 4\} = \{1, 2, 3, 4\}$.
\cap	„geschnitten mit“: $\{3, 4\} \cap \{1, 2, 4\} = \{4\}$.
\bigcup_j	Vereinigung über j : $\bigcup_{j=1}^n A_j = A_1 \cup \dots \cup A_n$.
\bigcap_j	Durchschnitt über j : $\bigcap_{j=1}^n A_j = A_1 \cap \dots \cap A_n$.
$z = a + ib$	komplexe Zahl z mit Realteil a und Imaginärteil b (a, b reelle Zahlen); i heißt hier imaginäre Einheit ($i = \sqrt{-1}$), vgl. Abschnitt 3.1 in Kap. XII.
\bar{z}	$\bar{z} = a - ib$: zu z komplexe konjugierte Zahl.
$ z $	$ z = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$: Betrag der komplexen Zahl z .
$z = z e^{i\varphi}$	Eulersche Relation mit $\varphi = \arctan(b/a)$.
$a = (a_1, \dots, a_n)$	reeller n -dimensionaler Zeilenvektor.
a^T	$a^T = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$; zu a transponierter Spaltenvektor, $(a^T)^T = a$.
X	$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}$; reelle $m \times n$ -Matrix.
X^T	zur Matrix X transponierte $n \times m$ -Matrix:
	$X^T = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ \vdots & \vdots & & \vdots \\ 1_{1n} & x_{2n} & \dots & x_{mn} \end{pmatrix}$.
I_n	$n \times n$ -Einheitsmatrix $\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$
X^{-1}	Inverse einer $n \times n$ -Matrix X : $XX^{-1} = I_n$.
X	die Matrix X ist symmetrische $n \times n$ -Matrix, falls $X = X^T$
X	die Matrix X ist eine positiv (negativ) definite $n \times n$ -Matrix, wenn für alle n -dimensionalen Spaltenvektoren $y \neq 0$ gilt: $y^T X y > (<) 0$.
$Cov(a)$	$n \times n$ -Kovarianzmatrix zum n -dimensionalen Zufallsvektor a :
	$Cov(a) = \begin{pmatrix} \text{Var}(a_1) & \text{Cov}(a_1, a_2) & \dots & \text{Cov}(a_1, a_n) \\ \text{Cov}(a_2, a_1) & \text{Var}(a_2) & \dots & \text{Cov}(a_2, a_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(a_n, a_1) & \text{Cov}(a_n, a_2) & \dots & \text{Var}(a_n) \end{pmatrix} = (Cov(a))^T$

Symbol	Bedeutung
$\text{Corr}(a)$	$n \times n$ -Korrelationsmatrix zum n -dimensionalen Zufallsvektor a :

$$\text{Corr}(a) = \begin{pmatrix} 1 & \varrho(a_1, a_2) & \dots & \varrho(a_1, a_n) \\ \varrho(a_2, a_1) & 1 & \dots & \varrho(a_2, a_n) \\ \vdots & \vdots & \ddots & \vdots \\ \varrho(a_n, a_1) & \varrho(a_n, a_2) & \dots & 1 \end{pmatrix} = (\text{Corr}(a))^T$$

Für eine elementare Darstellung mathematischer Grundlagen sei etwa auf die Bücher von B. L. van der Waerden (1975), Beckmann/Künzi (1973), Kall (1982, 1984), Ohse (1983, 1984) und Bosch (1984) verwiesen.

5. Literaturverzeichnis

- [1] Ahrens, J.H. / Dieter, U. / Grube, A. (1970): Pseudo-random numbers - A new proposal for the choice of Multiplicators. Computing, vol. 6, p. 121-138.
- [2] Aitken, A.C. (1935): On least squares and linear combinations of observations. Proceedings of the Royal Society of Edinburgh, vol. A 55, p. 42-48.
- [3] Akaike, H. (1974): Stochastic theory of minimal realizations. IEEE Transactions Automatic Control, vol. AC-19, p. 667-674.
- [4] Akaike, H. (1977): On entropy maximization principle. In: Krishnaiah, P.R. (ed.): Applications of statistics, North-Holland, Amsterdam, p. 27-41.
- [5] Allescher, H.-D. / Adler, G. / Hartung, J. / Manns, J.F. / Riemann, J.F. / Wienbeck, M. / Classen, M. / PRESTO-Studiengruppe (1999): Prospektive Epidemiologische Studie der Oberbauchbeschwerden (PRESTO) - Grundlagen und erste Ergebnisse. Deutsche Medizinische Wochenschrift, vol. 124, p. 443-450.
- [6] Allescher, H.-D. / Böckenhoff, A. / Knapp, G. / Wienbeck, M. / Hartung, J. (2001): Treatment of non-ulcer dyspepsia: A meta-analysis on placebo-controlled prospective studies. Scandinavian Journal of Gastroenterology, vol. 36, p. 934-941.
- [7] Allescher, H.-D. / Knapp, G. / Hartung, J. (2002): Treatment of non-ulcer dyspepsia: Prokinetics are effective. Scandinavian Journal of Gastroenterology, vol. 37, p. 618-620.
- [8] Anderson, O. (1914): Nochmals über: The elimination of spurious correlation due to position in time or space. Biometrika, vol. 10, p. 269-279.
- [9] Anderson, O. (1965): Probleme der statistischen Methodenlehre in den Sozialwissenschaften. Physica, Würzburg.
- [10] Anderson, R.L. / Bancroft, T.A. (1952): Statistical theory in research. McGraw-Hill, New York.
- [11] Anderson, T.W. (1971): The statistical analysis of time series. Wiley, New York.
- [12] Andrews, D.F. / Bickel, P.J. / Hampel, F.R. / Huber, P.J. / Rogers, W.H. / Tukey, J.W. (1972): Robust estimation of location. Princeton University Press, Princeton.
- [13] Anscombe, F.J. (1948): The transformation of Poisson, Binomial and Negative-Binomial data. Biometrika, vol. 35, p. 246-254.
- [14] Baird, D.C. (1962): Experimentation: An introduction to measurement theory and experiment design. Prentice-Hall, Englewood Cliffs (New Jersey).
- [15] Bamberg, G. / Baur, F. (1980): Statistik. Oldenbourg, München.
- [16] Bamberg, G. / Baur, F. / Krapp, M. (2008): Statistik. Oldenbourg, München.
- [17] Bandelow, C. (1976): Einführung in die Wahrscheinlichkeitstheorie. Brockmeyer, Bochum.
- [18] Bargmann, R.E. (1979): Scaling of multidimensional contingency tables by union intersection. Paper presented at the conference on multivariate statistical inference, Dalhousie University, Halifax.
- [19] Barlow, R.E. (1968): Likelihood ratio tests for restricted families of probability distributions. Annals of Mathematical Statistics, vol. 39, p. 547-560.
- [20] Barlow, R.E. / Bartholomew, D.J. / Bremner, J.M. / Brunk, H.D. (1972): Statistical inference under order restrictions. Wiley, Chichester.
- [21] Barlow, R.E. / Hunter, L.C. (1960): System efficiency and reliability. Technometrics, vol. 2, p. 43-53.

- [22] *Barlow, R. E. / Proschan, F.* (1969): A note on tests for monotone failure rate based on incomplete data. *Annals of Mathematical Statistics*, vol. 40, p. 595-600.
- [23] *Barlow, R.E. / Proschan, F.* (1978): *Statistische Theorie der Zuverlässigkeit*. Harri Deutsch, Frankfurt/M.
- [24] *Barlow, R.E. / Wu, A.S.* (1978): Coherent systems with multi-state components. *Mathematics of Operations Research*, vol. 3, p. 275-281.
- [25] *Barnett, V. / Lewis, T.* (1978): *Outliers in statistical data*. Wiley, London.
- [26] *Beckmann, M.J. / Künzi, H.P.* (1973): *Mathematik für Ökonomen I und II*. Springer, Berlin.
- [27] *Beichelt, F.* (1976): *Prophylaktische Erneuerung von Systemen*. Akademie Verlag, Berlin.
- [28] *Berkson, J.* (1950): Are there two regressions? *Journal of the American Statistical Association*, vol. 45, p. 164-180.
- [29] *Bickel, P.J. / Doksum, K.* (1969): Tests for monotone failure rate based on normalized spacings. *Annals of Mathematical Statistics*, vol. 40, p. 1216-1235.
- [30] *Bishop, Y.M.M. / Fienberg, S.E. / Holland, P.W.* (1975): *Discrete multivariate analysis: Theory and practice*. The MIT Press, Cambridge (MA).
- [31] *Blackman, R.B. / Tukey, J.W.* (1959): *The measurement of power spectra*. Dover, New York.
- [32] *Bleymüller, J. / Gehlert, G. / Gülicher, H.* (1983): *Statistik für Wirtschaftswissenschaftler*. Vahlen, München.
- [33] *Blum, E. / Oettli, W.* (1975): *Mathematische Optimierung*. Springer, Berlin.
- [34] *Bock, J.* (1998): *Bestimmung des Stichprobenumfangs*. Oldenbourg, München.
- [35] *Böckenhoff, A. / Hartung, J.* (1998): Some corrections of the significance level in meta-analysis. *Biometrical Journal*, vol. 40, p. 937-947.
- [36] *Böltken, F.* (1976): *Auswahlverfahren*. Teubner, Stuttgart.
- [37] *Bosch, K.* (1984): *Mathematik für Wirtschaftswissenschaftler*. Oldenbourg, München.
- [38] *Bowker, A.H* (1948): A test for symmetry in contingency tables. *Journal of the American Statistical Association*, vol. 43, p. 572-574.
- [39] *Box, G.E.P / Cox, D.R.* (1964): An analysis of transformations. *Journal of the Royal Statistical Society*, vol. B26, p. 211-252.
- [40] *Box, G.E.P / Jenkins, G.M.* (1976): *Time series analysis, forecasting and control*. Holden-Day, San Francisco.
- [41] *Box, G.E.P. / Ljung, G.M.* (1978): On a measure of lack of fit in time series models. *Biometrika*, vol. 65, p. 297-303.
- [42] *Box, G.E.P / Pierce, D.A.* (1970): Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, vol. 65, p. 1509-1526.
- [43] *Boxhammer, J.* (1981): *Zeitraffende Prüfung von Klimaeinflüssen*. Qualität und Zuverlässigkeit, vol. 26, p. 2-5.
- [44] *Bradburn, M.J. / Deeks, J.J. / Berlin, J.A. / Localio, A.R.* (2007): Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine*, vol. 26, p. 53-77.
- [45] *Breslow, N.E.* (1981): Odds ratio estimators when the data are sparse. *Biometrika*, vol. 68, p. 73-84.
- [46] *Brillinger, D.R.* (1975): *Time series – data analysis and theory*. Holt, Rinehart & Winston, New York.

- [47] Brubacher, S.R. / Wilson, G.T. (1976): Interpolating time series with application to the estimation of holiday effects on electricity demand. *Applied Statistics*, vol. 25, p. 107-116.
- [48] Bryson, M.C. / Siddiqui, M.M. (1969): Some criteria for aging. *Journal of the American Statistical Association*, vol. 64, p. 1472-1483.
- [49] Casagrande, J.T. / Pike, M.C. / Smith, P.G. (1978): An improved approximative formula for calculating sample sizes for comparing two binomial distributions. *Biometrics*, vol. 34, p. 483-486.
- [50] Chakravarti, A.K. / Vasanta, B. / Krishan, A.S.A. / Dubash, R.K. (1998): Modified Delphi methodology for technology forecasting: Case study of electronics and information technology. *Technological Forecasting and Social Change*, vol. 55, p. 155-166.
- [51] Chambers, J.M. / Cleveland, W.S. / Kleiner, B. / Tukey, P.A. (1983): Graphical methods for data analysis. Duxbury Press, Boston.
- [52] Chatfield, C. (1975): The analysis of time series, theory and practice. Chapman & Hall, London.
- [53] Chatfield, C. (2003): Time-series forecasting. Chapman & Hall, London.
- [54] Chatterjee, P. (1975): Modularization of fault trees: A method to reduce the cost of analysis. In: Reliability and fault tree analysis. SIAM, Philadelphia, p. 101-126.
- [55] Chatterjee, S. / Price, B. (1977): Regression analysis by example. Wiley, New York.
- [56] Chernoff, H. / Lehmann, E.L. (1954): The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Annals of Mathematical Statistics*, vol. 25, p. 579-586.
- [57] Clemen, R.T. / Winkler, R.L. (1986): Combining economic forecasts. *Journal of Business and Economic Statistics*, vol. 4, p. 39-46.
- [58] Cleveland, W.S. / Devlin, S.J. (1980): Calendar effects in monthly time series: Detection by spectrum analysis and graphical methods. *Journal of the American Statistical Association*, vol. 75, p. 487-496.
- [59] Cleveland, W.S. / Kleiner, B. / Tukey, J.W. (1983): Data analysis. Wadsworth, Boston.
- [60] Cloi, I. (2001): Unit root test for panel data. *Journal of International Money and Finance*, vol. 20, p. 249-272.
- [61] Cochran, W.G. (1937): Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society (Suppl.)*, vol. 4, p. 102-118.
- [62] Cochran, W.G. (1954): The combination of estimates from different experiments. *Biometrics*, vol. 10, p. 101-129.
- [63] Cochran, W.G. (1972): Stichprobenverfahren. De Gruyter, Berlin.
- [64] Cochran, W.G. / Carroll, S.P. (1953): A sampling investigation of the efficiency of weighting inversely as the estimated variance. *Biometrics*, vol. 9, p. 447-459.
- [65] Cochran, W.G. / Cox, G.M. (1975): Experimental designs. Wiley, New York.
- [66] Cochrane, D. / Orcutt, G.H. (1949): Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, vol. 44, p. 32-61.
- [67] Cogburn, D. / Davis, H.T. (1974): Periodic splines and spectral estimation. *Annals of Statistics*, vol. 2, p. 1108-1126.

- [68] *Conover, W.J.* (1971): Practical nonparametric statistics. Wiley, New York.
- [69] *Cox, D.R.* (1958): Planning of experiments. Wiley, New York.
- [70] *Cox, D.R.* (1961): Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society, Series B*, vol. 23, p. 414-422.
- [71] *Cox, D.R.* (1972): Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, series B*, vol. 34, p. 187-220.
- [72] *Cummings, G.E.* (1975): Application of the fault tree technique to a nuclear reactor containment system. In: Reliability and fault tree analysis. SIAM, Philadelphia, p. 805-825.
- [73] *Dalkey, N.C.* (1972): Studies in the quality of life - Delphi and decision making. D.C. Heath & Co., Lexington (MA).
- [74] *David, F.N.* (1954): Tables of the correlation coefficient. Cambridge University Press, London.
- [75] *David, F.N. / Johnson, N.L.* (1948): The probability integral transformation when parameters are estimated from the sample. *Biometrika*, vol. 35, p. 182-190.
- [76] *David, H.A.* (1970): Order statistics. Wiley, New York.
- [77] *David, H.A. / Hartley, H.O. / Pearson, E.S.* (1954): The distribution of the ratio in a single normal sample, of range to standard deviation. *Biometrika*, vol. 41, p. 482-493.
- [78] *Demetrescu, M. / Hassler, U. / Tarcolea, A.-I.* (2006): Combining significance of correlated statistics with application to panel data. *Oxford Bulletin of Economics and Statistics*, vol. 68, p. 647-663.
- [79] *DerSimonian, R. / Laird, N.* (1986): Meta-Analysis in clinical trials. *Controlled Clinical Trials*, vol. 7, p. 177-188.
- [80] *Dickey, D.A. / Fuller, W.A.* (1979): Distributions of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, vol. 74, p. 427-481.
- [81] *Dickey, D.A. / Fuller, W.A.* (1981): Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, vol. 49, p. 1057-1072.
- [82] *Diebold, F.X.* (1988): Serial correlation and the combination of forecasts. *Journal of Business and Economic Statistics*, vol. 6, p. 105-111.
- [83] *Dieter, U.* (1971): Pseudo-random numbers - The exact distributions of pairs. *Mathematics of Computation*, vol. 25, p. 121-138.
- [84] *Dieter, U. / Ahrens, J.H.* (1973): A combinatorical method for the generation of normally distributed random numbers. *Computing*, vol. 11, p. 137-146.
- [85] *Dixon, W.J.* (1951): Ratio involving extreme values. *Annals of Mathematical Statistics*, vol. 22, p. 68-78.
- [86] *Documenta Geigy Tabellen* (1968): Wissenschaftliche Tabellen. Basel.
- [87] *Dodge, H.F.* (1943): A sampling inspection plan for continuous production. *Annals of Mathematical Statistics*, vol. 14, p. 264-279.
- [88] *Dodge, H.F. / Romig, H.G.* (1929): A method of sampling inspection. *Bell System Technical Journal*, vol. 8, p. 613-631.
- [89] *Dodge, H.F. / Romig, H.G.* (1959): Sampling inspection tables. Wiley, New York.
- [90] *Domrowski, E.* (1970): Einführung in die Zuverlässigkeit elektronischer Geräte und Systeme. AEG-Telefunken, Berlin.

- [91] Draper, D. / Gaver, D.P. jr. / Goel, P.K. / Greenhouse, J.B. / Hedges, L.V. / Morris, C.N. / Tucker, J.R. / Waternaux, C.M. (1992): Combining information: Statistical issues and opportunities for research. American Statistical Association, National Academy Press, Washington D.C.
- [92] Draper, N.R. / Smith, H. (1981): Applied regression analysis. Wiley, New York.
- [93] Durbin, J. (1975): Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika, vol. 62, p. 5-22.
- [94] Durbin, J. / Watson, G.S. (1951): Testing for serial correlation in least squares regression – II. Biometrika, vol. 38, p. 159-178.
- [95] Eicker, F. (1963): Über die Konsistenz von Parameterschätzfunktionen für ein gemischtes Zeitreihen-Regressionsmodell. Zeitschrift für Wahrscheinlichkeitstheorie, vol. 1, p. 456-477.
- [96] Eisenhart, C. (1947): The assumptions underlying the analysis of variance. Biometrics, vol. 3, p. 1-21.
- [97] El Newehi, E. / Proschan, F. / Sethuraman, J. (1978): Multistate coherent systems. Journal of Applied Probability, vol. 15, p. 675-688.
- [98] Elpelt, B. (1982): Invariante quadratische Schätzfunktionen für Varianzkomponenten. Diplomarbeit, Universität Dortmund.
- [99] Elpelt, B. (1983): Invariante generalisierte quadratische Schätzfunktionen in multivariaten Varianzkomponentenmodellen. Dissertation, Universität Dortmund.
- [100] Elpelt, B. (1984): Über verallgemeinerte MINQ-Schätzungen in multivariaten Varianzkomponentenmodellen. Zeitschrift für Angewandte Mathematik und Mechanik, vol. 64, p. 328-330.
- [101] Elpelt, B. (1989): On linear statistical models of commutative quadratic type. Communications in Statistics, Theory and Methods, vol. 18, p. 3407-3450.
- [102] Elpelt, B. / Hartung, J. (1981): Zur kostenoptimalen prophylaktischen Qualitätssicherung bei der Herstellung von Mischprodukten. Forschungsbericht, Fachbereich Statistik der Universität Dortmund, Dortmund (auszugsweise erschienen als: Zur kostenoptimalen Qualitätsplanung bei Mischprodukten. Qualität und Zuverlässigkeit, vol. 27 (1982), p. 148-150).
- [103] Elpelt, B. / Hartung, J. (1982): Zur Fehleranalyse bei der Probenentnahme aus Schüttgütern. Qualität und Zuverlässigkeit, vol. 27, p. 225-229.
- [104] Elpelt, B. / Hartung, J. (1983a): Zum Schätzen fester und zufälliger Effekte in Gemischten Linearen Modellen. EDV in Medizin und Biologie, vol. 14, p. 7-12.
- [105] Elpelt, B. / Hartung, J. (1983 b): Mixed exterior-interior penalty methods revisited. Methods of Operations Research, vol. 47, p. 29-37.
- [106] Elpelt, B. / Hartung, J. (1984): On the estimation of parameters in general linear regression models. Methods of Operations Research, vol. 52, p. 585-600.
- [107] Elpelt, B. / Hartung, J. (1985): Diskrete Regression mittels Skalierung. Zeitschrift für Angewandte Mathematik und Mechanik, vol. 65, p. 303-304.
- [108] Elpelt, B. / Hartung, J. (2004): Grundkurs Statistik: Lehr- und Übungsbuch der angewandten Statistik. Oldenbourg, München, 3. Aufl.
- [109] EMEA (Committee for Proprietary Medicinal Products) (2001): Points to consider on application with 1. Meta-Analyses; 2. One Pivotal Study. The European Agency for the Evaluation of Medicinal Products, CPMP/EWP/2330/99,

- an Agency for the Evaluation of Medicinal Products, CPMP/EWP/2330/99, London.
- [110] *Epstein, B.* (1960): Testing for the validity of the assumption that the underlying distribution of life is exponential. *Technometrics*, vol. 2, p. 83-101 und 167-183.
 - [111] *Esary, J.D. / Ziehms, H.* (1975): Reliability analysis of phased missions. In: *Reliability and fault tree analysis*. SIAM, Philadelphia, p. 213-236.
 - [112] *Esenwein-Rothe, I.* (1976): Die Methoden der Wirtschaftsstatistik 1 und 2. Vandenhoeck & Ruprecht, Göttingen.
 - [113] *Ezekiel, M. / Fox, K.A.* (1970): Methods of correlation and regression analysis. Wiley, New York.
 - [114] *Fechner, G.T.* (1860): Elemente der Psychophysik. Breitkopf & Härtel, Leipzig.
 - [115] *Feigl, P.* (1978): A graphical aid for determining sample size when comparing two independent proportions. *Biometrics*, vol. 34, p. 111-122.
 - [116] *Feller, W.* (1968): An introduction to probability theory and its applications I. Wiley, New York.
 - [117] *Ferschl, F.* (1978): Descriptive Statistik. Physica, Würzburg.
 - [118] *Fieller, E.C.* (1940): The biological standardization of insulin. *Journal of the Royal Statistical Society (Suppl.)*, vol. 7, p. 1-64.
 - [119] *Fienberg, S.E.* (1977): The analysis of cross-classified categorical data. The MIT Press, Cambridge (MA).
 - [120] *Finney, D.J. / Latscha, R. / Bennett, B.M. / Hsu, P.* (1963): Tables for testing significance in a 2 x 2 contingency table. Cambridge University Press, London.
 - [121] *Fisher, R.A.* (1932): Statistical Methods for research workers. Oliver & Boyd, London, 4th ed.
 - [122] *Fisher, R.A.* (1940): The precision of discriminant functions. *Annals of Eugenics*, vol. 10, p. 422-429.
 - [123] *Fisher, R.A.* (1966): The design of experiments. Oliver & Boyd, Edinburgh.
 - [124] *Fisher, R.A.* (1972): Statistical methods for research workers. Oliver & Boyd, Edinburgh.
 - [125] *Fisz, M.* (1970): Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin.
 - [126] *Fleming, T.R. / Harrington, D.P.* (2005): Counting processes and survival analysis. Wiley, New York.
 - [127] *Förster, W.* (1969): Zerlegung und Lösung diskreter ökonomischer Prozeßmodelle. Mohr, Tübingen.
 - [128] *Fomby, T.B. / Hill, R.C. / Johnson, S.R.* (1984): Advanced econometric methods. Springer, New York.
 - [129] *Fuller, W.A.* (1976): Introduction to statistical time series. Wiley, New York.
 - [130] *Gaede, K.-W.* (1977): Zuverlässigkeit: Mathematische Modelle. Hanser, München.
 - [131] *Gail, M.H. / Gastwirth, J.L.* (1978): A scale free goodness-of-fit-test for the exponential distribution based on the Lorenz curve. *Journal of the American Statistical Association*, vol. 73, p. 787-793.
 - [132] *Galton, F.* (1889): Natural inheritance. Macmillan, London.
 - [133] *Gastwirth, J.L.* (1966): On robust procedures. *Journal of the American Statistical Association*, vol. 61, p. 929-948.

- [134] *Gastwirth, J.L. / Cohen, M.L.* (1970): Small sample behaviour of some robust linear estimators of location. *Journal of the American Statistical Association*, vol. 65, p. 946-973.
- [135] *Gaul, W. / Hartung, J.* (1979): A barrier method with arbitrary starting point. *Mathematische Operationsforschung und Statistik, Series Optimization*, vol. 10, p. 317-323.
- [136] *Gaul, W. / Hartung, J.* (1985): Multistate reliability problems for GSP-digraphs. In: *Neumann, K. / Pallaschke, D.* (eds.): Contributions to operations research. Lecture notes in economics and mathematical systems, vol. 240, Springer, Berlin, p. 41-53.
- [137] *Gehan, E.A.* (1965): A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika*, vol. 52, p. 203-223.
- [138] *Glass, G.V.* (1976): Primary, secondary, and meta-analysis of research. *Educational Researcher*, vol. 5, p. 3-8.
- [139] *Goldberg, M.F.* (1964): Comprehensive failure mechanisms theory. In: Physics of failure in electronics, p. 25-60.
- [140] *Goldsmith, P.L. / Boddy, R.* (1973): Critical analysis of factorial experiments and orthogonal fractions. *Applied Statistics*, vol. 22, p. 141-160.
- [141] *Goodman, L.B. / Kruskal, W.H.* (1954): Measures of association for crossclassifications. *Journal of the American Statistical Association*, vol 49, p. 732-764.
- [142] *Gosset, W.S.* (1914): The elimination of spurious correlation due to position in time or space. *Biometrika*, vol. 10, p. 179-180.
- [143] *Grafarend, E. / Schaffrin, B.* (1979): Variance-covariance-component estimation of Helmert type. Paper presented at the ACSM annual meeting, Washington DC.
- [144] *Gray, H.L. / Schucany, W.R.* (1969): Lower confidence limits for availability assuming lognormally distributed repair times. *IEEE-Transactions on Reliability*, vol. R-18, p. 157-162.
- [145] *Graybill, F.A.* (1961): An introduction to linear statistical models I. McGraw-Hill, New York.
- [146] *Graybill, F.A.* (1976): Theory and application of the linear model, Duxbury Press, North Scituate (MA).
- [147] *Greenland, S. / Robins, J.M.* (1985): Estimation of common effect parameter from sparse follow-up data. *Biometrics*, vol. 41, p. 55-68.
- [148] *Grenander, U. / Rosenblatt, M.* (1957): Statistical analysis of stationary time series. Wiley, New York.
- [149] *Grubbs, F.E.* (1973): Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics*, vol. 15, p. 53-66.
- [150] *Grubbs, F.E. / Beck, G.* (1972): Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, vol. 14, p. 847-854.
- [151] *Gutjahr, W.* (1971): Die Messung psychologischer Eigenschaften. VEB Deutscher Verlag der Wissenschaften, Berlin.
- [152] *Guttman, L.* (1941): An outline of the statistical theory of prediction, Supplementary Study B-1, In: *Horst, P.* (ed.): The prediction of personal adjustment. Bulletin 48, Social Science Research Council, New York, p. 253-318.
- [153] *Haberman, S.J.* (1978): Analysis of qualitative data I: introductory topics. Academic Press, New York.

- [154] *Haberman, S.J.* (1979): Analysis of qualitative data II: New developments. Academic Press, New York.
- [155] *Häder, M. / Häder, S.* (2000): Die Delphi-Technik in den Sozialwissenschaften. Medizinische Forschungen und innovative Anwendungen. Westdeutscher Verlag, Opladen.
- [156] *Hájek, J. / Šidák, Z.* (1967): Theory of rank tests. Academic Press, New York.
- [157] *Hald, A.* (1965): Statistical theory with engineering applications. Wiley, New York.
- [158] *Hamaker, H.C. / Strik, R.v.* (1955): The efficiency of double sampling for attributes. *Journal of the American Statistical Association*, vol. 50, p. 830-849.
- [159] *Hamilton, J.D.* (1994): Time series analysis. Princeton University Press, Princeton (CA).
- [160] *Hanau, A.* (1930): Die Prognose der Schweinepreise. *Vierteljahrshefte zur Konjunkturforschung*, Sonderheft 18.
- [161] *Hannan, E.J.* (1970): Multiple time series. Wiley, New York.
- [162] *Hannan, E.J. / Quinn, B.G.* (1979): The determination of the order of an auto-regression. *Journal of the Royal Statistical Society*, vol. B 41, p. 190-195.
- [163] *Harman, H.H.* (1967): Modern factor analysis. University of Chicago Press, Chicago.
- [164] *Hartung, J.* (1971): Verallgemeinerte Geometrische Optimierung. Diplomarbeit, Universität Bonn.
- [165] *Hartung, J.* (1972a): Separation bei Nicht-Linearen Programmen und Vektorminimierung. Dissertation, Universität Bonn.
- [166] *Hartung, J.* (1972b): Dualität und Sattelpunkte. *Operations Research Verfahren*, vol. XII, p. 194-200.
- [167] *Hartung, J.* (1974a): Über Ersatzprogramme bei der Vektoroptimierung. *Operations Research Verfahren*, vol. XVIII, p. 129-133.
- [168] *Hartung, J.* (1974b): A sequential regularization method for finding saddle points. *Operations Research Verfahren*, vol. IXX, p. 99-102.
- [169] *Hartung, J.* (1975a): Einige Kriterien für die Konvexität einer Funktion. *Operations Research Verfahren*, vol. XX, p. 37-43.
- [170] *Hartung, J.* (1975b): Penalty Methoden für Kontrollprobleme und Open-Loop-Differential-Spiele. In: *Bulirsch, R. / Oettli, W. / Stoer, J.* (eds.): Optimization and optimal control. Lecture Notes in Mathematics, vol. 477, Springer, Berlin-Heidelberg-New York, p. 127-144.
- [171] *Hartung, J.* (1976a): Some inequalities for a random function. *Teorija Verojatnostei i ee Primenenja*, vol. 21, p. 661-665 (und: Theory of Probability and its Applications, SIAM, 1977, vol. 21).
- [172] *Hartung, J.* (1976b): On a method for computing pseudoinverses. In: *Oettli, W. / Ritter, K.* (eds.): Optimization and operations research, Lecture notes in economics and mathematical systems, vol. 117, Springer, Berlin, p. 116-125.
- [173] *Hartung, J.* (1976c): A barrier method for finding saddle points. *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 56, p. 345-346.
- [174] *Hartung, J.* (1976d): On the marginal value of an antagonistic game. In: *Cea, J.* (ed.): Optimization techniques. Lecture Notes in Computer Science, vol. 41, Springer, Berlin-Heidelberg-New York, p. 1-8.
- [175] *Hartung, J.* (1976e): Über lexikographische Spiele. In: *Kohlas, J. / Seifert, O. / Zimmermann, H. J.* (eds.): Proceedings in Operations Research 5. Physica, Würzburg, p. 77-83.

- [176] Hartung, J. (1977a): Zur Darstellung pseudoinverser Operatoren. Archiv der Mathematik, vol. XXVIII, p. 200-208.
- [177] Hartung, J. (1977b): Zur Optimalität von Schätzungen in linearen Regressionsmodellen. Zeitschrift für Angewandte Mathematik und Mechanik, vol. 57, p. 328-329.
- [178] Hartung, J. (1977c): Zur Methode des steilsten Abstiegs für singuläre lineare Operatoren-Gleichungen. Operations Research Verfahren, vol. XXV, p. 50-57.
- [179] Hartung, J. (1977d): On the algorithm of Allran and Johnsen for solving nonlinear programming problems. In: Walukiewicz, S. / Wierzbicki, A.P. (eds.): Methods of mathematical programming. Proceedings of the International Conference held in Zakopane, Poland, p. 119-122.
- [180] Hartung, J. (1978a): Zur Verwendung von Vorinformation in der Regressionsanalyse. Operations Research Verfahren, vol. IXXX, p. 550-558.
- [181] Hartung, J. (1978 b): Über ein duales Verfahren für lineare Approximationen und dessen statistische Anwendung. Zeitschrift für Angewandte Mathematik und Mechanik, vol. 58, p. 475-477.
- [182] Hartung, J. (1978c): A stable interior penalty method for convex extremal problems. Numerische Mathematik, vol. 29, p. 149-158.
- [183] Hartung, J. (1978d): Minimum norm solutions of convex programs. In: Henn, R. / Korte, B. / Oettli, W. (eds.): Optimization and operations research. Lecture notes in economics and mathematical systems, vol. 157, Springer, Berlin, p. 127-135.
- [184] Hartung, J. (1979a): A note on restricted pseudoinverses. SIAM Journal on Mathematical Analysis, vol. 10, p. 266-273.
- [185] Hartung, J. (1979b): On the fixed endpoint quadratic regulator problem in the presence of noise. Methods of Operations Research, vol. 36, p. 121-135.
- [186] Hartung, J. (1980): On exponential penalty function methods. Mathematische Operationsforschung und Statistik, Series Optimization, vol. 11, p. 71-84.
- [187] Hartung, J. (1981): Non-negative minimum biased invariant estimation in variance component models. Annals of Statistics, vol. 9, p. 278-292.
- [188] Hartung, J. (1982a): On two-stage minimax problems. Pacific Journal of Mathematics, vol. 102, p. 355-368.
- [189] Hartung, J. (1982b): An extension of SION's minimax theorem with an application to a method for constrained games. Pacific Journal of Mathematics, vol. 103, p. 401-408.
- [190] Hartung, J. (1985): A Hadamard-stable extension of Courant's sequential method for convex extremal problems. Numerical Functional Analysis and Optimization, vol. 7, p. 363-378.
- [191] Hartung, J. (1985b): Studienmodell Datenanalytik - Datenanalytik als Studium neben dem Beruf. Zeitschrift für Hochschuldidaktik (Beiträge zu Studium, Wissenschaft und Beruf), vol. 9, p. 277-284.
- [192] Hartung, J. (1999a): An alternative method for meta-analysis. Biometrical Journal, vol. 41, p. 901-916.
- [193] Hartung, J. (1999b): A note on combining dependent tests of significance. Biometrical Journal, vol. 41, p. 849-855.
- [194] Hartung, J. (1999c): A short-cut method for computing positive variance component estimates. In: Krumm, F. / Schwarze, V.S. (eds.): Quo vadis geodesia ...? Festschrift for Erik W. Grafarend on the occasion of his 60th birthday, Part 1.

- Schriftenreihe der Institute des Studiengangs Geodäsie und Geoinformatik, Universität Stuttgart, Technical Report Nr. 1999.6-1, p. 151-154.
- [195] Hartung, J. (1999d): Ordnungserhaltende positive Varianzschätzer bei gepaarten Messungen ohne Wiederholungen. Allgemeines Statistisches Archiv, vol. 83, p. 230-247.
- [196] Hartung, J. (2000): A new class of self-designing clinical trials. In: Hasman, A. / Prokosch, H.U. / Blobel, B. / Dudeck, J. / Gell, G. / Engelbrecht, R. (eds.): Medical Infobahn for Europe. Proceedings of MIE 2000 and GMDS 2000. IOS Press, Amsterdam, p. 310-314.
- [197] Hartung, J. (2001): A self-designing rule for clinical trials with arbitrary response variables. Controlled Clinical Trials, vol. 22, p. 111-116.
- [198] Hartung, J. (2003): A short-cut method for computing positive variance component estimates. In: Grafarend, E.W. / Krumm, F.W. / Schwarze, V.S. (eds.): Geodesy - The challenge of the third millennium. Springer, Berlin, p. 329-331.
- [199] Hartung, J. (2006): Flexible designs by adaptive plans of generalized Pocock- and O'Brien-Fleming-type and by self-designing clinical trials. Biometrical Journal, vol. 48, p. 521-536.
- [200] Hartung, J. / Argac, D. (2001): Testing for homogeneity in combining of two-armed trials with normally distributed responses. Sankhya - The Indian Journal of Statistics, vol. B63, p. 298-310.
- [201] Hartung, J. / Argac, D. (2002a): Confidence intervals on the among group variance component in an unbalanced and heteroscedastic one way random-effects model. Statistics & Decisions, vol. 20, p. 331-353.
- [202] Hartung, J. / Argac, D. (2002b): Generalizing the Welch test to non-zero hypotheses on the variance component in the one-way random effects model under variance heterogeneity. Statistics, vol. 36, p. 89-99.
- [203] Hartung, J. / Argac, D. (2009): Confidence intervals for combined univariate economic forecasts. TU Dortmund, Eldorado, <http://hdl.handle.net/2003/26012>.
- [204] Hartung, J. / Argac, D. / Makambi, K.H. (2001): A note on testing the nullity of the between group variance in the one-way random effects model under variance heterogeneity. Journal of Applied Statistics, vol. 28, p. 215-222.
- [205] Hartung, J. / Argac, D. / Makambi, K.H. (2002): Small sample properties of tests on homogeneity in one-way ANOVA and meta-analysis. Statistical Papers, vol. 43, p. 197-235.
- [206] Hartung, J. / Argac, D. / Makambi, K.H. (2003): Homogeneity tests in meta-analysis. In: Schulze, R. / Holling, H. / Böhning, D. (eds.): Meta-analysis. New developments and applications in medical and social sciences. Hogrefe & Huber, Göttingen, p. 3-20.
- [207] Hartung, J. / Böckenhoff, A. (2000): Meta-analysis. Different methods - different conclusions? In: Hasman, A. / Prokosch, H.U. / Blobel, B. / Dudeck, J. / Gell, G. / Engelbrecht, R. (eds.): Medical Infobahn for Europe. Proceedings of MIE 2000 and GMDS 2000. IOS Press, Amsterdam, p. 39-43.
- [208] Hartung, J. / Böckenhoff, A. (2003): Zweistufige Schätzung in der Meta-Analyse. In: Christmann, A. / Weihs, C. (eds.): Data Mining und Statistik in Hochschule und Wirtschaft. Proceedings der 6. Konferenz der SAS-Anwender in Forschung und Entwicklung (KSFE). Shaker, Aachen, p. 35-46.
- [209] Hartung, J. / Böckenhoff, A. / Knapp, G. (1995): Zuverlässigkeitssnachweis von Bauteilarten. Uni-Report - Berichte aus der Forschung der Universität Dortmund, vol. 21, p. 25-29.

- [210] Hartung, J. / Böckenhoff, A. / Knapp, G. (2003): Generalized Cochran-Wald statistics in combining of experiments. *Journal of Statistical Planning and Inference*, vol. 113, p. 215-237.
- [211] Hartung, J. / Dugošija, D.P. (1974a): Über Stabilität in Zwei-Personen Null-Summen-Spielen. *Mathematica Balkanica*, vol. 4.26, p. 151-154.
- [212] Hartung, J. / Dugošija, D.P. (1974b): Zweistufige antagonistische Spiele. *Mathematica Balkanica*, vol. 4.27, p. 155-158.
- [213] Hartung, J. / Dugošija, D.P. (1975): Eine Penalty-Methode für antagonistische Spiele. *Operations Research Verfahren*, vol. XX, p. 23-25.
- [214] Hartung, J. / Dugošija, D.P. (1976): Über Gâteaux-Differenzierbarkeit der Minimum-Funktion. *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 56, p. 243-246.
- [215] Hartung, J. / Elpelt, B. (1984): *Multivariate Statistik - Lehr- und Handbuch der angewandten Statistik*. Oldenbourg, München.
- [216] Hartung, J. / Elpelt, B. (1986a): Zuverlässigkeitssanalyse elektronischer Baulemente. *Uni-Report - Berichte aus der Forschung der Universität Dortmund*, vol. 3, p. 8-10.
- [217] Hartung, J. / Elpelt, B. (1986b): On scaling of qualitative data. In: Gaul, W. / Schader, M. (eds.): *Classification as a tool of research*. Springer Berlin-Heidelberg-New York, p. 113-120.
- [218] Hartung, J. / Elpelt, B. (2007): *Multivariate Statistik: Lehr- und Handbuch der angewandten Statistik*. Oldenbourg, München, 7. Aufl.
- [219] Hartung, J. / Elpelt, B. / Schwaiger, B. (1991): *Multivariate Verfahren - Erweiterte Neufassung*. FernUniversität Hagen.
- [220] Hartung, J. / Elpelt, B. / Voet, B. (1997): *Modellkatalog Varianzanalyse*. Buch mit CD-ROM. Oldenbourg, München.
- [221] Hartung, J. / Hartung, J.E. (2001): Personalisieren. Ein strategisches Instrument im E-CRM. In: Schneider, D. / Pflaumer, P. (eds.): *Power-tools. Management-, Beratungs- und Controllinginstrumente*. Gabler, Wiesbaden., p. 169-186.
- [222] Hartung, J. / Heine, B. (1984): Schätzung der Varianz von GRUBBS' Schätzern für die Präzision von Meßinstrumenten. *Allgemeines Statistisches Archiv*, vol. 68, p. 257-272.
- [223] Hartung, J. / Heine, B. (1999): *Statistik-Übungen: Deskriptive Statistik*. Oldenbourg, München, 6. Aufl.
- [224] Hartung, J. / Heine, B. (2004): *Statistik-Übungen: Induktive Statistik*. Oldenbourg, München, 4. Aufl.
- [225] Hartung, J. / Jöckel, K.-H. (1982): Zuverlässigkeit- und Wirtschaftlichkeitsüberlegungen bei Straßenverkehrssignalanlagen. *Qualität und Zuverlässigkeit*, vol. 27, p. 65-68.
- [226] Hartung, J. / Kalin, D. (1980): Zur Zuverlässigkeit von Straßenverkehrssignalanlagen. *Qualität und Zuverlässigkeit*, vol. 25, p. 305-308.
- [227] Hartung, J. / Klösener, K.-H. (1985): Estimating precision of measurement and product variability. *Forschungsbericht, Fachbereich Statistik, Universität Dortmund*, Dortmund.
- [228] Hartung, J. / Klösener, K.-H. (1987): Zur Bestimmung der Präzision von Meßverfahren. In: Opitz, O. / Rauhut, B. (eds.): *Mathematik und Ökonomie. Rudolf Henn zum 65. Geburtstag*. Springer, Berlin-Heidelberg-New York, p. 286-296.

- [229] Hartung, J. / Knapp, G. (1998): Messen faserförmiger Partikel. Strategie zur Auswertung des ersten VDI-Ringversuchs zur Asbestfaserbestimmung nach Richtlinie VDI 3492 Blatt 1. Allgemeines Statistisches Archiv, vol. 82, p. 312-326.
- [230] Hartung, J. / Knapp, G. (1999): Eine Re-Analyse der Blue Mountains Eye Study zum Kataraktrisiko durch inhalative Kortikoide. Pneumologie, vol. 53, p. 411-416.
- [231] Hartung, J. / Knapp, G. (2000a): Combined test procedures in the meta-analysis of controlled clinical trials. In: Hasman, A. / Prokosch, H.U. / Blobel, B. / Dukeck, J. / Gell, G. / Engelbrecht, R. (eds.): Medical Infobahn for Europe. Proceedings of MIE 2000 and GMDS 2000. IOS Press, Amsterdam, p. 34-38.
- [232] Hartung, J. / Knapp, G. (2000b): Confidence intervals for the between group variance in the unbalanced one-way random effects model of analysis of variance. Journal of Statistical Computation and Simulation, vol. 65, p. 311-324.
- [233] Hartung, J. / Knapp, G. (2000c): Asymptotic tests for general linear hypotheses on variance components in models of commutative quadratic type. Statistics, vol. 35, p. 23-44.
- [234] Hartung, J. / Knapp, G. (2001a): On tests of the overall treatment effect in the meta-analysis with normally distributed responses. Statistics in Medicine, vol. 20, p. 1771-1782.
- [235] Hartung, J. / Knapp, G. (2001b): A refined method for the meta-analysis of controlled clinical trials with binary outcome. Statistics in Medicine, vol. 20, p. 3875-3889.
- [236] Hartung, J. / Knapp, G. (2001c): Strategien zur Beurteilung eines Behandlungseffektes mit Hilfe einer Meta-Analyse. Informatik, Biometrie und Epidemiologie in Medizin und Biologie, vol. 32, p. 44-59.
- [237] Hartung, J. / Knapp, G. (2001d): Data mining. In: Schneider, D. / Pflaumer, P. (eds.): Power-tools. Management- Beratungs- und Controllinginstrumente. Gabler, Wiesbaden, p. 189-201.
- [238] Hartung, J. / Knapp, G. (2001e): Komplexitätsreduktion im Data- und e-Mining: Dependente Q-Techniken. In: Schumacher, E. / Streichfuss, K. (eds.): Proceedings der 5. Konferenz der SAS-Anwender in Forschung und Entwicklung (KSFE), Hohenheim, p. 143-151.
- [239] Hartung, J. / Knapp, G. (2001f): On the assessment of heterogeneity in randomized controlled multi-centre trials with binary outcome. In: Kunert, J. / Trenkler, G. (eds.): Mathematical statistics with applications in biometry: Festschrift in honour of Siegfried Schach. Eul-Verlag, Lohmar, p. 343-354.
- [240] Hartung, J. / Knapp, G. (2003a): An alternative test procedure for meta-analysis. In: Schulze, R. / Holling, H. / Böhning, D. (eds.): Meta-Analysis. New developments and applications in medical and social sciences. Hogrefe & Huber, Göttingen, p. 53-69.
- [241] Hartung, J. / Knapp, G. (2003b): Confidence regions on variance components in an extended ANOVA model for combining information. Acta Applicandae Mathematicae, vol. 78, p. 207-221.
- [242] Hartung, J. / Knapp, G. (2003c): A new class of completely self-designing clinical trials. Biometrical Journal, vol. 45, p. 3-19.
- [243] Hartung, J. / Knapp, G. (2004): Improved tests of homogeneity in randomized controlled multi-center trials with binary outcome. Far East Journal of Theoretical Statistics, vol. 13, p. 101-126.

- [244] Hartung, J. / Knapp, G. (2005a): Models for combining results of different experiments: Retrospective and prospective. *American Journal of Mathematical and Management Sciences*, vol. 25, p. 149-188.
- [245] Hartung, J. / Knapp, G. (2005b): On confidence intervals for the among-group variance in the one-way random effects model with unequal error variances. *Journal of Statistical Planning and Inference*, vol. 127, p. 157-177.
- [246] Hartung, J. / Knapp, G. (2005c): Repeated confidence intervals in self-organizing studies. In: Baier, D. / Decker, R. / Schmidt-Thieme, L. (eds.): *Data analysis and decision support*. Springer, Heidelberg, p. 39-46.
- [247] Hartung, J. / Knapp, G. (2006): Repeated confidence intervals in self-designing clinical trials and switching between noninferiority and superiority. *Biometrical Journal*, vol. 48, p. 697-709.
- [248] Hartung, J. / Knapp, G. (2007): Nested repeated confidence intervals in adaptive group sequential clinical trials. *East-West Journal of Mathematics. Special volume (Proceedings of the International Conference in Mathematics and Applications 2007)*, p. 17-29.
- [249] Hartung, J. / Knapp, G. (2008): Exact and generalized confidence intervals in the common mean problem. In: Schipp, B. / Krämer, W. (eds.): *Statistical inference, econometric analysis and matrix algebra: Festschrift in honour of Götz Trenkler*. Physica, Heidelberg, p. 85-102.
- [250] Hartung, J. / Knapp, G. (2009a): Adaptive group sequential confidence intervals for the ratio of normal means. TU Dortmund, Eldorado, <http://hdl.handle.net/2003/26011>.
- [251] Hartung, J. / Knapp, G. (2009b): Standardized mean differences in adaptive group sequential trials. TU Dortmund, Eldorado, <http://hdl.handle.net/2003/26013>.
- [252] Hartung, J. / Knapp, G. (2009c): A confidence interval approach for difference and ratio of normal means in self-designing clinical trials. TU Dortmund, Eldorado, <http://hdl.handle.net/2003/26010>.
- [253] Hartung, J. / Knapp, G. (2009d): Adaptive controlled noninferiority group sequential trials. TU Dortmund, Eldorado, <http://hdl.handle.net/2003/26014>.
- [254] Hartung, J. / Knapp, G. (2009e): Adaptive confidence intervals of desired length and power for normal means. TU Dortmund, Eldorado, <http://hdl.handle.net/2003/26015>.
- [255] Hartung, J. / Knapp, G. / Sinha, B.K. (2008): *Statistical meta-analysis with applications*. Wiley, New York.
- [256] Hartung, J. / Maidhof, R. / Pirk, O. / Winkel, M. (1999): Es geht auch anders - Ein neues Instrument zur Begutachtung der Pflegestufen? *Das Gesundheitswesen*, vol. 61, p. 380-384.
- [257] Hartung, J. / Makambi, K.H. (2001): Simple non-iterative t-distribution based tests for meta-analysis. *South African Statistical Journal*, vol. 35, p. 1-17.
- [258] Hartung, J. / Makambi, K.H. (2002a): Positive estimation of the between-study variance in meta-analysis. *South African Statistical Journal*, vol. 36, p. 55-76.
- [259] Hartung, J. / Makambi, K.H. (2002b): A combined test for improved meta-analysis. *Allgemeines Statistisches Archiv*, vol. 86, p. 397-405.
- [260] Hartung, J. / Makambi, K.H. (2003): Reducing the number of unjustified significant results in meta-analysis. *Communications in Statistics, Simulation and Computation*, vol. 32, p. 1179-1190.

- [261] Hartung, J. / Makambi, K.H. / Argac, D. (2001): An extended ANOVA F-test with applications to the homogeneity problem in meta-analysis. *Biometrical Journal*, vol. 43, p. 135-146.
- [262] Hartung, J. / Schmitz, K.-J. (1979): Grundsätzliche Bemerkungen zur "energetischen Futterwertprüfung für Wiederkäuer" aus mathematisch-statistischer Sicht. *Kraftfutter*, vol. 62, p. 34-38 und 70-74.
- [263] Hartung, J. / Schmitz, K.-J. (1980): Zur Qualitätsbeurteilung bei einem nicht alleine verhüttbaren Erz. Qualität und Zuverlässigkeit, vol. 25, p. 65-68.
- [264] Hartung, J. / Stückler, M. / Memmel, U. / Hoffmann, M. / Altmeyer, P. (2001): Vitamin B12 cream containing avocado oil in the therapy of plaque psoriasis. *Dermatology*, vol. 203, p. 141-147.
- [265] Hartung, J. / Voet, B. (1984): Zur Gütebeurteilung von Varianzkomponentenschätzungen. *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 64, p. 332-334.
- [266] Hartung, J. / Voet, B. (1985): Modellkatalog "Varianzkomponenten". *Forschungsbericht, Fachbereich Statistik, Universität Dortmund*, Dortmund.
- [267] Hartung, J. / Voet, B. (1986): Best invariant unbiased estimators for the mean squared error of variance component estimators. *Journal of the American Statistical Association*, vol. 81, p. 689-691.
- [268] Hartung, J. / Voet, B. (1987): An asymptotic chi-square-test for variance components. In: *Sendler, W.* (ed.): Contributions to Stochastics. In honour of the 75th birthday of Walther Eberl, sr. Physica, Heidelberg, p. 153-163.
- [269] Hartung, J. / Voet, B. (1992): On planning and inference in linear models with general covariance structure. In: *Schach, S. / Trenkler, G.* (eds.): Data analysis and statistical inference. *Festschrift in honour of Friedhelm Eicker*. Eul-Verlag, Lohmar, p. 435-452.
- [270] Hartung, J. / Voet, B. (1999): Some variants for testing linear hypotheses on variance components. In: *Gaul, W. / Schader, M.* (eds.): *Mathematische Methoden der Wirtschaftswissenschaften. Festschrift für Otto Opitz*. Physica, Heidelberg, p. 130-140.
- [271] Hartung, J. / Werner, H.J. (1980): Zur Verwendung der restringierten Moore-Penrose-Inversen beim Testen von linearen Hypothesen. *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 60, p. 344-346.
- [272] Hartung, J. / Werner, H.J. (1983): Zum Auswerten unbalanzierter Daten. *EDV in Medizin und Biologie*, vol. 14, p. 13-18.
- [273] Hartung, J. / Werner, H.J. (1984): Hypothesenprüfung im Restringierten Linearen Modell - Theorie und Anwendungen. Vandenhoeck & Ruprecht, Göttingen.
- [274] Hartung, J. / Zehender, M. / Meinertz, T. / Hohenloser, S. / Geibel, A. / Seiler, K.-U. / Just, H. (1989): Incidence and clinical relevance of QT prolongation caused by the new selective Serotonin antagonist Ketanserin. *American Journal of Cardiology*, vol. 63, p. 826-832.
- [275] Hasemann, J.K. (1978): Exact sample sizes for use with the Fisher-Irwin test for 2 x 2 tables. *Biometrics*, vol. 34, p. 106-109.
- [276] Hassler, U. / Tarcolea, A.I. (2005): Combining multi-country evidence on unit roots: The case of long-term interest rates. *Applied Economics Quarterly*, vol. 51, p. 181-189.
- [277] Hauser, S. (1979): Daten, Datenanalyse und Datenbeschaffung in den Wirtschaftswissenschaften. Hain, Königsstein/Ts.

- [278] *Hebbel, H.* (1982): Lineare Systeme, Analyse, Schätzungen und Prognosen (unter Verwendung von Splinefunktionen). Habilitationsschrift. Universität Dortmund.
- [279] *Hebbel, H. / Heiler, S.* (1978): Die Verwendung von Spline-Funktionen bei der Schätzung von Spektraldichten. In: *Schäffer, K.A.* (ed.): Splinefunktionen in der Statistik. Sonderhefte zum Allgemeinen Statistischen Archiv, vol. 14, p. 66-85. Vandenhoeck & Ruprecht, Göttingen.
- [280] *Hebbel, H. / Heiler, S.* (1985): Zeitreihenglättung in einem Fehler-in-den-Variablen-Modell. In: *Vogel, F.* (ed.): Statistik zwischen Theorie und Praxis. Vandenhoeck & Ruprecht, Göttingen.
- [281] *Hebbel, H. / Kuhlmeier, N.* (1983): Eine Weiterentwicklung von Heiler's Berliner Verfahren. Forschungsbericht, Fachbereich Statistik, Universität Dortmund, Dortmund.
- [282] *Hedges, L.V.* (1983): A random effects model for effect sizes. *Psychological Bulletin*, vol. 93, p. 388-395.
- [283] *Hedges, L.V. / Gurevitch, I. / Curtis, P.S.* (1999): The meta-analysis of response ratios in experimental ecology. *Ecology*, vol. 80, p. 1150-1156.
- [284] *Hedges, L.V. / Olkin, I.* (1985): Statistical methods for meta-analysis. Academic Press, New York.
- [285] *Heiler, S.* (1966): Analyse der Struktur wirtschaftlicher Prozesse durch Zerlegen von Zeitreihen. Dissertation, Universität Tübingen.
- [286] *Heiler, S.* (1970): Theoretische Grundlagen des „Berliner Verfahrens“. In: *Wetzel, W.* (ed.): Neuere Entwicklungen auf dem Gebiet der Zeitreihenanalyse. Sonderhefte zum Allgemeinen Statistischen Archiv, vol. 1, p. 67-93, Vandenhoeck & Ruprecht, Göttingen.
- [287] *Heiler, S.* (1980): Robuste Schätzungen im linearen Modell. In: *Nowak, H. / Zentgraf, R.* (eds.): Medizinische Informatik und Statistik Band 20 - Robuste Verfahren. Springer, Berlin, p. 35-55.
- [288] *Heiler, S.* (1981): Zeitreihenanalyse. Handwörterbuch der Wirtschaftswissenschaften 36/37, gemeinsamer Verlag Fischer/Mohr/Vandenhoeck & Ruprecht, p. 582-599.
- [289] *Heiler, S. / Rinne, H.* (1971): Einführung in die Statistik. Hain, Meisenheim.
- [290] *Heine, B.* (1993): Nonnegative estimation of variance components in an unbalanced one-way random effects model. *Communications in Statistics, Theory and Methods*, vol. 32, p. 2351-2371.
- [291] *Helmbert, F.R.* (1876): Diskussion der Beobachtungsfehler in Koppes Vermessung für die Gotthardtunnelachse. *Zeitschrift für Vermessungswesen*, vol. V, p.129-155.
- [292] *Helmbert, F.R.* (1907): Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate. Teubner, Leipzig.
- [293] *Higgins, J.P.T. / Thompson, S.G.* (2002): Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, vol. 21, p. 1539-1558.
- [294] *Himmelblau, D.M.* (1972): Applied nonlinear programming. McGraw-Hill, New York.
- [295] *Hinderer, K.* (1975): Grundbegriffe der Wahrscheinlichkeitstheorie. Springer, Berlin.
- [296] *Hjorth, U.* (1980): A reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rates. *Technometrics*, vol. 22, p. 99-107.

- [297] *Hoaglin, D.C. / Mosteller, F. / Tukey, J.W.* (1983): Understanding robust and exploratory data analysis. Wiley, New York.
- [298] *Hollander, M. / Proschan, F.* (1972): Testing whether new is better than used. Annals of Mathematical Statistics, vol. 43, p. 1136-1146.
- [299] *Hollander, M. / Proschan, F.* (1975): Tests for mean residual life. Biometrika, vol. 62, p. 585-593.
- [300] *Hollander, M. / Wolfe, D.A.* (1973): Nonparametric statistical methods. Wiley, New York.
- [301] *Holm, S.* (1979): A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, vol. 6, p. 65-70.
- [302] *Holmes, R.B.* (1972): A course on optimization and best approximation. Springer, Berlin.
- [303] *Horvitz, D.G. / Thompson, D.J.* (1951): A generalization of sampling from a finite universe. Journal of the American Statistical Association, vol. 47, p. 663-685.
- [304] *Hotelling, H.* (1960): The behaviour of some standard statistical tests under non-standard conditions. Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. I, p. 319-360.
- [305] *Hovsepian, J.A. / Byar, D.P. and the Veterans Administration Cooperative Urological Research Group* (1975): Carcinome of prostatata - Correlation between radiologic quantitation of metastases and patient survival. Urology, vol. 6, p. 11-16.
- [306] *Hsu, J.C.* (1989): Sample size computation for designing multiple comparison experiments. Computational Statistics and Data Analysis, vol. 7, p. 79-91.
- [307] *Huber, P.J.* (1973): Robust regression: Asymptotics, conjectures and Monte Carlo. Annals of Statistics, vol. 1, p. 799-821.
- [308] *Huber, P.J.* (1981): Robust statistics. Wiley, New York.
- [309] *Irle, A. / Klösener, K.-H.* (1980): Note on the sign test in the presence of ties. Annals of Statistics, vol. 8, p. 1168-1170.
- [310] *Jaeckel, L.A.* (1972): Estimating regression coefficients by minimizing the dispersion of the residuals. Annals of Mathematical Statistics, vol. 43, p. 1449-1458.
- [311] *Jennison, C. / Turnbull, B.* (2000): Group sequential methods with applications to clinical trials. Chapman & Hall / CRC, Boca Raton (FL) and London.
- [312] *Jöckel, K.-H.* (1982): Iterierte Aitken-Schätzer. Allgemeines Statistisches Archiv, vol. 66, p. 361-375.
- [313] *Jurečková, J.* (1971): Nonparametric estimate of regression coefficients. Annals of Mathematical Statistics, vol. 42, p. 1328-1338.
- [314] *Kakwani, N.C.* (1967): The unbiasedness of Zellner's seemingly unrelated regression equations estimators. Journal of the American Statistical Association, vol. 62, p. 141-142.
- [315] *Kalbfleisch, J.D. / Prentice, R.L.* (1980): The statistical analysis of failure time data. Wiley, New York.
- [316] *Kall, P.* (1982): Analysis für Ökonomen. Teubner, Stuttgart.
- [317] *Kall, P.* (1984): Lineare Algebra für Ökonomen. Teubner, Stuttgart.
- [318] *Kalman, R.E.* (1960): A new approach to linear filtering and prediction problems. Journal of Basic Engineering, vol. 82, p. 35-45.

- [319] Kaplan, E.L. / Meier, P. (1958): Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, vol. 53, p. 457-481.
- [320] Kempthorne, O. (1952): Design and analysis of experiments. Wiley, New York.
- [321] Kendall, M.G. (1973): Time series. Griffin, London.
- [322] Kendall, M.G. / Plackett, R.L. (1977, eds.): Studies in the history of statistics and probability II. Griffin, London.
- [323] Ketellapper, R.H. / Ronner, A.E. (1984): Are robust estimation methods useful in the structural errors-in-variables models? *Metrika*, vol. 31, p. 33-41.
- [324] Klapper, M. (1999): Combination of forecasts using rank-based techniques. *ASA Proceedings of the Business and Economic Statistics Session*, p. 197-202.
- [325] Klapper, M. (2000): The combination of forecasts using rank-based techniques. Eul-Verlag, Lohmar.
- [326] Klein, J.P. / Moeschberger, M.L. (1997): Survival analysis. Springer, New York.
- [327] Klösener, K.-H. (1978): Der Vorzeichentest bei Zulassen von Bindungen. Diplomarbeit, Universität Münster.
- [328] Klösener, K.-H. (1983): Präzisionsbestimmung von Meßinstrumenten bei zerstörenden Prüfungen. Dissertation, Universität Dortmund.
- [329] Knapp, G. (2002): Variance estimation in the error components regression model. *Communications in Statistics, Theory and Methods*, vol. 31, p. 1499-1514.
- [330] Knapp, G. / Biggerstaff, B.J. / Hartung, J. (2006): Assessing the amount of heterogeneity in random effects meta-analysis. *Biometrical Journal*, vol. 48, p. 271-285.
- [331] Knapp, G. / Hartung, J. (2003a): Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, vol. 22, p. 2693-2710.
- [332] Knapp, G. / Hartung, J. (2003b): Data Mining in einer epidemiologischen Studie zu Oberbauchbeschwerden. In: Christmann, A. / Weih, C. (eds.): *Data Mining und Statistik in Hochschule und Wirtschaft*. Shaker Verlag, Aachen, p. 79-88.
- [333] König, H. / Wolter, J. (1972): Einführung in die Spektralanalyse ökonomischer Zeitreihen. Hain, Meisenheim.
- [334] Koller, S. (1967): Mathematisch-statistische Grundlagen der Diagnostik. *Klinische Wochenschrift*, vol. 45, p. 1065-1072.
- [335] Koller, S. (1971): Mögliche Aussagen bei Fragen der statistischen Ursachenforschung. *Metrika*, vol. 17, p. 30-42.
- [336] Kotowski, P. (1961): Erhöhung der Flugsicherheit durch elektronische Geräte. In: *Nachrichtentechnische Fachberichte Band 24 (Zuverlässigkeit von Bauelementen)*, Vieweg & Sohn, p. 248-269.
- [337] Koul, H.L. (1971): Asymptotic behaviour of a class of confidence regions based on ranks in regression. *Annals of Mathematical Statistics*, vol. 42, p. 466-476.
- [338] Krafft, O. (1977): Statistische Experimente: Ihre Planung und Analyse. *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 57, p. 17-23.
- [339] Krafft, O. (1978): Lineare statistische Modelle und optimale Versuchspläne. Vandenhoeck & Ruprecht, Göttingen.
- [340] Krauth, J. (1980): Skalierungsprobleme, In: Köpcke, W. / Überla, K. (eds.): *Biometrie - heute und morgen. Internationales Biometrisches Kolloquium in München*, Springer, Berlin, p. 202-233.

- [341] Kreienbrock, L. / Schach, S. (2005): Epidemiologische Methoden. Spektrum Akademischer Verlag, Heidelberg.
- [342] Krüger, H.-P. / Lehmacher, W. / Wall, K.-D. (1981): The fourfold table up to n = 80 – Die Vierfeldertafel bis n = 80. Fischer, Stuttgart.
- [343] Kruskal, J.B. / Wish, M. (1978): Multidimensional scaling. Sage Publications, Beverly Hills.
- [344] Kühlmeyer, M. (1970): Die nichtzentrale t-Verteilung: Grundlagen und Anwendungen mit Beispielen. Lecture notes in operations research and mathematical systems, vol. 31, Springer, Berlin.
- [345] Kundert, K.R. / Bargmann, R.E. (1972): Tools of analysis for pattern recognition. Themis Report No. 22, Technical Report No. 89, The University of Georgia, Department of Computer Sciences, Athens (Georgia).
- [346] Kuznets, S.S. (1930): Secular movements in production and prices. Houghton Mifflin, Boston.
- [347] Lancaster, H.O. (1957): Some properties of the bivariate normal distribution considered in the form of a contingency table. Biometrika, vol. 44, p. 289-292.
- [348] Lee, C.-F. (2008, ed.): Advances in quantitative analysis of finance and accounting, vol. 6. World Scientific Publ. Co., Singapore.
- [349] Lehmacher, W. / Wassmer, G. (1999): Adaptive sample size calculations in group sequential trials. Biometrics, vol. 55, p. 1286-1290.
- [350] Lehmann, E.L. (1959): Testing statistical hypotheses. Wiley, New York.
- [351] Lehmann, E.L. (1975): Nonparametrics: Statistical methods based on ranks. McGraw-Hill, San Francisco.
- [352] Lehmann, E.L. / Casella, G. (1998): Theory of point estimation. Springer, New York.
- [353] Li, Y. / Shi, L. / Roth, H.D. (1994): The bias of the commonly-used estimate of variance in meta-analysis. Communications in Statistics, Theory and Methods, vol. 23, p. 1063-1085.
- [354] Lienert, G.A. (1969): Testaufbau und Testanalyse. Beltz, Weinheim.
- [355] Lienert, G.A. (1973): Verteilungsfreie Methoden in der Biostatistik, Bd. 1. Hain, Meisenheim.
- [356] Lienert, G.A. (1976): Verteilungsfreie Methoden in der Biostatistik, Tafelband. Hain, Meisenheim.
- [357] Lienert, G.A. (1978): Verteilungsfreie Methoden in der Biostatistik, Bd. 2. Hain, Meisenheim.
- [358] Liew, C.K. (1976): Inequality constrained least-squares estimation. Journal of the American Statistical Association, vol. 71, p. 746-751.
- [359] Lilliefors, H.W. (1967): On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association, vol. 62, p. 399-402.
- [360] Lilliefors, H.W. (1969): On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. Journal of the American Statistical Association, vol. 64, p. 387-389.
- [361] MacKinnon, J.G. (1994): Approximate asymptotic distribution functions for unit-root and cointegration tests. Journal of Business and Economic Statistics, vol. 12, p. 167-176.
- [362] Maddala, G.S. / Wu, S. (1999): A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and Statistics, Special Issue, p. 631-652.

- [363] *Makambi, K.H. / Argac, D. / Hartung, J.* (2000): On the meta-analysis of treatment differences in heteroscedastic samples. *Biometrical Letters*, vol. 37, p. 1-11.
- [364] *Makridakis, S. / Wheelwright, S.C. / Hyndman, R.J.* (1998): *Forecasting: Methods and applications*. Wiley, New York.
- [365] *Mandel, J. / Paule, R.C.* (1970): Interlaboratory evaluation of a material with unequal number of replicates. *Analytical Chemistry*, vol. 42, p. 1194-1197.
- [366] *Mann, H.B.* (1945): Nonparametric tests against trend. *Econometrica*, vol. 13, p. 245-259.
- [367] *Mantel, N.* (1966): Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Research*, vol. 50, p. 163-170.
- [368] *Mantel, N. / Haenszel, W.* (1959): Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, vol. 22, p. 719-748.
- [369] *Margolin, B.H. / Light, R.J.* (1974): An analysis of variance for categorical data II: Small sample comparisons with chisquare and other competitors. *Journal of the American Statistical Association*, vol. 69, p. 755-764.
- [370] *Markowitz, H.M.* (2009, ed.): Harry Markowitz - Selected works. World Scientific Publ. Co., Singapore.
- [371] *Meier, P.* (1953): Variance of a weighted mean. *Biometrics*, vol. 9, p. 59-73.
- [372] *Menges, G. / Skala, H.J.* (1973): *Grundriß der Statistik*, Bd. 2: Daten. Westdeutscher Verlag, Opladen.
- [373] *Miller, I. / Freund, J.E.* (1965): *Probability and statistics for engineers*. Prentice Hall, Englewood Cliffs (New Jersey).
- [374] *Miller, R.G.* (1966): *Simultaneous statistical inference*. McGraw-Hill, New York.
- [375] *Minamide, N. / Nakamura, K.* (1970): A restricted pseudoinverse and its applications to constrained minima. *SIAM Journal on Applied Mathematics*, vol. 19, p. 167-177.
- [376] *Mises, R.v.* (1931): *Wahrscheinlichkeitsrechnung*. Teubner, Leipzig.
- [377] *Mitchell, W.C.* (1941): *Konjunkturzyklus*. Buske, Leipzig.
- [378] *Moore, G.F. / Shannon, C.E.* (1956): Reliable circuits using less reliable relays. *Journal of the Franklin Institute*, vol. 262, p. 191-208 und 281-297.
- [379] *Morgenstern, D.* (1968): *Einführung in die Wahrscheinlichkeitsrechnung und mathematische Statistik*. Springer, Berlin.
- [380] *Morrison, D.F.* (1967): *Multivariate statistical methods*. McGraw-Hill, New York.
- [381] *Müller, J.* (1979): Beurteilungsstichproben in der Investitionsgüter-Marktforschung. *Marktforschung*, vol. 23, p. 114-117.
- [382] *Müller, P.H.* (1980): *Lexikon der Stochastik*. Akademie-Verlag, Berlin.
- [383] *Myers, J.L.* (1979): *Fundamentals of experimental design*. Allyn & Bacon, Boston.
- [384] *Naeve, P.* (1969): Spektralanalytische Methoden zur Analyse von ökonomischen Zeitreihen. Physica, Würzburg.
- [385] *Noelle, E.* (1963): Umfragen in der Massengesellschaft. Rowohlt, Reinbeck.
- [386] *Nourney, M. / Söll, H.* (1976): Analyse von Zeitreihen nach dem Berliner Verfahren Version 3. In: *Schäffer, K.A.* (ed.): Beiträge zur Zeitreihenanalyse. Son-

- derhefte zum Allgemeinen Statistischen Archiv, vol. 9, p. 129-152, Vandenhoeck & Ruprecht, Göttingen.
- [387] Nullau, B. / Heiler, S. / Wäsch, P. / Meisner, B. / Filip, D. (1969): Das Berliner Verfahren: Ein Beitrag zur Zeireihenanalyse. DIW-Beiträge zur Strukturforschung, Heft 7, Berlin.
- [388] O'Brien, P.C. / Fleming, T.R. (1979): A multiple testing procedure for clinical trials. *Biometrics*, vol. 35, p. 549-556.
- [389] Ohse, D. (1983): Mathematik für Wirtschaftswissenschaftler I. Vahlen, München.
- [390] Ohse, D. (1984): Mathematik für Wirtschaftswissenschaftler II. Vahlen, München.
- [391] Osborne, J.K. / Simons, G.D. (1970): A more precise estimate of the cost of not knowing the variance when making a fixed-width confidence interval for the mean. University of North Carolina at Chapel Hill, Institute of Statistics, Memo Series No. 677.
- [392] Ott, W. (1972, ed.): Handbuch der praktischen Marktforschung, Moderne Industrie, München.
- [393] Owen, D.B. (1962): Handbook of statistical tables. Addison Wesley, Reading (Mass.).
- [394] Patnaik, P.B. (1949): The non-central χ^2 - and F-distributions and their applications. *Biometrika*, vol. 36, p. 202-232.
- [395] Pearl, R. / Reed, L.J. (1920): The rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Science*, vol. 6, p. 275-288.
- [396] Pearson, E.S. / Hartley, H.O. (1970): Biometrika tables for statisticians I. Cambridge University Press, London.
- [397] Pearson, E. S. / Hartley, H.O. (1972): Biometrika tables for statisticians II. Cambridge University Press, London.
- [398] Pearson, E.S. / Kendall, M.G. (1970, eds.): Studies in the history of statistics and probability I. Griffin, London.
- [399] Pearson, K. (1904): Report on certain enteric fever inoculation statistics. *British Medical Journal*, vol. 3, p. 1243-1246.
- [400] Pearson, K. / Lee, A. (1903): On the laws of inheritance in man. *Biometrika*, vol. 2, p. 357-462.
- [401] Peto, R. / Peto, J. (1972): Asymptotically efficient rank invariant procedures (with discussion). *Journal of the Royal Statistical Society, Series A*, vol. 135, p. 185-206.
- [402] Pfanzagl, J. (1958): Über die Zerlegung statistischer Reihen. *Metrika*, vol. 1, p. 130-147.
- [403] Pfanzagl, J. (1971): Theory of measurement. Physica, Würzburg.
- [404] Pfanzagl, J. (1972): Allgemeine Methodenlehre der Statistik I. Sammlung Göschen-de Gruyter, Berlin.
- [405] Pfanzagl, J. (1978): Allgemeine Methodenlehre der Statistik II. Sammlung Göschen-de Gruyter, Berlin.
- [406] Pflaumer, P. (2009): Grundwissen Investitionsrechnung. epubli, Berlin.
- [407] Pflaumer, P. / Heine, B. / Hartung, J. (2001): Induktive Statistik. Lehr- und Handbuch der angewandten Statistik. Oldenbourg, München.
- [408] Pflaumer, P. / Heine, B. / Hartung, J. (2009): Deskriptive Statistik. Lehr- und Handbuch der angewandten Statistik. Oldenbourg, München, 4. Aufl.

- [409] *Pflaumer, P. / Kohler, H.-P.* (2003): Investitionsrechnung. Oldenbourg, München.
- [410] *Pill, J.* (1971): The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Economical Planning Science*, vol. 5, p. 57-71.
- [411] *Plachky, D. / Baringhaus, L. / Schmitz, N.* (1978): Stochastik I. Akademische Verlagsgesellschaft, Wiesbaden.
- [412] *Plackett, R.L.* (1975): The analysis of categorical data. Griffin, London.
- [413] *Pocock, S.J.* (1977): Group sequential methods in the design and analysis of clinical trials. *Biometrika*, vol. 64, p. 191-199.
- [414] *Pöthke, R. / Schmitz, N.* (1980): Fixed-width confidence intervals for the mean of a normal distribution. *Schriftenreihe Rechenzentrum Münster*, Münster.
- [415] *Preuss, H.* (1978): Zuverlässigkeit elektronischer Einrichtungen. VEB Verlag Technik, Berlin.
- [416] *Proschan, F. / Pyke, R.* (1967): Tests for monotone failure rate. *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, III, p. 293-312.
- [417] *Pukelsheim, F.* (1981): On the existence of unbiased non-negative estimates of variance-covariance components. *Annals of Statistics*, vol. 9, p. 293-299.
- [418] *Raj, D.* (1968): Sampling theory. McGraw-Hill, Bombay.
- [419] *Rao, C.R.* (1970): Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, vol. 65, p. 161-172.
- [420] *Rao, C. R.* (1973): Linear statistical inference and its applications. Wiley, New York.
- [421] *Rasch, D.* (1976a): Einführung in die mathematische Statistik I. VEB Deutscher Verlag der Wissenschaften, Berlin.
- [422] *Rasch, D.* (1976b): Einführung in die mathematische Statistik II. VEB Deutscher Verlag der Wissenschaften, Berlin.
- [423] *Rasch, D. / Herrendörfer, G. / Bock, J. / Busch, K.* (1978): Verfahrensbibliothek: Versuchsplanung und Auswertung. VEB Deutscher Verlag der Wissenschaften, Berlin.
- [424] *REFA* (1971): Methodenlehre des Arbeitsstudiums Teil 2: Datenermittlung. Hanser, München.
- [425] *Rhodes, E.C.* (1940): Population mathematics (Part I, II and III). *Journal of the Royal Statistical Society*, vol. 103, p. 61-89, 218-245, 362-387.
- [426] *Richard, F.W.* (1979): Natürliche und künstliche Alterung von Bauelementen. Der Elektroniker, vol. 1, p. 31-34.
- [427] *Richter, H.* (1966): Wahrscheinlichkeitstheorie. Springer, Berlin.
- [428] *Rinne, H.* (1972): Strategien der Instandhaltung (Ein Beitrag zur statistischen Theorie der Zuverlässigkeit). Hain, Meisenheim.
- [429] *Robins, J.M. / Breslow, N.E. / Greenland, S.* (1986): Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics*, vol. 42, p. 311-323.
- [430] *Romig, H.G.* (1953): 50-100 binomial tables. Wiley, New York.
- [431] *Rosenthal, H.* (1977): Konfidenzintervalle vorgegebener Länge. Diplomarbeit, Universität Münster.
- [432] *Ross, S.* (1970): Applied probability models with optimization applications. Holden-Day, San Francisco.

- [433] *Roß, H.* (1981): Zum asymptotischen Verhalten von linearen Schätzungen für Erwartungswerte und quadratischen Schätzungen für Varianzen und Kovarianzen in linearen Regressionsmodellen. Diplomarbeit, Universität Münster.
- [434] *Rothe, G.* (1981): Some properties of the asymptotic relative Pitman efficiency. *Annals of Statistics*, vol. 9, p. 663-669.
- [435] *Rothe, G.* (1983): Block rank statistics. *Metrika*, vol. 30, p. 73-83.
- [436] *Ryan, T.M.* (1978): Theory of portfolio selection. Macmillan, London.
- [437] *Ryerson, C.M. / Webster, S.L. / Albright, F.G.* (1967): Reliability notebook. Rome Air Force Development Center, TR-67-108, I und II.
- [438] *Sachs, L.* (1974): Angewandte Statistik. Springer, Berlin.
- [439] *Satterthwaite, F.E.* (1946): An approximative distribution of estimates of variance components. *Biometrics Bulletin*, vol. 2, p. 110-114.
- [440] *Savin, N.E. / White, K.J.* (1977): The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica*, vol. 45, p. 1989-1996.
- [441] *Schaafsma, A.H. / Willemze, F.G.* (1970): Moderne Qualitätskontrolle. Philips Fachbücher, Deutsche Philips GmbH, Hamburg.
- [442] *Schach, S. / Schach, E.* (1978): Pseudoauswahlverfahren bei Personengesamtheiten I: Namensstichproben. *Allgemeines Statistisches Archiv*, vol. 62, p. 379-396.
- [443] *Schach, S. / Schach, E.* (1979): Pseudoauswahlverfahren bei Personengesamtheiten II: Geburtstagsstichproben. *Allgemeines Statistisches Archiv*, vol. 63, p. 108-122.
- [444] *Schach, S. / Schäfer, T.* (1978): Regressions- und Varianzanalyse. Springer, Berlin.
- [445] *Schäfer, E.* (1980): Burn-In – was ist das? Qualität und Zuverlässigkeit, vol. 25, p. 296-304.
- [446] *Schäffer, K.A.* (1970): Beurteilung einiger herkömmlicher Methoden zur Analyse von ökonomischen Zeitreihen. In: *Wetzel, W.* (ed.): Neuere Entwicklungen auf dem Gebiet der Zeitreihenanalyse. Sonderhefte zum Allgemeinen Statistischen Archiv, vol. 1, p. 131-164, Vandenhoeck & Ruprecht, Göttingen.
- [447] *Schäffer, K.A.* (1976, ed.): Beiträge zur Zeitreihenanalyse. Sonderhefte zum Allgemeinen Statistischen Archiv, vol. 9, Vandenhoeck & Ruprecht, Göttingen.
- [448] *Schäffer, K.A.* (1978, ed.): Splinefunktionen in der Statistik. Sonderhefte zum Allgemeinen Statistischen Archiv, vol. 14, Vandenhoeck & Ruprecht, Göttingen.
- [449] *Schaich, E. / Köhle, D. / Schweitzer, W. / Wegner, F.* (1979): Statistik für Volkswirte, Betriebswirte und Soziologen I. Vahlen, München.
- [450] *Schaich, E. / Köhle, D. / Schweitzer, W. / Wegner, F.* (1982): Statistik für Volkswirte, Betriebswirte und Soziologen II. Vahlen, München.
- [451] *Scheffé, H.* (1959): The analysis of variance. Wiley, New York.
- [452] *Scheffé, H.* (1973): A statistical theory of calibration. *Annals of Statistics*, vol. 1, p. 1-37.
- [453] *Schlittgen, R. / Streitberg, B.H.J.* (1984): Zeitreihenanalyse. Oldenbourg, München.
- [454] *Schmetterer, L.* (1966): Einführung in die mathematische Statistik. Springer, Berlin.
- [455] *Schmitz, K.-J.* (1979): R-Schätzer im allgemeinen linearen Modell. Dissertation, Universität Bonn.

- [456] Schmitz, N. / Lehmann, F. (1976): Monte-Carlo Methoden I: Erzeugen und Testen von Zufallszahlen. Hain, Meisenheim.
- [457] Schneeberger, H. (1971): Optimierung in der Stichprobentheorie durch Schichtung und Aufteilung. Unternehmensforschung, vol. 15, p. 240-252.
- [458] Schneeberger, H. (1984): Some characteristics of Neyman's allocation. Jahrbücher für Nationalökonomie und Statistik, vol. 199, p. 351-365.
- [459] Schneeberger, H. / Drefahl, D. (1982): Gain in precision by optimum stratification and optimum allocation in dependence on the sampling fraction. Statistische Hefte, vol. 23, p. 228-236.
- [460] Schneeweiß, H. (1976): Consistent estimation of a regression with errors in the variables. Metrika, vol. 23, p. 101-115.
- [461] Schneider, D. / Pflaumer, P. (2001, eds.): Power-tools. Management-, Beratungs- und Controllinginstrumente. Gabler, Wiesbaden.
- [462] Schönfeld, P. (1971): Methoden der Ökonometrie II. Vahlen, München.
- [463] Schumacher, M. (1984): Two-sample tests of Cramer-von Mises- and Kolmogorov-Smirnov-type for randomly censored data. International Statistical Review, vol. 52, p. 263-281.
- [464] Schwartz, D. / Flament, R. / Lellouch, J. (1980): Clinical trials. Academic Press, London.
- [465] Schwarz, H. (1975): Stichprobentheorie. Oldenbourg, München.
- [466] Searle, S.R. (1971): Linear models. Wiley, New York.
- [467] Seber, G.A.F. (1977): Linear regression analysis. Wiley, New York.
- [468] Seelbinder, B.M. (1953): On Stein's two-stage sampling scheme. Annals of Mathematical Statistics, vol. 24, p. 640-649.
- [469] Shapiro, S.S. / Wilk, M.B. / Chen, M.J. (1968): A comparative study of various tests for normality. Journal of the American Statistical Association, vol. 63, p. 1343-1372.
- [470] Shiskin, J. / Young, A.M. / Musgrave, J.C. (1967): The X-11 variant of the census method II: Seasonal adjustment program. Technical paper No. 15 (revised), U.S. Bureau of the Census, Washington D.C.
- [471] Sidik, K. / Jonkman, J.N. (2002): A simple confidence interval for meta-analysis. Statistics in Medicine, vol. 21, p. 3153-3159.
- [472] Sidik, K. / Jonkman J.N. (2005): A note on variance estimation in random effect meta-regression. Journal of Biopharmaceutical Statistics, vol. 15, p. 823-838.
- [473] Silcocks, P. (2005): An easy approach to the Robins-Breslow-Greenland variance estimator. Epidemiologic Perspectives & Innovations, vol. 2, p. 9(1-6).
- [474] Simonaitis, D.F. / Anderson, R.T. / Kaye, M.P. (1972): Reliability evaluation of a heart assist system. Proceedings of the 1972 Annual Reliability and Maintainability Symposium, New York, p. 233-241.
- [475] Singpurwalla, N.D. / Castellino, V.C. / Gozdschen, D.Y. (1975): Inference from accelerated life tests using Eyring type re-parameterizations. Naval Research Logistics Quarterly, vol. 22, p. 289-296.
- [476] Slakter, M.J. (1965): A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit-tests with respect to validity. Journal of the American Statistical Association, vol. 60, p. 854-858.
- [477] Slutsky, E. (1937): The summation of random causes as the source of cyclic processes. Econometrica, vol. 5, p. 105-146.
- [478] Snedecor, G.W. / Cochran, W.G. (1967): Statistical methods. The Iowa States University Press, Ames (IA).

- [479] *Stefansky, W.* (1971): Rejecting outliers by maximum normed residual. *Annals of Mathematical Statistics*, vol. 42, p. 35-45.
- [480] *Stefansky, W.* (1972): Rejecting outliers in factorial designs. *Technometrics*, vol. 14, p. 469-479.
- [481] *Stein, Ch.* (1945): A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics*, vol. 16, p. 243-258.
- [482] *Stenger, H.* (1971): *Stichprobentheorie*. Physica, Würzburg.
- [483] *Stier, W.* (1980): *Verfahren zur Analyse saisonaler Schwankungen in ökonomischen Zeitreihen*. Springer, Berlin.
- [484] *Stoer, J.* (1971): On the numerical solution of constrained least-squares problems. *SIAM Journal on Numerical Analysis*, vol. 8, p. 382-411.
- [485] *Stoer, J.* (1979): *Einführung in die numerische Mathematik*. Springer, Berlin.
- [486] *Störmer, H.* (1961): Über einen statistischen Test zur Bestimmung der Parameter einer Lebensdauerverteilung. *Metrika*, vol. 4, p. 63-77.
- [487] *Störmer, H.* (1970): *Mathematische Theorie der Zuverlässigkeit, Einführung und Anwendungen*. Oldenbourg, München.
- [488] *Stouffer, S.A. / Suchman, E.A. / DeVinney, L.C. / Star, S.A. / Williams, R.M. jr.* (1949): *The american soldier, Volume I. Adjustment during army life*. Princeton University Press, Princeton (NJ).
- [489] *Sweeting, M.J. / Sutton, A.J. / Lambert, P.C.* (2004): What to add to nothing? Use or avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*, vol. 23, p. 1351-1375.
- [490] *Swoboda, H.* (1971): *Knaurs Buch der modernen Statistik*. Droemer Knaur, München.
- [491] *Tarone, R.E. / Ware, J.* (1977): On distribution-free tests for equality of survival distributions. *Biometrika*, vol. 64, p. 156-160.
- [492] *Thomas, R.E.* (1965): When is a life test truly accelerated? *Electronic Design*, p. 64-70.
- [493] *Thomas, R.E. / Gorton H. C.* (1963): Research towards a physics of aging of electronic component parts. *Proceedings of the second Annual Symposium on the Physics of Failure in Electronics*, p. 25-26.
- [494] *Thompson, M.* (1966): Lower confidence limits and a test of hypotheses for system availability. *IEEE Transactions on Reliability*, vol. R-15, p. 32-36.
- [495] *Thompson, W.A. / Willke, T.A.* (1963): On an extreme rank sum test for outliers. *Biometrika*, vol. 50, p. 375-383.
- [496] *Tintner, G.* (1958): Eine neue Methode für die Schätzung der logistischen Funktion. *Metrika*, vol. 1, p. 154-157.
- [497] *Tintner, G.* (1960): *Handbuch der Ökonometrie*. Springer, Berlin.
- [498] *Tomášek, K.* (1971): Zur Problematik der zeitraffenden Zuverlässigkeitssprüfung an Si-Transistoren. *Nachrichtentechnische Zeitschrift*, vol. 1, p. 43-48.
- [499] *Trenkler, G.* (1981): Biased estimators in the linear regression model. Oelgeschlager, Gunn & Hain, Meisenheim.
- [500] *Troschke, S.-O.* (2002): Enhanced approaches to the combination of forecasts. Eul-Verlag, Lohmar.
- [501] *Tukey, J.W.* (1953): The problem of multiple comparisons. Unveröffentlichtes Manuskript, zitiert in Hsu (1989).
- [502] *Tukey, J.W.* (1977): *Exploratory data analysis*. Addison-Wesley, Reading (Mass.).

- [503] *Überla, K.* (1971): Faktorenanalyse. Springer, Berlin.
- [504] *Uhlmann, W.* (1966): Statistische Qualitätskontrolle. Teubner, Stuttgart, 2. Auflage 1982.
- [505] *Uhlmann, W.* (1969): Kostenoptimale Prüfpläne. Physica, Würzburg.
- [506] *Upton, G.J.G.* (1978): The analysis of cross-tabulated data. Wiley, Chichester.
- [507] *Velleman, P.D. / Hoaglin, D.C.* (1981): Applications, basics and computing of exploratory data analysis. Duxbury Press, Boston.
- [508] *Verhulst, P.-F.* (1938): Notice sur la loi que la population sent dans son accroissement. Correspondance Mathématique et Physique publiée par A. Quetelet, vol. 10, p. 113-121.
- [509] *Victor, N.* (1987): On clinically relevant differences and shifted nullhypotheses. Methods of Information in Medicine, vol. 26, p. 109-116.
- [510] *Viechtbauer, W.* (2007): Confidence intervals for the amount of heterogeneity in meta-analysis. Statistics in Medicine, vol. 26, p. 37-52.
- [511] *Vinče, I.* (1971): Mathematische Statistik mit industriellen Anwendungen. Akadémiai Kiadó, Budapest.
- [512] *Voet, B.* (1983): Schätz- und Testverfahren bei balancierten Varianzkomponentenmodellen. Dissertation, Universität Dortmund.
- [513] *Voet, B.* (1985): Iterierte Schätzer im allgemeinen linearen Modell. Zeitschrift für Angewandte Mathematik und Mechanik, vol. 65, p. 327-328.
- [514] *Vogel, W.* (1970): Wahrscheinlichkeitstheorie. Vandenhoeck & Ruprecht, Göttingen.
- [515] *Waerden, B.L.v.d.* (1960): Sampling inspection as a minimum loss problem. Annals of Mathematical Statistics, vol. 31, p. 369-384.
- [516] *Waerden, B.L.v.d.* (1965): Sequentielle Qualitätskontrolle als Minimumproblem. Zeitschrift für Wahrscheinlichkeitstheorie, vol. 4, p. 187-202.
- [517] *Waerden, B.L.v.d.* (1971): Mathematische Statistik. Springer, Berlin.
- [518] *Waerden, B.L.v.d.* (1975): Mathematik für Naturwissenschaftler. BI Hochschultaschenbücher, Mannheim.
- [519] *Wagner, G.* (1959): Abnahme mit Stichproben. Ausschuß für wirtschaftliche Fertigung e.V., Berlin.
- [520] *Wagner, S. / Warmuth-Metz, M. / Emser, A. / Gnekow, A.K. / Strater, R. / Rutkowski, S. / Jorch, N. / Schmid, H.J. / Berthold, F. / Graf, N. / Kortmann, R.D. / Pietsch, T. / Sorensen, N. / Peters, O. / Wolff, J.E.* (2006): Treatment options in childhood pontine gliomas. Journal of Neurooncology, vol. 79, p. 281-287.
- [521] *Wald, A.* (1940): A note on the analysis of variance with unequal class frequencies. Annals of Mathematical Statistics, vol. 11, p. 96-100.
- [522] *Wald, A.* (1945): Sequential tests for statistical hypotheses. Annals of Mathematical Statistics, vol. 16, p. 117-186.
- [523] *Wald, A.* (1947): Sequential analysis. Wiley, New York.
- [524] *Wald, A. / Wolfowitz, J.* (1945): Sampling inspection plans for continuous production which insure a prescribed limit on the outgoing quality. Annals of Mathematical Statistics, vol. 16, p. 30-49.
- [525] *Wallis, W.A. / Moore, G.H.* (1941): A significance test for time series analysis. Journal of the American Statistical Association, vol. 36, p. 401-409.
- [526] *Weber, E.* (1974): Einführung in die Faktorenanalyse. VEB G. Fischer-Verlag, Jena.
- [527] *Weber, M. / Argac, D. / Hallier, E. / Hartung, J.* (2001): Qualitätssicherung bei Faserstaubmessungen. Teil 2: Göttinger Ringversuch zur Faseranalytik. Zen-

- tralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie, mit Beiträgen zur Umweltmedizin, vol. 51, p. 13-24.
- [528] *Weibull, W.* (1939): A statistical theory of the strength of materials. Ing. Vetenskaps Akad. Handl., vol. 151, p. 1-45.
- [529] *Welch, B.L.* (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, vol. 38, p. 330-336.
- [530] *Wermuth, N.* (1979): Zusammenhangsanalysen medizinischer Daten. Springer, Berlin.
- [531] *Werner, H.J.* (1983): Zur besten linearen minimalverzerrten Schätzung im singulären restringierten linearen Modell. *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 63, p. 423-425.
- [532] *Werner, K.* (1980): Verbindungstechnik bei Halbleitern: mehr Zuverlässigkeit durch fehlerspezifische Prüfungen und technische Maßnahmen. *VDI-Nachrichten*, vol. 14, p. 6-7.
- [533] *Wetzel, W.* (1970, ed.): Neuere Entwicklungen auf dem Gebiet der Zeitreihenanalyse. Sonderhefte zum Allgemeinen Statistischen Archiv, vol. 1, Vandenhoeck & Rupprecht, Göttingen.
- [534] *Wetzel, W.* (1971): Statistische Grundausbildung für Wirtschaftswissenschaftler I. Springer, Berlin.
- [535] *Wetzel, W.* (1973): Statistische Grundausbildung für Wirtschaftswissenschaftler II. Springer, Berlin.
- [536] *Whitehead, A.* (1997): A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Statistics in Medicine*, vol. 16, p. 2901-2913.
- [537] *Whitehead, A.* (2001): Meta-analysis of controlled clinical trials. Wiley, Chichester.
- [538] *Wienhold, P.* (1978): Lagerinventur auf Stichprobenbasis. Diplomarbeit, Universität Dortmund.
- [539] *Williams, D.A.* (1973): Letter to the editors. *Applied Statistics*, vol. 22, p. 407-408.
- [540] *Wilson, E.B. / Hilferty, M.M.* (1931): The distribution of chi-square. *Proceedings of the National Academy of Science*, vol. 17, p. 684-688.
- [541] *Winkler, W.* (1951): Grundfragen der Ökonometrie. Springer, Wien.
- [542] *Witte, S. / Victor, N.* (2004): Some problems with the investigation of noninferiority in meta-analysis. *Methods of Information in Medicine*, vol. 43, p. 470-474.
- [543] *Witting, H.* (1979): Mathematische Statistik. Teubner, Stuttgart.
- [544] *Witting, H. / Nölle, G.* (1970): Angewandte Mathematische Statistik. Teubner, Stuttgart.
- [545] *Woll, A.* (2008): Wirtschaftslexikon. Oldenbourg, München, 10. Aufl.
- [546] *Yarnold, J.K.* (1970): The minimum expectation in χ^2 goodness of fit tests and the accuracy of approximations of the null distributions. *Journal of the American Statistical Association*, vol. 65, p. 864-886.
- [547] *Yates, F.* (1981): Sampling methods for censuses and surveys. Griffin, London.
- [548] *Yates, F. / Cochran, W.G.* (1938): The analysis of groups of experiments. *Journal of Agricultural Science*, vol. 28, p. 556-580.
- [549] *Youden, W.J.* (1960): The sample, the procedure, and the laboratory. *Analytical Chemistry*, vol. 13, p. 23A-37A.
- [550] *Youden, W.J.* (1972): Enduring values. *Technometrics*, vol. 14, p. 1-11.

- [551] *Yudkin, P.L. / Ellison, G.W. / Ghezzi, A. / Goodkin, D.E. / Hughes, R.A.C. / McPherson, K. / Merton, J. / Milanese, C.* (1991): Overview of azathioprine treatment in multiple sclerosis. *Lancet*, vol. 338, p. 1051-1055.
- [552] *Yusuf, S. / Peto, R. / Lewis, J. / Collins, R. / Sleight, P.* (1985): Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Disease*, vol. 27, p. 335-371.
- [553] *Ziegler, S. / Koch, A. / Victor, N.* (2001): Deficits and remedy of the standard random effects methods in meta-analysis. *Methods of Information in Medicine*, vol. 40, p. 148-155.

6. Literaturhinweise zu den einzelnen Kapiteln

Die Nummern beziehen sich auf das vorstehende Literaturverzeichnis in Abschnitt 5.

Einführung und Grundlagen

[9], [14], [191], [322], [338], [398], [490]

Kapitel I

[108], [112], [117], [223], [289], [404], [408], [449], [534]

Kapitel II

[16], [17], [108], [116], [125], [171], [224], [295], [379], [407], [411], [514]

Kapitel III

[108], [125], [224], [351], [352], [379], [407], [411], [421], [422], [454], [517], [543], [544]

Kapitel IV

[68], [76], [108], [156], [157], [218], [224], [229], [248], [249], [254], [300], [309], [327], [350], [355], [356], [357], [376], [379], [405], [511], [522], [523], [525], [544]

Kapitel V

Abschnitte 1-3

[36], [63], [277], [341], [392], [404], [418], [457], [458], [465], [482], [547]

Abschnitt 4

[13], [14], [85], [149], [150], [195], [228], [229], [262], [263], [405], [422], [549], [550]

Abschnitt 5

[10], [65], [69], [123], [124], [145], [146], [197], [199], [215], [242], [269], [273], [320], [338], [339], [506], [516], [517]

Abschnitt 6

[5], [79], [197], [199], [230], [236], [242], [247], [248], [252], [253], [256], [264], [274], [332], [334], [335], [464], [530]

Abschnitt 7

[107], [114], [151], [215], [217], [218], [354]

Kapitel VI

[87], [89], [102], [377], [504], [505], [524]

Kapitel VII

[30], [108], [119], [120], [141], [153], [154], [224], [229], [342], [412], [515], [516]

Kapitel VIII

[68], [108], [197], [199], [218], [224], [242], [247], [248], [249], [250], [251], [252], [253], [300], [350], [355], [356], [357], [405], [463]

Kapitel IX

[108], [113], [146], [215], [218], [224], [241], [420]

Kapitel X

[55], [92], [108], [113], [145], [146], [185], [215], [218], [224], [269], [329], [444], [467]

Kapitel XI

[10], [65], [68], [69], [101], [103], [123], [124], [145], [146], [187], [215], [220], [232], [233], [245], [266], [268], [270], [273], [300], [301], [320], [339], [355], [356], [357], [374], [383], [417], [435], [444], [451], [466]

Kapitel XII

[11], [40], [46], [52], [53], [127], [128], [129], [148], [159], [161], [223], [288], [321], [333], [384], [387], [453], [462], [483], [497]

Kapitel XIII

[20], [23], [27], [71], [126], [130], [177], [209], [216], [226], [296], [299], [315], [326], [415], [428]

Kapitel XIV

[12], [25], [51], [59], [108], [223], [287], [297], [308], [408], [502], [507]

Kapitel XV

[61], [62], [79], [91], [192], [193], [199], [210], [221], [234], [235], [237], [238], [241], [244], [245], [249], [255], [258], [284], [330], [331], [371], [537], [548], [553]

Anhang

[26], [37], [86], [316], [317], [389], [390], [393], [396], [397], [518]

7. Sach- und Namensregister

A – Alt
Additivität
-, endliche ... 94
-, σ ... 94
adjacent value 835
Adjustieren von Kontingenztafeln 501ff
adjustierter Varianzschätzer 908
Adler, G. 1048
Ähnlichkeitsskala 379
Änderungsausfall 746
äußerer Zaun 835
Ahrens, J.H. 194, 1048, 1051
Aitken, A.C. 608, 736, 1048
Aitkenschätzung 600, 736
Akaike, H. 687, 1048
Aktien-Portfolio 897, 965, 1004ff
-, automatische Überwachung eines ... 965
Albright, F.G. 1049
aleatorische Zuordnung 370
Allescher, H.-D. 917, 1048
allgemeine Dreiecksverteilung 196
-, Erwartungswert der ... 196
-, Varianz der ... 196
allometrische Funktion 642, 654ff
-, Prognose bei ... 659f
-, Schätzen der Parameter der ... 654ff
α -Fraktil 114
α -Gastwirth-Cohen-Mittel 882f
α -getrimmtes Mittel 880f
-, Bruchpunkt des ... 881
-, Einflußkurve des ... 881
α -Quantil 34, 114f, 118, 235ff, 826
-, ... bei gruppierten Daten 34
-, ... der χ^2 -Verteilung 153, 1029f
-, ... der F-Verteilung 157, 1031ff
-, ... der Normalverteilung 147
-, ... der Standardnormalverteilung 147, 1027
-, ... der studentisierten Spannweite 1038f
-, ... der t-Verteilung 154, 1028
-, nichtparametrische Konfidenzintervalle und Tests für ... 235ff
α -winsorisiertes Mittel 880ff
-, Bruchpunkt des ... 882
Al-Quorain 889, 918
Alternative 134
Alternativhypothese 133f
alternierende Zuordnung 370
altersabhängige Erneuerung 807f
Altersersatzstrategie 780, 807f, 823
Altmeyer, P. 1061

Am – As

- Amlodipine 995
- Amplitude einer Cosinusschwingung 699
- Amplitudendiagramm 703ff
 - , ... eines linearen Filters 703
- Analyse
 - , harmonische ... 639, 641, 717ff
 - , sequentielle ... 373
- Anderson, O. 8, 668, 1048
- Anderson, R.L. 363, 1048
- Anderson, R.T. 1070
- Anderson, T.W. 640, 1048
- Andrews, D.F. 827, 860, 872, 1048
- Andrews-Plot 860
- Andrews' wave 872ff
 - , ϕ -Funktion von ... 872
 - , ρ -Funktion von ... 872
- Angelpunkt 828
- Angina Pectoris 995f
- Anlage
 - , ...fenster 1008
 - , ... mit geringster Volatilität 1009
 - , ... mit minimalem, geringstem Risiko 899, 1006, 1009
 - , pareto-optimale ... 897, 1005ff
 - , stabile ... 897, 1008
- Annahmebereich eines Tests 134
- Annahmegrenze 384ff, 388f, 392, 397, 399f
- Annahmekennlinie 381, 385
- Annahmezahl 384ff, 391, 395
- ANOVA (s. auch Varianzanalyse)
 - , erweitertes ...-Modell 1013
 - , ...-Schätzer 630f, 633
 - , ...-Tafel 611ff
- Anpassungsgüte 686
- Anpassungstest 139, 825, 847
 - , χ^2 -... 182f
 - , ... auf Exponentialverteilung 225f
 - , ... auf Gleichverteilung 197f
 - , ... auf Normalverteilung 182f, 187
 - , ... auf Poissonverteilung 216f
 - , Kolmogoroff-Smirnov-... 183f
 - , ... auf Exponentialverteilung 226ff
 - , ... auf Normalverteilung 183ff
- Ansari 526ff, 1024
- Ansari-Bradley-Freund-Test 526ff
- Anscombe, F.J. 914, 920, 1048
- Anteilschätzer 293ff
 - , Varianz des ... 294
- Anteilswert 207 f
- Approximation
 - , ... der Binomialverteilung 201f
 - , ... der χ^2 -Verteilung 153
 - , ... der F-Verteilung 157
 - , ... der hypergeometrischen Verteilung 208
 - , ... der Poissonverteilung 213
 - , ... der t-Verteilung 154
 - , Hastings-... 1026f, 1040
 - , Peizer-Pratt-... 1028
 - , Wilson-Hilferty-... 955, 1030
- a-priori-Information 300
- Arbuthnot, J. 11
- Arcus-Sinus-Formel 420
- Arcus-Sinus-Transformation 352, 913
- AR(1)-Prozeß 644, 678ff
- Argac, D. 894f, 921, 945, 998, 1013, 1057, 1061, 1066, 1072
- ARIMA-Modell 684f
 - , multiplikativ saisonales ... 685
- ARIMA-Prozeß 684f
- Aristoteles 8f
- Arithmetik
 - , politische ... 12
- arithmetisches Mittel 11, 31f, 35ff, 44, 83ff
 - , ... bei gruppierten Daten 32
 - , gewogenes ... 32
 - , Minimumeigenschaft des ... 41, 324
 - , Robustheit des ... 861ff
- ARMA-Prozeß (s. auch gemischter Prozeß) 678, 682ff
- AR-Prozeß (s. auch autoregressiver Prozeß) 678ff, 962
- Arrhenius 745, 796, 799
- , ... Modell 745, 796, 799
- Arten von Hypothesen 138f
- ASA-II 639, 666
- Asbest-Belastung 896, 921ff
- Asbest-Faser-Zählung 896, 921ff
- Assoziationskoeffizient 12, 82f
 - , Yulescher ... 82f
- Assoziationsmaß 442ff, 545
 - , ... für die 2 x 2-Tafel 442ff
 - , ... für die r x s-Tafel 450ff
 - , Varianzen der ... 452
 - , Kombination von ... 906
 - , ... von Cramer 452
 - , ... von Tschuprow 451f
 - , Yulesches ... 443ff
 - , Eigenschaften des ... 446
 - , Konfidenzintervalle für das ... 444ff
 - , Schätzer für das ... 444ff
- Assoziiertheit
 - , stochastiche ... 759
- Asthma Bronchiale 979f, 982, 987, 1016ff
- Astronomische Einheit 887f

Auffüllen der Reihe der gleitenden Durchschnitte 664, 726f
 aufgehängtes Wurzeldiagramm 845
 Aufstellung des loglinearen Modells
 -, ... für 2×2 -Tafeln 425ff
 -, ... für $r \times s$ -Tafeln 429
 Aufteilung
 -, ... durch Schichtenbildung nach Auswahl der Stichprobe bei geschichteter Zufallsauswahl 286
 -, Genauigkeitsvergleiche für verschiedene ... bei geschichteter Zufallsauswahl 287f
 -, ... nach Auswahl der Stichprobe 286
 --, geschichtete Zufallsauswahl mit ... 286
 ---, Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit bei ... 286
 ----, Varianz des ... 286
 -----, Schätzer für die ... 286
 -, Neyman-Tschuprow-... bei geschichteter Zufallsauswahl 282ff, 313
 -, optimale ... bei geschichteter Zufallsauswahl 282ff, 313
 -, proportionale ... bei geschichteter Zufallsauswahl 282, 285, 287f, 313
 Augustinus 9
 Ausfallquote 795
 Ausfallrate 105, 111f, 119, 123f, 126, 218ff, 224, 231ff, 745, 775ff, 782, 788ff, 793ff, 809ff, 818, 824, 924
 -, ... der Erlang-n-Verteilung 235
 -, ... der Exponentialverteilung 220
 -, ... der Hjorth- (IDB-) Verteilung 232ff
 -, ... der Weibullverteilung 231
 -, empirische ... 19, 70ff, 795
 -, mittlere kumulierte ... 783
 -, Wertungsfaktoren für im Labor ermittelte ... 804
 Ausfallwahrscheinlichkeit eines wartbaren Systems 806
 Ausgangsereignis bei der Konstruktion von Fehlerbäumen 766
 Ausreißer 33, 41, 269, 321, 344ff, 586, 835
 -, Ermittlung von ... mittels Residualanalyse 585f
 Ausreißerempfindlichkeit 863
 Ausreißertest 343ff
 -, ... auf ein Ausreißerpaar 347ff
 -, ... von David-Hartley-Pearson 345f
 -, ... von Dixon 346f
 -, ... von Grubbs 344ff
 Ausschubanteil bei der Produktion 381ff, 385, 388f, 391, 394ff, 399

Au – Bah

-, durchschnittlich ausgelieferter ... 406
 Auswahl
 -, ... aufs Geratewohl 308, 316
 -, ... nach dem Konzentrationsprinzip 317
 -, ... nach Gudtünken 316
 -, ... nach Schichten 312, 316
 -, Quoten... 316, 318, 320
 -, systematische ... 308
 -, typische ... 316f
 -, ... von Versuchspersonen bei klinischen Versuchen 367ff
 Auswahlatz 275, 285, 287, 293, 305
 Auswahlverfahren
 -, ... bei Klumpenstichproben 288ff
 --, einstufiges ... 288f
 --, mehrstufiges ... 288f
 -, Pseudo... 309
 -, zweiphasiges ... 305
 Auswahlwahrscheinlichkeit
 -, Stichproben mit ungleichen ... 297ff
 Autokorrelation 675ff
 -, ... eines ARMA-Prozesses 683
 -, ... eines AR-Prozesses 679
 -, ... eines MA-Prozesses 681
 -, empirische ... 675f
 -, partielle ... 677f
 -, theoretische ... 677
 Autokovarianz 675f
 -, ... eines ARMA-Prozesses 682f
 -, ... eines AR-Prozesses 679
 -, ... eines MA-Prozesses 681
 -, empirische ... 675f
 -, theoretische ... 676
 autoregressiver Prozeß 639, 678ff
 -, Autokorrelationen eines ... 679
 -, Autokovarianzen eines ... 679
 -, ... der Ordnung p 679
 -, MA-Darstellung eines ... 682
 -, Ordnung eines ... 686
 -, partielle Autokorrelationen eines ... 679
 -, Spektraldichte eines ... 703f
 autoregressives Regressionsmodell 736, 744
 -, Fehlerprozeß beim ... 744
 Autosuggestion 371ff
 availability 820
 average outgoing quality 389
 average outgoing quality limit 389
 Axiome
 -, Kolmogoroffsche ... 94
 Azathioprine 902f

Bahrinhaus, L. 1068

Bai – Bez

- Baier, D. 1060
- Baird, D.C. 1048
- Bamberg, G. 95, 967, 1048
- Bancroft, T.A. 363, 1048
- Bandbreite eines Fensters 712
- Bandelow, C. 1048
- Bandpaß 674, 705ff
 - , idealer ... 706
 - , Bandsperrre 674, 705ff
 - , ideale ... 706
- Bargmann, R.E. 378, 1048, 1065
- Barlow, R.E. 746, 759, 763, 778, 783f, 788, 808, 823, 1048f
- Barnett, V. 827, 1049
- Bartholomew, D.J. 1048
- Bartlett 617f, 712ff
 - , ...-Fenster 712, 714
 - , Bandbreite des ... 712
 - , ...-Test 617f
- Basisperiode 57ff, 62f, 65ff, 70
- Basisquartal 58f
- Baum 860
- Baur, F. 95, 967, 1048
- Bayes, T. 10, 102f
- Bayessche Formel 102f
- Beck, G. 346ff, 1054
- Beckmann, M.J. 1047, 1049
- Bedeutung der Normalverteilung 143
- bedingte Gleichverteilungshypothese 434
- bedingte relative Häufigkeit 99
- bedingter Least-Squares-Ansatz 688
- bedingte Überlebenswahrscheinlichkeit 218f, 230, 233, 235, 775f, 779
 - , ... bei der Erlang-n-Verteilung 235
 - , ... bei der Exponentialverteilung 219
 - , ... bei der Hjorth- (IDB) Verteilung 233
 - , ... bei der Weibullverteilung 230
- bedingte Wahrscheinlichkeit 98ff, 148
 - , Multiplikationssatz für ... 99f
- Befragung
 - , Interview... 310f
 - , postalische ... 310f
- begleitende Variable 365
- Begriff der Zeitreihe 637ff
- Behandlung 356f, 363ff
 - , differentielle ... 358
 - , Schein... (s. auch Placebo) 371ff
- Behandlungseffekt 514, 619f, 628
 - , additiver ... 514
- Behrens 510, 983
- Behrens-Fisher-Problem 510, 983
- Bechelt, F. 808, 1049
- Bennett, B.M. 1053
- Beobachtung
 - , direkte ... 326
 - , diskrete ... 407ff, 913ff
 - , indirekte ... 326
- Beobachtungsgleichung 640
- Beobachtungsreihe 20ff, 826
- Berechnung von M-Schätzern 874ff
- Bereinigung von saisonaler Komponente 639, 660ff, 722ff
- Bereitschaftsstrategie 823
- Berichtsperiode 62ff
- Berkson, J. 604f, 1049
- Berkson-Modell 604f
- Berlin, J.A. 1049
- Berliner Verfahren 639, 666, 722ff
 - , Gewichtsmatrix des ... 724
 - , Gewichtsmatrix für die glatte Komponente beim ... 725
 - , Gewichtsmatrix für die saisonale Komponente beim ... 725
 - , Grundmodell des ... 723ff
- Bernoulli, J. 11, 200
- Bernoulli-Experiment 200
- Bernoulli-Verteilung 200
- Berry 122
- Berry-Esséen
 - , Satz von ... 122
- Berthold, F. 1072
- Beschädigungsmodell
 - , kumulatives ... 783
- beschleunigte Lebensdauerprüfung 745, 794ff
- Beschleunigungsfaktor 798f
- beschreibende Statistik 2, 12, 15ff
- Bestandsmasse 56
- Bestimmtheitsmaß 565f, 579f, 593ff, 646f
 - , ... der linearen Regression 579
 - , ... der multiplen Regression 595ff
 - , ... einer quadratischen Regressionsfunktion 593
 - , multiples ... 565ff, 595ff, 802
- Bestimmung von Schätzern
 - , Methoden zur ... 125ff
- Betafunktion 11, 156
 - , Eulersche ... 156
 - , vollständige ... 11
- Betaverteilung 11
- Betrag einer komplexen Zahl 700
- Beurteilungskriterien für Tests 137
- Beurteilungsstichprobe 315f, 318f
- Bewegungsmassen 56
- Bewertungssystem 376
- Beziehungszahl 55ff
- Beziehung zwischen
 - , ... χ^2 - und Poissonverteilung 213

-, ... F- und Binomialverteilung 201
 -, ... Klassen von Lebensdauerverteilungen 786f
 Bias eines Schätzers 125f, 300, 322
 Bildfrage 310
 Bickel, P.J. 778, 1048f
 Bienaymé 117
 -, Gleichung von ... 117
 Biggerstaff, B.J. 998, 1064
 binäre Daten 407ff, 913ff
 -, Test auf Homogenität für ... 1013
 Bindung 139f, 242, 244, 246, 340, 515,
 527, 529, 541f, 554, 559, 613ff, 622f
 Bindungsgruppe 139
 Binomialverteilung 11, 110, 115, 117,
 122f, 200ff, 208, 210, 213, 222f, 242f,
 251ff, 496ff
 -, Approximation der ... 201f
 -, Beziehung zwischen F-Verteilung und
 ... 201
 -, Erwartungswert der ... 200
 -, Konfidenzintervall für den Parameter p
 einer ... 203ff
 -, Meta-Analyse bei ... 913ff
 -, Schätzer für den Parameter p einer ...
 202
 -, sequentieller Quotiententest für den
 Parameter p einer ... 252ff
 -, Test über den Parameter p einer ... 205f
 --, benötigter Stichprobenumfang beim ...
 206f
 -, Varianz der ... 200
 -, Vergleich der Parameter mehrerer ...
 496ff
 BIP 893f, 945ff
 -, ...-Prognose 893f, 945ff
 -, ...-Vorhersage 893f
 bi-partielle Korrelation 545, 563f
 -, Schätzer für die ... 563
 -, Test über die ... 564
 Bi-Plot 861
 -, ...-Sonne 861
 Bishop, Y.M.M. 1049
 bivariate Normalverteilung 120
 Blackman, R.B. 711, 1049
 Bleymüller, J. 1049
 Blindversuch 343, 362, 364f, 372
 -, Doppel... 362, 365, 372f
 Blobel, B. 1057, 1059
 Block 356ff, 370, 619ff
 Blockbildung 355
 Blockeffekt 620
 Blockerneuerungsstrategie 807
 Blockexperiment 356ff, 609, 619ff
 -, einfaches ... 357, 609, 619ff

Bez – Bru

Blockplan
 -, unvollständiger ... 358, 361
 Blum, E. 642, 793, 1049
 Blumen 860
 Blumer 513
 Bock, J. 1016, 1049, 1068
 Boddy, R. 586, 1054
 Böckenhoff, A. 901, 943, 945, 1004, 1013,
 1048f, 1057f
 Böhning, D. 1057, 1059
 Böltken, F. 309, 317ff, 1049
 Boltzmann 796
 Bonferroni 628
 Bonitätspunkte 966
 Bonitätsschema 966
 Bonitur 966
 borrowing of strength 895, 950
 Bortkiewicz 213, 921
 Bosch, K. 1047, 1049
 Boscovich 11
 Bowker, A.H. 441f, 755, 1049
 Bowker
 -, Symmetrietest nach ... für die 2 x 2-
 Tafel 440f
 Box 835
 Boxcar-Funktion 842
 Box, G.E.P. 352, 639f, 678, 684ff, 690f,
 737, 1049
 Box-Cox-Transformation 352, 684f
 Boxhammer, J. 803, 1049
 Box-Jenkins-Verfahren (s. auch Methode
 von Box und Jenkins) 678
 Box-Plot 825, 827, 835ff, 889
 -, notched ... 835ff
 -, Vergleich von Beobachtungsreihen
 mittels ... 836ff
 Box und Jenkins
 -, Methode von ... 639f, 678, 686ff
 Bradburn, M.J. 918, 1049
 Bradley 526ff, 1024
 Brahe, T. 11, 899
 Bravais 73
 Bravais-Pearson
 -, Korrelationskoeffizient nach ... 73, 79f
 Bremner, J.M. 1048
 Breslow, N.E. 918f, 1049, 1068
 breakdown point 864f
 Breite eines Fensters 711
 Brillinger, D.R. 640, 1049
 Brouwer 888
 Brown, J.E. 602
 Brubacher, S.R. 638, 1050
 Bruchpunkt 864f

Bru – Cm

- , ... des α -getrimmten Mittels 881
- , ... des α -winsorisierten Mittels 882
- , ... eines Huber-k-Schätzers 870
- , ... eines M-Schätzers 868f
- , ... eines R-Schätzers 885f
- Brückenschaltung 750f, 759
- Brunk, H.D. 1048
- Bruttoinlandsprodukt (vgl. auch BIP) 893f, 945ff
- Bryson, M.C. 785, 1050
- Bürsten eines Scatter-Plots 860
- Bughi 530
- Bulirsch, R. 1055
- Burg 860
- Burn-In 745, 803
- Busch, K. 1068
- Byar, D.P. 1063
- Cal. Tech. 888
- Cantelli 121
- Cardano, G. 10
- Carroll, S.P. 1050
- Casagrande, J.T. 420f, 1050
- Casagrande/Pike/Smith
- , erforderliche Stichprobenumfänge beim Test auf Homogenität in der 2 x 2-Tafel nach ... 420f
- Casella, G. 1019, 1065
- cash flow 894, 947, 967
- , ... profit 894, 947
- , Steigerung des ... 947, 961
- , ... Prognose 967
- , ... ohne Zuverlässigkeitssangaben 967
- , ... Steigerung 967
- , Delphi-Kombination zur Prognose der ... 967
- Castellino, V.C. 1070
- Cauchy 113, 155
- Cauchy-Verteilung 113, 155
- Cea, J. 1055
- Census-X-11 639, 666
- Chakravarti, A.K. 965, 1050
- Chambers, J.M. 826, 1050
- Chatfield, C. 640, 894, 1050
- Chatterjee, P. 769, 1050
- Chatterjee, S. 608, 1050
- checking
- , observations for ... 365
- Chen, M.J. 1070
- Chemoff, H. 183, 860, 1070
- Chemoff-face 860
- χ^2 -Anpassungstest 182f
- , ... auf Exponentialverteilung 225f
- , ... auf Gleichverteilung 197f
- , ... auf Normalverteilung 182f
- , ... auf Poissonverteilung 216f
- χ^2 -Methode
- , inverse ... 955
- , Test über die globale Hypothese bei der Kombination unabhängiger Tests nach der ... 955
- χ^2 -Test 413, 419f, 435ff
- , ... auf Unabhängigkeit in der 2 x 2-Tafel 413f
- , ... im loglinearen Modell der r x s-Tafel 425ff
- , Scheffé's ... zum Vergleich der Dispersionen von p unabhängigen Meßreihen 617f
- χ^2 -Verteilung 152ff, 156f, 161f, 180, 214
- , Approximation der ... 153
- , Beziehung zwischen Poissonverteilung und ... 213
- , Dichte der ... 152
- , Erwartungswert der ... 153
- , Exzeß der ... 153
- , Quantile der ... 153, 737f
- , Schiefe der ... 153
- , Varianz der ... 153
- Christmann, A. 1057, 1064
- Chung 889, 918
- Cisapride 888f, 915ff
- Classen, M. 1048
- Clemen, R.T. 895, 1050
- Cleveland, W.S. 638, 826, 1050
- clinical research
- , non-therapeutic ... 363
- , therapeutic ... 363
- CloI 964, 1050
- Clopper 203f
- Cluster (s. auch Klumpen) 288, 910ff, 1014
- , ... analyse 374, 860
- , Erhebungs... 912
- , Studien... 910f
- , Konfidenzintervall für den Gesamteffekt bei ... 911
- , Schätzer für den Gesamteffekt bei ... 911
- , Varianz des ... 911
- CM-Problem (s. a. common mean problem) 900, 980
- , generalisierte Konfidenzintervalle für das gewöhnliche ... 980
- , generalized ... 941ff
- , Konfidenzintervall für das gemeinsame Mittel im ... 942

- | --, | Cm – Dec |
|---|-----------------|
| Schätzer für das gemeinsame Mittel im ... 942 | |
| Varianzschätzer bei Unkenntnis der Einzelvarianzen im ... 941ff | |
| direkter ... von Hartung-Böckenhoff-Knapp 942f | |
| C-Note nach Guilford 376 | |
| Cobb 654 | |
| Cobb-Douglas-Produktionsfunktion 654 | |
| Cochrane, D. 737f, 1050 | |
| Cochrane-Orcutt-Methode 738ff | |
| Cochrane-Orcutt-Transformationsmatrix 739 | |
| Cochran, W.G. 305, 363, 423f, 855, 900f, 909, 923, 1013, 1050, 1070, 1073 | |
| Cochran | |
| -, Homogenitätstest von ... 901, 1013 | |
| -, ... Homogenitätsstatistik 905 | |
| -, ... Quadratform 901, 906 | |
| Cochran's Q | |
| -, Test auf Symmetrie in der 2 x 2-Tafel mittels ... 423f | |
| Cogburn, D. 714, 1050 | |
| Cohen, M.L. 827, 880, 882, 1054 | |
| Collins, R. 1074 | |
| combined forecast 894f | |
| combining forecast (s. a. kombinerte Prognose) 947ff, 958ff | |
| combining information 887ff, 959 | |
| common mean problem 900, 980 | |
| -, generalisierte Konfidenzintervalle für das gewöhnliche ... 980 | |
| -, generalized ... 941ff | |
| --, Konfidenzintervall für das gemeinsame Mittel im ... 942 | |
| --, Schätzer für das gemeinsame Mittel im ... 942 | |
| --, Varianzschätzer bei Unkenntnis der Einzelvarianzen im ... 941ff | |
| --, direkter ... von Hartung-Böckenhoff-Knapp 942f | |
| concomitant variable 365 | |
| conditional least squares 688 | |
| confounding 941 | |
| Conover, W.J. 1051 | |
| continuous sampling plan 406 | |
| -, multi-level ... 406 | |
| Cosinoid 699 | |
| Cosinusfunktion 842 | |
| Costa 557 | |
| Cox. D.R. 247f, 352, 673, 684f, 824, 940f, 1049, 1051 | |
| Cox | |
| -, ... Proportional Hazard Regression Model 824, 940f | |
| Cox, G.M. 363, 1050 | |
| Cox und Stuart | |
| -, Test auf Trend von ... 247f, 643 | |
| Cramer 452 | |
| -, Assoziationsmaß von ... 452 | |
| Creyton 889, 918 | |
| Cromwell, O. 12 | |
| Cromwell, T. 12 | |
| cross-product-ratio 442f | |
| cttot-Test | |
| -, ... gegen DFRA-Lebensdauerverteilungen 784 | |
| -, ... gegen DFR-Lebensdauerverteilungen 778 | |
| -, ... gegen IFRA-Lebensdauerverteilungen 784 | |
| -, ... gegen IFR-Lebensdauerverteilungen 778 | |
| -, kritische Werte des ... 778 | |
| Cummings, G.E. 769, 1051 | |
| cumulative total time on test - Test (s. auch cttot-Test) 778, 784 | |
| Curtis, P.S. 981, 1062 | |
| cut-off-Verfahren 316f | |
| Dalenius 278 | |
| Dalenius und Hodges | |
| -, Kumulierungsregel von ... bei der geschichteten Zufallsauswahl 278 | |
| Dalkey, N.C. 965, 1051 | |
| Daniell 711f | |
| -, ...-Fenster 711f, 714 | |
| Daniels 588 | |
| Daten 15ff | |
| -, graphische Darstellung von ... 19, 22 | |
| -, gruppierte ... 33 | |
| -, Kenngrößen von ... 19 | |
| -, Original... 7 | |
| -, Symmetrisierung von ... 825, 827, 833ff | |
| -, tabellarische Darstellung von ... 19 | |
| Datenaufbereitung durch Skalierung 861 | |
| Datenmaterial 15 | |
| Datenmatrix 858 | |
| Datentransformation 825, 827, 832ff, 850 | |
| Dauerverfügbarkeit 820, 822 | |
| David, F.N. 185, 548, 1051 | |
| David, H.A. 236, 344, 1023, 1051 | |
| David-Hartley-Pearson | |
| -, Ausreißertest von ... 344 | |
| --, Quantile des ... 344 | |
| Davis, H.T. 714, 1050 | |
| Dean 571 | |
| Decker, R. 1060 | |

Dec – Dif

- decreasing failure rate (s. auch DFR) 776
- decreasing failure rate average (s. auch DFRA) 783
- Deeks, J.J. 1049
- Deflationierung 66
- Delphi-Kombination 967
 - , ... zur Prognose der Cash Flow Steigerung 967
- Delphi-Konfidenzintervall 966
- , kombiniertes ... 966
- Delphi-Methode 95, 965ff
 - , Anwendungsbereiche der ... 965
 - , ... für Prognosen und Vorhersagen 965ff
 - , iterative ... 966
 - , Kombination von Einzelprognosen mittels ... 966
 - , Kombination von subjektiven Risikobeurteilungen mittels ... 966
 - , ... zur Gewinnung subjektiver Wahrscheinlichkeiten 965ff
- Delphi-Prognose 965f
- , kombinierte ... 966
- Demetrescu, M. 960, 964, 1051
- DeNute 889, 918
- DerSimonian, R. 906, 932, 972, 984, 995, 1051
- DerSimonian/Laird-Schätzer (s. a. DSL-Schätzer) 906f
- Deruyttere 889, 918
- Descartes, R. 8f
- descriptive level 137
- deskriptive Kenngröße 15
- deskriptive Methoden der Zeitreihenanalyse 638, 640ff
- deskriptive Statistik 2, 15ff, 112, 124
- DeVinney, L.C. 1071
- Devlin, S.J. 638, 1050
- DFRA-Lebensdauerverteilung 782ff, 786
 - , cttot-Test gegen ... 784
- DFR-Lebensdauerverteilung 776ff, 786, 803
 - , cttot-Test nach Epstein gegen ... 778
 - , Proschan-Pyke-Test gegen ... 777
- DGQ-Qualitätsregelkarte 405
- Diagramm
 - , Flächen... 22
 - , Kreis... 22f
 - , Mengen... 92
 - , Stab... 21f
- Diamant 860
- Dichte 106f, 112ff, 118, 120
 - , ... der χ^2 -Verteilung 152
- , ... der Dreiecksverteilung 196
- , ... der Erlang-n-Verteilung 235
- , ... der Exponentialverteilung 219
- , ... der F-Verteilung 156
- , ... der Gleichverteilung 192
- , ... der Hjorth- (IDB-)Verteilung 232
- , ... der Lognormalverteilung 151
- , ... der Normalverteilung 143
- , ... der Standardnormalverteilung 145
- , ... der t-Verteilung 154
- , ... der Weibullverteilung 230
- , ... der zweidimensionalen Gleichverteilung 194
- , ... der zweiseitig gestützen Normalverteilung 150
- , empirische ... 827, 840ff
- , gemeinsame ... 108
- , Verteilung mit ... 106
- Dichtefunktion (s. auch Dichte) 106
- Dichteschätzer 841ff
- Dickey, D.A. 895, 962ff, 1051
- Dickey-Fuller
 - , ... p-Werte 963
 - , Kombination der ... nach Hartung 964
 - , ... Test 744, 895, 962f
 - , ... Unit Root Test auf Stationarität einer Zeitreihe 744, 962f
 - , p-Werte des ... 963
 - , ... nach McKinnon 963
 - , Teststatistik des ... 963
 - , Grenzverteilung der ... 963
- Diebold, F.X. 895, 1051
- Dieter, U. 194, 1048, 1051
- differentielle Behandlung 358
- differentieller Effekt 619f, 625
- Differenz
 - , d-te ... 668ff
 - , erste ... 668ff
 - , saisonale ... 668ff
 - , zweite ... 668ff
- Differenz der Mittelwerte (Effektmaß) 978ff
 - , exakte Methoden im FE-Modell für ... 978f
 - , Hartung-Knapp-Verfahren für ... 978f
 - , exakter Schätzer nach dem ... 978
 - , exaktes Konfidenzintervall nach dem ... 978f
 - , MUPML Gesamt-Schätzer für ... 978f
 - , approximativer ... 979
 - , exakter ... 978
- Differenzenfilter 674
- Differenzenmethode 668ff
- Differenzenschätzer 303
 - , Varianz des ... 303

- ,- Schätzer für die ... 303
 Differenzenschätzung 303
 Dirac-Funktion 712
 direkte Beobachtung 326
 disjunkte Ereignisse 93f, 100, 103
 disjunktive Normalform 752f, 756f
 -, ... der Strukturfunktion eines monotonen Systems 752f
 -, ... der Zuverlässigkeitsfunktion eines monotonen Systems 756f
 diskrete Beobachtungen 407ff, 913ff
 diskrete Daten 407ff, 913ff
 diskrete Verteilung 106f, 126
 diskret-verteilter Zufallsvektor 107
 Diskriminanzanalyse 374
 Diskriminationsgüte 860
 Dispersion (s. Streuung)
 Dispersionsmaß 40, 116
 Dispersionstest 526
 Dispersionsvergleiche
 -, ... bei der einfachen Varianzanalyse 617f
 -, ... bei zwei abhängigen Meßreihen und normalverteilter Grundgesamtheit 538f
 -, ... bei zwei unabhängigen Meßreihen und normalverteilter Grundgesamtheit 524ff
 -, verteilungsfreie ... bei zwei unabhängigen Meßreihen 526ff
 Distanzindex 379
 Diversifikation 899
 -, ...entscheidung 899
 Dixon, W.J. 346f, 1023f, 1051
 Dixon
 -, Ausreißertests von ... 346f
 --, kritische Werte der ... 347
 Documenta Geigy Tabellen 417, 1051
 Dodge, H.F. 373, 389, 392, 406, 1051
 Doksum, K. 778, 1049
 Dombrowski, E. 805, 1051
 Dominanz einer Studie oder Prognose 1013
 -, Partial Weighting nach Yates-Cochran bei ... 1013
 -, verallgemeinertes RE-Modell bei ... 1013
 --, Methoden im ... 1013
 Doomsday Book 12
 Doppelblindversuch 362, 365, 372f
 doppelte Quadratsummenzerlegung 620
 doppelter Stichprobenplan für qualitative Merkmale 390ff
 -, durchschnittlicher Stichprobenumfang beim ... 390f
 Double-Root-Residuum 847
- Douglas 654
 Draftsman-Display 858
 Draper, D. 887, 950, 958f, 1052
 Draper, N.R. 608, 1052
 Drefahl, D. 1070
 Dreieck-Fenster 712, 714
 -, Bandbreite des ... 712
 Dreieckskern 842
 Dreiecksverteilung 195f, 273, 276
 -, allgemeine ... 196
 --, Erwartungswert der ... 196
 --, Varianz der ... 196
 -, einseitige ... 196
 -, gleichschenklige ... 195f
 --, Dichte der ... 196
 --, Erwartungswert der ... 196
 --, Modalwert der ... 196
 --, Varianz der ... 196
 Dreifachgitter 361
 DSL-Schätzer
 -, ... für den Heterogenitätsparameter im RE-Modell 906f
 d-spread 826, 829
 Dubash, R.K. 1050
 Dudeck, J. 1057, 1059
 Dugošija, Đ.P. 1058
 Duncan 628
 Durbin, J. 227, 640, 1052
 Durbin-Watson-Test 740f
 -, Entscheidungsregeln des ... 741
 -, kritische Werte des ... 740f
 --, obere und untere Schranken der ... 740
 -, Teststatistik des ... 740
 Durchschlupf
 -, mittlerer ... 389, 406
 Durchschnitt
 -, gewogener gleitender ... 642f, 646
 -, gleitender ... 640, 643, 646
 --, ... gerader Ordnung 640
 --, ... ungerader Ordnung 640
 durchschnittlich ausgelieferter Ausschßanteil 406
 durchschnittlicher Fehler 324
 durchschnittlicher Stichprobenumfang 390ff, 394f, 406
 Durchschnittsbildung bei Ereignissen 92
 Durchschnittswert (s. auch Mittelwert) 1
 Dwass 616
 Dyspepsie 888f, 915ff
 -, funktionelle ... 888f, 915ff
- EBM** 887, 910
EDA 825ff

Efe – Ein

- EF-E-Modell 895, 949
 - , Investitionsberatung im ... 950
 - , Kombination von Prognosen im ... 949
 - , Konfidenzintervall für den Gesamteffekt im ... 949
 - , Schätzer für den Gesamteffekt im ... 949
 - , Schätzer für die Varianz des ... 949
- Effekt
 - , Behandlungs... 619f, 628
 - , Block... 620
 - , differentieller ... 619f, 625
 - , ... eines Faktors 355
 - , fester ... 609, 629
 - , identifizierbarer ... 355
 - , Placebo... 371
 - , Wechselwirkungs... 625
 - , Wiederholungs... 624
 - , zufälliger ... 609, 629, 634
- Effektmaß 914
 - , Homogenitätstest für ... 917
 - , ... Odds Ratio OR 915
 - , ... relatives Risiko RR 915
 - , ... Risiko Differenz RD 914
 - , für normalverteilte Daten ... 976ff
 - , spezielle ... 976ff
 - , ... Differenz der Mittelwerte 978ff
 - , ... Quotient der Mittelwerte 981ff
 - , ... Standardisierte Mittlere Differenzen 991ff
- effiziente Portfolioauswahl 1006ff
- Eichung 322, 335
- Eicker, F. 744, 1052
- Eigenschaften der GM-Schätzer in der Meta-Analyse 1018ff
 - eightths 828ff
 - , ...-Abstand 829
 - einarmige Studie 909f, 921ff
 - Einbrennungsprozeß 803
 - eindimensionale Gleichverteilung 192ff
 - , Dichte der ... 192
 - , Erwartungswert der ... 193
 - , Simulation von Realisationen aus einer ... 193f
 - , Varianz der ... 193
 - , Verteilungsfunktion der ... 193
 - , zentrale Momente der ... 193
 - einelementige Nullhypothese 139
 - einfacher Stichprobenplan 383ff, 395ff
 - , ... für qualitative Merkmale 383ff
 - , Bestimmung eines ... durch Vorgabe des Indifferenzpunktes und der Steilheit 387ff
 - , Bestimmung eines ... durch Vorgabe zweier Punkte der Operationscharakteristik 348ff
 - , ... für quantitative Merkmale 395ff
 - , Bestimmung eines ... durch Vorgabe des Indifferenzpunktes und der Steilheit 398f
 - , Bestimmung eines ... durch Vorgabe zweier Punkte der Operationscharakteristik 395ff
- einfaches Blockexperiment 357, 609, 619ff
 - , multiple Vergleiche beim ... 620, 623
 - , Schätzer für die Parameter beim ... 620
 - , Testen auf signifikante Lokationsunterschiede beim ... 620, 622f
- einfache Streuungszerlegung 612
- einfache Varianzanalyse 355, 609ff
 - , Dispersionsvergleiche bei der ... 617 f
 - , Modellbetrachtung bei der ... 619
 - , Modell der ... 610
 - , multiple Vergleiche bei der ... 614ff
 - , Schätzen der Parameter bei der ... 611
 - , simultane Vergleiche bei der ... 614ff
 - , Testen auf signifikante Lokationsunterschiede bei der ... 611ff
 - , Varianzanalysetafel der ... 611
- einfache Zufallsstichprobe (s. auch einfache Zufallsauswahl) 124, 274ff, 286ff, 312, 316
- einfache Zufallsauswahl 274ff
 - , benötigter Stichprobenumfang bei der ... 275
 - , geschichtete Zufallsauswahl, bei der jede einzelne Auswahl eine ... ist 280
 - , Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit bei der ... 280
 - , Varianz des ... 280
 - , Schätzer für die ... 280
 - , Schätzer für die Varianz der Merkmalswerte in der Grundgesamtheit bei der ... 280
 - , Konfidenzintervall für den mittleren Merkmalswert in der Grundgesamtheit bei der ... 275
 - , Konfidenzintervall für die Summe der Merkmalswerte in der Grundgesamtheit bei der ... 275
 - , Schätzer für den mittleren Merkmalswert in der Grundgesamtheit bei der ... 274
 - , Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit bei der ... 274
 - , Varianz des ... 275f
 - , Schätzer für die ... 275

-, Schätzer für die Varianz der Merkmalswerte in der Grundgesamtheit bei der ... 274
 einfache hierarchische Klassifikation 630ff
 -, Modell der ... 630
 -, Schätzen der Varianzen der Effekte bei der ... 630f
 -, Schätzen des Intraklasskorrelationskoeffizienten bei der ... 631
 -, Varianzanalysetafel der ... 630
 Einflußfaktor 362f
 -, Haupt... 362
 -, kontrollierbarer ... 362
 -, nicht-kontrollierbarer ... 362f
 Einflußkurve 863f
 -, ... des α -getrimmten Mittels 881
 -, ... eines Huber-k-Schätzers 870
 -, ... eines M-Schätzers 867f
 Eingangssereignis bei der Konstruktion von Fehlerbäumen 766
 Eingangskontrolle 381, 406
 einseitige Hypothese 138
 einseitiger Test 139
 einseitiges Konfidenzintervall 130
 einseitiges Testproblem 139
 einseitig gestutzte Normalverteilung 150
 Einstichprobengaußtest 135ff
 Einstichprobenproblem 143
 einstufiges Auswahlverfahren bei der Klumpenstichprobe 288ff
 einstufige Strategie 823
 Eisenhart, C. 629, 1052
 Elementarereignis 91, 93, 97
 elementares System 746f
 -, Zuverlässigkeit eines ... 746f
 Elimination der glatten Komponente 639, 660ff
 Ellison, G.W. 1074
 EI Newehi, E. 746, 1052
 Elpelt, B. 120, 341, 358, 363, 378f, 450, 545, 549, 600, 608, 628, 636, 678, 826, 838, 847, 861, 895, 934, 944, 949f, 1004, 1010, 1052, 1058
 EMEA 1015, 1052
 Empfindlichkeit eines Meß- oder Analyseverfahrens 342
 empirische Ausfallrate (s. auch Ausfallquote) 19, 70ff, 219, 795
 empirische Dichte 827
 empirische Kenngrößen der EDA 825, 827ff
 -, ... für Einzeldaten 827ff
 -, ... für gruppierte Daten 827, 831f
 empirische Kovarianz 74
 empirischer Variationskoeffizient 174

Ein – Ere

empirisches Moment 126
 empirisches Quantil 825
 empirische Standardabweichung 324, 831f
 empirische Varianz 154
 empirische Verteilungsfunktion 29ff, 121, 825, 827, 839f
 -, stetige ... 827, 840
 Empirische Yule-Walker-Gleichungen 689f
 Emser, A. 1072
 Endkontrolle 381, 406
 endliche Additivität 94
 endliche Grundgesamtheit 305
 Endlichkeitskorrektur 208, 275, 307
 endogene Variable 744
 Engelbrecht, R. 1057, 1059
 Entscheidungsfenster 1008
 Entscheidungsraum 1007
 Entscheidungsregeln beim Testen 134
 Entsprechungszahl 55f
 Entwicklung des loglinearen Modells für 2 x 2-Tafeln 425
 Epstein, B. 778, 1053
 Epstein
 -, cttot-Test gegen DFR-
 Lebensdauerverteilungen nach ... 778
 -, cttot-Test gegen IFR-
 Lebensdauerverteilungen nach ... 778
 Erblichkeit eines Merkmals 892f
 Ereignis 91ff, 108, 121, 123
 -, Elementar... 91, 93, 97
 -, Haupt... 765f
 -, komplementäres ... 92
 -, sicheres ... 93
 -, unmögliches ... 93
 -, zufälliges ... 91
 Ereignisbaum 766
 Ereignisraum 91
 Ereignisse
 -, disjunkte ... 93f, 100, 103
 -, unvereinbare ... 93
 ERE-Modell 895, 951f
 -, Einbettung der Portfolioauswahl in das ... 1011f
 -, Heterogenitätsparameter im ... 951
 -, verallgemeinerter Hedges Schätzer für den ... 951
 -, Investitionsberatung im ... 951f
 -, Kombination von Prognosen im ... 951f
 -, Konfidenzintervall für den Gesamteffekt im ... 951f
 -, Schätzer für den Gesamteffekt im ... 951f

Ere – Exp

- , Schätzer für die Varianz des ... 951f
- , stabiler Hartung ... 952
- , Wertpapiermanagement im ... 1011f
- Erfolgsbaum 766
- Erhebung 6ff, 12, 18, 272ff
 - , geplante ... 6
 - , nichtgeplante ... 6
 - , primärstatistische ... 7
 - , sekundärstatistische ... 7
 - , Stichproben... 6, 13, 18f, 269ff
 - , Teil... 6, 18f
 - , tertiärstatistische ... 7
 - , Total... 6, 18f, 271, 300
 - , Voll... 6, 13, 18f, 271, 319
 - , Vor... 272, 275f, 282ff
- Erhebungskluster 912
- Erhebungstechnik 271
- Erkenntnis
 - , Kombination von ... 887ff
- erklärende Variable 365
- Erklärungsgrad (s. Bestimmtheitsmaß)
- Erlang-n-Verteilung 224f, 234f, 717
 - , Ausfallrate der ... 235
 - , bedingte Überlebenswahrscheinlichkeit bei der ... 235
 - , Dichte der ... 235
 - , Erwartungswert der ... 235
 - , Überlebenswahrscheinlichkeit bei der ... 235
 - , Varianz der ... 235
 - , Verteilungsfunktion der ... 235
- Erneuerung 807f
 - , altersabhängige ... 807f
 - , Block... 807
 - , Gruppen... 807f
 - , Havarie... 807
 - , prophylaktische ... 807
- Erneuerungsfunktion 808
 - , Schranken der ... 808
- Erneuerungsprozeß 807f
- Erneuerungstheorem 808
- Ersatzvariable 365
- erwartungstreuer Schätzer 125f
 - , ... mit gleichmäßig minimaler Varianz 126
 - Erwartungswert 112ff, 125
 - , Berechnung des ... 113f
 - , ... der allgemeinen Dreiecksverteilung 196
 - , ... der Binomialverteilung 200
 - , ... der χ^2 -Verteilung 153
 - , ... der eindimensionalen Gleichverteilung 193
- , ... der Erlang-n-Verteilung 235
- , ... der Exponentialverteilung 220
- , ... der Fehlerterme bei der linearen Regression 576
- , ... der F-Verteilung 156
- , ... der gleichschenkligen Dreiecksverteilung 196
- , ... der Hjorth- (IDB-)Verteilung 233
- , ... der hypergeometrischen Verteilung 207
- , ... der Lognormalverteilung 151
- , ... der Multinomialverteilung 210
- , ... der Normalverteilung 143
- , ... der Poissonverteilung 213
- , ... der Schätzer für die Parameter der multiplen Regression 597
- , ... der t-Verteilung 154
- , ... der Weibullverteilung 231
- , ... der zweiseitig gestützten Normalverteilung 149
- , ... des Schätzers für das Absolutglied einer Regressionsgeraden 576
- , ... des Schätzers für den Steigungsparameter einer Regressionsgeraden 576
- erweitertes ANOVA-Modell 1013
- Esary, J.D. 770, 1053
- Esenwein-Rothe, I. 317, 1053
- e-spread 829
- Esséen 122
- Euler 156, 730
- Eulersche Betafunktion 156
- Eulersche Relation 730
- evidence based medicine (s. a. EBM) 888, 910
- exakter Test 414ff
 - , ... auf Homogenität in der 2 x 2-Tafel 421f
 - , ... auf Unabhängigkeit in der 2 x 2-Tafel 414ff
 - , ... von Fisher 416ff
- ex-ante Prognosen 693, 741
- exogene Variable 744
- Experiment 3ff, 91, 359ff, 609
 - , höheres Split-Plot-... 360
 - , Kombination von 887ff
 - , kontrolliertes ... 364
 - , p x q-faktorielles ... in vollständig randomisierten Blöcken 359
 - , Split-Plot-... 359
 - , Split-Plot-in-strips-... 359
 - , Zufalls... 92ff, 103, 121, 123
- experimentelle Gruppe 914
- exploratory observations 365
- explizite Schätzer für den Heterogenitätsparameter im RE-Modell 996

explorative Datenanalyse 825ff
 explorative Regressionsfunktion 825,
 855ff
 explorative Verfahren 825ff
 -, multivariate ... 825f, 857ff
 Exponentenleiter 833f
 exponential smoothing 672f
 Exponentialverteilung 105f, 111f, 115,
 117f, 123f, 126, 219ff, 230ff, 251, 261,
 776f, 780, 786, 788ff, 809, 821
 -, Ausfallrate der ... 220
 -, bedingte Überlebenswahrscheinlichkeit
 bei der ... 219
 -, χ^2 -Anpassungstest auf ... 225f
 -, Dichte der ... 219
 -, Erwartungswert der ... 220
 -, Exzeß der ... 220
 -, Kolmogoroff-Smirnov-Anpassungstest
 auf ... 226ff
 -, Konfidenzintervall für den Parameter λ
 einer ... 220ff, 788ff
 -, Median der ... 220
 -, Schätzer für den Parameter λ einer ...
 220f, 788ff
 -, Schätzer und Konfidenzintervalle für
 den Parameter λ einer ... 220ff, 788ff
 --, ... bei vorgegebener Beobachtungsdauer
 788f
 ---, ... mit Ersetzen ausgefallener Objekte
 788f
 ---, ... ohne Ersetzen ausgefallener Objekte
 789f
 ---, ... bei vorgegebener Zahl von Ausfällen
 789f
 ---, ... mit Ersetzen ausgefallener Objekte
 789f
 ---, ... ohne Ersetzen ausgefallener Objekte
 790
 -, Schiefe der ... 220
 -, sequentieller Quotiententest für den
 Parameter λ einer ... 261ff
 -, Simulation von Realisationen aus einer
 ... 220
 -, Test über den Parameter λ einer ... 223ff
 -, Überlebenswahrscheinlichkeit bei der ...
 219
 -, Varianz der ... 220
 -, Verteilung der Restlebensdauer bei der
 ... 219f
 -, Verteilungsfunktion der ... 219
 exponentielles Glätten 672f
 -, Prognose mittels ... 672f
 ex-post Prognose 693, 741
 extended Fixed-Effects Model (s. a. EFE-
 Modell) 949

Exp – Feh

extended Random-Effects Model (s. a.
 ERE-Modell) 951
 externe geschlossene Frage 309
 extremer Rangsummentest von Youden
 339f
 Extremwert einer Beobachtungsreihe 829
 Exzeß 47, 49, 118
 -, ... der χ^2 -Verteilung 153
 -, ... der Exponentialverteilung 220
 -, ... der Normalverteilung 144
 -, Normalverteilungstest auf Schiefe und
 ... 189
 Eyring 745, 796, 799
 Eyring-Modell 745, 796, 799
 -, verallgemeinertes ... 745, 799
 Ezekiel, M. 608, 1053

 Facette 860
 Faktor
 -, Effekt eines ... 355
 Faktorenanalyse 375, 379, 860
 faktorieller Versuchsplan 628
 Faltung 110
 fault tree analysis 745, 765ff
 Fechner, G.T. 78f, 375, 450, 1053
 Fechner
 -, marginale Normalisierung nach ... 375f
 Fechnerscher Korrelationskoeffizient 78f
 Fechnersche Skalierung 375f
 Fehler
 -, durchschnittlicher ... 324
 -, ... 1. Art 133, 135, 137, 251
 -, ... in den Variablen 601ff
 --, Regression mit ... 601ff
 ---, Schätzer für die Parameter einer Reg-
 ressionsgeraden bei der ... 601ff
 -, Meß... 269, 320ff
 -, mittlerer ... 324
 -, mittlerer quadratischer ... 125f, 291,
 300f
 -, relativer ... 271f, 276f, 285, 324
 -, statistischer ... 321
 -, systematischer ... 141, 308, 321f, 325,
 332f, 335, 337, 340ff, 354, 362, 366
 -, wahrscheinlicher ... 323
 -, zufälliger ... 272, 275, 285, 287, 321ff,
 337, 341f, 574, 638
 --, relativer ... bei Stichprobenerhebungen
 269f
 -, ... 2. Art 133ff, 138f, 181ff, 251
 Fehlerbaum 765ff
 -, modulare Repräsentation eines ... 769
 Fehlerbaumanalyse 745, 765ff, 965

Feh – Fes

- Fehlerfortpflanzungsgesetz
 - , Gaußsches ... 173, 323, 326, 328ff, 832f, 969, 984f
- Fehlerprozeß 639, 736ff
 - , AR(1)-... 736ff
 - , ... beim autoregressiven Regressionsmodell 744
- Fehlerterme bei der linearen Regression 574ff
 - , Erwartungswert der ... 576
 - , Varianz der ... 576
 - , Konfidenzintervall für die ... 581
 - , Schätzer für die ... 578
 - , Test über die ... 580
 - , Verteilung der ... 577
- Feigl, P. 421, 1053
- Feller, W. 1053
- FE-Modell 890, 899ff
 - , ... bei Normalverteilung 976ff
 - , Effektmaß Differenz der Mittelwerte im ... 978ff
 - , exakte Methoden für das ... 978ff
 - , Hartung-Knapp-Verfahren für das ... 978f
 - , exakter Schätzer nach dem ... 978
 - , exaktes Konfidenzintervall nach dem ... 978f
 - , MUPML Gesamt-Schätzer für das ... 978f
 - , approximativer ... 979
 - , exakter ... 978
 - , Effektmaß Quotient der Mittelwerte im ... 981ff
 - , alternatives Modell für das ... 986f
 - , Konfidenzintervall im ... 987
 - , MUPML Gesamt-Schätzer im ... 987
 - , Schätzer für die Varianz des ... 987
 - , exakte Methoden für das ... 981ff
 - , graphische Ermittlung der Konfidenzgrenzen und Punktschätzer für das ... 982f
 - , Hartung- Verfahren für das ... 981ff
 - , exakter Schätzer nach dem ... 981
 - , exaktes Konfidenzintervall nach dem ... 981
 - , MUPML Gesamt-Schätzer für das ... 981
 - , approximativer ... 982
 - , exakter ... 981
 - , Effektmaß Standardisierte Mittlere Differenzen im ... 991ff
 - , exakte Methoden für das ... 991ff
 - , Hartung-Verfahren für das ... 991ff
- , approximatives Konfidenzintervall nach dem ... 993
- , exakter Schätzer nach dem ... 992
- , exaktes Konfidenzintervall nach dem ... 992
- , MUPML Gesamt-Schätzer für das ... 992
- , approximativer ... 992
- , exakter ... 992
- , generalisierte Konfidenzintervalle für das ... 980
- , ... bei Poissonverteilung 920ff
- , ... bei vergleichenden Time-to-Event Studien 932, 936f
- , ... bei vergleichenden Überlebenszeitstudien 931f, 936f
- , erweitertes ... (vgl. auch EFE-Modell) 895
- , homogenes ... 900f
- , Gesamteffekt im ... 900
- , Schätzer für den ... 900
- , Gauß-Markov ... 900
- , Varianz des ... 900
- , inhomogenes ... 900f
- , Gesamteffekt im ... 900
- , Konfidenzintervall für den ... 901
- , Schätzer für den ... 900
- , unverzerrter ... mit minimaler Varianz 900
- , Varianz des ... 900f
- , Mischung aus ... und RE-Modell 910f
- , Konfidenzintervall für den Gesamteffekt bei der ... 911
- , Schätzer für den Gesamteffekt bei der ... 911
- , Varianz des ... 911
- , Investitionsberatung im ... 947f
- , Kombination von Kaplan-Meier Kurven im ... 936f
- , Prognose des Gesamteffekts bei geschätzten Zuverlässigkeiten im ... 947
- Fenster 711ff, 734
 - , Bartlett-... 712, 714
 - , Daniell-... 711f, 714
 - , Dreieck-... 712, 714
 - , lag-... 711ff, 734
 - , negatives Seiten... 714
 - , Parzen-... 713ff
 - , Rechteck-... 711f, 714
 - , Spektral-... 711ff, 734
 - , Spline-... 714
 - , Tukey-... 712, 714
- Fermat, P. de 10
- Ferschl, F. 1053
- fester Effekt 629

- Fieller, E.C. 983, 1053
 Fienberg, S.E. 1049, 1053
 fiktive Grundgesamtheit 306f
 Filip, D. 1067
 Filter 660, 673ff, 705ff
 -, Differenzen... 674
 --, nichtsaisonaler ... 674
 -, Hochpaß... 674, 705ff
 -, Klassifizierung von ... 674, 705ff
 -, linearer ... 660, 673ff
 -, Serienschaltung von ... 674f
 -, symmetrischer ... 673f
 -, Tiefpaß... 674, 705ff
 Filterung (s. auch Glättung) 673
 Finney, D.J. 417, 1053
 Fisher, I. 70
 Fisher, R.A. 12, 354, 363, 365, 370, 378,
 414, 416, 450, 521, 548, 609, 612, 628f,
 891, 904, 955, 957f, 960f, 964, 983,
 1024, 1044, 1053
 Fisher
 -, exakter Test von ... auf Unabhängigkeit
 in der 2 x 2-Tafel 416ff
 -, ...-Methode 955
 --, Test über die globale Hypothese bei der
 Kombination unabhängiger Tests nach
 der ... 955
 -, ...-Probit 955
 --, Kombination der Dickey-Fuller p-
 Werte nach Hartung mit ... 964
 --, kombinierter p-Wert für ... 960
 Fisher-Lancaster-Skalierung 450
 Fisherscher Idealindex 70
 Fishersche Z-Transformation 354, 548
 Fisichelli 557
 Fisz, M. 1053
 fixed effects model (s. a. FE-Modell) 890,
 899ff
 -, erweitertes ... (s. a. EFE-Modell) 895
 Flächendiagramm 22
 Flament, R. 1070
 Fleming, T.R. 928, 1016, 1053, 1067
 Flensburger Verkehrskartei 7
 flexibler Gruppenumfang 1016
 Flury 860
 Flury-Riedwyl-face 860
 Förster, W. 641, 1053
 Fomby, T.B. 744, 1053
 Forest Plot 889
 Formierprozeß 803
 Forth 572
 Foumival, R. de 10
 Fox, K.A. 608, 1053
 Frage
 -, Bild... 310
 Fie – Gai
 -, externe geschlossene ... 309
 -, geschlossene ... 309
 -, interne geschlossene ... 309
 -, offene ... 309f
 Fragebogen 309ff
 Fragestellung
 -, suggestive ... 309
 Fraktil 114
 Francois 889, 918
 freie Hochrechnung 301
 freier Fall
 -, Gesetz vom ... 4
 freiwillige Basis
 -, Zuordnung auf ... 369
 Frequenz 705
 -, ...bereich 639, 699, 728
 -, ...responsefunktion 703
 Freund 526ff, 1024
 Freund, J.E. 826, 1066
 Friedman 622f, 628, 1023
 Friedman-Rangsummentest für multiple
 Vergleiche im einfachen Blockexperiment
 623
 Friedman-Test auf signifikante Lokations-
 unterschiede
 -, ... bei der zweifachen Varianzanalyse
 ohne Wechselwirkungen zwischen den
 Faktoren 628
 -, ... im einfachen Blockexperiment 622f
 -, kritische Werte des ... 623
 Fruchtbarkeitsrate 56f, 60
 F-Test 362, 609, 611f, 620, 625, 627f
 Fuller, W.A. 640, 895, 962ff, 1051, 1053
 full analysis set 910
 Funktion
 -, positiv homogene ... 1021
 -, quasi-konkave ... 1021
 -, quasi-konvexe ... 1021
 funktionelle Dyspepsie 888f, 915ff
 Funktionszeit
 -, mittlere ... 820ff
 F-Verteilung 12, 156f, 201, 204
 -, Approximation der ... 157
 -, Beziehung zwischen Binomialverteilung
 und ... 201
 -, Dichte der ... 156
 -, Erwartungswert der ... 156
 -, Quantile der ... 157, 1031ff
 -, Varianz der ... 156
 Gaede, K.-W. 756, 762, 811, 1053
 Gail, M.H. 781, 1053
 Gain 730f, 735

Gal – Ges

- Galilei, G. 9f
 Galton, F. 11, 569f, 1053
 Gammafunktion 152f, 156, 1040
 Gammaverteilung 11
 Gastropokinetikum 917
 Gastwirth, J.L. 781, 827, 880, 882f, 1053f
 Gastwirth-Mittel 883
 Gaul, W. 608, 1054, 1058, 1061
 Gauß, C.F. 109, 135, 144, 152, 178, 190,
 224, 326, 507, 535, 549, 900, 968f, 984,
 1018ff, 1022, 1040
 Gaußklammer 142, 828
 Gauß-Markov Schätzer 900, 906
 -, Eigenschaften der ... in der Meta-
 Analyse 1018ff
 Gaußsche Multiplikationsformel für die
 Gammafunktion 1040
 Gaußsches Fehlerfortpflanzungsgesetz
 173, 323, 326, 328ff, 832f, 969, 984f
 Gaußsches Integral 190
 Gaußsche Verdoppelungsformel für die
 Gammafunktion 1040
 Gaußsche Verdreibefachungsformel für die
 Gammafunktion 1040
 Gauß-Test 135ff, 144, 152, 178, 215, 224,
 246, 507, 527, 535, 542, 549
 Gaußverteilung (s. auch Normalverteilung)
 11, 109
 Gaver, D.P. jr. 1052
 GCM Problem 941ff
 -, Konfidenzintervall für das gemeinsame
 Mittel im ... 942
 -, Schätzer für das gemeinsame Mittel im
 ... 942
 -, Varianzschatzter bei Unkenntnis der
 Einzelvarianzen im ... 941ff
 --, direkter ... von Hartung-Böckenhoff-
 Knapp 942f
 gebundene Hochrechnung 301
 Geburtenziffer 56
 Geburtstagsstichprobe 370
 gedächtnislose Verteilung 219
 gefilterte Zeitreihe 673
 geglättetes Stichprobenkreuzspektrum 734
 geglättetes Stichprobenspektrum 711ff
 Gehan, E.A. 824, 928f, 1054
 Gehlert, G. 1049
 Geibel, A. 1061
 Gell, G. 1057, 1059
 gemeinsame Dichte 108
 gemeinsame Verteilung 107, 120, 195
 gemeinsame Verteilungsfunktion 107
 gemischter Prozeß 639, 678, 682ff
 -, Autokorrelationen eines ... 683
 -, Autokovarianzen eines ... 682f
 -, Darstellung ... durch AR- bzw. MA-
 Prozesse 684
 -, ... der Ordnungen p und q 682ff
 -, Ordnung eines ... 687
 -, Spektraldichte eines ... 704f
 gemischtes Modell der Varianzanalyse 629
 Genauigkeit eines Meß- oder Analysever-
 fahrens 342
 Genauigkeitsvergleiche bei Aufteilungs-
 verfahren für die geschichtete Zu-
 fallsauswahl 287f
 generalisierter Least-Squares-Schätzer 736
 generalisiertes Konfidenzintervall 980
 -, ... für das gewöhnliche Common Mean
 Problem 980
 -, ... für das FE-Modell mit normalvertei-
 ten Daten 980
 generalized common mean (s. a. GCM)
 problem 941ff
 -, Konfidenzintervall für das gemeinsame
 Mittel im ... 942
 -, Schätzer für das gemeinsame Mittel im
 ... 942
 -, Varianzschatzter bei Unkenntnis der
 Einzelvarianzen im ... 941ff
 --, direkter ... von Hartung-Böckenhoff-
 Knapp 942f
 Geodäsie
 -, Varianzanalysemmodell aus der ... 634ff
 geometrisches Mittel 35ff
 -, gewogenes ... 36
 gepaarte Differenz 534
 geplante Erhebung 6
 gepoolte Varianz 977
 Gesamteffekt
 -, ... im EFE-Modell 949
 --, Konfidenzintervall für den ... 949
 --, Schätzer für den ... 949
 -, ... im ERE-Modell 951f
 --, Konfidenzintervall für den ... 951f
 --, Schätzer für den ... 951f
 -, ... im GCM Problem 941ff
 --, Konfidenzintervall für den ... 942
 --, Schätzer für den ... 942
 -, ... im homogenen FE-Modell 900f
 --, Schätzer für den ... 900
 ---, Varianz des ... 900
 -, ... im inhomogenen FE-Modell 900f
 --, Konfidenzintervall für den ... 901
 --, Schätzer für den ... 900f
 ---, Varianz des ... 900f
 -, ... im RE-Modell 906f
 --, Schätzer für den ... 907

- Ges – Gle
- , Varianz des ... 907
 - , Konfidenzintervalle für den ... 907f
 - , Prognose des ... bei unbekannten Zuverlässigkeiten 947f
 - , geschätzte Varianz der ... 947
 - , geschätzte Zuverlässigkeiten der ... 947
 - , ... im FE-Modell 947
 - , ... im RE-Modell 948
 - Gesamtfehler
 - , relativer ... 325, 327
 - Gesamtmittel 619f, 625, 630
 - geschichtete Stichprobe (s. auch geschichtete Zufallsauswahl) 355
 - geschichtete Zufallsauswahl 278ff
 - , ... bei der jede einzelne Auswahl eine einfache Zufallsauswahl ist 280
 - , Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit bei der ... 280
 - , Varianz des ... 280
 - , Schätzer für die ... 280
 - , Genauigkeitsvergleiche bei Aufteilungsverfahren für die ... 287f
 - , Kumulierungsregel von Dalenius und Hodges bei der ... 278
 - , optimale Aufteilung (Neyman-Tschuprow-Aufteilung) bei der ... 282ff
 - , Stichprobenumfang bei ... 282f
 - , Varianz des Schätzers für die Summe der Merkmalswerte in der Grundgesamtheit bei ... 284
 - , Schätzer für die ... 284
 - , proportionale Aufteilung bei der ... 285
 - , Stichprobenumfang bei ... 285
 - , Varianz des Schätzers für die Summe der Merkmalswerte in der Grundgesamtheit bei ... 285
 - , Schätzer für die ... 285
 - , Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit bei der ... 280
 - , Varianz des ... 280
 - , Schätzer für die ... 280
 - , Schichtenbildung nach Auswahl der Stichprobe bei der ... 286
 - , Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit bei ... 286
 - , Varianz des ... 286
 - , Schätzer für die ... 286
 - , Schichtungsvariable bei der ... 278
 - geschlossene Frage 309
 - , externe ... 309
 - , interne ... 309
- Gesetz der großen Zahlen 11, 121, 125
 - Gesetz der kleinen Zahlen 213
 - Gesetz der universalen Regression 569
 - Gesetz vom freien Fall 4
 - Gesichter 860
 - gestutzte Normalverteilung 148ff
 - , einseitig ... 150
 - , linksseitig ... 150
 - , rechtsseitig ... 150
 - , zweiseitig ... 149f
 - , Dichte der ... 150
 - , Erwartungswert der ... 149
 - , Varianz der ... 149
 - , Verteilungsfunktion der ... 149
 - gestutzte Verteilung 150
 - gewichtete Inverse Normalmethode 955
 - , Test über die globale Hypothese bei der Kombination unabhängiger Tests nach der ... 955
 - Gewichtsfunktion 841ff
 - Gewichtsmatrix beim Berliner Verfahren 724f
 - , ... für die glatte Komponente 725
 - , ... für die saisonale Komponente 725
 - gewogener gleitender Durchschnitt 667
 - gewogenes arithmetisches Mittel 32
 - gewogenes geometrisches Mittel 36
 - gewogenes harmonisches Mittel 37
 - Ghezzi, A. 1074
 - Gini 52f
 - Gini-Koeffizient 52f
 - Glättung 639, 660ff, 725, 825, 849ff
 - Glass, G.V. 887, 1054
 - glatte Komponente einer Zeitreihe 639ff, 660ff, 725
 - , Elimination der ... 639, 660ff
 - , Gewichtsmatrix der ... beim Berliner Verfahren 725
 - , Schätzung der ... 639, 660ff, 725
 - , ... beim Berliner Verfahren 725
 - gleichmäßig bester Test 138
 - gleichmäßig minimale Varianz
 - , erwartungstreuer Schätzer mit ... 126
 - gleichschenklige Dreiecksverteilung 195f
 - , Dichte der ... 196
 - , Erwartungswert der ... 196
 - , Modalwert der ... 196
 - , Varianz der ... 196
 - Gleichung von Bienaymé 117
 - Gleichverteilung 116, 128, 192ff
 - , eindimensionale ... 116, 128, 192ff
 - , χ^2 -Anpassungstest auf ... 198f
 - , Dichte der ... 192

Gle – Haen

- , Erwartungswert der ... 193
- , Konfidenzintervall für die Differenz der Parameter einer ... 197f
- , Schätzer für die Parameter einer ... 128, 197
- , Simulation von Realisationen aus einer ... 193f
- , Varianz der ... 193
- , Verteilungsfunktion der ... 193
- , zentrale Momente der ... 193
- , zweidimensionale ... 194f
- , Dichte der ... 194
- Gleichverteilungshypothese im loglinearen Modell für $r \times s$ -Tafeln
 - , bedingte ... 434
 - , totale ... 434
- gleitender Durchschnitt 660ff, 726f
 - , Auffüllen der Reihe der ... 664, 726f
 - , ... gerader Ordnung 660ff
 - , gewogener ... 667, 726f
 - , Rekursionsformeln für ... 660f
 - , ... ungerader Ordnung 660f
- gleitender Median 849ff
 - , ... gerader Ordnung 850
 - , ... ungerader Ordnung 850
- Gliederungszahl 55
- Glivenko 121
- Glivenko-Cantelli
 - , Satz von ... 121
- Glücksspiel 9ff, 95
- Glyph 860
- GM-Schätzer 900, 906
 - , Eigenschaften der ... in der Meta-Analyse 1018ff
- Gnekow, A.K. 1072
- Goel, P.K. 1052
- Goldberg, M.F. 799, 1054
- Goldschen, D.Y. 925
- Goldsmith, P.L. 586, 1054
- Gompertz 642, 648ff
 - , ...-Kurve 642, 648ff
- Goodkin, D.E. 1074
- Goodman, L.B. 455, 1054
- Gorton, H.C. 799, 1071
- Gosset, W.S. 668, 1054
- Gozdschen, D.Y. 1070
- Graf, N. 1072
- Grafarend, E. 634f, 1054, 1057
- graphische Darstellung von Daten 19, 22, 835ff
- Gravitationsgesetz 4
- Gray, H.L. 821, 1054
- Graybill, F.A. 335, 567, 595, 608, 628, 631, 634, 1054
- Greenhouse, J.B. 1052
- Greenland, S. 919f, 1054, 1068
- Grenander, U. 640, 1054
- Grenzwertsatz
 - , Poisson'scher ... 122, 213
 - , zentraler ... 122, 138, 143, 147, 203, 272, 322, 899, 943
- Griechisches Alphabet 1043
- griechisch-lateinisches Quadrat 361, 628
- gross-error sensitivity 863
- Grubbs, F.E. 341, 345ff, 731, 758, 944, 1054
- Grubbs
 - , Ausreißertest von ... 344f
 - , Grubbsches Maßmodell 944
- Grube, A. 1048
- Grundgesamtheit 18, 269ff, 305ff
 - , Abgrenzung der ... 305ff
 - , endliche ... 305
 - , fiktive ... 306f
 - , konkrete ... 306f
 - , unendliche ... 305
- Grundraum 91ff, 97, 103
- Gruppenerneuerung 807f
- Gruppeneratzstrategie 780
- gruppensequentieller Versuch 1016
 - , self-designing ... 1016
 - , self-organizing ... 1016
- Gruppenumfang
 - , flexibler ... 1016
- gruppierte Daten
 - , arithmetisches Mittel bei ... 32
 - , explorative Kenngrößen für ... 831f
 - , Median bei ... 33
 - , mittlere absolute Abweichung vom ... 43
 - , Quantile bei ... 34
 - , Varianz bei ... 44ff
- Gülicher, H. 1049
- Güteeigenschaften von Schätzern 125f
- Gütfunktion eines Tests 135ff, 254, 381
- Güteprüfung für Zeitreihenmodelle 639, 678, 690f
- Guilford 376
 - , C-Note nach ... 376
- Gurevitch, I. 981, 1062
- Gutjahr, W. 376, 1054
- Guttman, L. 455, 1054
- Habermann, S.J.** 1054f
- Häder, M. 965, 1055
- Häder, S. 965, 1055
- Haenszel, W. 918ff, 1066

- | Häufigkeit | Häu – Har |
|--|--|
| -, absolute ... 20f, 26, 93 | |
| -, bedingte relative ... 99 | |
| -, Klassen... 27 | |
| -, relative ... 20f, 24, 26ff, 93f, 121, 123, 409f | 1000, 1004, 1006f, 1010, 1012ff, 1021f, 1048f, 1052, 1054ff, 1064, 1066f, 1072 |
| Häufigkeitspolygon 21 | Hartung |
| Häufigkeitsverteilung 21, 24, 37, 40f, 72, 112, 143 | -, Kombination der Dickey-Fuller p-Werte nach ... 964 |
| -, absolute ... 21 | --, ... mit Fisher Probits 964 |
| -, eingleipflige ... 37f, 47, 49 | -, nichtnegativer minimalverzerrter Schätzer nach ... 906f |
| -, linksschiefe ... 38f, 47 | --, ... für den Heterogenitätsparameter im RE-Modell 906f |
| -, linkssteile ... 38 | -, ...-Methode 959ff |
| -, mehrgipflige ... 37f | --, ... zur Kombination abhängiger p-Werte oder Tests 959ff |
| -, rechtsschiefe ... 38f, 47f | --, kombinierter p-Wert nach der ... 960 |
| -, rechtssteile ... 38 | --, ... für Probits 960 |
| -, relative ... 21 | --, ... für Fisher-Probits 960 |
| -, symmetrische ... 38f, 47 | -, ...-Konfidenzintervall für Heterogenitätsmaße 973f |
| -, U-förmige ... 37f | -, Quadratformen nach ... 908, 996f |
| häufigster Wert 35 | -, ...-Schätzer für den Heterogenitätsparameter im RE-Modell 972 |
| Haiger 904, 566 | --, median unverzerrter ... 1000 |
| Hájek, J. 242, 1055 | --, Pseudo-Maximum-Likelihood ... 1000 |
| Halbsteigung 854 | -, ...-Schätzer für Heterogenitätsmaße 973f |
| Hald, A. 214, 1055 | --, median unverzerrte ... 973f |
| half-slope 854 | -, stabiler Schätzer für die Varianz des Schätzers für den Gesamteffekt im ERE-Modell von ... 951 |
| Halley, E. 12 | -, ...-Verfahren 981ff, 991ff |
| Hallier, E. 895, 921, 1072 | --, ... für das Effektmaß Quotient der Mittelwerte im FE- und RE-Modell mit normalverteilten Daten 981ff |
| Hamaker, H.C. 391, 1055 | --, exakter Schätzer nach dem ... 981 |
| Hamilton, J.D. 963, 1055 | --, exaktes Konfidenzintervall nach dem ... 981 |
| Hampel, F.R. 872, 1048 | --, ... für das Effektmaß Standardisierte Mittlere Differenz im FE-Modell mit normalverteilten Daten 991ff |
| Hampel-Schätzer 872 | --, approximatives Konfidenzintervall nach dem ... 993 |
| Hanau, A. 637, 1055 | --, exakter Schätzer nach dem ... 992 |
| Hanauscher Schweinezyklus 637 | --, exaktes Konfidenzintervall nach dem ... 992 |
| Hannan, E.J. 640, 687, 1055 | Hartung-Böckenhoff-Knapp |
| Hannon 889, 918 | -, direkter Varianzschätzer von ... 942f, 945 |
| Harden 379 | Hartung-Knapp |
| Harman, H.H. 379, 1055 | -, ...-Konfidenzintervall für den Heterogenitätsparameter im RE-Modell 996f |
| Harmonische 716 | -, ... -Methode 956 |
| harmonische Analyse 639, 641, 717ff | -, ... Studien-Cluster 910f |
| harmonischer Prozeß 699, 701f | |
| -, schwach stationärer ... 699 | |
| -, Spektraldichte des ... 701f | |
| harmonische Schwingung 641 | |
| harmonisches Mittel 36f | |
| -, gewogenes ... 37 | |
| Harrington, D.P. 928, 1053 | |
| Hartley, H.O. 185, 189, 227, 344, 1023, 1051, 1067 | |
| Hartung, J. 120, 314, 328, 332, 337, 341, 358, 363, 378f, 450, 545, 549, 600, 608, 628, 631, 634, 635, 678, 763, 809, 820, 826, 838, 847, 861, 889, 891, 893ff, 897, 901, 906ff, 910, 917f, 921, 934, 943ff, 949f, 952, 954ff, 964, 969, 971, 973, 976, 978, 980, 991, 994ff, 998, | |

Har – Het

- , Test über die globale Hypothese bei der Kombination unabhängiger Tests nach der ... 956
- , ...-Verfahren 978f
- , ... für das Effektmaß Differenz der Mittelwerte im FE-Modell mit normalverteilten Daten 978f
- , exakter Schätzer nach dem ... 978
- , exaktes Konfidenzintervall nach dem ... 978f
- Hartung, J.E. 1058
- Hasemann, J.K. 421, 1061
- Hasman, A. 1057, 1059
- Hassler, U. 960, 964f, 1051, 1061
- Hastings 930, 1026f, 1040
- Hastings-Approximation 1026f, 1040
- Haupteinflußfaktor 362
- Hauptereignis 765f
- Hauser, S. 298, 315, 1061
- Hausken 889, 918
- Havarie-Erneuerung 807
- Hazardquotient 927
 - , Kombination von ... 931
 - , kombiniertes Konfidenzintervall für den ... 927
 - , Konfidenzintervall für den ... 927
 - , konstanter ... 927
 - , logarithmierter ... 927
 - , Konfidenzintervall für den ... 927
 - , Schätzer für den ... 927
 - , Schätzer für den ... 927
- Hazardrate 218, 775, 889f, 924ff, 940f
- , Cox Proportional Hazard Regression Model für die ... 824, 940f
- , Quotient von ... (s. a. Hazardquotient) 889f, 927ff
- Hazard Ratio (s. a. Hazardquotient) 890, 925ff
 - , logarithmierter ... 890
- Hazard Regression Model 940f
 - , Proportional ... 824, 940f
 - , Cox ... 824, 940f
- Hebbel, H. 666, 714, 722, 1062
- Hedges, L.V. 906, 911f, 931, 948, 951, 955f, 960, 972f, 981, 984, 1052, 1062
- Hedges-Schätzer
 - , ... für den Heterogenitätsparameter im RE-Modell 906f, 948
 - , verallgemeinerter ... 911, 951
- Heiler, S. 640, 714, 722, 827, 1062, 1067
- Heine, B. 341, 636, 1004, 1007, 1058, 1062, 1067
- Heisenberg 2

- Heisenbergsche Unschärferelation 2
- heisse Reserve
 - , System mit ... 762f, 809f, 812
- Helmert, F.R. 11, 629, 634ff, 1062
- Henn, R. 1056
- Henschler 572
- Heritabilitätskoeffizient 631, 892, 968, 970
- , verallgemeinerter ... 892f
- Herrendörfer, G. 1068
- Heterogenität
 - , ...maße 892
 - , ...parameter 906, 948, 951ff, 996ff
 - , ... im RE-Modell 906f, 948, 996ff
 - , DerSimonian/Laird-Schätzer für den ... (s.a. DSL-Schätzer) 906f
 - , explizite Schätzer für den ... 996
 - , Hartung-Knapp-Konfidenzintervall für den ... 996f
 - , Hedges-Schätzer für den ... 906f, 948
 - , implizite Schätzer für den ... 999ff
 - , iterativ definiertes Konfidenzintervall für den ... 998
 - , iterative Schätzer für den ... 999ff
 - , Konfidenzintervalle vom Waldschen Typ für den ... 999
 - , Likelihood-Methoden für den ... 998f
 - , ... vom Waldschen Typ 998f
 - , Mandel-Paule-Schätzer für den ... 999
 - , Maximum-Likelihood-Schätzer für den ... 1000f
 - , restringierter ... 1001
 - , Pseudo-... 1000
 - , median unverzerrter Schätzer nach Hartung für den ... 1000
 - , ML-Schätzer für den ... 1000f
 - , nichtnegativer minimalverzerrter Schätzer für den... 906f
 - , positiver Schätzer für den ... 907
 - , Pseudo-Maximum-Likelihood-Schätzer nach Hartung für den ... 1000
 - , REML-Schätzer für den ... 1001
 - , restringierter Maximum-Likelihood-Schätzer für den ... 1001
 - , verallgemeinerter Welch Test für den ... 998
 - , Schätzung des ... 906, 996ff
 - , ... im ERE-Modell 951
 - , verallgemeinerter Hedges Schätzer für den ... 951
 - , Test über die globale Hypothese bei der Kombination unabhängiger Tests nach ... 956
 - , ... von Studien 890
 - , extreme ... 891
 - Heterosuggestion 371f

hierarchische Klassifikation

- , einfache ... 630
- hierarchisches Modell 434, 465, 477, 483ff
- Higgins, J.P.T. 892f, 969, 971, 976, 1062
- Higgins-Thompson
 - , Maße für die Inhomogenität nach ... 969
- Hilferty, M.M. 955, 960f, 964, 1030, 1073
- Hill, R.C. 1053
- Himmelblau, D.M. 642, 793, 1062
- Hinderer, K. 1062
- hinge 828ff
 - , ...-Abstand 829, 831
 - , mid... 829
 - , oberer ... 828ff
 - , unterer ... 828ff
- Hinks 888
- Histogramm 21ff, 27ff, 825, 827, 838ff
 - , nicht-normiertes ... 840, 844
- Hjorth, U. 232f, 794, 1062
- Hjorth-Verteilung (s. auch IDB-Verteilung) 232ff, 794
 - , Ausfallrate der ... 233
 - , bedingte Überlebenswahrscheinlichkeit bei der ... 233
 - , Dichte der ... 232
 - , Erwartungswert der ... 233
 - , Schätzer für die Parameter der ... 234
 - , Überlebenswahrscheinlichkeit bei der ... 233
 - , Varianz der ... 233
 - , Verteilungsfunktion der ... 232
- Hoaglin, D.C. 826f, 837, 1063, 1072
- Hochpaß-Filter 664, 705ff
 - , idealer ... 706
- Hochrechnung
 - , freie ... 301
 - , gebundene ... 301
- Hodges 278
- Hoffmann, M. 1061
- Hohenloser, S. 1061
- Holland, P.W. 1049
- Hollander, M. 249, 527, 614, 623, 780f, 784, 1023, 1063
- Hollander-Proshan-Test
 - , ... gegen NBUE-Lebensdauerverteilungen 784f
 - , ... gegen NBU-Lebensdauerverteilungen 780f
 - , kritische Werte des ... 781
 - , ... gegen NWUE-Lebensdauerverteilungen 784f
 - , ... gegen NWU-Lebensdauerverteilungen 780f
 - , kritische Werte des ... 781
- Holling, H. 1057, 1059

Hie – Hug

- Holm, S. 567f, 628, 1063
- Holm
 - , Prinzip von ... 567f
 - , Test auf Unabhängigkeit von p Meßreihen mittels multipler Vergleiche nach ... 567f
- Holmes, R.B. 600, 1063
- homogenes FE-Modell 900f
 - , Gesamteffekt im ... 900
 - , Schätzer für den ... 900
 - , Gauß-Markov ... 900
 - , Varianz des ... 900
- homogene Studien und Prognosen 970
 - , signifikant ... 970, 997
- Homogenität
 - , ...analyse 970
 - , Cochran's ...statistik 905
 - , Test auf ... in der 2 x 2-Tafel 418ff
 - , erforderlicher Stichprobenumfang beim ... 419ff
 - , exakter ... 421f
 - , ... gemäß der Approximation der Gütfunktion des χ^2 -Tests 419f
 - , ... nach Casagrande/Pike/Smith 420f
 - , ... nach der Arcus-Sinus-Formel 420
 - , ... von Studien 890
 - , Test auf ... 900f
 - , ... für Effektmaße 917
 - , ... von Cochran 901
 - , ...hypothese 412, 418, 433, 901
 - , ...test
 - , ... für Effektmaße 917
 - , ... von Cochran 900
- Homoskedastizität 678
- Horst, P. 1054
- Horvitz, D.G. 299, 1063
- Horvitz-Thompson-Varianzformel 299
- Hotelling, H. 362, 556f, 899, 943, 1023, 1063
- Hotelling-Pabst-Statistik 556f
- Hovsepian, J.A. 781, 1063
- h-spread 826, 829
- Hsu, J.C. 1063
- Hsu, P. 1053
- Huber, P.J. 827, 872, 1048, 1063
- Huber-k-Schätzer 869ff
 - , nicht-skaleninvarianter ... 869f
 - , Bruchpunkt des ... 870
 - , Einflußkurve des ... 870
 - , ψ -Funktion des ... 869
 - , ρ -Funktion des ... 869
 - , skaleninvarianter ... 870ff
- Hughes, R.A.C. 1074

Hum – Int

- Hume 9
- Hunter, L.C. 823, 1048
- Huygens, C. 10
- Hyndman, R.J. 895, 1066
- hypergeometrische Verteilung 207ff
 - , Approximation der ... 208
 - , Erwartungswert der ... 207
 - , multidimensionale ... 210
 - , Schätzer für den Anteilswert der ... 208
 - , Schätzer für den Umfang der Grundgesamtheit bei der ... 209
 - , Varianz der ... 207
- Hypothese 3f, 6, 9, 133f, 138f
 - , Alternativ... 133f
 - , einseitige ... 138
 - , Homogenitäts... 901
 - , Null... 133f, 138f
 - , zweiseitige ... 139, 178
- Hypothesenprüfung 15
- IDB-Verteilung (s. auch Hjorth-Verteilung) 232ff, 794
 - , Ausfallrate der ... 233
 - , bedingte Überlebenswahrscheinlichkeit bei der ... 233
 - , Dichte der ... 232
 - , Erwartungswert der ... 233
 - , Schätzer für die Parameter der ... 234
 - , Überlebenswahrscheinlichkeit bei der ... 233
 - , Varianz der ... 233
 - , Verteilungsfunktion der ... 232
- identifizierbarer Effekt 355
- IFRA-Lebensdauerverteilungen 782ff, 786f
 - , cttot-Test gegen ... 784
- IFR-Lebensdauerverteilungen 776ff, 780, 782, 786f, 808f
 - , cttot-Test nach Epstein gegen ... 778
 - , Proschan-Pyke-Test gegen ... 777
- imaginäre Einheit 700
- implizit definierte Schätzer für den Heterogenitätsparameter im RE-Modell 99ff
- Imaginärteil 700
- Impulsresponsefunktion 703
- increasing failure rate (s. auch IFR) 776
- increasing failure rate average (s. auch IFRA) 783
- Index 19, 55, 57ff, 62f, 66
 - , Distanz... 379
 - , einfacher ... 55, 57f
 - , Fisherscher Ideal... 70
 - , Mengen... 65 f
- , Preis... 63, 66ff
- , Sub... 66ff
- , Wert... 62
- Indexzahl 55, 57, 62
- , Zeitreihe einfacher ... 57ff
 - , Umbasierung einer ... 59f
 - , Verkettung einer ... 60
- , zusammengesetzte ... 55, 57, 62, 65
- Indifferenzpunkt 387f, 391f, 398
- indirekte Beobachtung 326
- induktive Statistik 2f, 12, 124
- ineffiziente Portfolioauswahl 1008
- influence curve 863f
- Information
 - , a-priori ... 300
- inhomogenes FE-Modell 900f
 - , Gesamteffekt im ... 900
 - , Konfidenzintervall für den ... 901
 - , Schätzer für den ... 900
 - , unverzerrter ... mit minimaler Varianz 900
 - , Varianz des ... 900f
 - , Mischung aus ... und RE-Modell 910f
 - , Konfidenzintervall für den Gesamteffekt bei der ... 911
 - , Schätzer für den Gesamteffekt bei der ... 911
 - , Varianz des ... 911
- inhomogenes Modell 740
- inhomogene Studien und Prognosen
 - , signifikant ... 970, 997
- Inhomogenität (s. a. Heterogenität)
 - , ... von Studien 890, 968ff
 - , ...maß 892, 968ff
 - , Maße für die ... des RE-Modells 968ff
 - , ... nach Hartung 969
 - , ... nach Higgins-Thompson 969
- innerer Zaun 835
- inner-fence 835
- Inspektionszeitraum bei Straßenverkehrs-signalanlagen 817f
- Instandhaltung 806
- Instandhaltungsstrategie 745
- Instandsetzung 806, 823
 - , vorbeugende ... 824
- instationärer Prozeß 684f
- Intaktwahrscheinlichkeit 806
- Intelligenztest 269, 376f
- Intelligenzquotient 376f, 558f
 - ... nach Wechsler 376f
 - , Stanford-Binet... 558f
- Intensität 783
- Intent-To-Treat Population 910
- Interaktion (s. auch Wechselwirkung) 360f
- interlaboratorieller Vergleich 337

interne geschlossene Frage 309
 Interpolation nach Laubscher 1037
 interquartile range 866
 Intervall
 -, Stichproben... 296
 Intervallschätzung (s. auch Konfidenz-, Prognose- und Toleranzintervall) 157, 166ff, 203ff, 214, 220ff
 -, benötigter Stichprobenumfang unter Normalverteilungsannahme bei ... 166ff
 --, ... zur Einhaltung absoluter Genauigkeiten 166ff
 --, ... zur Einhaltung prozentualer Genauigkeiten 173 ff
 Interview-Umfrage 310f
 Intraklasskorrelationskoeffizient 290, 631, 892, 968
 -, verallgemeinerter ... 968
 --, ... für heteroskedastische Fehlervarianzen 968
 intralaboratorieller Vergleich 337
 Inventur auf Stichprobensbasis 269ff
 inverse χ^2 -Methode 955
 -, Test über die globale Hypothese bei der Kombination unabhängiger Tests nach der ... 955
 inverse Normal-Methode 955
 -, ... für unabhängige Zeitreihen 964
 -, gewichtete ... 955
 -, Test über die globale Hypothese bei der Kombination unabhängiger Tests nach der ... 955
 inverse Regression 335
 inverse Verteilungsfunktion
 -, verallgemeinerte ... 114
 inverse Wilson-Hilferty Transformation 960
 invertierbarer Prozeß 678
 Investitionsberatung 947ff, 960f, 967
 -, ... mit der Delphi Methode 967
 -, ... im EFE-Modell 950
 -, ... im ERE-Modell 951f
 -, ... im FE-Modell 947f
 -, ... im RE-Modell 948f
 Investitionsgewinn 947
 Investitionsprognose 960f
 Irle, A. 1063
 irreguläre Komponente einer Zeitreihe 639ff
 Irrtumswahrscheinlichkeit 129
 iterativ definierte Konfidenzintervalle für den Heterogenitätsparameter im RE-Modell 998
 iterative Delphi Methode 967

Int – Kau

iterative Schätzer für den Heterogenitätsparameter im RE-Modell 999ff
 iteratives Verfahren zur Schätzung erwarteter Häufigkeiten unter einer Hypothese im saturierten loglinearen Modell für mehrdimensionale Kontingenztafeln 477ff
 Jaeckel, L.A. 827, 1063
 jährliche Volatilität 1005
 Jenkins, G.M. 639f, 678, 685ff, 737, 1049
 Jennison, C. 1016, 1063
 Jensen 113, 1020f
 Jensensche Ungleichung 113, 1020f
 Jodrell Bank 888
 Jöckel, K.-H. 608, 809, 820, 1058, 1063
 Johnson, N.L. 185, 1051
 Johnson, S.R. 1053
 Johnston 588
 Jonkman, J.N. 908, 1014, 1070
 Jorch, N. 1072
 judgement sampling 316
 Jurečková, J. 827, 1063
 Just, H. 1061
 Justierung 322, 335
 Kaiser 343
 Kaisersche Nachweisgrenze 343
 Kakwani, N.C. 608, 1063
 Kalbfleisch, J.D. 1063
 Kalendereffekt 638
 Kalibration 322, 335
 Kalin, D. 763, 809, 1058
 Kall, P. 1047, 1063
 Kalman, R.E. 640, 1063
 Kalman-Filter 640
 kalte Macro-Reserve 763, 809, 811ff
 kalte Micro-Reserve 763, 809, 812ff
 kalte Reserve 762f, 811
 Kant, I. 8f
 Kaplan, E.L. 824, 930, 933, 935ff, 1064
 Kaplan-Meier
 -, ... Kurven 930, 934
 --, Kombination von ... 936f
 ---, ... im FE-Modell 936f
 ---, ... im RE-Modell 936f
 -, ... Schätzer für die Survivalfunktion 933
 -, ... Survival Analysis 744, 930ff
 Karelitz 557
 Karl der Große 12
 Kastenzeichnung 835
 kategorielles Merkmal 861
 k-aus-n-System 752

Kay – Kol

- Kaye, M.P. 1070
 Kellow 889, 915, 918
 Kelvin 796
 Kempthorne, O. 5, 363, 1064
 Kendall, M.G. 81, 249, 251, 553, 559f,
 563, 640, 1023, 1064, 1067
 Kendall
 -, Schätzer für die partielle Korrelation
 nach ... 363
 Kendallsche K-Statistik 249, 251, 560
 Kendallscher Rangkorrelationskoeffizient
 81, 559ff, 563
 -, Test über den ... 560f
 Kendall's τ 559, 563
 Kenngrößen
 -, deskriptive ... 15
 -, empirische ... in der EDA 825, 827ff
 -, ... für den Zusammenhang von zwei
 Zeitreihen 639, 727ff
 -, ... von Verteilungen 112
 Kepler, J. 11
 Kerbe 835
 Kern 841ff
 -, Dreiecks... 842
 -, Normal... 843
 -, Rechtecks... 842
 Kernschätzer 827, 841ff
 Kernschätzung 825, 827, 841ff
 Ketellapper, R.H. 827, 1064
 Keuls 628
 Klapper, M. 895, 945, 1064
 Klassen von Lebensdauerverteilungen
 774ff
 Klassenbildung 26, 831
 -, empirische Verteilungsfunktion bei ...
 29
 -, ... von Kovariablen 910
 Klassenbreite 27f, 33ff
 Klasseneinteilung 27
 Klassengrenze 26, 33
 Klassengröße
 -, natürliche ... 357
 Klassenhäufigkeit 26ff, 31, 183
 -, absolute ... 26ff
 -, relative ... 26ff
 Klassenmitte 26, 29ff
 Klassifikation
 -, einfache hierarchische ... 359, 630ff
 -, ... der Zuverlässigkeitsskala eines
 monotonen Systems 759f
 Klassifizierung von Filtern 674, 705ff
 Klein, J.P. 928, 941, 1064
 Kleiner, B. 1050
- Kleinste-Quadrat-Methode 128f, 574ff,
 587, 590, 596, 604, 640, 642ff, 666ff,
 723ff, 736ff, 791
 Kleinste-Quadrat-Schätzer 575f, 578,
 590, 592, 595, 599
 klinische Prüfung 888
 klinische Relevanz 994ff
 klinischer Versuch 363ff
 -, Auswahl und Zuordnung von Versuchs-
 personen beim ... 367f
 -, ethische Probleme beim ... 366
 -, Vergleichbarkeit der Versuchsergebnisse
 beim ... 370f
 Klösener, K.-H. 341, 636, 944, 1058,
 1063f
 Klumpen 288ff, 357
 Klumpenstichprobe 288ff, 308, 316
 Klumpenstichprobenverfahren
 -, ... als einstufiges Verfahren 289ff
 -, Schätzer für die Summe der Merk-
 malswerte in der Grundgesamtheit beim
 ... 289ff
 ---, Varianz des ... 289ff
 -, ... als mehrstufiges Verfahren 291f
 -, Schätzer für die Summe der Merk-
 malswerte in der Grundgesamtheit beim
 ... 291
 ---, Varianz des ... 291
 ---, Schätzer für die ... 292
 Knapp, G. 889, 891, 895, 901, 908, 910,
 917f, 921, 943, 945, 954ff, 960, 978,
 980, 994, 996, 998, 1004, 1013ff, 1018,
 1022, 1048, 1057ff, 1064
 Koch, A. 1074
 Koeffizient
 -, Heritabilitäts... 892
 --, verallgemeinerter ... 892f
 -, Korrelations... 904
 --, Pearsonscher ... 904
 ---, Kombination von ... 904
 Köhle, D. 1069
 König, H. 640, 1064
 Köpcke, W. 1064
 Kohlas, J. 1055
 Kohler, H.-P. 894, 1068
 Kohärenz 731, 735
 Koller, S. 374, 1064
 Kolmogoroff, A.N. 11, 94, 183ff, 199,
 226ff, 240, 520ff, 526, 1023
 Kolmogoroffsche Axiome 94
 Kolmogoroff-Smirnov-Anpassungstest
 183ff, 199, 220, 226ff, 240
 -, ... auf Exponentialverteilung 226ff
 --, ... bei nichtspezifiziertem Parameter λ
 226ff

- Kol – Kom**
- , kritische Werte des ... 227
 - , ... bei spezifiziertem Parameter λ 226ff
 - , ... auf Normalverteilung 183ff
 - , ... bei nichtspezifizierten Parametern 184ff
 - , kritische Werte des ... 185
 - , ... bei spezifizierten Parametern 183f
 - , kritische Werte des ... bei spezifizierten Parametern 184
 - Kolmogoroff-Smirnov-Test zum Vergleich zweier unabhängiger Meßreihen 513, 520ff
 - , kritische Werte des ... 521
 - Kombination 96f
 - , ... abhängiger p-Werte 895, 930f, 958ff
 - , Hartung-Methode zur ... 959ff
 - , kombinierter p-Wert nach der ... 960
 - , ... für Probits 960
 - , ... für Fisher-Probits 960
 - , ... abhängiger Tests 958ff
 - , Hartung-Methode zur ... 959ff
 - , kombinierter p-Wert nach der ... 960
 - , ... für Probits 960
 - , ... für Fisher-Probits 960
 - , ... unabhängiger p-Werte 954ff
 - , ... unabhängiger Tests 954ff
 - , globale Hypothese bei der ... 954
 - , Test über die ... 954f
 - , ... bei Heterogenität 956
 - , ... nach der Fisher-Methode 955
 - , ... nach der gewichteten Inversen Normalmethode 955
 - , ... nach der Hartung-Knapp-Methode 956
 - , ... nach der inversen χ^2 -Methode 955
 - , ... nach der Stouffer-Methode 954f
 - , ... von Assoziationsmaßen 906
 - , ... von Hazardquotienten 931
 - , ... von Kaplen-Meier Kurven 936f
 - , ... von Korrelationskoeffizienten 904
 - , ... von kreuzkorrelierten Zeitreihen (s. a. kreuzkorrigiertes Zeitreihenpanel) 962ff
 - , ... von medianen Überlebenszeiten 937f
 - , ... von Prognosen 944ff
 - , ... bei geschätzten Korrelationen und heterogenen Prognosen (s. a. ERE-Modell) 951f
 - , ... bei geschätzten Korrelationen und homogenen Prognosen (s. a. EFE-Modell) 949
 - , ... bei unbekannten Korrelationen 958ff
 - , ... im EFE-Modell 949
 - , ... im ERE-Modell 951f
 - , ... mit geschätzten Zuverlässigkeit und Korrelationen 947ff
 - , ... mit Scoringpunkten 944f
 - , ... von Time-to-Event Studien 930ff
 - , ... von Überlebenswahrscheinlichkeiten 938
 - , ... von Überlebenszeiten 935ff
 - , ... von vergleichenden Überlebenszeitstudien 930ff
 - Kombinatorik 96
 - kombinierte Delphi-Prognose 966
 - kombinierte Prognosen
 - , ... bei unbekannten Korrelationen 958ff
 - , ... mit Scoringpunkten 944f
 - , direkte Varianzformal von Hartung-Böckenhoff-Knapp bei ... 945
 - , Konfidenzintervall für den Gesamteffekt bei ... 945
 - , Schätzer für den Gesamteffekt bei ... 944f
 - , ... mittels Delphi Methode 966
 - kombinierte Quote 319
 - kombinierter Odds Ratio 918ff
 - , Konfidenzintervall für den ... 919f
 - , ... von Mantel-Haenszel 919
 - , ... von Peto 920
 - , Schätzer für den ... 918ff
 - , ... von Mantel-Haenszel 918f
 - , logarithmierter ... 918f
 - , Varianz des ... 919
 - , Schätzer für die ... 919
 - , ... von Peto 918ff
 - , logarithmierter ... 918f
 - , Varianz des ... 919
 - , Schätzer für die ... 919
 - kombiniertes Delphi-Konfidenzintervall 966
 - kombinierte subjektive Wahrscheinlichkeit 966
 - , ... mittels Delphi Methode 966
 - komplexe Zahl 639, 700
 - , Betrag einer ... 700
 - , Imaginärteil einer ... 700
 - , Produkt von ... 700
 - , Realteil einer ... 700
 - , Summe von ... 700
 - komplexe Zufallsvariable 700
 - komplex konjugierte Zahl 700
 - Komponente einer Zeitreihe 640f
 - , glatte ... 639ff
 - , irreguläre ... 639ff
 - , saisonale ... 639ff
 - , Trend... 639, 641
 - , zyklische ... 641

Kom – Kon

- Komponentenzuverlässigkeit 747, 756, 809
- Konfidenzbereich 129
- , ... für die Parameter der Multinomialverteilung 211f
- , ... für eine quadratische Regressionsfunktion 590
- Konfidenzintervall 129ff
- , approximativer ... für das Effektmaß Standardisierte Mittlere Differenzen im FE-Modell mit normalverteilten Daten nach dem Hartung-Verfahren 993
- , asymptotisches ... 131
- , einseitiges ... 130f
- , exaktes ... für das Effektmaß Differenz der Mittelwerte im FE-Modell mit normalverteilten Daten nach dem Hartung-Knapp-Verfahren 978f
- , exaktes ... für das Effektmaß Quotient der Mittelwerte im alternativen FE-Modell mit normalverteilten Daten nach dem Hartung-Verfahren 987
- , exaktes ... für das Effektmaß Quotient der Mittelwerte im FE-Modell mit normalverteilten Daten nach dem Hartung-Verfahren 981
- , graphische Ermittlung des ... 982f
- , exaktes ... für das Effektmaß Quotient der Mittelwerte im RE-Modell mit normalverteilten Daten nach dem Hartung-Verfahren 983
- , exaktes ... für das Effektmaß Standardisierte Mittlere Differenzen im FE-Modell mit normalverteilten Daten nach dem Hartung-Verfahren 992
- , ... für das Absolutglied einer Regressionsgeraden 581
- , ... für das gemeinsame Mittel im GCM Problem 942
- , ... für das logarithmierte Relative Risiko RR 916
- , ... für das Maximum (Minimum) einer quadratischen Regressionsfunktion 591
- , ... für das Relative Risiko RR 915f
- , ... für den Erwartungswert des Regressanden bei vorgegebenem Wert der Regressoren der multiplen Regression 597f
- , ... für den Erwartungswert des Regressanden bei vorgegebenem Wert des Regressors
- , ... einer linearen Regression 582
- , simultanes ... 582f

- , ... einer quadratischen Regressionsfunktion 590
- , ... für den Gesamteffekt bei gescored kombinierten Prognosen 945
- , ... für den Gesamteffekt im EFE-Modell 949
- , ... für den Gesamteffekt im ERE-Modell 951f
- , ... für den Gesamteffekt im inhomogenen FE-Modell 901
- , ... für den Gesamteffekt im RE-Modell 907f
- , ... für den Hazardquotienten 927
- , ... für den Heterogenitätsparameter im RE-Modell 996ff
- , Hartung-Knapp... 996f
- , iterativ definiertes ... 998
- , ... vom Waldschen Typ 999
- , ... für den kombinierten Odds ratio 919f
- , ... für den logarithmierten Hazardquotienten 927
- , ... von Mantel-Haenszel 919
- , ... von Peto 920
- , ... für den Median 835f
- , ... für den mittleren Merkmalswert in der Grundgesamtheit bei einfacher Zufallsauswahl 275
- , ... für den Parameter einer Binomialverteilung 203ff
- , ... für den Parameter einer Exponentialverteilung 220ff, 788ff
- , ... für den Parameter einer Poissonverteilung 214
- , ... für den Pearsonschen Korrelationskoeffizienten 549f
- , ... für den Quotienten der Varianzen zweier unabhängiger Meßreihen aus einer Normalverteilung 524ff
- , ... für den Steigungsparameter einer Regressionsgeraden 581
- , ... für die Differenz der Mittelwerte zweier abhängiger Meßreihen aus einer Normalverteilung 534ff
- , ... für die Differenz der Mittelwerte zweier unabhängiger Meßreihen aus einer Normalverteilung 505ff
- , Stichprobenumfang beim ... 511ff
- , ... für die Differenz der Parameter einer Gleichverteilung 197ff
- , ... für die Kombination von Überlebenswahrscheinlichkeiten 938
- , ... für die Maße für die Heterogenität im RE-Modell von Hartung 969f
- , ... für die Parameter einer multiplen Regression 597

- ,-, simultane ... 597
- ,-, ... für die Parameter einer Normalverteilung 130f, 160ff
- , benötigter Stichprobenumfang beim ... 166ff
- ,-, ... für die Parameter einer Weibullverteilung 791ff
- ,-, ... für die Parameter eines saturierten loglinearen Modells für mehrdimensionale Kontingenztafeln 471ff
- ,-, ... für die Summe der Merkmalswerte in der Grundgesamtheit bei einfacher Zufallsauswahl 275
- ,-, ... für die Überlebenswahrscheinlichkeit 932
- ,-, ... für die Überlebenszeit 933f
- ,-, ... für die Varianz der Differenz zweier abhängiger Meßreihen aus einer Normalverteilung 538
- ,-, ... für die Varianz der Fehlerterme bei der linearen Regression 581
- ,-, ... für die Yuleschen Assoziationsmaße in der 2 x 2-Tafel 444ff
- ,-, ... für Quantile 236f
- ,-, generalisiertes ... 980
- , ... für das gewöhnliche Common Mean Problem 980
- , ... für das FE-Modell mit normalverteilten Daten 980
- ,-, kombiniertes Delphi-... 966
- ,-, ... vorgegebener Breite 1016
- ,-, simultanes ...
- , ... für das Absolutglied und den Steigungsparameter einer Regressionsgeraden 581
- , ... für den Erwartungswert des Regressanden bei vorgegebenem Wert des Regressors einer linearen Regression 582f
- , ... für die Parameter einer multiplen Regression 597
- Konfidenzstreifen 240f, 582ff, 593f, 933
- ,-, ... für die Survivalsfunktion 933
- ,-, ... für eine Regressionsgerade 582
- ,-, ... für eine unbekannte Verteilungsfunktion 240f
- ,-, simultaner ... für eine Regressionsgerade 582f
- Kongruenzmethode
- ,-, multiplikative ... 194
- Konjunkturzyklus 641
- konkrete Grundgesamtheit 306f
- konservativer Test 138
- konservatives Verhalten 138
- konsistenter Schätzer 125

Kon – Kor

- konsistenter Test 138
- Konsistenz 125, 131, 138
- konstanter Hazardquotient 927
- Konsumentenrisiko 384, 387, 393, 397
- Kontingenzkoeffizient
- ,-, Pearsonscher ... 449ff
- , ... für die r x s-Tafel 451f
- , ... korrigierter ... 451f
- , ... für die 2 x 2-Tafel 449f
- Kontingenztafel 72f, 82, 378, 408ff, 609
- ,-, Adjustieren von ... 501ff
- ,-, Logit-Modell bei ... 498ff
- ,-, saturiertes loglineares Modell für mehrdimensionale ... 464ff
- ,-, Skalierung in mehrdimensionalen ... 378
- ,-, Verteilungsannahmen bei ... 492ff
- ,-, zweidimensionale ... 425ff, 492ff
- ,-, 2 x 2-... 12, 82f, 411ff, 442ff
- kontinuierlicher Stichprobenplan 406
- Kontrolle
- ,-, laufende ... der Produktion 401ff
- , ... bei qualitativen Merkmalen 404ff
- , ... bei quantitativen Merkmalen 401ff
- ,-, Eingangs... 381,406
- ,-, End... 381,406
- ,-, Qualitäts... 381ff
- ,-, Stichproben ... 381,406
- ,-, Total... 381, 389, 392, 406
- ,-, Zuordnung durch historische ... 368
- Kontrollgrenze 335ff, 402ff
- Kontrollgruppe 914
- kontrollierbarer Einflußfaktor 362
- kontrolliertes Experiment 364
- Kontrollkarte 335ff, 402ff
- Kontrollvariable 365
- Konvergenz
- ,-, stochastische ... 121
- Konzentration 19, 50, 52
- Konzentrationsprinzip
- ,-, Auswahl nach dem ... 316f
- Korrektur
- ,-, Endlichkeits... 275, 307
- Korrelation 77ff, 119f, 299, 301, 545ff, 578f, 587, 590, 593
- ,-, bi-partielle ... 545, 563f
- ,-, ... der Schätzer für das Absolutglied und den Steigungsparameter einer Regressionsgeraden 578
- ,-, ... der Schätzer für die Parameter einer quadratischen Regressionsfunktion 590
- ,-, kanonische ... 545
- ,-, kombinierte Prognosen bei ... 958ff

Kor – Kri

- , multiple ... 545, 564ff
- , Nonsense... 77, 545
- , partielle ... 561ff
- , Schätzer für die durchschnittliche ... abhängiger p-Werte 959
- , ... von Wertpapierkursen 1010
- , Schein... 77, 545, 561
- , schwach stationäre ... 727
- , Stichproben... 546
- , Zusammenhang zwischen Regression und ... 578f
- , ... zweier Prozesse 728
- , empirische ... 728
- , schwach stationäre ... 727
- Korrelationsanalyse
- , multivariate ... 545
- Korrelationskoeffizient 11, 73f, 78ff, 301, 407, 442, 446ff, 450f, 547
- , Intraklass... 290, 631, 892
- , verallgemeinerter ... 968
- , ... für heteroskedastische Fehlervari- anzen 968
- , Kendallscher Rang... 559ff
- , multipler ... 596
- , ... nach Bravais-Pearson 73f, 79f
- , ... nach Fechner 78f
- , partieller ... 545
- , Pearsonscher ... 546ff, 561, 563, 565, 567, 578, 904
- , Kombination von ... 904
- , Rang... 545, 553ff
- , Spearmanscher Rang... 80, 553ff
- Korrelationsmatrix 550
- korrelierte Fehler
- , Regression bei ... 608
- Korrelogramm 676
- , Kreuz... 728
- korrigierende Wartung 807
- korrigierter Pearsonscher
- Kontingenzkoeffizient für die r x s-Tafel 451f
- Korte, B. 1056
- Kortmann, R.D. 1072
- Kospektrum 729, 734
- Kotowski, P. 804, 1064
- Koul, H.L. 827,920
- Kovariablen 361f
- , Klassenbildung von ... 910
- Kovarianz 117, 119
- , ... der Multinomialverteilung 210
- , ... der Schätzer für das Absolutglied und den Steigungsparameter einer Regressi- onsgeraden 578
- , ... der Schätzer für die Parameter der multiplen Regression 597
- , empirische ... 74
- , ... zweier Prozesse 728
- Kovarianzanalyse 361, 628
- Kovarianzmatrix 120, 640, 949
- Krämer, W. 1060
- Krafft, O. 1, 5, 363, 608, 1064
- Krapp, M. 967, 1048
- Krauth, J. 379, 1064
- Kreienbrock, L. 918, 1065
- Kreisdiagramm 22 f
- Kreuzamplitudenspektrum 730
- Kreuzklassifikation 358
- , zweifache ... 609, 624
- kreuzkorreliert
- , ... Zeitreihenpanel 744, 895, 960, 962ff
- , Prüfung auf Stationarität eines ... 744, 964
- , Kombination der Dickey-Fuller p- Werte nach Hartung zur ... 964
- , ... mit Fisher Probits 964
- , kombinierter Test auf Stationarität eines ... 964f
- , Interpretation des ... 965
- , Stationarität eines ... 744, 895, 964
- , kombinierter Test auf ... 964f
- , Interpretation des ... 965
- , Prüfung auf ... 964
- Kreuzkorrelogramm 728
- Kreuzspektralanalyse 639, 728ff
- Kreuzspektraldichte 729, 734
- Kreuzspektrum 729ff
- , Stichproben... 734
- , geglättetes ... 734
- Krishan, A.S.A. 1050
- Krishnaiah, P.R. 1048
- Kriterien für die Ordnungen eines ARMA- Prozesses 687
- kritischer Pfadvektor eines Multi-State- Systems 764f
- kritische Werte 314
- , ... der cttot-Teststatistik 778
- , ... der Dixon-Tests auf Ausreißer 347
- , ... der Hotelling-Pabst-Statistik 557
- , ... der Kendallschen K-Statistik 249
- , ... des Durbin-Watson-Tests 740
- , ... des Friedman-Tests auf signifikante Lokationsunterschiede 623
- , ... des Hollander-Proschan-Tests gegen NBU- (NWU-) Lebensdauerverteilun- gen 781
- , ... des Kruskal-Wallis-Tests 615
- , ... des Wilcoxon-Rangsummentests 518

- , ... des Wilcoxon-Vorzeichenrangtests 245
 -, ... für den Kolmogoroff-Smirnov-Anpassungstest
 --, ... auf Exponentialverteilung bei nicht-spezifiziertem Parameter 227
 --, ... bei nichtspezifizierten Parametern 185
 --, ... auf Normalverteilung bei spezifizierten Parametern 184
 -, für den Kolmogoroff-Smirnov-Test zum Vergleich zweier unabhängiger Meßreihen 521f
 -, ... für den Test auf ein Ausreißerpaaar 348
 -, ... für Youden's extremen Rangsummentest 339
 Krüger, H.-P. 414, 1065
 Krumm, F.W. 58f
 Kruskal, J.B. 379, 1065
 Kruskal, W.H. 455, 613ff, 617, 1023, 1054
 Kruskal-Wallis-Test 613ff, 617
 -, kritische Werte des ... 615
 Kühlmeyer, M. 155, 1065
 Künzi, H.P. 1047, 1049
 Kuhlmeier, N. 722, 1062
 kumulatives Beschädigungsmodell 783
 Kumulierungsregel von Dalenius und Hodges bei der geschichteten Zufallsauswahl 278
 Kundert, K.R. 378, 1065
 Kunert, J. 1059
 Kurtosis (s. auch Schiefe) 49
 Kuznets, S.S. 643, 1065
- ladder of power 833f, 850, 854ff
 lag 675, 744
 -, ...-Fenster 711ff, 734
 Lagemaß 31ff, 114, 116
 Lageparameter 19, 112
 lagged dependent variable model 736, 744
 Laird, N. 906, 932, 972, 984, 995, 1051
 λ-Maß 455ff
 Lambert, P.C. 918, 1071
 Lancaster, H.O. 378, 450, 1065
 Laplace, P.S. de 11f, 95ff
 Laplace-Wahrscheinlichkeit 95f
 Laspeyres 63ff
 lateinisches Quadrat 361, 628
 lateinisches Rechteck 361
 Latscha, R. 1053
 Laubscher 1037
 -, Interpolation nach ... 1037
 laufende Kontrolle der Produktion 401ff
 -, ... bei qualitativen Merkmalen 404ff
- Kri – Lel
- , ... bei quantitativen Merkmalen 401ff
 Laur 573
 Least-Squares-Ansatz (s. auch Methode der kleinsten Quadrate) 688f
 -, bedingter ... 688
 -, unbedingter ... 688f
 least squares estimator 575
 Least-Squares-Schätzer 736f
 -, generalisierter ... 736
 -, gewöhnlicher ... 737
 Lebensdauer 5, 70, 91, 103f, 111f, 124, 126, 150, 193, 218ff, 241, 261, 264ff, 554, 745, 774ff, 809ff, 817f, 820, 822
 -, mittlere ... 219f, 261, 745, 791ff, 797f, 803, 810ff
 Lebensdauerprüfung 221, 262, 267, 381, 787, 791, 794ff
 -, beschleunigte ... 745, 794ff
 -, vergleichende ... 824
 -, zensierte ... 787f
 Lebensdauerverteilung 218ff, 351, 745, 774ff, 794ff, 808, 810, 812
 -, DFR-... 776ff, 786f
 --, cttot-Test nach Epstein gegen ... 778
 --, Proschan-Pyke-Test gegen ... 777
 -, DFRA-... 782f, 786f
 --, cttot-Test gegen ... 784
 -, IFR-... 776ff, 786f
 --, cttot-Test nach Epstein gegen ... 778
 --, Proschan-Pyke-Test gegen ... 777
 -, IFRA-... 782ff, 786f
 -, ctott-Test gegen ... 784
 -, Klassen von ... 774ff
 --, Beziehungen zwischen ... 786f
 -, NBU-... 779ff, 786f
 --, Hollander-Proshan-Test gegen ... 780f
 -, NBUE-... 784ff
 --, Hollander-Proshan-Test gegen ... 784f
 -, NWU-... 779ff, 786f
 --, Hollander-Proshan-Test gegen ... 780f
 -, NWUE-... 784ff
 --, Hollander-Proshan-Test gegen ... 784f
 -, Schätzten der ... 824
 Lebenshaltungsindex 66f
 Lee, A. 569, 1067
 Lee, C.-F. 899, 1065
 Legendre, A.M. 11
 Lehacher, W. 993, 1065
 Lehmann, E.L. 183, 242, 1019, 1050, 1065
 Lehmann, F. 194, 1070
 Leibniz 9
 Leitertafel 174ff, 325
 Lellouch, J. 1070

Ler – Log

Lernstichprobe 377
 level attained 137, 927
 Levene 531, 617f
 Levene-Test für Dispersionsvergleiche bei der einfachen Varianzanalyse 531, 617f
 Lewis, J. 1074
 Lewis, T. 827, 1049
 Li, Y. 901, 995, 1065
 Lienert, G.A. 376f, 1065
 Lienert
 -, Schulnote nach ... 376f
 Liew, C.K. 508, 1065
 Light, R.J. 439, 1066
 Likelihood 127f
 -, log-... 127
 Likelihood-Methoden für den Heterogenitätsparameter im RE-Modell 998f
 -, ... vom Waldschen Typ 998f
 Likelihood-Quotienten-Test im loglinearen Modell für r x s-Tafeln 435ff
 Lilliefors, H.W. 185, 227, 1065
 limiting quality 384
 lineare Regression (s. auch Regressionsgerade) 573ff
 -, ... bei unterschiedlicher Varianz der Beobachtungen 584f
 -, Bestimmtheitsmaß der ... 579
 -, ... durch einen vorgegebenen Punkt 584
 --, Schätzer für das Absolutglied bei der ... 584
 --, Schätzer für den Steigungsparameter bei der ... 584
 --, Fehlerterme bei der ... 574ff
 --, Erwartungswert der ... 576
 --, Varianz der ... 576ff
 ---, Konfidenzintervall für die ... 581
 ---, Schätzer für die ... 578
 ---, Test über die ... 580
 --, Verteilung der ... 577
 -, Konfidenzintervall für den Erwartungswert des Regressanden bei vorgegebenem Wert des Regressors bei der ... 582
 --, simultanes ... 582f
 -, ... ohne Absolutglied 584
 --, Schätzer für den Steigungsparameter bei der ... 584
 -, Prognoseintervall für den Regressanden bei vorgegebenem Wert des Regressors bei der ... 582
 -, Unbestimmtheitsmaß der ... 579
 -, Zusammenhang zwischen Korrelation und ... 578f

linearer Filter 660, 673ff, 703ff
 -, Amplitudendiagramm eines ... 703
 -, Wirkungsweise eines ... 705ff
 lineare Transformation 85, 108f, 119
 Linearisierung einer Punkteschar 825, 849f, 854ff
 Linearität
 -, Transformation auf ... 387ff
 linksschiefe Häufigkeitsverteilung 38f, 833
 linksseitig gestützte Normalverteilung 150
 linkssteile Häufigkeitsverteilung 38
 List, P.H. 1
 Ljung, G.M. 691, 1049
 Localio, A.R. 1049
 local shift sensitivity 864
 logarithmierter Hazardquotient 927
 --, Konfidenzintervall für den ... 927
 --, Schätzer für den ... 927
 logarithmierter Odds Ratio OR 915
 --, Schätzer für den ... 915
 --, Schätzer für die Varianz des ... 915
 logarithmiertes Relatives Risiko RR 915
 --, Konfidenzintervall für das ... 916
 --, Schätzer für das ... 915
 --, Schätzer für die Varianz des ... 915
 logarithmische Transformation 351f
 logistische Funktion 642ff, 653f
 -, Prognose mittels ... 644
 --, Schätzen der Parameter einer ... 643ff
 logistische Kurve 642ff
 Logit-Modell bei Kontingenztafeln 409f, 492, 498ff
 log-Likelihood 127
 loglineares Modell 409, 425ff
 -, ... für r x s-Tafeln 425ff
 --, Aufstellung des ... 429
 --, bedingte Gleichverteilungshypothese im ... 434
 --, χ^2 -Test im ... 435ff
 --, Homogenitätshypothese im ... 433f
 --, Likelihood-Quotienten-Test im ... 435ff
 --, Schätzer für die Parameter des ... 430
 ---, approximative Varianz der ... 430
 --, Symmetrietest nach Bowker im ... 440f
 --, Test mittels der Statistiken τ_A und τ_B im ... 439f
 --, totale Gleichverteilungshypothese im ... 434f
 --, Unabhängigkeitshypothese im ... 433f
 --, ... für 2 x 2-Tafeln 425ff
 --, Aufstellung des ... 425ff
 --, Entwicklung des ... 425
 --, Schätzer für die Parameter des ... 427ff
 ---, approximative Varianz der ... 428f

- | - | Log – McC |
|---|---|
| -, saturiertes ... für mehrdimensionale Kontingenztafeln 464ff | Mantel, N. 918ff, 926, 1066 |
| --, iteratives Verfahren zur Schätzung erwarteter Häufigkeiten unter einer Hypothese im ... 477ff | Mantel-Haenszel |
| --, Parameter des ... 465ff | -, Konfidenzintervall für den kombinierten Odds Ratio von ... 919 |
| ---, Konfidenzintervalle für die ... 471ff | -, Schätzer für den kombinierten Odds Ratio von ... 918f |
| ---, Schätzer für die ... 466ff | --, logarithmierter ... 918f |
| ----, Schätzer für die Varianz der ... 471ff | ---, Varianz des ... 919 |
| ---, Testen von Hypothesen über die ... 477ff | ----, Schätzer für die ... 919 |
| ----, Bestimmung der Freiheitsgrade beim ... 485ff | -, Schätzer vom ... Typ |
| ----, Partitionierung der Teststatistiken beim ... 488ff | --, ... für das Relative Risiko 920 |
| Lognormalverteilung 151f, 273, 276, 351, 794 | --, ... für die Risiko Differenz 920 |
| -, Dichte der ... 151 | MA-Prozeß (s. auch moving average Prozeß) 678, 681f |
| -, Erwartungswert der ... 151 | marginale Normalisierung 375 |
| -, Median der ... 152 | Margolin, B.H. 439, 1066 |
| -, Schiefe der ... 152 | Markov 900, 968f, 1018ff, 1022 |
| -, Varianz der ... 152 | Markowitz, H.M. 899, 1066 |
| -, Verteilungsfunktion der ... 151 | Marktforschung 311, 314ff |
| Log-Rank Test 824, 890, 926ff | -, primärstatistische ... 314 |
| -, weighted ... 928 | -, sekundärstatistische ... 314 |
| Lokationsmodell | -, Stichprobenverfahren in der ... 314ff |
| -, additives ... 514 | Maß |
| Lokationsparameter 514 | -, ... für die Heterogenität im RE-Modell 968ff |
| Lokationsverschiebung 833 | --, Hartung-Konfidenzintervalle für die ... 969f |
| Lorenz 50ff, 273 | -, ... für die Inhomogenität des RE-Modells 968ff |
| Lorenzkurve 50ff, 273 | --, ... nach Hartung 969 |
| Lorenzsches Konzentrationsmaß 52f | --, ... nach Higgins-Thompson 969 |
| Lotto 91, 93, 97f | -, Heterogenitäts... 892, 968ff |
| Lowe 70 | --, ... für Meta-Analysen im RE-Modell 968ff |
| L-Schätzer 880ff | --, Hartung-Schätzer für ... 973f |
| Lustig 323 | ---, median unverzerrte ... 973ff |
| MacKinnon, J.G. 963, 1065 | -, Inhomogenitäts... 892f, 968ff |
| Macro-Reserve | Maßzahlen 55f |
| -, kalte ... 763, 809, 811, 813f | -, statistische ... 55 |
| MAD 865 | maximale Verfügbarkeit 730 |
| Maddala, G.S. 964, 1065 | Maximum einer Häufigkeitsverteilung 49 |
| Maidhoff, R. 1060 | Maximum-Likelihood-Schätzer (s. a. ML-Schätzer) 126, 1000f |
| maintainability 806 | -, ... für den Heterogenitätsparameter im RE-Modell 1000f |
| Makambi, K.H. 907, 1004, 1013, 1057, 1060f, 1066 | --, Pseudo-... nach Hartung 1000 |
| Makridakis, S. 895, 1066 | --, restriktiver ... 1001 |
| Mandel, J. 602, 999, 1066 | Maximum-Likelihood-Verfahren 435ff, 464ff |
| Mandel-Paule | Maximum von Zufallsvariablen 108, 111 |
| -, ...-Schätzer für den Heterogenitätsparameter im RE-Modell 999 | Maxwell 5 |
| Mann, H.B. 249f, 513, 520, 669, 1066 | McCall 376 |
| Mann | -, T-Note nach ... 376 |
| -, Test auf Trend nach ... 249ff, 669 | |
| Manns, J.F. 1048 | |

Mck – Mer

- McKinnon 963
- , p-Werte des Dickey-Fuller Unit Root Tests nach ... 963
- McNemar 423f, 441
- , Test auf Symmetrie in der 2 x 2-Tafel von ... 423f, 441
- McPherson, K. 1074
- MDS 860
- Mean squared error (MSE) eines Schätzers 125f
- Median 32, 42, 87, 114, 826ff, 880
 - , ... bei gruppierten Daten 33
 - , mittlere absolute Abweichung vom ... 43
 - , ... der absoluten Abweichungen vom Median 865
 - , ... der Exponentialverteilung 220
 - , ... der Lognormalverteilung 152
 - , ... der Normalverteilung 144
 - , ... der t-Verteilung 154
 - , ... der Weibullverteilung 231
 - , gleitender ... 849ff
 - , Konfidenzintervall für den ... 835f
 - , Minimumseigenschaft des ... 42, 324
 - , mittlere absolute Abweichung vom ... 42ff, 87ff
 - , Robustheit des ... 861ff
 - , Varianz des ... 685
- median absolute deviation 865
- mediane Überlebenszeit 933f
- , Kombination von ... 937f
- Medianpunkt 853, 855
- median survival time (s. a. mediane Überlebenszeit) 934
- Median unverzerrter Pseudo Maximum Likelihood Gesamt-Schätzer (s. a. MUPML Gesamt-Schätzer) 978f, 981ff, 987, 992
 - , approximierender ... 979, 982, 992
 - , ... für das Effektmaß Differenz der Mittelwerte im FE-Modell für normalverteilte Daten 979
 - , ... für das Effektmaß Quotient der Mittelwerte im FE-Modell für normalverteilte Daten 982
 - , ... für das Effektmaß Standardisierte Mittlere Differenzen im FE-Modell für normalverteilte Daten 992
 - , exakter ... 978, 981ff, 992
 - , ... für das Effektmaß Differenz der Mittelwerte im FE-Modell für normalverteilte Daten 978
- , ... für das Effektmaß Quotient der Mittelwerte im FE-Modell für normalverteilte Daten 981
- , graphische Ermittlung des ... 982f
- , ... für das Effektmaß Standardisierte Mittlere Differenzen im FE-Modell für normalverteilte Daten 992
- , ... für das Effektmaß Quotient der Mittelwerte im alternativen FE-Modell für normalverteilte Daten 987
- , Varianz des ... 987
- , ... für das Effektmaß Quotient der Mittelwerte im RE-Modell für normalverteilte Daten 983f
- , Varianz des ... 984
- median unverzerrter Schätzer 973f, 1000
- , ... für den Heterogenitätsparameter im RE-Modell nach Hartung 1000
- , ... für Heterogenitätsmaße von Hartung 973f
- mehrdimensionale hypergeometrische Verteilung 210
- mehrdimensionale Kontingenztafel 464ff
- , saturiertes loglineares Modell für ... 464ff
- mehrdimensionales Datenmaterial 858
- mehrdimensionale Skalierung in Kontingenztafeln 378
- mehrdimensionales Merkmal 19, 72
- mehrfacher Stichprobenplan
 - , ... für qualitative Merkmale 389f
 - , ... für quantitative Merkmale 399ff
- Mehrfachnennung 309
- mehrphasige Mission 769ff
 - , Mindestzuverlässigkeit einer ... 770ff
 - , Transformation einer ... 772ff
- mehrstufiges Auswahlverfahren bei der Klumpenstichprobe 291f
- Meier, P. 824, 901, 930, 933, 935ff, 1022, 1064, 1066
- Meinertz, T. 1061
- Meisner, B. 1067
- Memmel, U. 1061
- Mengenindex 65f
 - , ... nach Laspeyres 65f
 - , ... nach Paasche 65
- Mengenoperationen 92ff
- Mengenschema (s. auch Warenkorb) 65
- Menges, G. 315, 1066
- Méré, de 10
- Merkmal 16ff
 - , diskretes ... 17
 - , eindimensionales ... 19
 - , Erblichkeit eines ... 892
 - , Klassifizierung von ... 16f

- Konzentration von ... 50ff
 -, mehrdimensionales ... 19, 72
 -, metrisches ... 17, 21, 23, 26, 31, 40, 50
 73
 -, nominales ... 17, 35, 82f
 -, ordinale ... 17, 21, 23, 26, 33, 79
 -, qualitativen ... 16
 -, quantitativen ... 16
 -, Quotierungs... 311, 318
 -, stetiges ... 17, 22
 -, zweidimensionales ... 72ff, 79ff
 Merkmalsausprägung 16, 20, 24, 26, 32,
 35
 Merkmalswert 16, 20, 270
 Merton, J. 1074
 Meßfehler 269, 320ff
 Meßreihe
 -, Test auf Unabhängigkeit von p ... mit-
 tels multipler Vergleiche nach Holm
 567f
 -, Vergleich von p unabhängigen ... 609ff
 -, Vergleich zweier ... 505ff
 -, Vergleich zweier abhängiger ... 533ff
 -, Vergleich zweier unabhängiger ... 505ff
 Meßzahl 58ff
 -, Preis... 66
 Meta-Analyse 887ff
 -, Eigenschaften der Schätzer in der ...
 1018ff
 -, ... für Delphi-Prognosen 965ff
 -, ... für diskrete Daten 913ff
 -, ... für die Binomialverteilung 913ff
 -, ... für die Poissonverteilung 920ff
 -, ... für normalverteilte Daten 976ff
 --, exakte Verfahren und spezielle Effekt-
 maße der ... 976ff
 --, Modell der ... 976f
 -, ... für Time-to-Event Daten 924ff
 -, ... für Überlebenswahrscheinlichkeiten
 938
 -, ... für Überlebenszeiten 924ff
 -, Heterogenitätsmaße für ... 968ff
 -, prospektive ... 1014ff
 --, adaptive Versuchsanlage zur ... 1016
 --, self-designing Versuchsanlage zur ...
 1016
 Meta-Regression 1013f
 -, RE-Modell der ... 1014
 Methode der kleinsten Quadrate 11, 128,
 574ff, 590ff, 595ff, 640, 642ff, 666ff,
 723ff, 736ff
 Methode von Box und Jenkins 639f, 678,
 686ff
 Methode von Verhagen 649
 metrische Skala 17
 metrisches Merkmal 17, 21, 23, 26, 31, 40,
 50,73
 Mettler 573
 Micro-Reserve
 -, kalte ... 763, 809, 812ff
 mideighth 829
 midextreme 829
 midhinge 829, 831
 midquartile 826, 880
 midrank 140, 244, 516, 554
 midsummary 829, 832
 Mikrozensus 19
 Milanese, C. 1074
 Miller, I. 826, 1066
 Miller, R.G. 628, 1066
 Milo 889, 918
 Milstone Hill 888
 Minamide, N. 608, 1066
 minimale Parallelschnittstruktur 750
 minimale Pfadmenge 750, 754, 757, 760,
 763
 minimale Pfadmengendarstellung der
 Strukturfunktion eines monotonen Sys-
 tems 754f
 minimale Pfadserienstruktur 750
 minimale Schnittmenge 750, 754, 760,
 763, 773
 minimale Schnittmengendarstellung der
 Strukturfunktion eines monotonen Sys-
 tems 754f
 Minimumeigenschaft des arithmetischen
 Mittels 44, 324
 Minimumeigenschaft des Medians 42, 324
 Minimum von Zufallsvariablen 108, 111f
 Mises, R.v. 94, 1066
 Mission
 -, mehrphasige ... 745, 769ff
 --, Mindestzuverlässigkeit einer ... 770ff
 --, Transformation einer ... 772ff
 Missionszuverlässigkeit 770ff
 Mitchell, W.C. 643, 1066
 Mitscherlich 642, 648ff
 --, ...-Funktion 642, 648ff
 --, Schätzen der Parameter der ... 648ff
 Mittel
 -, α -Gastwirth-Cohen-... 880, 882f
 -, α -getrimmtes ... 880f
 -, α -winsoriertes ... 881f
 -, arithmetisches ... 31f, 35f, 43, 83ff
 --, ... bei gruppierten Daten 32
 --, Minimumeigenschaft des ... 44, 324
 -, Gastwirth-... 883
 --, geometrisches ... 35ff

Mit – Mul

- , harmonisches ... 36f, 900
- Mittelwert 1, 31, 112
- mittlere absolute Abweichung vom Median 43, 87ff
- , ... bei gruppierten Daten 43
- mittlere Funktionszeit 820f
- mittlere Intaktzeit 808
- mittlere kumulierte Ausfallrate 783
- mittlere Lebensdauer 745, 791f, 794, 798, 803, 809ff, 817
- mittlerer Durchschlupf 389
- mittlere Rendite 897, 1005
- mittlere Reparaturzeit 820ff
- mittlerer Fehler (s. auch Standardabweichung) 324
- mittlerer Merkmalswert in der Grundgesamtheit 274f
- , Konfidenzintervall für den ... bei einfacher Zufallsauswahl 275
- , Schätzer für den ... bei einfacher Zufallsauswahl 274
- mittlerer quadratischer Fehler (s. auch Varianz) 125f, 301
- ML-Schätzer für den Heterogenitätsparameter im RE-Modell 1000f
- Modalwert 35, 116, 196
- , ... der gleichschenkligen Dreiecksverteilung 196
- Modell
 - , Berkson... 604f
 - , gemischtes ... 629
- Modellbetrachtung bei der einfachen Varianzanalyse 619
- Modellbildung
 - , statistische ... 123
- Modellidentifikation für Zeitreihen 639, 678, 686f
- Modellüberprüfung für Zeitreihen 690f
- model mixing 959
- modulare Repräsentation eines Fehlerbaums 769
- Modus (s. auch Modalwert) 35, 116
- Moeschberger, M.L. 928, 941, 1064
- Moivre, A.de 11
- Moment 118, 126
 - , empirisches ... 126
 - , Existenz von ... 142
 - , k-tes empirisches ... 126
 - , zentrales ... 118, 144, 193
 - , ... der eindimensionalen Dreiecksverteilung 193
 - , ... der Normalverteilung 144
 - momentane Verfügbarkeit 820ff

- Momentenmethode 127, 906, 983
- monotones System 750ff, 756ff
- , minimale Pfadmengendarstellung eines ... 750f
- , minimale Schnittmengendarstellung eines ... 750f
- , Strukturfunktion eines ... 751f
- , Zuverlässigkeitfunktion eines ... 756ff
- Moore, G.F. 759, 1066
- Moore, G.H. 1072
- Morgenstern, D. 210, 1066
- Morris, C.N. 1052
- Morrison, D.F. 628, 1066
- Moses 529
 - , Test von ... 529ff
- Mosteller, F. 1063
- moving average 660ff
- moving average Prozeß 640
 - , AR-Darstellung eines invertierbaren ... 682
 - , Autokorrelationen eines ... 681
 - , Autokovarianzen eines ... 681
 - , ... der Ordnung q 681f
 - , Invertierbarkeit eines ... 682
 - , Ordnung eines ... 687
 - , Spektraldichte eines ... 704
- M-Schätzer 866ff
 - , Berechnung von ... 874ff
 - , nicht-skaleninvarianter ... 866ff
 - , Bruchpunkt des ... 868
 - , Einflußkurve des ... 867
 - , skaleninvarianter ... 868ff
 - , Bruchpunkt des ... 869
 - , Einflußkurve des ... 868
- Müller, J. 317, 1066
- Müller, P.H. 978, 983, 1066
- Münzwurf 91, 103, 122
- multidimensionale Skalierung 379, 860
- Multikollinearität 598
- multi-level continuous sampling plan 406
- Multinomialverteilung 209ff, 498
 - , Erwartungswert der ... 210
 - , Kovarianz der ... 210
 - , Randverteilungen der ... 210
 - , Vergleich mehrerer ... 498
- multiple Korrelation 564ff
 - , Schätzer für die ... 565
 - , Test über die ... 566
- multiple Regression 595ff
 - , Bestimmtheitsmaß der ... 595ff
 - , Konfidenzintervalle für die Parameter der ... 597
 - , simultane ... 597
 - , Konfidenzintervall für den Erwartungswert des Regressanden bei vorgegebener

- Mul – Ney**
- nen Werten der Regressoren bei der ...
 - , 597f
 - , Modellreduktion bei der ... 598
 - , Reduktionstest bei der ... 598
 - , Schätzer für die Parameter der ... 597
 - , Erwartungswert der ... 597
 - , Kovarianz der ... 597
 - , Schätzer für die Varianz der Fehlerterme der ... 597f
 - , simultane Konfidenzintervalle für die Parameter der ... 597
 - , Test über die Parameter der ... 596
 - Multiple Regressionsrechnung 595ff
 - multipler Korrelationskoeffizient 565f, 596
 - , Schätzer für den ... 565f
 - , Test über den ... 566
 - multiples Bestimmtheitsmaß 565ff, 595f
 - , Schätzer für das ... 565ff
 - multiple Sklerose (MS) 902f, 908f, 972, 974ff, 997, 1000f
 - multiple Vergleiche 567, 614ff, 620, 623, 625, 628
 - , ... bei der einfachen Varianzanalyse 614ff
 - , ... bei der zweifachen Varianzanalyse mit Wechselwirkungen zwischen den Faktoren 625
 - , ... beim einfachen Blockexperiment 620
 - , Test auf Unabhängigkeit von p Meßreihen mittels ... nach Holm 567
 - Multiplikationssatz für bedingte Wahrscheinlichkeiten 99ff
 - multiplikative Kongruenzmethode 194
 - multiplikatives Zeitreihenmodell 641
 - multiplikativ saisonales ARIMA-Modell 685
 - Multi-State-System 746, 763ff
 - , Bestimmung des Systemzustandes eines ... 763f
 - , kritische Pfadvektoren eines ... 764f
 - multivariate Analyse 628
 - multivariate Korrelationsanalyse 545
 - multivariate Regressionsanalyse 374
 - multivariate Statistik 826, 857ff
 - multivariate Varianzanalyse 358, 374
 - MUPML Gesamt-Schätzer 978f, 981ff, 987, 992
 - , approximierender ... 979, 982, 992
 - , ... für das Effektmaß Differenz der Mittelwerte im FE-Modell für normalverteilte Daten 979
 - , ... für das Effektmaß Quotient der Mittelwerte im FE-Modell für normalverteilte Daten 982
 - , ... für das Effektmaß Standardisierte Mittlere Differenzen im FE-Modell für normalverteilte Daten 992
 - , exakter ... 978, 981ff, 992
 - , ... für das Effektmaß Differenz der Mittelwerte im FE-Modell für normalverteilte Daten 978
 - , ... für das Effektmaß Quotient der Mittelwerte im FE-Modell für normalverteilte Daten 981
 - , graphische Ermittlung des ... 982f
 - , ... für das Effektmaß Standardisierte Mittlere Differenzen im FE-Modell für normalverteilte Daten 992
 - , ... für das Effektmaß Quotient der Mittelwerte im alternativen FE-Modell für normalverteilte Daten 987
 - , Varianz des ... 987
 - , ... für das Effektmaß Quotient der Mittelwerte im RE-Modell für normalverteilte Daten 983f
 - , Varianz des ... 984
 - Musgrave, J.C. 1070
 - Myers, J.L. 608, 1066
 - Nachweisgrenze
 - , Kaisersche ... 343
 - Naeve, P. 640, 1066
 - Nakamura, K. 608, 1066
 - Namensstichprobe 370
 - natürliche Klassengröße 357
 - NBUE-Lebensdauerverteilung 784ff, 808, 823
 - , Hollander-Proschan-Test gegen ... 784ff
 - NBU-Lebensdauerverteilung 779ff, 786f
 - , Hollander-Proschan-Test gegen ... 780f
 - negatives Seitenfenster 714
 - negativ stochastische Abhängigkeit 416
 - Neumann, K. 1054
 - new better than used (s. auch NBU) 780
 - new better than used in expectation (s. auch NBUE) 784
 - Newcomb 888
 - Newman 628
 - Newton, I. 4, 232, 642, 874
 - Newton-Raphson-Verfahren 874ff
 - Newton-Verfahren 232, 642
 - new worse than used (s. auch NWU) 780
 - new worse than used in expectation (s. auch NWUE) 784
 - Neyman, J. 12, 282, 284f, 287
 - Neyman-Tschuprow-Aufteilung (s. auch optimale Aufteilung) 282ff

Ney – Nor

- , Stichprobenumfang bei geschichteter Zufallsauswahl mit ... 282ff
- , Varianz des Schätzers für die Summe der Merkmalswerte in der Grundgesamtheit bei geschichteter Zufallsauswahl mit ... 284
- , Schätzer für die ... 284
- Nichtbeantwortung 311ff, 317
- nichtgeplante Erhebung 6
- nicht-kontrollierbarer Einflußfaktor 362f
- nichtlineare Regression 642ff
 - , Trendschätzung mittels ... 642ff
 - nichtlineare Regressionsgleichung 600
 - nichtlineare Regressionsfunktion 589ff
 - nichtnegativer minimalverzerrter Schätzer
 - , ... für den Heterogenitätsparameter im RE-Modell 906f
- nicht-normiertes Histogramm 840, 844
- nichtparametrische Toleranzintervalle 238ff
- nichtparametrische Verfahren 139ff
- nicht-skaleninvarianter M-Schätzer 866ff
- Nichtunterlegenheit (s. a. Noninferiority) 916
- Nichtunterlegenheitsstudie 1016ff
 - , Placebo-kontrollierte ... 1016ff
 - , adaptive ... 1016ff
 - , adaptive prospektive ... 1016ff
 - , prospektive ... 1016ff
- Nichtzentralitätsparameter 152, 157
- Nitrogenpentoxid 588
- Niveau
 - , ... einer Skala 16
 - , ... eines Tests 129, 132f, 138
- Niveau- α -Test (s. auch Test zum Niveau α) 133
- Niveauangleichung bei Skalen 409
- Noelle, E. 319, 1066
- Nölle, G. 242, 1073
- noisy sciences 5
- nominales Merkmal 17, 82f, 407, 417
- Nominalskala 16
- nominal skalierte Größe 17, 82f, 407, 417
- Noninferiority 910, 915
 - , ... Parameter 915
 - , Test über den ... 915
 - , ... Schranke 916f
 - , ... Studie 910, 915, 1016ff
- Nonresponder 913
- Nonsens-Korrelation 77, 545
- non-therapeutic clinical research 363
- Normalengleichungen 574, 595
- Normalform
 - , disjunktive ... 752f, 756f
 - , ... für die Strukturfunktion eines monotonen Systems 752f
 - , ... für die Zuverlässigkeitfunktion eines monotonen Systems 756f
- normalisierter Abstand 777, 779, 785
- Normalisierung
 - , marginale ... 375
- Normalkern 843
- Normalmethode
 - , gewichtete Inverse ... 955
 - , Inverse ... 964
 - , ... für unabhängige Zeitreihen 964
 - , Test über die globale Hypothese bei der Kombination unabhängiger Tests nach ... 955
- Normal-Probit 955
- , kombinierter p-Wert für ... 960
- Normal-Score 885
- Normalverteilung 11, 49, 109, 111, 113, 115, 118, 124ff, 128, 130, 132, 135ff, 143ff, 259ff
 - , Bedeutung der ... 143
 - , bivariate ... 120
 - , χ^2 -Anpassungstest auf ... 182f
 - , Dichte der ... 143
 - , Dispersionsvergleiche bei zwei ... 524ff, 538f
 - , einseitig gestutzte ... 150
 - , Erwartungswert der ... 143
 - , Exzeß der ... 144
 - , gestutzte ... 148ff
 - , einseitig ... 150
 - , linksseitig ... 150
 - , rechtsseitig ... 150
 - , zweiseitig ... 149f
 - , Kolmogoroff-Smirnov-Anpassungstest auf ... 183ff
 - , ... bei nichtspezifizierten Parametern 185f
 - , ... bei spezifizierten Parametern 183ff
 - , Konfidenzintervall für den Erwartungswert einer ... 130f, 160ff
 - , asymptotisches ... bei unbekannter Varianz 131
 - , ... bei bekannter Varianz 130f, 160
 - , Absicherung der Breite d eines ... 166
 - , ... bei unbekannter Varianz 161
 - , Absicherung der Breite d eines ... 166ff
 - , mehrstufiges Verfahren zur ... 167f
 - , sequentielles Verfahren zur ... 168ff
 - , zweistufiges Verfahren zur ... 166f
 - , Einhaltung einer prozentualen Genauigkeit eines ... 173ff

- | --- | Nor – Odd |
|---|------------------|
| -, Streuzahlverfahren zur ... 175ff | |
| -, Variationszahlverfahren zur ... 173f | |
| -, Konfidenzintervall für die Standardabweichung einer ... 161f | |
| --, ... bei bekanntem Erwartungswert 162 | |
| --, ... bei unbekanntem Erwartungswert 161f | |
| -, Konfidenzintervall für die Varianz einer ... 161 | |
| --, ... bei bekanntem Erwartungswert 161 | |
| --, ... bei unbekanntem Erwartungswert 161 | |
| -, linksseitig gestutzte ... 150 | |
| -, Log... 151f | |
| -, Lokationsvergleiche bei zwei ... 505ff, 534ff, 539ff | |
| -, Median der ... 144 | |
| -, Meta-Analyse bei ... 976ff | |
| --, exakte Verfahren und spezielle Effektmaße der ... 976ff | |
| --, Modell der ... 976f | |
| -, Prognoseintervall für den Mittelwert einer zukünftigen Stichprobe aus einer ... 132f, 163f | |
| --, Absicherung der Breite d eines ... 170f | |
| -, Quantile der ... 147 | |
| -, rechtsseitig gestutzte ... 150 | |
| -, Regression bei ... 569ff | |
| -, Schätzer für den Erwartungswert einer ... 125f, 157ff | |
| -, Schätzer für die Standardabweichung einer ... 157ff | |
| -, Schätzer für die Varianz einer ... 153f, 157ff | |
| --, ... bei bekanntem Erwartungswert 154, 157 | |
| --, ... bei unbekanntem Erwartungswert 153, 157 | |
| -, Schiefe der ... 144 | |
| -, sequentieller Quotiententest für den Erwartungswert einer ... 259ff | |
| -, Simulation von Realisationen aus einer ... 145 | |
| -, Standard... 109ff, 115, 117, 119, 122, 144 | |
| -, Test auf Schiefe und Exzeß für die ... 189 | |
| -, Test über den Erwartungswert einer ... 135ff, 178ff | |
| --, ... bei bekannter Varianz 135ff, 178f | |
| --, ... bei unbekannter Varianz 179 | |
| --, benötigter Stichprobenumfang beim ... 181f | |
| -, Test über die Varianz einer ... 179ff | |
| -, Toleranzintervall für eine zukünftige Stichprobe aus einer ... 163ff | |
| --, Absicherung der Breite d eines ... 171ff | |
| --, ... bei bekannter Varianz 171f | |
| --, ... bei unbekannter Varianz 172f | |
| --, ... bei bekannter Varianz 164f | |
| --, ... bei unbekannter Varianz 163f | |
| -, Überprüfung der ... mittels Wahrscheinlichkeitspapier 189f | |
| -, Varianzanalyse bei ... 609ff | |
| -, Varianz der ... 143 | |
| -, Verteilungsfunktion der ... 145 | |
| -, zentrale Momente der ... 144 | |
| -, zweiseitig gestutzte ... 149f | |
| -, normierte Residuen 585ff | |
| Normiertheit 94 | |
| Notation von Zufallsvariablen 142 | |
| notch 835 | |
| notched box-plot 835ff | |
| Noteboom 888 | |
| Nourney, M. 722, 1066 | |
| Nowak, H. 1062 | |
| Nullau, B. 722, 1067 | |
| Nullhypothese 133ff | |
| -, eindeutige ... 139 | |
| -, zusammengesetzte ... 139 | |
| NWUE-Lebensdauerverteilung 784ff | |
| -, Hollander-Prochan-Test gegen ... 784f | |
| NWU-Lebensdauerverteilung 779ff, 786f | |
| -, Hollander-Prochan-Test gegen ... 780f | |
| oberes Quartil 34, 41, 114 | |
| Oberschwingung 716f | |
| Objekt 4 | |
| O'Brien, P.C. 1016, 1067 | |
| observation | |
| -, explanatory ... 365 | |
| -, ... for checking 365 | |
| -, substitute primary ... 365 | |
| -, supplementary ... 365 | |
| Odds ratio OR 442, 915 | |
| -, Effektmaß ... 915 | |
| -, kombinierter ... 918ff | |
| --, Konfidenzintervall für den ... 919f | |
| --, ... von Mantel-Haenszel 919 | |
| --, ... von Peto 920 | |
| --, Schätzer für den ... 918ff | |
| --, ... von Mantel-Haenszel 918f | |
| --, ... von Peto 918f | |
| --, logarithmierter ... 918f | |
| --, Varianz des ... 919 | |
| --, Schätzer für die ... 919 | |
| --, ... von Peto 918ff | |

Odd – Pea

- , logarithmierter ... 918f
- , Varianz des ... 919
- , Schätzer für die ... 919
- , logarithmierter ... 915
- , Schätzer für den ... 915
- , Schätzer für die Varianz des ... 915
- ODER-Glied 766, 770
- Oettli, W. 642, 793, 1049, 1055f
- offene Frage 309
- Ohse, D. 1047, 1067
- Olkov, I. 955f, 960, 1062
- Omnibus-Test 524
- Operationscharakteristik 136, 254, 257f, 381, 387, 390, 396f, 399, 402, 405
- Opitz, O. 1058
- optimale Aufteilung (s. auch Neyman-Tschuprow-Aufteilung) 282ff
- , Stichprobenumfang bei geschichteter Zufallsauswahl mit ... 282f
- , Varianz des Schätzers für die Summe der Merkmalswerte in der Grundgesamtheit bei geschichteter Zufallsauswahl mit ... 284
- , Schätzer für die ... 284
- optimale h-Schrittprognose 691ff
- , Konfidenzintervall für ... 693
- , Prognosefehler einer ... 691f
- optionale Strategie 823
- Orcutt, G.H. 737f, 1050
- ordinales Merkmal 17, 21, 23, 26, 33, 79, 417
- Ordinalskala 17
- ordinal skalierte Größe 17, 21, 23, 26, 33, 79, 417
- Ordnung
 - , ... eines ARMA-Prozesses 687
 - , Kriterien für die ... 687
 - , ... eines AR-Prozesses 686
 - , ... eines gleitenden Durchschnitts 661
 - , ... eines MA-Prozesses 687
- ordnungserhaltende Varianzschätzung 908
- Ordnungsstatistik 140, 237
- Originaldaten 7
- Osborne, J.K. 169, 1067
- Ott, W. 314, 1067
- outer-fence 835
- Owen, D.B. 614, 1067
- P**aasche 63ff
- Pabst 556f, 1023
- Pallaschke, D. 1054
- Panel kreuzkorrelierter Zeitreihen (s. a. kreuzkorreliertes Zeitreihenpanel) 744, 962ff
- Panel-Untersuchung 311
- Parallelredundanz 761
- Parallelschnittstruktur
 - , minimale ... 750
- Parallelsystem 102, 111, 746f, 752, 759
- , Zuverlässigkeit eines ... 747
- Parameter 123f
 - , Heterogenitäts... 906, 948, 951, 956, 996ff
 - , Noninferiority ... 915
 - , Test über den ... 915
- pareto-optimal 897, 1005ff
 - , ... Anlage 897, 1005ff
 - , ... Portfolioselektion 897, 1005ff
 - , ... im Wertpapiermanagement 897, 1005ff
- Paretooptimierung 897, 1005ff
- Parsinomitätsprinzip 686
- Partial Weighting nach Yates-Cochran bei Dominanz einer Studie oder Prognose 1013
- partielle Autokorrelation 677f
 - , ... eines AR-Prozesses 679
 - , empirische ... 678
 - , theoretische ... 677f
- partielle Korrelation 561ff
 - , Schätzer für die ... 561ff
 - , ... nach Kendall 563
 - , Test über die ... 562f
- partieller Korrelationskoeffizient 562f
- partieller Regressionskoeffizient 595
- Partitionierung der Teststatistiken für Tests im saturierten loglinearen Modell für mehrdimensionale Kontingenztafeln 464ff
- Parzen 713ff
 - , ...-Fenster 713ff
 - , Bandbreite des ... 714
- Pascal, B. 10
- Paterson, R. 369
- Patnaik, P.B. 969, 1067
- Paule, R.C. 999, 1066
- Pearl, R. 643, 1067
- Pearson, E.S. 12, 185, 189, 203f, 227, 344, 373, 1023, 1051, 1067
- Pearson, K. 11, 73, 413, 449f, 546, 561, 563, 565, 567, 569f, 578, 903, 1067
- Pearson
 - , Korrelationskoeffizient nach Bravais und ... 73ff, 79
- Pearson-Clopper-Werte 203ff

- Pearson'scher Kontingenzkoeffizient 449ff,
904
 -, ... für die $r \times s$ -Tafel 451f
 --, korrigierter ... 451f
 -, ... für die 2×2 -Tafel 449f
 Pearson'scher Korrelationskoeffizient 546ff
 -, approximativer Test über den ... 548ff
 -, Konfidenzintervall für den ... 549f
 -, Schätzer für den ... 546f
 -, Test über den ... 547f
 -, Vergleich mehrerer ... 552f
 -, Vergleich zweier ... 551f
 Peizer 1028
 Peizer-Pratt-Approximation 1028
 Penalty-Methode 1010
 -, sequentielle ... 1010
 -, stabile innere ... 1010
 -, stabile äußere ... 1010
 -, gemischte äußere-innere ... 1010
 Penalty-Verfahren 1010
 -, sequentielles ... 1010
 -, stabiles inneres ... 1010
 -, stabiles äußeres ... 1010
 -, gemischtes äußeres-inneres ... 1010
 periodische Strategie 823f
 -, streng... 824
 -, zufällig... 824
 Periodogramm 709ff
 Permutation 96
 per protocol population 910
 Peters, O. 1072
 Peto, J. 926, 1067
 Peto, R. 918ff, 926, 1067, 1074
 Peto
 -, Konfidenzintervall für den kombinierten Odds Ratio von ... 920
 -, Schätzer für den kombinierten Odds Ratio von ... 918ff
 --, logarithmierter ... 919f
 ---, Varianz des ... 920
 ----, Schätzer für die ... 920
 Pfadmenge
 -, minimale ... 750, 754, 757, 760, 763
 Pfadserienstruktur
 -, minimale ... 750
 Pfadvektor
 -, kritischer ... 764f
 Pfanzagl, J. 313, 323, 379, 531, 1067
 Pflanzagl-Test 531
 Pflaumer, P. 894, 947, 1007, 1058f, 1067f, 1070
 Phase 730
 phased mission 770ff
 Phasenspektrum 730, 735
 Phasenwinkel 699
- Pea – Por
- Phasenuverlässigkeit bei mehrphasigen Missionen 770ff
 Phi-Koeffizient 446ff
 Philips-Standard-Stichprobenplan 387
 Pierce, D.A. 690, 1049
 Pietsch, T. 1072
 Pike, M.C. 420, 1050
 Pill, J. 965, 1068
 Pirk, O. 1060
 Placebo 371ff
 Placebo-Effekt 371
 Placebo-kontrollierte Nichtunterlegeneitsstudie 1016ff
 -, adaptive ... 1016ff
 -, adaptive prospektive ... 1016ff
 -, prospektive ... 1016ff
 Plachky, D. 1068
 Plackett, R.L. 1064, 1068
 Plato 8
 Pocock, S.J. 1016, 1068
 Pöthke, R. 169, 1068
 Poisson 113, 122f, 212ff, 349f, 381, 383ff, 387, 390, 493, 495f, 783, 1044
 -, Grenzwertsatz von ... 122, 213
 Poisson-Prozeß 783
 Poissonverteilung 113, 122, 202, 212, 495f, 920ff
 -, Approximation der ... 213
 -, Beziehung zwischen χ^2 -Verteilung und ... 213
 -, χ^2 -Anpassungstest auf ... 216f
 -, FE-Modell bei ... 920ff
 -, Erwartungswert der ... 213
 -, Konfidenzintervall für den Parameter der ... 214
 -, Meta-Analyse bei ... 920ff
 -, RE-Modell bei ... 920ff
 -, Schätzer für den Parameter der ... 214
 -, Test über den Parameter der ... 214ff
 -, Varianz der ... 213
 -, Vergleich der Parameter mehrerer ... 495f
 politische Arithmetik 12
 Polygonzug 860
 Polynomapproximation einer Zeitreihe 666ff
 Polynomialverteilung 209ff
 Portfolio 897, 1004ff
 -, Aktien-... 965, 1004ff
 --, automatische Überwachung eines ... 965
 -, ...selektion 897, 1004f
 --, effiziente ... 1006ff

Por – Pro

- , Einbettung der ... in das ERE-Modell 1011f
- , Einbettung der ... in des RE-Modell 1011f
- , ... im Wertpapiermanagement 897, 1004f
- , ineffiziente ... 1008
- , pareto-optimale ... 897, 1005ff
- , Wertpapier... 897, 1004f
- Portmantau-lack-of-fit-test 690f
- , Prüfgröße des ... 691
- positiver Schätzer
- , ... für den Heterogenitätsparameter im RE-Modell 907
- positiv homogene Funktion 1021
- Positivität 94
- positiv stochastische Abhängigkeit 416
- Potenzmenge 93
- Potenzreihe 1022
- , ...entwicklung 1022
- Powerspektrum 702
- Power-Transformation 833f, 854
- PPAS-Verfahren 302
- PPS-Verfahren 298f
- , Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit beim ... 298
- , Varianz des ... 298
- Präventivstrategie 823
- Präzision eines Meß- oder Analyseverfahrens 341
- Präzision von Meßinstrumenten 944
- Pratt 1028
- Preisbereinigung 66
- Preisindex 63ff
- , ... für die Lebenshaltung 66
- , ... nach Laspeyres 63f, 66
- , ... nach Paasche 63f, 67f
- Preismeßzahl 66 .
- Prentice, R.L. 1063
- PRESTO study group 1048
- Preuß, H. 804, 806, 1068
- Pribilla 573
- Price, B. 608, 1050
- primärstatistische Erhebung 7
- primärstatistische Marktforschung 314
- principle of parsimony 686
- Prinzip von Holm 567f
- probability proportional to size Verfahren (s. auch PPS-Verfahren) 298f
- Probit 955
- , Fisher-... 955
- , Kombination der Dickey-Fuller p-Werte nach Hartung mit ... 964
- , kombinierter p-Wert für ... 960
- , Normal-... 955
- Produzentenrisiko 384, 387, 392, 397
- Profil 860
- Prognose 410, 691ff
- , betriebswirtschaftliche ... 894f
- , ... nach der Delphi Methode 965ff
- , kombinierte ... 966
- , ... des BIP 893f, 945ff
- , ... des Gesamteffekts bei unbekannten Zuverlässigen 947f
- , geschätzte Varianz der ... 947
- , geschätzte Zuverlässigen der ... 947
- , ... im FE-Modell 947
- , ... im RE-Modell 948
- , ...-Differenz 944
- , Dominanz einer ... 1013
- , Partial Weighting nach Yates-Cochran bei ... 1013
- , verallgemeinertes RE-Modell bei ... 1013
- , Methoden im ... 1013
- , ex-ante ... 693, 741
- , ex-post ... 693, 741
- , ...fehler 943
- , gesamtwirtschaftliche ... 893f
- , ...horizont 943
- , Kombination von ... 887ff, 947ff
- , ... im EFE-Modell 949
- , ... im ERE-Modell 951f
- , kombinierte ... 894, 944ff
- , ... nach der Delphi Methode 966
- , ... mit Scoringpunkten 944f
- , direkte Varianzformal von Hartung-Böckenhoff-Knapp bei ... 945
- , Konfidenzintervall für den Gesamteffekt bei ... 945
- , Schätzer für den Gesamteffekt bei ... 944f
- , optimale h-Schritt... 691ff
- , Wahl... 2
- , Wirtschafts... 943ff
- , ... mit Scoringssystem 943ff
- Prognosfähigkeit 691, 693
- Prognosefehler einer optimalen h-Schrittprognose 691f
- Prognosegüte 691ff
- Prognosehorizont 691
- Prognoseintervall 132f, 163ff, 170f, 582
- , ... für den Mittelwert einer zukünftigen Stichprobe aus einer Normalverteilung 132f, 163ff

- , benötigter Stichprobenumfang für ein ... 170f
 -, ... für den Wert des Regressanden bei vorgegebenem Wert des Regressors einer linearen Regression 582
Prognosestreifen für eine Regressionsgerade 582ff
Prognosewert 597
progression free survival 928, 931f, 934f, 938ff
progressionsfreies Überleben 928, 931f, 934f, 938ff
Prokosch; H.U. 1057, 1059
prophylaktische Erneuerung 807
prophylaktische Wartung 807
proportionale Aufteilung 282, 285f, 287f
 -, Stichprobenumfang bei geschichteter Zufallsauswahl mit ... 285
 -, Varianz des Schätzers für die Summe der Merkmalswerte in der Grundgesamtheit bei geschichteter Zufallsauswahl mit ... 285
 --, Schätzer für die ... 285
Proportional Hazard Regression Model 824, 940f
 --, Cox ... 824, 940f
Proschan, F. 759, 777, 780, 783f, 788, 808, 1023, 1049, 1052, 1063, 1068
Proschan-Pyke-Test auf DFR- (IFR-)
 Lebensdauerverteilungen 777
 prospektive Meta-Analyse 1014ff
 --, adaptive Versuchsanlage zur ... 1016
 --, self-designing Versuchsanlage zur ... 1016
 prospektive Placebo-kontrollierte Nichtunterlegenheitsstudie 1016ff
Prozentpunkt 58
Prozentrang 376f
Prozentrangverfahren 376f
 prozentuale Genauigkeit
 -, Einhaltung einer ... bei Konfidenzintervallen für den Erwartungswert einer Normalverteilung 173ff
 --, Bestimmung des benötigten Stichprobenfangs zur ... 173ff
 ---, Streuzahlverfahren zur ... 175ff
 ---, Variationszahlverfahren zur ... 173f
Prozeß
 -, AR-... 678ff
 -, ARIMA-... 684f
 -, ARMA-... 678, 682ff
 -, autoregressiver ... 639, 678ff
 -, Fehler... 694
 -, gemischter ... 639, 678, 682ff
 --, integrierter ... 684f
Pro – Pwe
 --, saisonaler ... 684f
 -, harmonischer ... 699, 701
 -, homoskedastischer ... 686
 -, Identifikation eines ... 678
 -, instationärer ... 684
 -, invertierbarer ... 678
 -, Korrelation zweier ... 728
 -, Kovarianz zweier ... 728
 -, ... mit stetiger Spektraldichte 702
 -, MA-... 678, 681f
 -, mittelwertbereinigter ... 678ff
 -, moving average ... 639, 678, 681f
 -, saisonbereinigter ... 686
 -, SARIMA-... 684f
 -, SARMA-... 684f
 -, Stärke des linearen Zusammenhangs zwischen zwei ... 728
 -, stochastischer ... 639, 678ff
 --, schwach stationärer ... 678ff
 -, trendbereinigter ... 686
Prüfgröße 134, 152
Prüfkosten 384
Prüfverteilung 152
pseudaleatorische Zuordnung 370
Pseudoauswahlverfahren 309
pseudo-inverse Matrix 950
Pseudo-Maximum-Likelihood-Schätzer für den Heterogenitätsparameter im RE-Modell nach Hartung 1000
Pseudozufallszahlen 194
psychologischer Test 269
Pukelsheim, F. 636, 1068
Punkteschar
 -, Glättung einer ... 825, 849ff
 --, Linearisierung einer ... 825, 849f, 854ff
Punktschätzer (s. auch Schätzer) 130
p-Wert 137, 927
 -, ... des Dickey-Fuller Unit Root Test 963
 --, Kombination der ... nach Hartung 964
 --, ... nach McKinnon 963
 --, Kombination abhängiger ... 895, 930f, 958ff
 --, Hartung-Methode zur ... 959ff
 ---, kombinierter p-Wert nach der ... 960
 ----, ... für Probits 960
 ----, ... für Fisher-Probits 960
 -, Kombination unabhängiger ... (s. a. Kombination unabhängiger Tests) 954ff
 --, Kombination von ... 930f
 --, Stouffer-Methode zur ... 930f
 --, kombinierter ... nach der Hartung-Methode 960
 --, ... für Probits 960

Pwe – Quo

- , ... für Fisher-Probits 960
- , Schätzer für die durchschnittliche Korrelation abhängiger ... 959
- Pyke, R. 777, 1068
- Q**-Koeffizient von Yule 443ff
- Q-Q-Plot 190, 587, 825, 827, 847ff
- Quader 860
- Quadrat
 - , griechisch-lateinisches ... 361
 - , lateinisches ... 361
- Quadratform
 - , Cochrans ... 901
 - , ... nach Hartung 908, 996f
- quadratische Abweichung 642
- quadratischer Abstand 43
- quadratische Regressionsfunktion 589ff
 - , Bestimmtheitsmaß einer ... 593
 - , Konfidenzbereich für eine ... 590
 - , Konfidenzintervall für das Maximum (Minimum) einer ... 591
 - , Konfidenzintervall für den Erwartungswert des Regressanden bei vorgegebenem Wert des Regressors einer ... 590
 - , Schätzer für das Maximum (Minimum) einer ... 591
 - , Schätzer für die Parameter einer ... 590
 - , Korrelation der ... 590
 - , Varianz der ... 590
 - , Schätzer für die Varianz der Fehlerterme einer ... 590
 - , Tests über die Parameter einer ... 590f
- Quadratsummenzerlegung 615, 625
 - , doppelte ... 620
- Quadraturspektrum 729, 735
- Quadredundanz 762
- Qualitätskontrolle 38 ff
- Qualitätsregelkarte (s. auch Kontrollkarte) 405
 - qualitatives Merkmal 16
 - , doppelter Stichprobenplan für ... 390ff
 - , einfacher Stichprobenplan für ... 383ff
 - , laufende Kontrolle der Produktion bei ... 404f
 - , mehrfacher Stichprobenplan für ... 389f
 - , sequentieller Stichprobenplan für ... 392ff
 - quality of life 913
 - Quantil 34, 114ff, 236f, 825
 - , α -... 114
 - , ... bei gruppierten Daten 34
 - , ... der χ^2 -Verteilung 153, 1029f
 - , ... der F-Verteilung 157, 1031ff
 - , ... der Normalverteilung 147
 - , ... der Standardnormalverteilung 147, 1027
 - , ... der Studentisierten Spannweite 1038f
 - , ... der t-Verteilung 154, 1028
 - , Konfidenzbereich für ... 236f
 - , Test über ... 237
 - Quantile-Quantile-Plot 827, 847ff
 - quantitatives Merkmal 16
 - , einfacher Stichprobenplan für ... 395ff
 - , laufende Kontrolle der Produktion bei ... 401ff
 - , sequentieller Stichprobenplan für ... 399ff
 - Quartil 34f
 - , oberes ... 34, 41, 114, 118
 - , unteres ... 34, 41, 114, 118
 - Quartilsabstand 41, 118, 866
 - quasi-konkave Funktion 1021
 - quasi-konvexe Funktion 1021
 - Quasi-Rangtest von Moses 529ff
 - Quetelet, A. 11
 - Quinn, B.G. 687, 1055
 - Quote 318ff
 - , kombinierte ... 319
 - , Rücklauf ... 310
 - Quotenauswahl 316, 318ff
 - Quotenplan 318ff
 - Quotenverfahren 311
 - Quotient der Mittelwerte (Effektmaß) 981ff
 - , alternative Methoden mit expliziten Lösungen für ... 982, 986f
 - , alternatives FE-Modell für ... 986f
 - , Konfidenzintervall im ... 987
 - , MUPML Gesamt-Schätzer im ... 987
 - , Schätzer für die Varianz des ... 987
 - , exakte Methoden im FE-Modell für ... 981ff
 - , graphische Ermittlung der Konfidenzgrenzen und Punktschätzer für ... 982f
 - , Hartung-Verfahren für ... 981ff
 - , exakter Schätzer nach dem ... 981
 - , exaktes Konfidenzintervall nach dem ... 981
 - , ... im RE-Modell 982ff
 - , Konfidenzintervall im ... 983
 - , MUPML Gesamt-Schätzer im ... 983f
 - , Schätzer für die Varianz des ... 984
 - , MUPML Gesamt-Schätzer für ... 981f
 - , approximativer ... 981
 - , exakter ... 982
 - Quotiententest
 - , sequentieller ... 251 ff

- | --, ... für den Erwartungswert einer Normalverteilung 259ff | Quo – Reg |
|---|--|
| --, ... für den Parameter λ einer Exponentialverteilung 261ff | Rechteckverteilung (s. auch Gleichverteilung) 116, 192ff |
| --, ... für den Parameter p einer Binomialverteilung 252ff | -, zweidimensionale ... 194f |
| Quotient von Hazardraten 889f | rechtsschiefe Häufigkeitsverteilung 38f, 47, 833 |
| Quotierungsmerkmal 311, 318 | rechtsseitig gestutzte Normalverteilung 150 |
| Rabe 888 | rechtssteile Häufigkeitsverteilung 38 |
| Raj, D. 303, 305, 1068 | redescending M-estimators 872 |
| Randhäufigkeit | Reduktionstest bei der multiplen Regression 598 |
| -, absolute ... 72 | Redundanz 760ff |
| -, relative ... 72 | -, Erhöhung der Zuverlässigkeit durch ... 760ff |
| random effects model (s. a. RE-Modell) | -, Parallel... 761 |
| 890 | -, Quad... 762 |
| -, erweitertes ... (s. a. ERE-Modell) 895 | -, Serien... 762 |
| Randomisation 141 | Redundanzplanung 822 |
| randomisierter Versuchsplan | Reed, L.J. 643, 1067 |
| -, vollständig ... 355 | re-expression 833 |
| Randomisierung 355, 362, 368, 370 | REFA 173, 1068 |
| random sampling | Regelgrenze 402ff |
| -, simple ... 274 | Regressand 569, 595, 640 |
| random walk 965 | Regression 335, 374, 569ff |
| Randverteilung 72, 107, 120 | -, ... bei Fehlern in den Variablen 601ff |
| -, ... der bivariaten Normalverteilung 120 | -, ... bei konvexen Nebenbedingungen 608 |
| -, ... der Multinomialverteilung 210 | -, ... bei korrelierten Fehlern, 608, 736ff |
| range 40 | -, ... bei Vorinformation über den Regressanden 608 |
| Rangkorrelation 79ff, 553ff | -, inverse ... 335 |
| Rangkorrelationskoeffizient 79ff, 553ff | -, lineare ... 573ff |
| -, Kendallscher ... 81f, 559ff | --, ... bei ungleicher Varianz der Beobachtungen 584f, 600 |
| -, Spearmanscher ... 79f, 553ff | --, ... durch einen vorgegebenen Punkt 584 |
| Rangstatistik 140 | --, ... ohne Absolutglied 584 |
| Rangsummentest | -, Meta-... 1013f |
| -, extremer ... von Youden 339f | --, RE-Modell der ... 1014 |
| -, Friedman-... 623f | -, ... mit Orthogonalpolynomen 594f |
| -, ... von Wilcoxon 884f | -, multiple ... 595ff |
| Rangzahl 79ff, 140f | -, multivariate ... 374 |
| -, mittlere ... 140 | -, nichtlineare ... 589ff |
| Rao, C.R. 634, 1068 | -, polynomiale ... 595 |
| Raphson 874 | -, quadratische ... 589ff |
| Rasch, D. 376, 1068 | Regressionsanalyse (s. auch Regression) 374, 407 |
| Rauhut, B. 1058 | -, multivariate ... 374 |
| Rayleigh 5, 231 | Regressionsfunktion (s. auch Regression) |
| Rayleigh Verteilung 231 | -, explorative ... 825, 855ff |
| Reaktion 1. Ordnung 587f | -, nichtlineare ... 589ff |
| Realisation 104, 124 | -, quadratische ... 589ff |
| Realteil einer komplexen Zahl 700 | Regressionsgerade (s. auch lineare Regression) 573ff, 825, 849 |
| Rechenregeln für Erwartungswerte 113 | -, Absolutglied einer ... 574ff |
| Rechenregeln für Kovarianzen 119 | |
| Rechenregeln für Varianzen 117 | |
| Rechteck | |
| -, ...-Fenster 711f, 714 | |
| -, lateinisches ... 361 | |
| Rechteckskern 842 | |

Reg – Rem

- , Konfidenzintervall für das ... 581
- , Schätzer für das ... 575
- , Erwartungswert des ... 576
- , Korrelation des Schätzers für den Steigungsparameter und des ... 578
- , Kovarianz des Schätzers für den Steigungsparameter und des ... 578
- , Varianz des ... 576
- , Verteilung des ... 577
- , simultane Konfidenzintervalle für den Steigungsparameter und das ... 581
- , ... bei Fehlern in den Variablen 601ff
- , explorative ... 852ff
- , explorative Bestimmung einer ... 849, 852ff
- , Konfidenzstreifen für eine ... 582
- , simultaner ... 582f
- , ... nach Wald 605
- , Schätzer für die Parameter einer ... 606
- , Prognosestreifen für eine ... 582
- , Schätzer für die Parameter einer ... 575ff
- , ... bei Fehlern in den Variablen 601ff
- , simultaner Konfidenzstreifen für eine ... 582f
- , Steigungsparameter einer ... 574ff
- , Konfidenzintervall für den ... 581
- , Schätzer für den ... 575
- , Erwartungswert des ... 576
- , Korrelation des Schätzers für das Absolutglied und des ... 578
- , Kovarianz des Schätzers für das Absolutglied und des ... 578
- , Varianz des ... 576
- , Verteilung des ... 577
- , simultane Konfidenzintervalle für das Absolutglied und den ... 581
- , Test über den ... 580
- Regressionsgleichung**
- , nichtlineare ... 600
- Regressionsmodell 128, 640
- , autogressives ... 736ff, 744
- , inhomogenes ... 740
- , ... mit AR(1)-Fehlerprozeß 737ff
- Regressionsparabel** (s. auch quadratische Regressionsfunktion) 589ff
- Regressionsrechnung** (s. auch Regression) 11, 72
- Regressionsschätzer** 304
- , Varianz des ... 304
- , Schätzer für die ... 304
- Regressions-Zeitreihenmodelle**
- , gemischte ... 639, 735ff
- , Prognose und Modellüberprüfung bei ... 737
- , Prognose und Prognosegüte bei ... 741
- Regressor** 569, 595, 640
- regula falsi** 650, 997
- Rekursionsformeln** für gleitende Durchschnitte 660f
- relative Häufigkeit** 20ff, 24, 26, 50, 55, 93, 98
- , bedingte ... 99
- relative Häufigkeitsverteilung** 21
- relative Klassenhäufigkeit** 26ff, 28
- relative Randhäufigkeit** 72
- relativer Fehler** 324f
- relativer Gesamtfehler** 325f
- relativer Zufallsfehler** 269ff
- Relatives Risiko RR** 915
- , Effektmaß 915
- , Konfidenzintervall für das ... 915f
- , logarithmiertes ... 915
- , Konfidenzintervall für das ... 916
- , Schätzer für das ... 915
- , Schätzer für die Varianz des ... 915
- , Schätzer für das ... vom Mantel-Haenszel Typ 920
- relative Summenhäufigkeit** 23
- REML-Schätzer** für den Heterogenitätsparameter im RE-Modell 1001
- RE-Modell** 890, 899, 906ff, 968ff
- , ... bei normalverteilten Daten 982ff, 994
- , exakte Methoden im ... 982ff, 994
- , Effektmaß Quotient der Mittelwerte im ... 982ff
- , Konfidenzintervall für das ... 983
- , MUPML Gesamt-Schätzer für das 983f
- , Schätzer für die Varianz des ... 984
- , Effektmaß Standardisierte Mittlere Differenzen im ... 994
- , ... bei Poissonverteilung 920ff
- , ... bei vergleichenden Time-to-Event Studien 932, 936f
- , ... bei vergleichenden Überlebenszeitstudien 931f, 936f
- , ... der Meta-Regression 1014
- , Heterogenitätsparameter im ... 906f, 996ff
- , DSL-Schätzer für den ... 906ff
- , explizite Schätzer für den ... 996
- , Hartung-Knapp-Konfidenzintervall für den ... 996f
- , Hartung-Schätzer für den ... 906ff
- , Hedges-Schätzer für den ... 906ff
- , implizite Schätzer für den ... 999ff

- | --, | Rem – Res |
|--|--|
| iterativ definiertes Konfidenzintervall für den ... 998 | |
| iterative Schätzer für den ... 999ff | --, iterative ... 999ff |
| Konfidenzintervalle vom Waldschen Typ für den ... 999 | --, Mandel-Paule-... 999 |
| Likelihood-Methoden für den ... 998f | --, Maximum-Likelihood-... 1000f |
| ... vom Waldschen Typ 998f | --, restringierter ... 1001 |
| Mandel-Paule-Schätzer für den ... 999 | --, Pseudo-... 1000 |
| Maximum-Likelihood-Schätzer für den ... 1000f | --, median unverzerrter ... nach Hartung 1000 |
| restringierter ... 1001 | --, ML-... 1000f |
| Pseudo-... 1000 | --, nichtnegativer minimalverzerrter ... 906f |
| median unverzerrter Schätzer nach Hartung für den ... 1000 | --, positiver ... 907 |
| ML-Schätzer für den ... 1000f | --, Pseudo-Maximum-Likelihood-... nach Hartung 1000 |
| Pseudo-Maximum-Likelihood-Schätzer nach Hartung für den ... 1000 | --, REML-... 1001 |
| REML-Schätzer für den ... 1001 | --, restringierter Maximum-Likelihood-... 1001 |
| restringierter Maximum-Likelihood-Schätzer für den ... 1001 | --, erweitertes ... (s. a. ERE-Modell) 895 |
| verallgemeinerter Welch Test für den ... 998 | -, Prognose des Gesamteffekts bei geschätzten Zuverlässigkeiten im ... 948 |
| Investitionsberatung im ... 948f | -, totale Varianz im ... 968 |
| Kombination von Kaplan-Meier Kurven im ... 936f | -, Varianzzerlegung im ... 968 |
| Konfidenzintervalle für den Gesamteffekt im ... 907 | -, verallgemeinertes ... bei Dominanz einer Studie oder Prognose 1013 |
| Konfidenzintervall für den Heterogenitätsparameter im ... 996ff | --, Methoden im ... 1013 |
| Hartung-Knapp-... 996f | Rendite 897, 1004f |
| iterativ definiertes ... 998 | --, mittlere ... 897, 1005 |
| ... vom Waldschen Typ 999 | --, ...-schwelle 1008f |
| Maße für die Heterogenität des ... 968ff | --, ...-Volatilitäts-Projektionsebene 1007 |
| Hartung-Konfidenzintervall für die ... 969f | Reparametrisierungsbedingung 619f, 625 |
| Hartung-Schätzer für die ... 969f, 973f | Reparaturzeit |
| median unverzerrte ... 973f | --, mittlere ... 820 |
| Maße für die Inhomogenität des ... 968ff | repräsentative Stichprobe 315 |
| ... nach Hartung 969 | Repräsentativität 315f, 318, 320 |
| ... nach Higgins-Thompson 969 | Reserve |
| Mischung aus inhomogenem FE-Modell und ... 910f | --, heiße ... 762, 809f, 814 |
| Konfidenzintervall für den Gesamteffekt bei der ... 911 | --, kalte ... 762, 809, 811, 814 |
| Schätzer für den Gesamteffekt bei der ... 911 | --, kalte Macro... 763, 711ff |
| Varianz des ... 911 | --, kalte Micro... 763, 712ff |
| Schätzer für den Gesamteffekt im ... 907 | Residualanalyse 585ff |
| Varianz des ... 907 | --, Ermittlung von Ausreißern mittels ... 585 |
| Schätzer für den Heterogenitätsparameter im ... 906f | Residual-Quadratsumme 574, 595 |
| DSL-... 906f | Residuum 578, 585 |
| explizite ... 996 | --, Double-Root-... 847 |
| Hedges-... 906f | --, normiertes ... 585 |
| implizite ... 999ff | --, Wurzeldiagramm-... 845 |

Ret – Sca

- Retestmethode 83
- retrospektive Zuordnung 368f
- reziproke Transformation 349
- Rhodes, E.C. 644f, 647, 1068
- Richard, F.W. 797, 1068
- Richter, H. 1068
- Richtigkeit eines Meß- oder Analyseverfahrens 342
 - , totale ... 342
- Riedwyl 860
- Riemann, J.F. 1048
- Ringversuch 337ff
- Rinne, H. 823, 1062, 1068
- Risiko
 - , Anlage mit geringstem ... 1006, 1009
 - , Konsumenten... 384, 387, 393, 397
 - , Produzenten... 384, 387, 393, 397
 - , relatives ... RR 915ff, 1014f
- Risikobeurteilung 966
 - , subjektive ... 966
 - , Kombination von ... 966
- Risiko Differenz RD 914
 - , Effektmaß ... 914
 - , Schätzer für die ... 914
 - , Schätzer für die Varianz des ... 914
 - , ... vom Mantel-Haenszel Typ 920
- Risikofunktion im Wertpapiermanagement 1006
- Riskostreuung 899
- risk shift Phänomen 967
- Ritter, K. 1055
- r x s-Tafel (s. zweidimensionale Kontingenztafel)
- Robins, J.M. 919f, 1054, 1068
- robust 144, 861ff
- robuste Schätzfunktion 861ff
- robuste Schätzung 861ff
- robuste Statistik 826, 861ff
- robuste Verfahren 825f, 861ff
- Robustheit
 - , globale Kenngröße der ... 864
 - , lokale Kenngröße der ... 863
- Robustheitseigenschaften 861ff
- Roesch 889, 918
- Rogers, W.H. 1048
- Romig, H.G. 201, 373, 389, 392, 1051, 1068
- Ronner, A.E. 827, 1064
- rootogram 844
- Rosenblatt, M. 640, 1054
- Rosenfeld 557
- Rosenthal, H. 169, 1068
- Roß, H. 600, 1068
- Ross, S. 808, 1069
- Roth, H.D. 901, 995, 1065
- Rothe, G. 1069
- R-Schätzer 883ff
 - , Bruchpunkt eines ... 885f
- r-Statistiken von Dixon für Tests auf Ausreißer 346ff
- Rücklaufquote 310
- Rumme 572
- Rundungsempfindlichkeit 864
- running median 850
- Russel, M.H. 369
- Rutkowski, S. 1072
- Ryan, T.M. 899, 1069
- Ryerson, C.M. 805, 1069
- Sachs, L. 1069
- saisonale Komponente 639ff
 - , Bereinigung von ... 639, 660ff
 - , Schätzer für die ... 665f, 725
- saisonbereinigte Zeitreihe 665f, 725
- Saisonbereinigung 665f, 722ff
 - Saisonbereinigungsverfahren 639, 660ff, 722ff
- Saisonfigur
 - , konstante ... 641, 661, 668ff
 - , ... mit variabler Amplitude 641
 - , variable ... 641
- Saisonschwankung 6
- Salk 373f
- Salk-Vakzine 373f
- sampling
 - , judgement ... 316
 - , simple random ... 124, 274
- Sansoviso, F. 12
- SARIMA-Prozeß 684f
- SARMA-Prozeß 684f
- Satterthwaite, F.E. 763, 977, 1069
- Sättigungsgrenze 643ff
- saturiertes loglineares Modell 432, 439, 464ff
 - , ... für mehrdimensionale Kontingenztafeln 464ff
 - , Parameter des ... 465ff
 - , Konfidenzintervalle für die ... 471ff
 - , Schätzer für die ... 466ff
 - , Schätzer für die Varianz der ... 471ff
 - , Test über die ... 477ff
- Satz von Berry-Esséen 122
- Satz von der totalen Wahrscheinlichkeit 99f, 102
- Satz von Glivenko-Cantelli 121
- Satz von Steiner 117
- Savin, N.E. 741, 1069
- Scatter-Plot 226, 858ff

- , Bürsten eines ... 860
 -, ...-Matrix 858
 Schaafsma, A.H. 387, 389, 392, 398, 1069
 Schach, E. 309, 370, 1069
 Schach, S. 309, 370, 585, 608, 918, 1061,
 1065, 1069
 Schader, M. 1058, 1061
 Schäfer, E. 796, 803, 1069
 Schäfer, T. 585, 608, 1069
 Schäffer, K.-A. 666, 1062, 1066, 1069
 Schätzer 124ff
 -, Anteil... 293ff
 --, Varianz des ... 294
 -, Bias eines ... 125
 -, Differenzen... 303
 --, Varianz des ... 303
 ---, Schätzer für die ... 303
 -, direkter ... für die Varianz von Hartung-
 Böckenhoff-Knapp 942f, 945
 -, erwartungstreuer ... 125
 --, ... mit gleichmäßig minimaler Varianz
 126
 -, exakter ... nach dem Hartung-Knapp-
 Verfahren für das Effektmaß Differenz
 der Mittelwerte im FE-Modell für nor-
 malverteilte Daten 978
 -, exakter ... nach dem Hartung-Verfahren
 für das Effektmaß Quotient der Mittel-
 werte im FE-Modell für normalverteilte
 Daten 981
 -, exakter ... nach dem Hartung-Verfahren
 für das Effektmaß Standardisierte Mitt-
 lere Differenzen im FE-Modell für
 normalverteilte Daten 992
 -, ... für das Absolutglied einer Regressi-
 onsgeraden 575, 584, 601ff
 --, ... bei Fehlern in den Variablen 601ff
 --, ... durch einen vorgegebenen Punkt 584
 --, ... nach Wald 606
 --, ... für das Maximum (Minimum) einer
 quadratischen Regressionsfunktion 591
 -, ... für das gemeinsame Mittel im GCM
 Problem 942
 -, ... für das logarithmierte relative Risiko
 RR 915
 --, Schätzer für die Varianz des ... 915
 -, ... für das multiple Bestimmtheitsmaß
 565ff
 -, ... für den Erwartungswert 826
 -, ... für den Erwartungswert einer Nor-
 malverteilung 125f, 157ff
 -, ... für den Gesamteffekt bei gescort
 kombinierten Prognosen 944f
 -, ... für den Gesamteffekt bei unbekann-
 ten Zuverlässigkeiten 947f
- Sca – Schä**
- , geschätzte Varianz des ... 947
 --, geschätzte Zuverlässigkeiten des ... 947
 --, ... im FE-Modell 947
 --, ... im RE-Modell 948
 - , ... für den Gesamteffekt im EFE-Modell
 949
 --, Schätzer für die Varianz des ... 949
 - , ... für den Gesamteffekt im ERE-Modell
 951f
 --, Schätzer für die Varianz des ... 951f
 --, stabiler Hartung ... 952
 - , ... für den Gesamteffekt im homogenen
 FE-Modell 900
 --, Varianz des ... 900f
 - , ... für den Gesamteffekt im inhomoge-
 nen FE-Modell 900f
 - , ... für den Gesamteffekt im RE-Modell
 907
 --, Varianz des ... 907
 - , ... für den Hazardquotienten 927
 - , ... für den Heterogenitätsparameter im
 ERE-Modell 951f
 - , ... für den Heterogenitätsparameter im
 RE-Modell 906f, 972
 --, DSL-... 906f
 --, explizite ... 996
 --, Hedges-... 906f, 948
 --, verallgemeinerter ... 911, 951
 --, implizite ... 999ff
 --, iterative ... 999ff
 --, Mandel-Paule-... 999
 --, Maximum-Likelihood-... 1000f
 --, restriktiver ... 1001
 --, Pseudo-... 1000
 --, median unverzerrter ... nach Hartung
 1000
 --, ML-... 1000f
 --, nichtnegativer minimalverzerrter ...
 906f
 --, positiver ... 907
 --, Pseudo-Maximum-Likelihood-... nach
 Hartung 1000
 --, REML-... 1001
 --, restriktiver Maximum-Likelihood-...
 1001
 -, ... für den kombinierten Odds Ratio
 918ff
 -, ... für den logarithmierten Hazardquo-
 tienten 927
 --, ... von Mantel-Haenszel 918f
 ---, logarithmierter ... 918f
 ----, Varianz des ... 919
 -----, Schätzer für die ... 919

Schä – Schä

- , ... von Peto 918ff
- , logarithmierter ... 918f
- , Varianz des ... 919
- , Schätzer für die ... 919
- , ... für den logarithmierten Odds Ratio OR 915
- , Schätzer für die Varianz des ... 915
- , ... für den mittleren Merkmalswert in der Grundgesamtheit bei einfacher Zufallsauswahl 274
- , ... für den Parameter λ einer Exponenti-alverteilung 126, 220f, 694ff
- , ... für den Parameter λ einer Poissonver-teilung 214
- , ... für den Parameter p einer Binomial-verteilung 127, 202
- , ... für den Anteilswert einer hypergeo-metrischen Verteilung 208
- , ... für den Pearsonschen Korrelationsko-effizienten 546ff
- , ... für den Steigungsparameter einer Regressionsgeraden 575, 584, 601ff
- , ... bei Fehlern in den Variablen 601ff
- , ... durch einen vorgegebenen Punkt 584
- , ... nach Wald 606
- , ... ohne Absolutglied 584
- , ... für den Umfang einer Grundgesamt-heit mit hypergeometrischer Verteilung 209
- , ... für die bi-partielle Korrelation 563
- , ... für die durchschnittliche Korrelation abhängiger p-Werte 959
- , ... für die Kombination von Überle-benswahrscheinlichkeiten 938
- , ... für die Maße der Heterogenität 969ff
- , ... nach Hartung 973f
- , ... nach Higgins-Thompson 969
- , ... für die multiple Regression 565f
- , ... für die Parameter der einfachen Varianzanalyse 611
- , ... für die Parameter der zweifachen Varianzanalyse
- , ... mit Wechselwirkungen zwischen den Faktoren 625
- , ... ohne Wechselwirkungen zwischen den Faktoren 627
- , ... für die Parameter des einfachen Blockexperiments 620
- , ... für die Parameter des loglinearen Modells
- , ... für die $r \times s$ -Tafel 430
- , approximative Varianz der ... 430
- , ... für die 2×2 -Tafel 427f
- , approximative Varianz der ... 428f
- , ... für die Parameter des saturierten log-linearen Modells für mehrdimensionale Kontingenztafeln 466f
- , Schätzer für die Varianz der ... 471ff
- , ... für die Parameter einer eindimensio-nalen Gleichverteilung 128, 197
- , ... für die Parameter einer Hjorth- (IDB) Verteilung 234
- , ... für die Parameter einer multiplen Regression 595ff
- , ... für die Parameter einer Normalvertei-lung 125f, 153f, 157ff
- , ... für die Parameter einer quadratischen Regressionsfunktion 590
- , ... für die Parameter einer Weibullver-teilung 232
- , ... für die partielle Korrelation 561ff
- , ... nach Kendall 563
- , ... für die Risiko Differenz RD 914
- , Schätzer für die Varianz des ... 914
- , ... für die Standardabweichung 826
- , ... für die Standardabweichung einer Normalverteilung 157ff
- , ... für die Streuung 826
- , ... für die Summe der Merkmalswerte in der Grundgesamtheit
- , ... bei einfacher Zufallsauswahl 274
- , Varianz des ... 275f
- , Schätzer für die ... 275
- , ... bei geschichteter Zufallsauswahl 280
- , Varianz des ... 280
- , Schätzer für die ... 280
- , ... mit Schichtenbildung nach Auswahl der Stichprobe 286
- , Varianz des ... 286
- , Schätzer für die ... 286
- , ... wenn jede einzelne Auswahl eine einfache Zufallsauswahl ist 280
- , Varianz des ... 280
- , Schätzer für die ... 280
- , ... beim Klumpenstichprobenverfahren als einstufiges Verfahren 289ff
- , Varianz des ... 289f
- , ... beim Klumpenstichprobenverfahren als mehrstufiges Verfahren 291f
- , Varianz des ... 291
- , Schätzer für die ... 292
- , ... beim PPS-Verfahren 298
- , Varianz des ... 298
- , ... bei systematischer Stichprobe 295ff
- , Varianz des ... 296
- , ... für die Survivalfunktion 933
- , Kaplan-Meier ... 933

- , ... für die Überlebenswahrscheinlichkeit
932
- , ... für die Überlebenszeit 933f
- , ... für die Varianz der Merkmalswerte in
der Grundgesamtheit
- , ... bei einfacher Zufallsauswahl 274
- , ... bei geschichteter Zufallsauswahl,
wenn jede einzelne Auswahl eine einfache
Zufallsauswahl ist 280
- , ... für die Varianz der Fehlerterme
- , ... bei der linearen Regression 578
- , ... bei der multiplen Regression 597f
- , ... einer quadratischen Regressions-
funktion 590
- , ... für die Varianz der Schätzer für die
Parameter des saturierten loglinearen
Modells für mehrdimensionale Kontingen-
zenztafeln 471ff
- , ... für die Varianz des Differenzenschät-
zers 303
- , ... für die Varianz des Regressionsschät-
zers 304
- , ... für die Varianz des Schätzers für die
Summe der Merkmalswerte in der
Grundgesamtheit
- , ... bei einfacher Zufallsauswahl 275
- , ... bei geschichteter Zufallsauswahl 280
- , ... mit Neyman-Tschuprow-Aufteilung
284
- , ... mit optimaler Aufteilung 284
- , ... mit proportionaler Aufteilung 285
- , ... mit Schichtenbildung nach Auswahl
der Stichprobe 286
- , ... wenn jede einzelne Auswahl eine
einfache Zufallsauswahl ist 280
- , ... beim mehrstufigen Klumpenstich-
probenverfahren 292
- , ... für die Varianz des Verhältnisschät-
zers 302
- , ... für die Varianz einer Normalvertei-
lung 153f, 157ff
- , ... bei bekanntem Erwartungswert 154,
157
- , ... bei unbekanntem Erwartungswert,
153, 157
- , ... für die Varianzen der Effekte bei der
einfach hierarchischen Klassifikation
630f
- , ... für die Varianz im GCM Problem
941ff
- , ... für die Yuleschen Assoziationsmaße
444ff
- , Gauß-Markov ... 900, 907
- , Eigenschaften der ... in der Metaanaly-
se 1018ff
- Schä – Schä
- , Güteeigenschaften von ... 125f
- , ... in der Meta-Analyse
- , Eigenschaften der ... 1018ff
- , Kaplan-Meier ... für die Survivalfunktio-
n 933
- , konsistenter ... 125
- , L-... 826, 880ff
- , M-... 826, 866ff
- , Maximum-Likelihood-... 126ff
- , Mean squared error (MSE) eines ... 125f
- , Methoden zur Bestimmung eines ...
125ff
- , MUPML Gesamt-... 978f, 981f, 992,
983f, 987
- , approximierender ... 979, 982, 992
- , ... für das Effektmaß Differenz der
Mittelwerte im FE-Modell für normal-
verteilte Daten 979
- , ... für das Effektmaß Quotient der
Mittelwerte im FE-Modell für normal-
verteilte Daten 982
- , ... für das Effektmaß Standardisierte
Mittlere Differenzen im FE-Modell für
normalverteilte Daten 992
- , exakter ... 978, 981ff, 992
- , ... für das Effektmaß Differenz der
Mittelwerte im FE-Modell für normal-
verteilte Daten 978
- , ... für das Effektmaß Quotient der
Mittelwerte im FE-Modell für normal-
verteilte Daten 981
- , graphische Ermittlung des ... 982f
- , ... für das Effektmaß Standardisierte
Mittlere Differenzen im FE-Modell für
normalverteilte Daten 992
- , ... für das Effektmaß Quotient der
Mittelwerte im alternativen FE-Modell
für normalverteilte Daten 987
- , Varianz des ... 987
- , ... für das Effektmaß Quotient der
Mittelwerte im RE-Modell für normal-
verteilte Daten 983f
- , Varianz des ... 984
- , ... nach der Methode der kleinsten
Quadrat 128f
- , ... nach der Momentenmethode 126
- , R-... 826, 883ff
- , Regressions... 304
- , Varianz des ... 304
- , Schätzer für die ... 304
- , robuste Lokations... 826, 866ff
- , robuste Skalen... 826, 864ff

Schä – Sel

- , unverzerrter ... mit minimaler Varianz 900
- , Varianz...
 - , Eigenschaften der ... in der Meta-Analyse 1018ff
 - , Verhältnis... 302
 - , Varianz des ... 302
 - , Schätzer für die ... 302
 - , Verzerrung eines ... 125
 - , ... vom Mantel-Haenszel Typ
 - , ... für das Relative Risiko 920
 - , ... für die Risiko Differenz 920
- Schätzfunktion (s. auch Schätzer) 124ff
- Schätzmatrix 640
- Schätzung (s. auch Schätzer) 123ff
 - , ... der glatten Komponente 639, 642ff, 660ff
 - , ... der Spektraldichte 709ff
 - , geglättete ... 711ff
 - , ... kreuzspektraler Größen 735
 - , statistische ... 123ff
- Schaffrin, B. 634f, 1054
- Schaich, E. 1069
- Scheffé H. 335, 616f, 628, 634, 1069
- Scheffé's χ^2 -Test bei der einfachen Varianzanalyse 617f
- Scheffé-Test bei der einfachen Varianzanalyse 616
- Scheinbehandlung 362, 364
- Scheinkorrelation 77, 545
- Scheinmedikament 371
- Schicht einer Grundgesamtheit 278ff, 286ff, 312ff, 355, 370
- Schichtungsvariable 278
- Schiefe 47f, 49, 118
 - , ... der χ^2 -Verteilung 153
 - , ... der Exponentialverteilung 220
 - , ... der Lognormalverteilung 152
 - , ... der Normalverteilung 144
 - , Test auf ... und Exzeß für die Normalverteilung 189
- Schipp, B. 1060
- schließende Statistik 2, 15, 124ff
- Schlittgen, R. 640, 693, 1069
- Schmetterer, L. 1069
- Schmid, H.J. 1072
- Schmidt-Thieme, L. 1060
- Schmitz, K.-J. 328, 332, 827, 1061, 1069
- Schmitz, N. 169, 194, 1068, 1070
- Schneeberger, H. 282, 1070
- Schneeweiß, W. 608, 1070
- Schneider, D. 1058f, 1070
- Schnittmenge
- , minimale ... 750, 754, 760, 763, 773
- Schockmodell 783
- Schönenfeld, P. 744, 1070
- Schopfer 513
- Schranken der Erneuerungsfunktion 808
- Schranke
 - , Noninferiority ... 916f
- Schucany, W.R. 821, 1054
- Schulnote nach Lienert 376f
- Schulze, R. 1057, 1059
- Schumacher, E. 1059
- Schumacher, M. 1070
- schwach stationärer stochastischer Prozeß 678ff
- schwach stationär korreliert 727
- Schwaiger, B. 895, 1058
- Schwankung 705
- Schwartz, D. 373, 1070
- Schwarz, H. 276, 1070
- Schwarze, V.S. 1056f
- Schweinezyklus
- , Hanaußer ... 637
- Schweitzer, W. 1069
- Schwingungskomponente 711
- Score 883ff
 - , Normal-... 885
 - , van der Waerden-... 885
- Scorefunktion 883ff
- scoring point (s. a. Scoringpunkt) 944
- Scoringpunkt 944
 - , kombinierte Prognosen mit ... 944f
 - , ... für den Gesamteffekt 944f
 - , direkter Varianzschatz von Hartung-Böckenhoff-Knapp bei ... 945
- Scoringsystem 943ff, 966
 - , Entwicklung eines ... 943f
 - , ... bei bekannten Referenzwerten 943f
 - , ... mit Gruppenmitteln als Referenzen 944
 - , Wirtschaftspragnosen mit ... 943ff
- Screening-Test 803
- Searle, S.R. 631, 634, 1070
- Seber, G.A. F. 608, 1070
- Seckendorf, V.L.v. 12f
- Seelbinder, B.M. 167, 1016, 1070
- Segment 296
- Seifert, O. 1055
- Seiler, K.-U. 1061
- Seitz, H. 3
- sekundärstatistische Erhebung 7
- sekundärstatistische Marktforschung 314
- self-designing
 - , ... Versuchsanlage 1016
 - , ... gruppensequentieller Versuch 1016
- self-organizing

- , ... gruppensequentieller Versuch 1016
 Senderl, W. 1061
 Sensitivitätskurve 862f
 sensitivity curve 862f
 sequential probability ratio test 251
 sequentielle Analyse 372f
 sequentieller Quotiententest 251ff
 -, ... für den Erwartungswert einer Normalverteilung 259ff
 -, ... für den Parameter λ einer Exponentialverteilung 261ff
 -, ... für den Parameter p einer Binomialverteilung 252ff
 sequentieller Stichprobenplan 392ff, 399ff
 -, ... für qualitative Merkmale 392ff
 --, durchschnittlicher Stichprobenumfang beim ... 394
 -, ... für quantitative Merkmale 399ff
 --, durchschnittlicher Stichprobenumfang beim ... 399ff
 sequentielle Strategie 823
 Serienredundanz 762
 Serienschaltung von Filtern 674f
 Seriensystem 101, 111f, 746f, 751f, 759, 773, 809f
 -, Zuverlässigkeit eines ... 747
 Sethuraman, J. 1052
 Shannon, C.E. 759, 1066
 Shapiro, S.S. 190, 1070
 Sheppard 46, 86
 Sheppardsche Korrektur 46, 86f
 Shi, L. 901, 995, 1065
 Shiskin, Y. 666, 1070
 short-cut methods 826
 sicheres Ereignis 93
 Sicherheitswahrscheinlichkeit 132
 Šidák, Z. 242, 1055
 Siddiqui, M.M. 785, 1050
 Sidik, K. 1070, 908, 1014
 Siegel 526, 529
 Siegel-Tukey-Test zum Vergleich der Dispersionen zweier unabhängiger Meßreihen 529
 σ -Additivität 94
 σ -Intervalle 147
 signifikant homogene Studien und Prognosen 970, 997
 signifikant inhomogene Studien und Prognosen 970, 997
 Signifikanzniveau 134
 Signifikanztest 139
 Silcocks, P. 919, 1070
 Simonaitis, D.F. 767, 1070
 Simons, G.D. 169, 1067
 Simpson 195f
 Simpson-Verteilung (s. auch gleichschenlige Dreiecksverteilung) 196
 Simulation von Realisationen aus
 -, ... einer eindimensionalen Gleichverteilung 193f
 -, ... einer Exponentialverteilung 220
 -, ... einer Normalverteilung 145
 -, ... einer Weibullverteilung 231
 Simulationsstudie 147, 169, 202, 220, 231
 simultane Konfidenzintervalle
 -, ... für die Erwartungswerte der Regressanden bei vorgegebenen Werten des Regressors einer linearen Regression 582f
 -, ... für den Steigungsparameter und das Absolutglied einer Regressionsgeraden 581
 -, ... für die Parameter einer multiplen Regression 597
 simultaner Konfidenzstreifen für eine Regressionsgerade 582f
 simultane Vergleiche bei der einfachen Varianzanalyse 614ff
 Singpurwalla, N.D. 799, 801, 1070
 Sinha, B.K. 917, 954f, 960, 980, 994, 998, 1004, 1013f, 1060
 sixteenth 828ff
 Skala 16f, 36
 -, Ähnlichkeits... 379
 -, logarithmische ... 36
 -, metrische ... 17
 -, Niveaualgleichung bei einer ... 409
 -, Niveau einer ... 16
 -, Nominal... 16
 -, Ordinal... 17
 Skala, H.J. 315, 761, 1066
 skaleninvariante M-Schätzer 868ff
 Skalenniveau 408
 Skalierung 374ff
 -, Fechnersche ... 450
 -, Fisher-Lancaster... 450
 -, mehrdimensionale ... in Kontingenztafeln 378
 -, multidimensionale ... 379
 Slakter, M.J. 183, 1070
 Sleight, P. 1074
 Slutsky, E. 664, 1070
 Slutsky-Jule-Effekt 664
 Smirnov 183ff, 199, 226ff, 240, 520ff, 526, 1023
 Smith, H. 608, 1052
 Smith, P.G. 420, 1050
 smoothing (s. auch Glättung) 849

Sne – Sta

- Snedecor, G.W. 12, 855, 909, 1070
 Sodium-Ionen-Gehalt 530f
 Söll, H. 722, 1066
 Sonne 860
 Sorensen, N. 1072
 Spannweite 40f, 158f, 175, 197f
 Sparsamkeitsprinzip 686
 sparse data 918
 Spearman 79f, 553ff
 Spearmanscher Rangkorrelationskoeffizient 79f, 553ff
 -, Test über den ... 556ff
 Spektralanalyse 639, 699ff
 -, Kreuz... 639, 729ff
 Spektraldichte 639, 701ff
 -, diskrete ... des harmonischen Prozesses 70lf
 -, ... eines ARMA-Prozesses 704f
 -, ... eines AR-Prozesses 703f
 -, ... eines MA-Prozesses 704
 -, ... eines weißen Rauschens 702
 -, ... gefilterter Prozesse 703ff
 -, Kreuz... 729
 -, Schätzer für die ... 709ff
 --, erwartungstreuer ... 710f
 --, geglätteter ... 711ff
 --, verzerrter ... 710f
 Spektralfenster 711ff, 734
 Spektrum 639, 701ff
 -, ... eines weißen Rauschens 703
 -, Ko... 729
 -, Kreuz... 729ff
 -, Kreuzamplituden... 730
 -, Phasen... 730
 -, Quadratur... 729
 -, Stichproben... 709ff
 --, geglättetes ... 711ff
 Spencer Jones 888
 Spezifität eines Meß- oder Analyseverfahrens 341
 Splineapproximation 666
 Splinefenster 714
 Split-Plot-Experiment 359f, 628
 -, höheres ... 360
 Split-Plot-in-Strips-Experiment 359
 Spread 826, 829ff
 Stabdiagramm 21
 stabile Anlage 897, 1008
 Stabilisierung von Erträgen 899
 Stamm-und-Blätter-Darstellung 825, 827, 838f
 Standardabweichung 116f, 157ff, 864f
 -, empirische ... 46, 74, 83ff, 324, 831f
 Standardaltersaufbau 61
 standardisierte Zufallsvariable 117, 130, 144
 Standardisierte Mittlere Differenzen (Effektmaß) 991ff
 -, exakte Methoden im FE-Modell für ... 991ff
 -, exakte Methoden im RE-Modell für ... 994
 -, Hartung- Verfahren für ... 991ff
 --, approximatives Konfidenzintervall nach dem ... 993
 --, exakter Schätzer nach dem ... 992
 --, exaktes Konfidenzintervall nach dem ... 992
 -, MUPML Gesamt-Schätzer für ... 992
 --, approximativer ... 992
 --, exakter ... 992
 Standardisierung 60ff, 144
 Standardnormalverteilung 11, 109, 111ff, 115, 117, 119, 122, 144, 1026f
 -, Quantile der ... 131, 147, 1027
 -, Tabelle der Verteilungsfunktion der ... 1026
 Stanford-Binet-Intelligenzquotient 558
 Star, S.A. 1071
 stationärer Prozeß 644
 stationäre Verfügbarkeit 820, 824
 stationäre Zeitreihe 648, 962
 Stationarität
 -, ... einer Zeitreihe 744, 962f
 -- Dickey-Fuller (Unit Root Test) auf ... 744, 962f
 ---, p-Werte des ... 963
 ---, ... nach McKinnon 963
 ---, Teststatistik für den ... 963
 ---, Grenzverteilung der ... 963
 -, ... eines kreuzkorrelierten Zeitreihenpaares 744, 895
 --, kombinierter Test auf ... 964f
 ---, Interpretation des ... 965
 --, Prüfung auf ... 964
 Statistik 1ff, 8, 11, 12f
 -, beschreibende ... 2, 12, 15ff
 -, Bevölkerungs... 12
 -, deskriptive ... 2, 15ff
 -, Hotelling-Pabst... 556f
 -, induktive ... 2, 12, 124
 -, Kendallsche K-... 560f
 -, Ordnungs... 140
 -, Rang... 140
 -, schließende ... 2, 15, 124
 statistische Maßzahlen 55
 statistische Modellbildung 123
 statistischer Fehler 321

- Sta – Sti
- statistischer Schatten 8
 - statistischer Test 123
 - statistische Schätzung 123f
 - statistisches Experiment 4
 - Steel 616
 - Steel und Dwass
 - , Test von ... bei der einfachen Varianzanalyse 616
 - Stefansky, W. 586, 764, 1071
 - Steigerung des cash flow 947, 961
 - Steigungsparameter einer Regressionsgeraden 574ff
 - , Konfidenzintervall für den ... 581
 - , Schätzer für den ... 575
 - , Erwartungswert des ... 576
 - , Korrelation des Schätzers für das Absolutglied und des ... 578
 - , Kovarianz des Schätzers für das Absolutglied und des ... 578
 - , Varianz des ... 576
 - , Verteilung des ... 577
 - , simultane Konfidenzintervalle für das Absolutglied und den ... 581
 - , Test über den ... 580
 - Steilheit 387ff, 391f, 398
 - Stein, Ch. 166, 1016, 1071
 - Steiner 117, 920, 971
 - , Satz von ... 117, 920, 971
 - , Verschiebungssatz von ... 117, 920, 971
 - stem-and-leaves 838 f
 - Stenger, H. 1071
 - Stephens 185
 - Sterbewahrscheinlichkeit 12
 - Sterbeziffer 60ff
 - , altersspezifische ... 61f
 - , globale ... 60
 - , mittlere ... 60
 - , standardisierte ... 61f
 - , Vergleich von ... 61f
 - sterile Teilmasse 56, 60
 - Stern 860
 - stetige empirische Verteilungsfunktion 827, 840
 - stetiges Merkmal 17, 22
 - stetige Verteilung 106f, 112
 - stetig verteilter Zufallsvektor 108
 - Stichprobe 124, 141f, 267ff
 - , Beurteilungs... 315f
 - , einfache Zufalls... (s. auch einfache Zufallsauswahl) 124, 274
 - , Geburtstags... 370
 - , geschichtete ... (s. auch geschichtete Zufallsauswahl) 278ff, 355
 - , Lern... 377
 - , ... mit ungleichen Auswahlwahrscheinlichkeiten 297ff
 - , Namens... 370
 - , repräsentative ... 315
 - , systematische ... 295ff, 308
 - Stichprobenbasis
 - , Inventur auf ... 269ff
 - Stichprobenerhebung 6, 18, 269ff
 - , relativer Zufallsfehler bei der ... 269f
 - , Verteilungsannahmen bei ... 272ff
 - Stichprobenintervall 296
 - Stichprobenkontrolle 381, 406
 - Stichprobenkorrelation 546
 - Stichprobenkreuzspektrum 734
 - , geglättetes ... 734
 - Stichprobenplan 2, 6f, 316, 381ff
 - , doppelter ... 390ff
 - , einfacher ... 383ff, 395ff
 - , kontinuierlicher ... 406
 - , mehrfacher ... 389ff
 - , sequentieller ... 392ff, 399ff
 - Stichprobenspektrum 709ff
 - , geglättetes ... 711ff
 - Stichprobensystem
 - , Philips-Standard ... 387
 - Stichprobentheorie 6, 269ff
 - Stichprobenumfang 2
 - , benötigter ...
 - , ... bei der einfachen Zufallsauswahl 275
 - , ... bei der geschichteten Zufallsauswahl
 - , ... mit Neyman-Tschuprow-Aufteilung 282f
 - , ... mit optimaler Aufteilung 282f
 - , ... mit proportionaler Aufteilung 285
 - , ... beim Test auf Homogenität in der 2 x 2-Tafel 419ff
 - , ... beim Test über den Erwartungswert einer Normalverteilung 181f
 - , ... beim Test über den Parameter p einer Binomialverteilung 206f
 - , ... beim Test über die Differenz der Erwartungswerte zweier unabhängiger Normalverteilungen 537f
 - , ... zur Einhaltung absoluter Genauigkeiten von Intervallabschätzungen bei Normalverteilungsannahme 166ff
 - , ... zur Einhaltung prozentualer Genauigkeiten von Konfidenzintervallen für den Erwartungswert einer Normalverteilung 173ff
 - , durchschnittlicher ...

Sti – Stu

- , ... beim doppelten Stichprobenplan für qualitative Merkmale 390f
- , ... beim sequentiellen Stichprobenplan
- , ... für qualitative Merkmale 394
- , ... für quantitative Merkmale 399ff
- Stichprobenverfahren 269ff, 381f, 387
- , ... in der Marktforschung 314ff
- , Klumpen... 288ff
- , ... als einstufiges Verfahren 289ff
- , ... als mehrstufiges Verfahren 291f
- Stier, W. 640, 666, 722, 1071
- Stirling 1040
- Stirlingsche Formel 1040
- stochastisch 11
- stochastische Assoziiertheit 759
- stochastische Konvergenz 121
- stochastischer Prozeß 639, 678ff
- , schwach stationärer ... 678ff
- stochastisch größer 514
- stochastisch unabhängig 101f, 108, 119
- Stoer, J. 608, 642, 650, 793, 1055, 1071
- Störmer, H. 791, 1071
- Störvariable 361
- Stoll, R.C. 555
- Stouffer, S.A. 891, 930, 954, 957f, 1071
- Stouffer
- , ...-Methode 930f, 954f
- , Test über die globale Hypothese bei der Kombination unabhängiger Tests nach der ... 954f
- , zur Kombination von p-Werten 930f
- Straßenverkehrssignalanlagen 808ff
- , Inspektionszeiträume bei ... 817f
- , Vergleich von ... 812ff
- , Wirtschaftlichkeit von ... 815ff
- , Zuverlässigkeit von ... 808ff
- Strategie
- , Altersersatz... 780, 824
- , Bereitschafts... 823
- , einstufige ... 823
- , Gruppenersatz... 780
- , Instandhaltungs... 745
- , optionale ... 823
- , periodische ... 823
- , Präventiv... 823
- , sequentielle ... 823
- , streng-periodische ... 824
- Strater, R. 1072
- Streichfuss, K. 1059
- Streitberg, B.H.J. 640, 693, 1069
- Streuung (Dispersion) 40ff, 116ff
- , ... einer Beobachtungsreihe 826
- Streubereich 40
- Streudiagramm 338
- Streuungsmaß 40
- Streuungsparameter 19, 112, 116ff
- Streuungszerlegung
- , einfache ... 612
- Streuzahl 325
- Streuzahlverfahren zur Einhaltung prozentualer Genauigkeiten von Konfidenzintervallen für den Erwartungswert einer Normalverteilung 175ff
- Strichliste 21, 838
- Strik, R.v. 391, 1055
- Strukturanalyse 745
- Strukturfunktion eines monotonen Systems 746, 751ff
- , disjunktive Normalform der ... 752f
- , Gewinnung des Zuverlässigkeitsschaltbilds zu einer ... 755
- , minimale Pfadmengendarstellung der ... 754f
- , minimale Schnittmengendarstellung der ... 754f
- Stuart 247f
- , Test auf Trend von Cox und ... 247f
- Student 12, 154, 179, 215, 350, 1044
- Studentisierte Spannweite
- , Quantile der ... 1038f
- Studentscher t-Test 179
- Studentsche t-Verteilung (s. auch t-Verteilung) 154f
- Studie
- , Cluster-... 910f
- , Konfidenzintervall für den Gesamteffekt bei ... 911
- , Schätzer für den Gesamteffekt bei ... 911
- , Varianz des ... 911
- , Dominanz einer ... 1013
- , Partial Weighting nach Yates-Cochran bei ... 1013
- , verallgemeinertes RE-Modell bei ... 1013
- , Methoden im ... 1013
- , ...effekt 890
- , einarmige ... 909f, 921ff
- , heterogene ... 890
- , homogene ... 890
- , signifikant ... 997
- , inhomogene ... 890
- , signifikant ... 997
- , Kombination von ... 887ff
- , Nichtunterlegenheits... 1016ff
- , Placebo-kontrollierte ... 1016ff
- , adaptive ... 1016ff
- , adaptive prospektive ... 1016ff

- Stu – Test
- , prospektive ... 1016ff
 - , Noninferiority ... 910, 915
 - , Superiority ... 993
 - , Time-to-Event ... 920ff
 - , Kombination von vergleichenden ... 930ff
 - , Überlebenszeit... 920ff
 - , Kombination von vergleichenden ... 930ff
 - , zweiarmige ... 902f, 917, 974ff
 - Stüber, M. 1061
 - Subindex 67ff
 - subjektive Risikobeurteilung 966
 - , Kombination von ... 966
 - subjektive Wahrscheinlichkeit 95, 965ff
 - , Delphi Methode zur Gewinnung ... 965ff
 - , kombinierte ... 966
 - substitute primary observation 365
 - Suchman, E.A. 1071
 - summary point 828, 853
 - Süßmilch, J.P. 12
 - Suggestion
 - , Auto... 371f
 - , Hetero... 371f
 - suggestive Fragestellung 309
 - Summe der Abweichungssquarete 574
 - Summenhäufigkeit
 - , absolute ... 23
 - , relative ... 23f
 - Summenhäufigkeitsfunktion 24
 - Summe von Zufallsvariablen 108ff, 114f, 146
 - Superiority-Studie 993
 - supplementary observation 365
 - Surrogat-Variable 1014
 - Survival Analysis 744
 - , Kaplan-Meier ... 744, 925ff
 - Survivalsfunktion 925ff, 933
 - , Kaplan-Meier Schätzer für die ... 933
 - , Konfidenzstreifen für die ... 933
 - , Test auf Gleichheit zweier ... 925ff
 - , erweiterter Wilcoxon ... für zensierte Daten 928
 - , Log-Rank ... 926ff
 - , weighted ... 928
 - , Tarone-Ware ... 928
 - , Wilcoxon-Gehan ... 928
 - Sutton, A.J. 918, 1071
 - Swanson 562
 - Sweeting, M.J. 918, 1071
 - Swoboda, H. 1071
 - Symmetrie
 - , Test auf ... in der r x s-Tafel nach Bowker 440f
 - , Test auf ... in der 2 x 2-Tafel 422ff
 - , ... mittels Cochran's Q 423f
 - , ... von McNemar 423
 - symmetrische Häufigkeitsverteilung 38f, 45, 47
 - symmetrische Verteilung 115, 144f, 154, 243ff
 - Symmetrisierung
 - , ... einer Beobachtungsreihe 833
 - , ... von Daten 825, 827, 833ff
 - System
 - , elementares ... 746f
 - , ...gleichung 640
 - , k-aus-n-... 752
 - , ... mit heißer Reserve 762
 - , ... mit kalter Macro-Reserve 763
 - , ... mit kalter Micro-Reserve 763
 - , ... mit kalter Reserve 762
 - , ... mit mehr als zwei Zuständen 763ff
 - , monotones ... 750, 759, 763, 773, 782
 - , Multi-State-... 763ff
 - , Parallel... 102, 111, 746f, 752, 759
 - , redundantes ... 760, 762
 - , Serien... 101, 111, 746f, 751, 759, 773, 809f
 - , wartbares ... 806
 - systematische Auswahl 308
 - systematischer Fehler 141, 321f, 325, 332ff, 337, 342, 362
 - systematische Stichprobe 295ff, 308
 - , Schätzer für die Summe der Merkmalswerte in der Grundgesamtheit bei der ... 296
 - , Varianz des ... 296
 - Systemausfall 766, 768
 - Systemzuverlässigkeit 745, 747, 763, 771f
 - tabellarische Darstellung von Daten 19
 - Tarcolea, A.I. 960, 964f, 1051, 1061
 - Tarone, R.E. 824, 928, 1071
 - Tarone-Ware
 - , ... Test für Survivalfunktionen 824, 928
 - t-Maß 455, 459ff
 - Teilerhebung 6, 18
 - Teilmasse
 - , sterile ... 56, 60
 - tertiärstatistische Erhebung 7
 - Test 123, 133ff
 - , Anpassungs... 139
 - , χ^2 -... 182f
 - , ... auf Exponentialverteilung 225ff
 - , ... auf Gleichverteilung 198f
 - , ... auf Normalverteilung 182ff

Test – Test

- , ... auf Poissonverteilung 216f
- , Kolmogoroff-Smirnov-... 183ff
- , ... auf Exponentialverteilung 226ff
- , ... bei nichtspezifiziertem Parameter 226ff
- , ... bei spezifiziertem Parameter 226ff
- , ... auf Normalverteilung 183ff
- , ... bei nichtspezifizierten Parametern 185f
- , ... bei spezifizierten Parametern 183ff
- , Ansari-Bradley-Freund-... 526ff
- , ... auf Ausreißer 343ff
- , ... von David-Hartley-Pearson 345f
- , ... von Dixon 346f
- , ... von Gubbs 344f
- , ... auf ein Ausreißerpaar 347ff
- , ... auf Gleichheit zweier Survivalfunktionen 925ff
- , erweiterter Wilcoxon ... für zensierte Daten 928
- , ... auf Homogenität
- , ... für binäre Daten 1013
- , ... für zweiarmige Studien 1013
- , ... in der 2 x 2-Tafel 418ff
- , ... von Cochran 901, 1013
- , ... von Effektmäßen 917
- , ... von Studien 900ff
- , ... von Welch 1013
- , ... auf Normalität 178ff
- , ... auf Schiefe und Exzeß 189
- , ... auf Stationarität einer Zeitreihe 744, 962ff
- , Dickey-Fuller Unit Root ... 744, 962f
- , p-Werte des ... 963
- , ... nach McKinnon 963
- , Teststatistik des ... 963
- , Grenzverteilung der ... 963
- , ... auf Stationarität eines kreuzkorrelierten Zeitreihenpanels 744, 964
- , Kombination der Dickey-Fuller p-Werte nach Hartung für den ... 964
- , ... mit Fisher Probits 964
- , ... auf Symmetrie
- , ... in der 2 x 2-Tafel 422ff
- , ... mittels Cochran's Q 423f
- , ... von McNemar 423
- , ... nach Bowker im loglinearen Modell für die r x s-Tafel 440f
- , ... auf Trend 247ff, 641, 669f
- , ... nach Mann 249ff, 669f
- , ... von Cox und Stuart 247f
- , ... auf Unabhängigkeit
- , ... in der 2 x 2-Tafel 413ff
- , ... von p Meßreihen mittels multipler Vergleiche nach Holm 567f
- , ... von zwei Meßreihen 547ff
- , Bartlett-... 617
- , ... bei der zweifachen Varianzanalyse
- , ... mit Wechselwirkungen zwischen den Faktoren 625f
- , ... ohne Wechselwirkungen zwischen den Faktoren 628
- , ... beim einfachen Blockexperiment 620, 622f
- , Beurteilungskriterien für ... 137
- , χ^2 -...
- , ... auf Unabhängigkeit in der 2 x 2-Tafel 413f
- , ... im loglinearen Modell für die r x s-Tafel 435ff
- , χ^2 -Anpassungs... 182f
- , ... auf Exponentialverteilung 225ff
- , ... auf Gleichverteilung 198f
- , ... auf Normalverteilung 182ff
- , ... auf Pössonverteilung 216f
- , Cochran'scher Homogenitäts... 901, 1013
- , Log-Rank ... 824, 926ff
- , weighted ... 928
- , Tarone-Ware ... 824, 928
- , Wilcoxon-Gehan ... 824, 928
- , cttot -...
- , ... gegen DFRA-Lebensdauerverteilungen 784
- , ... gegen DFR-Lebensdauerverteilungen nach Epstein 778
- , ... gegen IFRA-Lebensdauerverteilungen 784
- , ... gegen IFR-Lebensdauerverteilungen nach Epstein 778
- , Dickey-Fuller ... 744, 895
- , Dickey-Fuller Unit Root ... 744, 962f
- , ... auf Stationarität einer Zeitreihe 744, 962ff
- , p-Werte des ... 963
- , ... nach McKinnon 963
- , Teststatistik des ... 963
- , Grenzverteilung der ... 963
- , Durbin-Watson-... 740f
- , einseitiger ... 133 ff
- , Einstichprobengauß... 135ff, 152, 178f, 215f, 223f, 507, 535, 549
- , erweiterter Wilcoxon ... für zensierte Daten 928
- , exakter ... auf Unabhängigkeit in der 2 x 2-Tafel 414ff
- , ... von Fisher 416ff

- | Test – Test | Test – Test |
|---|---|
| -, F-... 611f, 620f, 625f, 628 | -, Kombination abhängiger ... 895, 930f, 958ff |
| -, Friedman-... | --, Hartung-Methode zur ... 959ff |
| --, ... bei der zweifachen Varianzanalyse ohne Wechselwirkung zwischen den Faktoren 628 | ---, kombinierter p-Wert nach der ... 960 |
| --, ... beim einfachen Blockexperiment 622f | ----, ... für Probits 960 |
| -, Friedman-Rangsummen... 623 | ----, ... für Fisher-Probits 960 |
| --, ... für Quantile 237 | -, Kombination unabhängiger ... 954ff |
| -, Gauß... 135ff | --, globale Hypothese bei der ... 954 |
| --, ... gegen DFRA-
Lebensdauerverteilungen 784 | --, Test über die ... 954f |
| --, ... gegen DFR-Lebensdauerverteilungen 777f | ----, ... bei Heterogenität 956 |
| --, ... gegen IFRA-
Lebensdauerverteilungen 784 | ----, ... nach der Fisher-Methode 955 |
| --, ... gegen IFR-Lebensdauerverteilungen 777f | ----, ... nach der gewichteten Inversen Normalmethode 955 |
| --, ... gegen NBUE-
Lebensdauerverteilungen 784f | ----, ... nach der Hartung-Knapp-Methode 956 |
| --, ... gegen NBU-Lebensdauerverteilungen 780f | ----, ... nach der inversen χ^2 -Methode 955 |
| --, ... gegen NWUE-Lebensdauerverteilun-
gen 784f | ----, ... nach der Stouffer-Methode 954f |
| --, ... gegen NWU-
Lebensdauerverteilungen 780f | -, kombinierter ... auf Stationarität eines kreuzkorrelierten Zeitreihenpanels 744, 964f |
| -, gleichmäßig bester ... 138 | --, Interpretation des ... 965 |
| -, Hollander-Proschan -... | -, konservativer ... 138 |
| --, ... gegen NBUE-Lebensdauerverteilun-
gen 784f | -, konsistenter ... 138 |
| --, ... gegen NBU-
Lebensdauerverteilungen 780f | -, Kruskal-Wallis... 613ff |
| --, ... gegen NWUE-
Lebensdauerverteilungen 784f | -, Levene-... 617 |
| --, ... gegen NWU-
Lebensdauerverteilungen 780f | -, Likelihood-Quotienten-... im loglinea-
ren Modell für die $r \times s$ -Tafel 435ff |
| -, Homogenitäts... | -, Log-Rank ... 890, 926ff |
| --, ... für binäre Daten 1013 | --, weighted ... 928 |
| --, ... für zweiarmige Studien 1013 | --, ... auf Gleichheit zweier Survivalfunk-
tionen 926ff |
| --, ... von Cochran 901, 1013 | --, weighted ... 928 |
| --, ... von Welch 1013 | -, ... mittels der Statistiken τ_A und τ_B im
loglinearen Modell für die $r \times s$ -Tafel
439f |
| -, Intelligenz... 269, 376f | -, Niveau- α -... 133 |
| -, Kolmogoroff-Smirnov-Anpassungs... | -, Omnibus... 524 |
| --, ... auf Exponentialverteilung 226ff | -, Pflanzagl... 531 |
| --, ... bei nichtspezifiziertem Parameter 226ff | -, Proschan-Pyke-... |
| --, ... bei spezifiziertem Parameter 226ff | --, ... gegen DFR-Lebensdauerverteilun-
gen 777 |
| --, ... auf Normalverteilung 183ff | --, ... gegen IFR-Lebensdauerverteilungen
777 |
| --, ... bei nichtspezifizierten Parametern 185f | -, psychologischer ... 269 |
| --, ... bei spezifizierten Parametern 183ff | -, Reduktions... bei der multiplen Regres-
sion 598 |
| -, Kolmogoroff-Smirnov-... zum Ver-
gleich zweier unabhängiger Meßreihen 520ff | -, Scheffé-... 616 |
| | -, Scheffé's χ^2 -... 617f |
| | -, Screening... 803 |
| | -, sequentieller Quotienten... 251ff |
| | --, ... für den Erwartungswert einer Nor-
malverteilung 259ff |

Test – Test

- , ... für den Parameter λ einer Exponentialverteilung 261ff
- , ... für den Parameter p einer Binomialverteilung 252ff
- , Siegel-Tukey-... 529
- , Signifikanz... 133f
- , statistischer ... 123, 133ff
- , Studentscher t-... 179, 536ff
- , Symmetrie...
- , ... in der 2 x 2-Tafel 422ff
- , ... mittels Cochran's Q 423f
- , ... von McNemar 423
- , ... nach Bowker im loglinearen Modell für die r x s-Tafel 440f
- , t-... 179, 536ff
- , Tarone-Ware ... 928
- , Tukey-... 616, 620
- , U-... von Mann Whitney 519f
- , ... über das Absolutglied einer Regressionsgeraden 580
- , ... über den Erwartungswert einer Normalverteilung 135ff, 178f
- , ... bei bekannter Varianz 135ff, 178f
- , ... bei unbekannter Varianz 179
- , benötigter Stichprobenumfang beim ... 181f
- , ... über den Heterogenitätsparameter im RE-Modell 998
- , ... über den Kendallschen Rangkorrelationskoeffizienten 560f
- , ... über den Noninferiority Parameter 915
- , ... über den Parameter λ einer Exponentialverteilung 223ff
- , ... über den Parameter λ einer Poissonverteilung 214ff
- , ... über den Parameter p einer Binomialverteilung 205f
- , benötigter Stichprobenumfang beim ... 206f
- , ... über den Pearsonschen Korrelationskoeffizienten 547f
- , approximativer ... 548ff
- , ... über den Quotienten der Varianzen zweier unabhängiger Normalverteilungen 524ff
- , ... über den Spearmanschen Rangkorrelationskoeffizienten 556ff
- , ... über den Steigungsparameter einer Regressionsgeraden 580
- , ... über die bi-partielle Korrelation 564
- , ... über die Differenz der Erwartungswerte zweier abhängiger Normalverteilungen 535ff
- , ... über die Differenz der Erwartungswerte zweier unabhängiger Normalverteilungen 505ff
- , benötigter Stichprobenumfang beim ... 511ff
- , ... über die globale Hypothese bei der Kombination unabhängiger Tests 954ff
- , ... bei Heterogenität 956
- , ... nach der Fisher-Methode 955
- , ... nach der gewichteten Inversen Normalmethode 955
- , ... nach der Hartung-Knapp-Methode 956
- , ... nach der inversen χ^2 -Methode 955
- , ... nach der Stouffer-Methode 954f
- , ... über die multiple Korrelation 566
- , ... über die Parameter einer multiplen Regression 597
- , ... über die Parameter einer Normalverteilung 135ff, 178ff
- , ... über die Parameter einer quadratischen Regressionsfunktion 590f
- , ... über die partielle Korrelation 562f
- , ... über die Varianz der Fehlerterme einer linearen Regression 580
- , ... über die Varianz einer Normalverteilung 179ff, 539
- , ... über Dispersionsunterschiede bei der einfachen Varianzanalyse 617f
- , ... über Lokationsunterschiede bei der einfachen Varianzanalyse 611ff
- , unverfälschter ... 137
- , van der Waerden-... 885
- , verallgemeinerter Welch-... über den Heterogenitätsparameter im RE-Modell 998
- , ... von Hypothesen über die Parameter des saturierten loglinearen Modells für mehrdimensionale Kontingenztafeln 477ff
- , ... von Moses 529ff
- , ... von Steel und Dwass 616
- , Vorzeichen... 242f, 540
- , Vorzeichenrang... von Wilcoxon 243ff, 541f
- , Welch-... auf Homogenität 1013
- , Wilcoxon-Gehan ... 928
- , Wilcoxon-Rangsummen-... 513ff
- , Wilcoxon-signed-rank-... 243ff, 541f
- , X- ... 885
- , Zeichen... 242f, 540
- , zeitraffender ... 795ff

- Test – Tuk
- , ... zum Niveau α 133
 - , ... zum Vergleich der Dispersionen
 - zweier unabhängiger Meßreihen 524ff
 - , ... zum Vergleich der Lokationen zweier unabhängiger Meßreihen 505ff, 513ff
 - , ... zum Vergleich der Parameter mehrerer Binomialverteilungen 496ff
 - , ... zum Vergleich der Parameter mehrerer Multinomialverteilungen 498
 - , ... zum Vergleich der Parameter mehrerer Poissonverteilungen 495f
 - , zweiseitiger ... 139
 - Testproblem 133ff
 - einseitiges ... 133ff
 - zweiseitiges ... 139
 - Teststatistik 134
 - Testtheorie 12, 133ff
 - therapeutic clinical research 363
 - Thermodynamik
 - zweites Gesetz der ... 5
 - thirtyseconds 828ff
 - Thomas, R.E. 799, 1071
 - Thompson, D.J. 299, 1063
 - Thompson, M. 821, 1071
 - Thompson, S.G. 892f, 969, 971, 976, 1062
 - Thompson, W.A. 340, 1071
 - Threshold Value 959, 961
 - tie (s. auch Bindung) 515
 - Tiefe 828
 - Tiefe-Index 828ff
 - Tiefpaß-Filter 664, 705ff
 - idealer ... 706f
 - Time-to-Event Studie 924ff
 - Kombination von vergleichenden ... 930ff
 - Meta-Analyse für ... 924ff
 - Tintner, G. 643f, 668, 1071
 - T-Note nach McCall 376
 - Toleranzfaktor 164, 171f
 - Toleranzintervall 132f, 163ff, 238ff
 - , ... für eine zukünftige Stichprobe aus einer Normalverteilung 163ff
 - , ... bei bekannter Varianz 163f
 - , benötigter Stichprobenumfang für ein ... 171f
 - , ... bei unbekannter Varianz 164f
 - , benötigter Stichprobenumfang für ein ... 172f
 - , nichtparametrisches ... 283ff
 - Tomášek, K. 795, 1071
 - totale Gleichverteilungshypothese 434
 - Totalerhebung 6, 18, 271
 - totale Richtigkeit eines Meß- oder Analyseverfahrens 342
 - totale Varianz im RE-Modell 968
 - Totalkontrolle 381f, 389, 392, 406
 - Trägermenge 106
 - Transformation 109, 144, 193, 321, 349ff
 - Arcus-Sinus-... 352f
 - , ... auf Linearität 587ff
 - , Box-Cox-... 352, 684f
 - , Fishersche Z ... 354, 548ff
 - , inverse Wilson-Hilferty ... 960
 - , lineare ... 85, 109
 - , logarithmische ... 351f
 - , ... mehrphasiger Missionen 772ff
 - , reziproke ... 349
 - , Wurzel... 349ff
 - Transformationsmatrix
 - , Cochrane-Orcutt-... 739
 - Trefferwahrscheinlichkeit 132f, 163ff, 170, 582
 - Trend 247ff, 531ff, 641, 669f
 - , Test auf ... 247ff, 531ff, 641, 669f
 - , ... nach Mann 249ff, 669f
 - , ... von Cox und Stuart 247f
 - , ... zweier unabhängiger Meßreihen 531ff
 - Trendkomponente 639, 641ff
 - Trendmodell
 - , nichtlineares ... 642ff
 - Trendschätzung 642ff
 - Trenkler, G. 608, 1059, 1061, 1071
 - Trennschärfe 387
 - Trichophyton 513
 - trigonometrische Funktion 860
 - trimean 883
 - Troschke, S.-O. 895, 1071
 - truncated distribution 150
 - Tryon, M. 602
 - Tschebyscheff 116, 121, 147, 323, 943, 966
 - Tschebyscheff(sche) Ungleichung 116, 121, 147, 323, 943
 - Tschuprow 282ff, 451
 - , Assoziationsmaß von ... 451f
 - t-Test 179, 536ff
 - Tuberkulose TB 1013ff
 - Tucker, J.R. 1052
 - Tukey, J.W. 526, 529, 616, 620, 711ff, 826, 872, 950, 1016, 1048ff, 1063, 1071
 - Tukey, P.A. 1050
 - Tukey-Fenster 712, 714
 - , Bandbreite des ... 712
 - Tukey's biweight 872ff
 - , ψ -Funktion von ... 872
 - , ρ -Funktion von ... 872

Tuk – Var

- Tukey-Test 616, 620
- Turnbull, B. 1016, 1063
- t-Verteilung 12, 131, 154f
 - , Approximation der ... 154
 - , Dichte der ... 154
 - , Erwartungswert der ... 154
 - , Median der ... 154
 - , Quantile der ... 154, 736
 - , Varianz der ... 154
- typische Auswahl 316f

- Überla, K. 379, 1064, 1072
- Überlebensdauerverteilung 824
 - , Schätzen der ... 824
 - Überlebenswahrscheinlichkeit 218ff, 230,
 - 233, 235, 775, 779, 782f, 786, 791, 809, 820, 925, 932
 - , bedingte ... 218, 230, 233, 235, 775, 779
 - , ... bei der Erlang-n Verteilung 235
 - , ... bei der Exponentialverteilung 219
 - , ... bei der Hjorth- (IDB-) Verteilung 233
 - , ... bei der Weibullverteilung 230
 - , ... bei der Erlang-n Verteilung 235
 - , ... bei der Exponentialverteilung 219
 - , ... bei der Hjorth- (IDB-) Verteilung 233
 - , Kombination von ... 938
 - , Konfidenzintervalle für die ... 938
 - , Schätzer für die ... 938
 - , Konfidenzintervall für die ... 932
 - , Schätzer für die ... 932
 - , ... bei der Weibullverteilung 230
 - Überlebenszeit
 - , Kombination von ... 935ff
 - , Konfidenzintervall für die ... 933f
 - , mediane ... 933f
 - , Kombination von ... 937f
 - , Schätzer für die ... 933f
 - , ...studie 889f, 924ff
 - , Kombination von vergleichenden ... 930ff
 - Meta-Analyse für ... 924ff
 - Überwachung 806
 - Uhlmann, W. 391, 1072
 - Umbasierung einer Zeitreihe einfacher Indizes 59
 - Umbullquote 892
 - Unabhängigkeit 101, 108, 119
 - , Test auf ... in der 2 x 2-Tafel 413ff
 - , Test auf ... von p Meßreihen mittels multipler Vergleiche nach Holm 567ff
 - , Test auf... von zwei Meßreihen 547ff
 - Unabhängigkeitshypothese im loglinearen Modell für die r x s-Tafel 433ff
 - unbedingter Least-Squares-Ansatz 688f
 - Unbestimmtheitsmaß 579f
 - unconditional least squares 688f
 - UND-Glied 766
 - unendliche Grundgesamtheit 305f
 - unendlischer Grundraum 91
 - Ungleichung von Tschebyscheff 116, 121, 147, 323
 - Ungleichung von Jensen 113
 - Unit Root Test
 - , Dickey-Fuller ... 744, 962f
 - Unkorreliertheit 76, 119f, 546
 - Unpräzision 341
 - unteres Quartil 34, 41
 - Untersuchungseinheit 4, 16, 18, 20, 72
 - unvereinbare Ereignisse 93
 - unverfälschter Test 137
 - Unverfälschtheit 137
 - unvollständige Betafunktion 11
 - unvollständiger Blockplan 358
 - Upton, G.J.G. 1072
 - Urliste 20f, 34, 40, 44
 - US National Research Council 887, 958
 - U-Test von Mann-Whitney 519f

 - van der Waerden-Score 885
 - van der Waerden-Test 885
 - van Outryve 889, 918
 - Variable
 - , begleitende ... 365
 - , concomitant ... 365
 - , endogene ... 744
 - , erklärende ... 365
 - , Ersatz... 365
 - , exogene ... 744
 - , Kontroll... 365
 - , verzögerte ... 744
 - , Ziel... 365
 - , Zufalls... 103ff, 112ff
 - Varianz 44ff, 116f, 125f
 - , ... der allgemeinen Dreiecksverteilung 196
 - , ... der Binomialverteilung 200
 - , ... der χ^2 -Verteilung 153
 - , ... der eindimensionalen Gleichverteilung 193
 - , ... der Erlang-n-Verteilung 235
 - , ... der Exponentialverteilung 220
 - , ... der F-Verteilung 156
 - , ... der gleichschenkligen Dreiecksverteilung 196
 - , ... der Hjorth- (IDB-) Verteilung 233

- | Var – Verg | Sach- und Namensregister |
|---|--------------------------|
| -,... der hypergeometrischen Verteilung
207 | |
| -,... der Lognormalverteilung 152 | |
| -,... der Normalverteilung 143 | |
| -,... der Poissonverteilung 213 | |
| -,... der t-Verteilung 154 | |
| -,... der Weibullverteilung 231 | |
| -,... der zweiseitig gestützten Normalverteilung 149 | |
| -,... empirische ... 44ff | |
| -,... bei gruppierten Daten 44ff | |
| -,... gepoolte ... 977 | |
| -,... Rechenregeln für ... 117 | |
| -,... totale ... im RE-Modell 968 | |
| Varianzanalyse 12, 128, 358, 374,
407, 409, 609ff | |
| -,... einfache ... 610ff | |
| -,... multivariate ... 358, 374 | |
| -,... zweifache ... 358, 624ff | |
| --,... mit Wechselwirkungen zwischen den Faktoren 625ff | |
| --,... ohne Wechselwirkungen zwischen den Faktoren 627f | |
| Varianzanalysemethode 128 | |
| Varianzanalysemodell aus der Geodäsie
634ff | |
| Varianzanalysetafel | |
| -,... der einfachen Varianzanalyse 611 | |
| -,... der einfach hierarchischen Klassifikation 630 | |
| -,... der zweifachen Varianzanalyse | |
| --,... mit Wechselwirkungen zwischen den Faktoren 626 | |
| --,... ohne Wechselwirkungen zwischen den Faktoren 628 | |
| -,... des einfachen Blockexperiments 621 | |
| Varianzaufblähungsfaktor 290 | |
| Varianzkomponente 612, 629, 631, 634 | |
| Varianzschatzter | |
| -,... adjustierter 908 | |
| -,... direkter ... von Hartung-Böckenhoff-Knapp 942f, 945 | |
| -,... Eigenschaften der ... in der Meta-Analyse 1018ff | |
| Varianzschätzung | |
| -,... bei Unkenntnis der Einzelvarianzen im GCM Problem 941ff | |
| -,... ordnungserhaltende ... 908 | |
| Varianzspektrum 702 | |
| Varianzzerlegung im RE-Modell 968 | |
| variate differences 668ff | |
| Variationskoeffizient 47, 118, 174, 244,
313, 324 | |
| Variationszahlverfahren zur Einhaltung prozentualer Genauigkeiten von Konfi | |
| denzintervallen für den Erwartungswert einer Normalverteilung 173f | |
| Vasanta, B. 1050 | |
| Vektoroptimierung (s. a. Paretooptimierung) 897 | |
| Velleman, P.D. 826, 837, 1072 | |
| verallgemeinerte inverse Verteilungsfunktion 114 | |
| verallgemeinerter Hedges-Schätzer 911,
951 | |
| verallgemeinerter Intraklass-Korrelationskoeffizient 968 | |
| -,... für heteroskedastische Fehlervarianzen 968 | |
| verallgemeinerter Welch-Test für den Heterogenitätsparameter im RE-Modell 998 | |
| verallgemeinertes RE-Modell 1013 | |
| -,... bei Dominanz einer Studie oder Prognose 1013 | |
| --,... Methoden im ... 1013 | |
| Verbundenheitskoeffizient von Yule 444ff | |
| Verdoppelungsformel bei der Gammafunktion 1040 | |
| Verdreifachungsformel bei der Gammafunktion 1040 | |
| Vereinigungsbildung bei Mengen 92 | |
| Verfügbarkeit 745, 820ff | |
| -,... Dauer... 820, 822 | |
| -,... maximale ... 824 | |
| -,... momentane ... 820 | |
| -,... stationäre ... 820, 824 | |
| Vergleich | |
| -,... der Parameter mehrerer Binomialverteilungen 496ff | |
| -,... der Parameter mehrerer Multinomialverteilungen 498 | |
| -,... der Parameter mehrerer Poissonverteilungen 495f | |
| -,... interlaboratorieller ... 337 | |
| -,... intralaboratorieller ... 337 | |
| -,... mehrerer Pearsonsscher Korrelationskoeffizienten 552f | |
| -,... multiple ... bei der einfachen Varianzanalyse 614ff | |
| -,... multiple ... bei der zweifachen Varianzanalyse 625 | |
| -,... multiple ... beim einfachen Blockexperiment 620, 623 | |
| -,... multiple ... nach Holm zum Test auf Unabhängigkeit von p Meßreihen 567f | |
| -,... simultane ... bei der einfachen Varianzanalyse 614ff | |

Verg – Vert

- , ... von Beobachtungsreihen mittels Box-Plots 836ff
- , ... von p unabhängigen Meßreihen 609ff
- , ... von Sterbeziffern 61f
- , ... von Straßenverkehrssignalanlagen 812ff
- , ... von zwei abhängigen Meßreihen 533ff
- , ... von zwei Pearsonschen Korrelationskoeffizienten 551f
- , ... von zwei unabhängigen Meßreihen 505ff
- , Methode von ... 649
- Verhältnisschätzer 300, 302
- , Varianz des ... 302
- , Schätzer für die ... 302
- Verhältniszahl 19, 55
- Verhagen 649
- Verhulst, P.-F. 643, 1072
- Verkettung von Zeitreihen einfacher Indizes 60
- Vermengung (s. a. confounding) 941
- Vermutung
- , Kombination von ... 887ff
- Verschiebungssatz von Steiner 117, 920, 971
- Verschleiß 807
- Versuch
- , Blind... 372
- , Doppelblind... 372f
- , gruppensequentieller ... 1016
- , self-designing ... 1016
- , self-organizing ... 1016
- , klinischer ... 363ff
- Versuchsanlage
- , ... zur prospektiven Meta-Analyse 1014ff
- , adaptive ... 1016
- , self-designing ... 1016
- Versuchsplan 364f
- , faktorieller ... 628
- , vollständig randomisierter ... 355f, 365, 609f, 619
- Versuchsplanung 12, 354ff
- Verteilung 39, 104ff
- , allgemeine Dreiecks... 195f
- , Bernoulli... 200
- , Binomial... 100, 113ff, 122f, 127, 199ff, 252ff, 496ff
- , bivariate Normal... 120
- , Cauchy... 113
- , χ^2 ... 152f
- , ... der Lebensdauer 218ff, 774ff, 780ff, 824
- , Schätzen der ... 824
- , ... der Restlebensdauer 218ff
- , DFR-... 776ff, 786f, 803
- , DFRA-... 783f, 786f
- , diskrete ... 106, 108, 126f
- , Dreiecks... 195f
- , einseitige Dreiecks... 196
- , Erlang-n... 234f
- , Exponential... 105ff, 111f, 115, 117f, 124, 126, 193, 219ff, 261 ff, 788ff
- , F-... 12, 156f
- , Gauß... (s. auch Normalverteilung) 11, 109
- , gedächtnislose ... 219
- , gemeinsame ... 108, 120, 194f
- , gestutzte Normal... 148ff
- , Gleich... 116, 128, 147, 192ff
- , gleichschenklige Dreiecks... 195f
- , Hjorth-... 232ff, 794
- , hypergeometrische ... 207ff
- , IDB-... 232ff, 794
- , IFR-... 779f, 786f, 808, 823
- , IFRA-... 783f, 786f
- , Kenngrößen von ... 112
- , Laplace... 192
- , Lebensdauer ... 218ff, 774ff, 780ff, 824
- , Schätzen der ... 824
- , linksseitige Dreiecks... 196
- , linksseitig gestutzte Normal... 150
- , Lognormal... 151f, 793f
- , ... mit Dichte 106
- , multidimensionale hypergeometrische ... 210
- , Multinomial... 209ff, 498
- , NBU-... 779ff, 786f
- , NBUE-... 784ff, 808, 823
- , Normal... 109, 111ff, 115, 117ff, 124, 128, 130, 132f, 143ff, 259ff
- , NWU-... 779ff, 786f
- , NWUE-... 784ff
- , Poisson... 113, 122, 212ff, 495f
- , Polynomial... 210
- , Rand... 107, 120
- , Rayleigh... 231
- , Rechteck... 116, 192ff
- , rechtsschiefe ... 151
- , rechtsseitige Dreiecks... 196
- , rechtsseitig gestutzte Normal... 150
- , Simpson... 196
- , Standardnormal... 106, 112, 127
- , Summe von ... 111, 113f
- , symmetrische ... 115
- , t-... 12, 154f

- , Überlebensdauer ... 824
 --, Schätzen der ... 824
 -, Weibull ... 230ff, 791ff
 -, zweidimensionale ... 72ff, 194f
 -, zweiseitig gestützte Normal... 150
 Verteilungsannahmen bei Stichprobenerhebungen 272ff
 Verteilungsfunktion 11, 104ff, 109, 121
 -, ... der eindimensionalen Gleichverteilung 193
 -, ... der Erlang-n-Verteilung 235
 -, ... der Exponentialverteilung 219
 -, ... der Hjorth- (IDB-) Verteilung 232
 -, ... der Lognormalverteilung 151
 -, ... der Normalverteilung 145
 -, ... der Standardnormalverteilung 109, 1026
 -, ... der Weibullverteilung 230
 -, ... der zweiseitig gestützten Normalverteilung 149
 -, Eigenschaften der ... 106
 -, empirische ... (s. auch Häufigkeitsverteilung) 11, 29, 121, 825, 827, 839f
 --, stetige ... 827, 840
 -, gemeinsame ... 108
 -, Konfidenzstreifen für eine unbekannte ... 240f
 -, mehrdimensionale ... 107
 -, multivariate ... 107
 -, verallgemeinerte inverse ... 114
 Vertrauensbereich 130
 Vertrauensintervall 130
 Verursachungszahl 56
 Verzerrung eines Schätzers 125, 300, 362
 Verzerrungskorrektur 906
 verzögerte Variable 744
 Victor, N. 910, 994, 1072f, 1074
 Viechtbauer, W. 998, 1072
 Vierfeldertafel (s. auch 2 x 2-Kontingenztafel) 12, 82f, 411ff, 442ff
 Vinçe, I. 1072
 Voet, B. 608, 621, 634, 636, 971, 1004, 1058, 1061, 1072
 Vogel, F. 1062
 Vogel, W. 1072
 Volatilität 897, 1004ff
 -, Anlage mit geringster ... 1009
 -, jährliche ... 1005
 -, Rendite-...-Projektionsebene 1007
 -, ...schranke 1008f
 Volkszählung 12, 19
 Vollerhebung 6, 18, 271
 vollständiges Ignorieren 864
 vollständig randomisierter Versuchsplan 355f, 364f, 609f, 619
- Vert – War
- Voralterungsmethode 803
 vorbeugende Instandsetzung 823
 Vorerhebung 272, 276, 282
 Vorstichprobe 166f, 170f, 176
 Vorzeichenrangtest von Wilcoxon 243ff, 541f
 Vorzeichentest (s. Zeichentest) 242f, 540
- W**
 Wachstumskurve 642ff
 Waerden, B.L.v.d. (s. a. van der Warden) 885, 1047, 1072
 Wäsch, P. 1067
 Wagner, G. 392, 1072
 Wagner, S. 928, 1072
 Wahlprognose 2
 wahrscheinlicher Fehler 323
 Wahrscheinlichkeit 93ff, 104
 -, bedingte ... 98ff, 104
 -, Laplace ... 95, 97f
 -, Multiplikationssatz für bedingte ... 99f
 -, Satz von der totalen ... 99f, 102
 -, Sicherheits... 271, 276
 -, subjektive ... 95, 965ff
 --, Delphi Methode zur Gewinnung ... 965ff
 --, kombinierte ... 966
 Wahrscheinlichkeitsdichte (s. auch Dichte) 107
 Wahrscheinlichkeitspapier 189, 825, 827, 847
 -, Überprüfung der Normalverteilungsannahme mittels ... 189f
 Wahrscheinlichkeitsrechnung 91ff
 Wahrscheinlichkeitsverteilung 104
 Wald, A. 12, 373, 406, 605, 607, 999, 1013, 1072
 Wald
 -, Likelihood-Methoden für den Heterogenitätsparameter im RE-Modell vom ... Typ 998f
 -, Konfidenzintervall für den Heterogenitätsparameter im RE-Modell vom ... Typ 999
 -, Regressionsgerade nach ... 605ff
 --, Schätzer für die Parameter der ... 606
 Waldsches Verfahren 605ff
 Walker 679, 689
 Wall, K.-D. 1065
 Wallenstein 573
 Wallis, W.A. 613ff, 617, 1023, 1072
 Walukiewicz, S. 1056
 Ware, J. 824, 928, 1071
 Warenkorb 62f

War – Wie

- , Wert eines ... 63
- Warmuth-Metz, M. 1072
- Warngrenze 336f, 402ff
- wartbares System 806f
- , Ausfallwahrscheinlichkeit eines ... 806
- Wartbarkeit 806f
- Wartung 806f
 - , korrigierende ... 807
 - , prophylaktische ... 807
- Wassmer, G. 993, 1065
- Waternaux, C.M. 1052
- Watson, G.S. 740, 1052
- Weber, E. 379, 1072
- Weber, M. 895, 921, 1072
- Webster, S.L. 1069
- Wechselwirkung 360, 429, 442, 465, 624ff, 890, 906
- Wechselwirkungseffekt 625
- Wechsler 376f
- , Intelligenztest nach ... 376f
- Weelwright, S.C. 895, 1066
- Wegner, F. 1069
- Weibull, W. 230ff, 791ff, 1073
- Weibullverteilung 230ff, 791ff
 - , Ausfallrate einer ... 231
 - , bedingte Überlebenswahrscheinlichkeit bei einer ... 230
 - , Dichte einer ... 230
 - , Erwartungswert einer ... 231
 - , Konfidenzintervalle für die Parameter einer ... 791ff
 - , ... bei unzensierten Lebensdauerprüfungen 793
 - , ... bei vorgegebener Zahl von Ausfällen ohne Ersetzung 791ff
 - , Median einer ... 231
 - , Schätzer für die Parameter einer ... 232, 791ff
 - , ... bei unzensierten Lebensdauerprüfungen 793
 - , ... bei vorgegebener Zahl von Ausfällen ohne Ersetzung 791ff
 - , Simulation von Realisationen aus einer ... 231
 - , Überlebenswahrscheinlichkeit bei einer ... 230
 - , Varianz einer ... 231
 - , Verteilungsfunktion einer ... 230
- weighted Log-Rank Test 928
- Weihs, C. 1057, 1064
- weißes Rauschen 639, 678ff
 - , Spektraldichte eines ... 702
 - , Spektrum eines ... 703
- Welch, B.L. 998, 1013, 1073
- Welch
 - , ...-Test auf Homogenität 1013
 - , verallgemeinerter ...-Test über den Heterogenitätsparameter im RE-Modell 998
- Wermuth, N. 1073
- Wemer, H.J. 363, 600, 608, 628, 1061, 1073
- Wemer, K. 796, 1073
- Wertindex 62f
- Wertpapierkurs 1010
 - , Korrelation der ... 1010
- Wertpapiermanagement 897, 1004ff
 - , Anlagefenster im ... 1008
 - , Anlage mit geringstem Risiko im ... 1006, 1009
 - , Anlage mit geringster Volatilität im ... 1009
 - , Entscheidungsfenster im ... 1008
 - , Entscheidungsraum im ... 1007
 - , jährliche Volatilität im ... 1005
 - , mittlere Rendite im ... 1005
 - , Portfolio im ... 1004ff
 - , Portfolioselektion im ... 897, 1004f
 - , effiziente ... 1006ff
 - , Einbettung der ... 1011f
 - , ... in das ERE-Modell 1011f
 - , ... in das RE-Modell 1011f
 - , ineffiziente ... 1008
 - , pareto-optimale ... 897, 1005ff
 - , Rendite im ... 1004f
 - , Rendite-Schwelle im ... 1008f
 - , Rendite-Volatilitäts-Projektionsebene im ... 1007
 - , Risikofunktion im ... 1006
 - , stabile Anlage im ... 1008
 - , Volatilitätschranke im ... 1008f
 - , Volatilität im ... 1004ff
 - , Wertpapierkurs im ... 1010
 - , Korrelation der ... 1010
 - , Wertpapiermanagement im ... 1004ff
 - , Wertpapier-Portfolio im ... 1004f
- Wertpapier-Portfolio 1004f
- Wertungsfaktoren für im Labor ermittelte Ausfallraten 804
- Wetzel, W. 1062, 1069, 1073
- White, K.J. 741, 1069
- Whitehead, A. 1015, 1073
- white noise 678
- Whitney 513, 520
- Wiederholbarkeit 370
- Wiederholungseffekt 624
- Wienbeck, M. 1048
- Wienhold, P. 273, 285, 1073

- Wie – Zei**
- Wierzbicki, A.P. 1056
 - Wilcoxon 243, 245f, 513ff, 517ff, 530f, 539, 541f, 824, 884f, 928f, 1023
 - , erweiterter ... Test für Survivalsfunktionen 928
 - , ...-Rangsummentest 513ff, 531, 884f
 - , ...-signed-rank-Test 243ff, 541f
 - , Vorzeichenrangtest von ... 243ff, 541f
 - Wilcoxon-Gehan
 - , ... Test für Survivalsfunktionen 824, 928
 - Wilhelm der Eroberer 12
 - Wilk, M.B. 1070
 - Willemze, F.G. 387, 389, 392, 398, 1069
 - Williams, D.A. 586, 1073
 - Willke, T.A. 340, 1071
 - Wilson, E.B. 955, 960f, 964, 1030, 1073
 - Wilson, G.T. 638, 1050
 - Wilson-Hilferty
 - , Approximation nach ... 955, 1030
 - , inverse ... Transformation 960
 - Winkel, M. 1060
 - Winkler, R.L. 895, 1050
 - Winkler, W. 643, 1073
 - Wirtschaftlichkeit von Straßenverkehrs-signalanlagen 815ff
 - Wirtschaftsprägnose 943ff
 - , ... mit Scoringssystem 943ff
 - Wish, M. 379, 1065
 - Wissen
 - , Kombination von ... 887ff
 - Witt 888
 - Witte, S. 910, 1073
 - Witting, H. 242, 1073
 - Wölbung (s. auch Schiefe) 49, 118
 - Wolf 371
 - Wolfe, D.A. 249, 527, 614, 623, 1063
 - Wolff, J.E. 1072
 - Wolfowitz, J. 406, 1072
 - Woll, A. 965, 967, 1073
 - Wolter, J. 640, 1064
 - Woyzeck 332
 - Woyzeck-Symptom 332
 - Wu, A.S. 746, 763, 1049
 - Wu, S. 964, 1065
 - Würfel 10f, 91, 93ff, 97, 103, 105, 107f
 - Wurzeldiagramm 825, 827, 844ff
 - , aufgehängtes ... 845ff
 - , ...-Residuum 845
 - Wurzeltransformation 349ff

 - X-Test 885

 - Yarnold, J.K. 439, 1073
 - Yates, F. 414, 900, 1073
 - Yates-Cochran
 - , Partial Weighting nch ... 1013
 - Yeo 889, 915f, 918
 - Yntema 572
 - Youden, W.J. 320f, 339f, 731, 1073
 - Youden's extremer Rangsummentest 339ff
 - Young, A.M. 1070
 - Yudkin, P.L. 902, 1074
 - Yule, G.U. 12, 82, 443f, 446, 664, 679, 689
 - Yule
 - , Q-Koeffizient von ... 443ff
 - , Verbundeneitskoeffizient von ... 444ff
 - Yulescher Assoziationskoeffizient 82f, 443ff, 458
 - Yulesche Assoziationsmaße 443ff
 - , Eigenschaften der ... 446
 - , Konfidenzintervalle für die ... 444ff
 - , Schätzer für die ... 444ff
 - Yule-Walker-Gleichung 679
 - , empirische ... 689f
 - Yusuf, S. 918f, 1074
 - Yusuf-Peto (s. a. Peto)
 - , Konfidenzintervall für den kombinierten Odds Ratio von ... 920
 - , Schätzer für den kombinierten Odds Ratio von ... 918ff
 - , logarithmierter ... 919f
 - , Varianz des ... 920
 - , Schätzer für die ... 920

 - Zaun
 - , äußerer ... 835
 - , innerer ... 835
 - Zehender, M. 1061
 - Zeichentest 242f, 540f
 - Zeitbereich 639, 699, 728ff
 - zeitraffender Test 795ff
 - Zeitreihen 57ff, 249, 637ff
 - , Begriff der ... 637ff
 - , ... einfacher Indizes 57ff
 - , Umbasierung einer ... 59f
 - , Verkettung einer ... 60
 - , formale Zerlegung von ... 640f
 - , Komponenten einer ... 640f
 - , saisonbereinigte ... 665f, 725
 - , stationäre ... 962
 - , Stationarität einer ... 744, 962f
 - Dickey-Fuller (Unit Root Test) auf ... 744, 962f
 - , p-Werte des ... 963
 - , ... nach McKinnon 963
 - , Teststatistik für den ... 963
 - , Grenzverteilung der ... 963

Zei – Zuv

- , Zusammenhang zweier ... 639, 727ff
- Zeitreihenanalyse 362, 637ff
- Zeitreihenmodell
 - , additives ... 640f
 - , multiplikatives ... 641
- Zeitreihenpanel 744, 895
- , kreuzkorreliertes ... 744, 895, 960, 962ff
- , Prüfung auf Stationarität eines ... 744, 964
 - , Kombination der Dickey-Fuller p-Werte nach Hartung zur ... 964
 - , ... mit Fisher Probits 964
- , kombinierter Test auf Stationarität eines ... 964f
 - , Interpretation des ... 965
 - , Stationarität eines ... 744, 895, 964
 - , kombinierter Test auf ... 964f
 - , Interpretation des ... 965
 - , Prüfung auf ... 744, 964
- Zeitverschiebung 675
- Zeitverzögerung 744
- zeit-zensierte Lebensdauerprüfung 787
- zensierte Daten 926
 - , erweiterter Wilcoxon Test für ... 928
 - , Log-Rank Test für ... 926ff
 - , Tarone-Ware Test für ... 928
 - , Wilcoxon-Gehan Test für ... 928
- zensierte Lebensdauerprüfung 787
- Zensierung 926
- Zensierungsmechanismus 151
- Zentgraf, R. 1062
- zentraler Grenzwertsatz 122, 138, 899, 943
- zentrales Moment 118
- Zentralwert 32
- Ziegler, S. 1074
- Ziehen mit Zurücklegen 292, 294
- Ziehen ohne Zurücklegen 292, 294
- Ziehms, H. 770, 1053
- Zielvariable 365
- Zimmermann, H.J. 1055
- zufälliger Effekt 609, 629ff
- zufälliger Fehler 321ff, 337, 342, 574
- zufälliges Ereignis 91
- zufällig-periodische Strategie 824
- Zufallsauswahl (s. auch Stichprobe, Stichprobenverfahren) 141, 274ff
 - , einfache ... 274ff
 - , geschichtete ... 278ff
- Zufallsexperiment 4, 91ff, 103ff, 124, 129
- Zufallsfehler
 - , relativer ... bei Stichprobenerhebungen 269f
- Zufallsstichprobe (s. auch Zufallsauswahl, Stichprobe, Stichprobenverfahren) 124, 274ff, 307, 315
 - , einfache ... (s. auch einfache Zufallsauswahl) 124, 274ff, 287, 289, 307
 - Zufallsvariable 103ff, 112ff
 - , komplexe ... 700
 - , lineare Transformation einer ... 109
 - , Maximum von ... 108, 111
 - , Minimum von ... 108, 111
 - , Momente einer ... 118f
 - , standardisierte ... 117, 130
 - , Summe von ... 108ff
 - Zufallsvektor 104, 107f, 112
 - , diskret-verteilter ... 107f
 - , stetig-verteilter ... 108, 112
 - Zufallszahlen 141, 193
 - , gleichverteilte ... 193, 749f
 - , Pseudo ... 193
 - Zufallszahlengenerator 193
 - Zuordnung
 - , ... von Versuchspersonen bei klinischen Versuchen 367ff
 - , aleatorische ... 370
 - , alternierende ... 370
 - , ... auf freiwilliger Basis 369
 - , ... durch historische Kontrolle 368
 - , pseudaleatorische ... 370
 - , retrospektive ... 368f
 - zusammengesetzte Indexzahl 57, 62
 - zusammengesetzte Nullhypothese 139
 - Zusammenhang zweier Zeitreihen 639, 727ff
 - , ... im Frequenzbereich 639, 728ff
 - , ... im Zeitbereich 639, 728
 - Zustandsraummodell 640
 - Zuverlässigkeit 745ff
 - , ... eines Parallelsystems 747
 - , ... eines Seriensystems 747
 - , ... elementarer Systeme 746f
 - , Erhöhung der ... durch Redundanz 760ff
 - , Komponenten... 747, 756, 809
 - , ... mehrphasiger Missionen 770ff
 - , Missions... 770 ff
 - , Phasen... 770ff
 - , System... 746ff, 756ff, 772ff, 809ff.
 - , ... von Straßenverkehrssignalanlagen 808ff
 - Zuverlässigkeitsanalyse 745ff
 - Zuverlässigkeitsfunktion eines monotonen Systems 756ff
 - , disjunktive Normalform für die ... 756ff
 - , Klassifikation der ... 759f
 - Zuverlässigkeitsschaltbild 748ff, 755, 772f
 - , Gewinnung des ... zu einer Strukturfunktion 755

- 20-80-Regel 273, 278
 zweiarmige Studie 902f, 917, 974ff, 1013
 -, Test auf Homogenität für ... 1013
 zweidimensionale Gleichverteilung 194f
 -, Dichte der ... 194
 zweidimensionale Kontingenztafel (s. auch
 2 x 2-Kontingenztafel) 425ff, 492ff
 -, Adjustieren von ... 501ff
 -, Assoziationsmaße für die ... 450ff
 -, Logit-Modell bei ... 498ff
 -, loglineares Modell für die ... 425ff
 --, Aufstellung des ... 429
 --, bedingte Gleichverteilungshypothese im
 ... 434
 --, χ^2 -Test im ... 435ff
 --, Homogenitätshypothese im ... 433f
 --, Likelihood-Quotiententest im ... 435ff
 --, Schätzer für die Parameter des ... 430
 ---, approximative Varianz der ... 430
 --, Symmetrietest nach Bowker im ... 440f
 --, Test mittels der Statistiken τ_A und τ_B im
 ... 439f
 --, totale Gleichverteilungshypothese im
 ... 434
 --, Unabhängigkeitshypothese im ... 433f
 -, Verteilungsannahmen bei ... 492ff
 zweidimensionales Merkmal 72ff, 79ff
 zweifache Kreuzklassifikation (s. auch
 zweifache Varianzanalyse) 609, 624
 zweifacher Stichprobenplan (s. auch dop-
 pelter Stichprobenplan) 389ff
 zweifache Varianzanalyse 358, 609, 624ff
 -, ... mit Wechselwirkungen zwischen den
 Faktoren 625ff
 --, F-Tests auf signifikante Lokationsun-
 terschiede bei der ... 625f
 --, multiple Vergleiche bei der ... 625
 --, Schätzer für die Parameter der ... 625
 --, Varianzanalysetafel der ... 626
 --, ... ohne Wechselwirkungen zwischen
 den Faktoren 627f

Zwa – Zyk

- , Friedman-Test auf signifikante Lokati-
 onsunterschiede bei der ... 628
 --, F-Tests auf signifikante Lokationsun-
 terschiede bei der ... 627
 --, Schätzer für die Parameter der ... 627
 --, Varianzanalysetafel der ... 628
 2 x 2-Kontingenztafel 12, 82f, 411ff, 442ff
 --, Assoziationsmaße für die ... 442f
 --, loglineares Modell für die ... 425ff
 --, Aufstellung des ... 425f
 --, Entwicklung des ... 425
 --, Schätzer für die Parameter des ... 427f
 --, approximative Varianz der ... 428f
 -, Test auf Homogenität in der ... 418ff
 --, erforderlicher Stichprobenumfang beim
 ... 419ff
 --, exakter ... 421f
 --, ... gemäß der Approximation der
 Gütfunktion des χ^2 -Tests 419f
 ---, ... nach Casagrande/Pike/Smith 420f
 ---, ... nach der Arcus-Sinus-Formel 420
 -, Test auf Symmetrie in der ... 422ff
 --, ... mittels Cochran's Q 423f
 --, ... von McNemar 423
 -, Test auf Unabhängigkeit in der ... 413ff
 --, χ^2 -... 413f
 --, exakter ... 414ff
 --, ... von Fisher 416ff
 2 x 2-Tafel (s. 2 x 2-Kontingenztafel)
 Zweisatzgitter 361
 zweiseitige Hypothese 139
 zweiseitiges Testproblem 139
 zweiseitig gestützte Normalverteilung 149f
 -, Dichte der ... 150
 -, Erwartungswert der ... 149
 -, Varianz der ... 149
 -, Verteilungsfunktion der ... 149
 Zwilling
 -, statistischer ... 311
 zyklische Komponente 641

8. Weiterführende und ergänzende Hinweise

Am Schluss dieses Buches möchten wir noch auf einige weiterführende bzw. ergänzende Lehrbücher hinweisen, die ebenfalls im Oldenbourg-Verlag erschienen sind.

Zunächst wäre hier das Buch

Hartung/Elpelt: Multivariate Statistik
Lehr- und Handbuch der angewandten Statistik

zu nennen, in dem die hier behandelten Verfahren auf den multivariaten Fall erweitert werden und in dem speziell auf diesen Fall zugeschnittene Methoden dargestellt werden. Unter Methoden der multivariaten Statistik versteht man ganz allgemein Verfahren, die es erlauben, nicht isoliert Aussagen über nur ein Merkmal zu machen, sondern gleichzeitig mehrere, an Objekten aus einer interessierenden Grundgesamtheit beobachtete Merkmale zu berücksichtigen. Zu erwähnen sind hier natürlich zunächst die multivariaten Analoga der im vorliegenden Buch ausführlich behandelten Ein-, Zwei- und Mehrstichprobenprobleme, wobei insbesondere noch die Diskriminanzanalyse hervorgehoben werden soll, mit deren Hilfe ein Objekt aufgrund der an ihm beobachteten Merkmalsausprägungen verschiedener Merkmale einer der in Frage stehenden Grundgesamtheiten von Objekten zugeordnet werden kann; die Wichtigkeit der einzelnen Merkmale für die Güte einer solchen Zuordnung wird hierbei anhand so genannter Trennmaße gemessen. Weiterhin werden die in diesem Buch nur relativ knapp dargestellten Methoden der Korrelations-, Regressions- und Varianzanalyse ausführlich auch für den multivariaten Fall behandelt. Ein ausführliches Kapitel des Buches ist speziellen graphischen Verfahren gewidmet. Sowohl Methoden zur Darstellung von Objekten und/oder Merkmalen in einem gemeinsamen Koordinatensystem derart, dass die z. B. euklidischen Abstände in diesem System die Ähnlichkeit von Objekten bzw. Merkmalen widerspiegeln, als auch spezielle Methoden zur Darstellung von Objekten (oder Merkmalen) in Einzelgraphiken werden behandelt. Objekte werden beispielsweise anhand der an ihnen beobachteten Merkmalsausprägungen derart dargestellt, dass beim Vergleich der Einzelgraphiken miteinander direkt deutlich wird, worin sich die zugrunde liegenden Objekte unterscheiden. Intensiv werden selbstverständlich auch diejenigen multivariaten statistischen Methoden abgehandelt, die klassischerweise der psychologischen oder sozialwissenschaftlichen Statistik zugeordnet werden, in zunehmendem Maße aber auch in anderen Bereichen eingesetzt werden. Als erstes ist hier die multidimensionale Skalierung (MDS) zu nennen, bei der basierend auf Informationen über die Ähnlichkeit respektive Verschiedenheit von Objekten eine Skala derart bestimmt wird, dass die Distanzen zwischen den Objektvektoren bzgl. dieser Skala die Verschiedenheiten möglichst gut widerspiegeln. Bei der Clusteralanalyse werden die interessierenden Objekte so in Klassen eingeteilt, dass die Objekte in einer Klasse sich möglichst ähnlich sind und die Objekte aus verschiedenen Klassen sich möglichst stark unterscheiden; verschiedene mögliche Klassifikationstypen, von denen es im wesentlichen vier Stück gibt, differieren in der Struktur ihrer Klassen: Die Klassen einer Partition sind disjunkt, d. h. sie überschneiden sich nicht, wohingegen sich die Klassen einer Überdeckung überlappen können, keine Klasse aber vollständig in einer anderen enthalten ist. Hierarchien und Quasihierarchien entstehen als Folgen von Partitionen bzw. Überdeckungen, indem z. B. ausgehend von einer feinen Partition in der nächsten Stufe Klassen der feinen Partition vereinigt werden usw., wobei darauf zu achten ist, dass jede Stufe selbst wieder eine Partition

darstellt. Die Faktorenanalyse schließlich basiert auf der Vorstellung, dass gewissermaßen hinter den an Objekten beobachtbaren Merkmalen wenige latente, nicht beobachtbare Faktoren stehen, mit deren Hilfe die Objekte beschrieben werden können. Diese Faktoren werden bei der Faktorenanalyse zunächst extrahiert, den Objekten werden Werte bzgl. der wenigen, wesentlichen Faktoren zugeordnet und mit diesen Faktorenwerten kann dann weitergearbeitet werden, ohne viel Informationsverlust hinnehmen zu müssen. Als letztes, aber wesentliches inhaltliches Kapitel des Buches über multivariate Statistik sei die Skalierung von Merkmalsausprägungen, die im Kapitel V dieses Buches bereits kurz angerissen wurde, erwähnt. Ausgehend von Merkmalen mit kategorialen bzw. qualitativen Ausprägungen werden hierbei diese Ausprägungen quantifiziert, so dass alle statistischen Verfahren für quantitative, stetige Merkmale auf ein derartiges Datenmaterial angewandt werden können. Die konkrete Form einer solchen Skalierung hängt von der statistischen Methode ab, die später auf die Daten angewandt werden soll.

Insbesondere an Studenten und Praktiker aller Fachrichtungen, die eine statistische Grundausbildung nachweisen bzw. absolvieren müssen, richtet sich das Buch

**Elpelt/Hartung: Grundkurs Statistik
Lehr- und Übungsbuch der angewandten Statistik.**

Die Grundlagen der Wahrscheinlichkeitsrechnung und Statistik werden dort ausführlicher als hier in den Kapiteln II und III dargelegt und erläutert. Hinzu kommt eine explizite Darstellung der wichtigsten speziellen statistischen Verfahren sowie ein Überblick über weitere relevante Gebiete der Statistik. Für den Lernenden ist das Buch auch dadurch überaus geeignet, dass jedes Kapitel anhand von Übungsaufgaben mit ausführlichen Lösungen selbstständig aufgearbeitet werden kann. Schließlich enthält das Buch noch einen Klausurteil mit ‚echten‘ Klausuren (inklusive Punktierung und Benotung) und zugehörigen detaillierten Lösungen zu den Klausuraufgaben, die mit Bewertungshinweisen ausgestattet sind.

Sowohl eigenständig als auch begleitend und vertiefend zum vorliegenden Buch können die Bücher

**Hartung/Heine: Statistik – Übungen
Deskriptive Statistik**

und

**Hartung/Heine: Statistik – Übungen
Induktive Statistik**

verwandt werden. Zu quasi allen Kapiteln und Methoden des hier vorliegenden Buches findet man in den Statistik-Übungen Aufgaben mit ausführlichen Musterlösungen. Dadurch, dass in den Lösungen der benötigte Stoff zunächst noch einmal in knapper Form allgemein dargestellt wird, sind die Bücher auch eigenständig nutzbar und eignen sich als vorlesungsbegleitendes Material sowie als Klausurvorbereitung. Letzteres wird noch durch die Klausurteile beider Statistik-Übungen unterstrichen, in denen typische Klausuraufgaben nebst detaillierten Lösungen zusammengestellt sind.



Die einzige deutschsprachige Einführung



Christian FG Schendera

Regressionsanalyse mit SPSS

2008 | 466 S. | Broschur | € 34,80

ISBN 978-3-486-58692-3

Die Regressionsanalyse gehört zu den am häufigsten eingesetzten statistischen Verfahren. Dieses Buch führt ein in die Ansätze: Korrelation, Regression (linear, multipel, nichtlinear), logistische (binär, multinomial) und ordinale Regression, sowie die Überlebenszeitanalyse (Sterbetafel-Methode, Kaplan-Meier-Ansatz, sowie Regressionen nach Cox).

Weitere Abschnitte behandeln zusätzliche regressionsanalytische Ansätze und Modelle (z.B. Partial Least Squares-Regression, Ridge-Regression, Modellierung individueller Wachstumskurven). Zahlreiche Rechenbeispiele werden von der Fragestellung, der Anforderung der einzelnen Statistiken (per Maus, per Syntax) bis hin zur Interpretation des SPSS Ausgaben systematisch durchgespielt. Auch auf mögliche Fallstricke und häufig begangene Fehler wird eingegangen. Separate Abschnitte stellen die diversen Voraussetzungen für die Durchführung der jeweiligen Analyse, sowie Ansätze zu ihrer Überprüfung zusammen.

Dieses Buch ist angenehm verständlich und anwendungsorientiert geschrieben, ohne jedoch die Komplexität und damit erforderliche Tiefe bei der Vorstellung der Verfahren zu vernachlässigen.

Dieses Buch ist für Einsteiger in die Regressionsanalyse, Studierende sowie fortgeschrittene Wissenschaftler in den Wirtschafts-, Bio- und Sozialwissenschaften gleichermaßen geeignet.

Oldenbourg



150 Jahre
Wissen für die Zukunft
Oldenbourg Verlag

Bestellen Sie in Ihrer Fachbuchhandlung oder
direkt bei uns: Tel: 089/45051-248, Fax: 089/45051-333
verkauf@oldenbourg.de



Die einzige deutschsprachige Einführung



Ulrich Kohler, Frauke Kreuter

Datenanalyse mit Stata

Allgemeine Konzepte der Datenanalyse und
ihre praktische Anwendung

3., aktualisierte und überarbeitete Auflage 2008

398 S. | Broschur

€ 34,80 | ISBN 978-3-486-58456-1

Dieses Buch ist die bislang einzige deutschsprachige Einführung in das Statistikprogramm Stata und zugleich das einzige Buch über Stata, das auch Anfängern eine ausreichende Erklärung der Datenanalysetechnik liefert.

Die Neuauflage enthält Anpassungen an die neueste Stata Programmversion (Version 10) und beschreibt den neu entwickelten »Stata Graph Editor«.

Das Buch richtet sich an Studierende der Sozialwissenschaften.

Über die Autoren:

Dr. phil. Ulrich Kohler ist am Wissenschaftszentrum Berlin tätig.

Asst. Prof. Dr. Frauke Kreuter ist an der University of Maryland tätig.

Oldenbourg



150 Jahre
Wissen für die Zukunft
Oldenbourg Verlag

Bestellen Sie in Ihrer Fachbuchhandlung oder
direkt bei uns: Tel: 089/45051-248, Fax: 089/45051-333
verkauf@oldenbourg.de



Das Original: Wirtschaftswissen komplett



Artur Woll

Wirtschaftslexikon

10., vollständig neubearbeitete Auflage 2008

863 S. | gebunden

€ 29,80 | ISBN 978-3-486-25492-1

Der Name »Woll« sagt bereits alles über dieses Lexikon. Das Wollsche Wirtschaftslexikon erfüllt das verbreitete Bedürfnis nach zuverlässiger Wirtschaftsinformation in vorbildlicher Weise. Längst ist der »Woll« das Standardlexikon im Ausbildungsbereich. Es umfasst die Kernbereiche Betriebswirtschaftslehre, Volkswirtschaftslehre und die Grundlagen der Statistik, aber auch die wirtschaftlich bedeutsamen Teile der Rechtswissenschaft. Besonderer Wert wurde auf eine möglichst knappe, jedoch zuverlässige Stichwortabhandlung gelegt.

Das Wirtschaftslexikon eignet sich nicht nur für den akademischen Gebrauch, sondern richtet sich auch an Praktiker in Wirtschaft und Verwaltung.



Prof. Dr. Dr. h. c. mult. Artur Woll
lehrte Volkswirtschaftslehre an der
Universität Siegen.

Oldenbourg

150 Jahre
Wissen für die Zukunft
Oldenbourg Verlag

Bestellen Sie in Ihrer Fachbuchhandlung oder
direkt bei uns: Tel: 089/45051-248, Fax: 089/45051-333
verkauf@oldenbourg.de



Umfassend. Aktuell. Fundiert.



Axel Noack
Business Essentials:
Fachwörterbuch Deutsch-Englisch Englisch-Deutsch
2007, VII, 811 Seiten, gebunden
€ 59,80
ISBN 978-3-486-58261-1

Das Wörterbuch gibt dem Nutzer das Fachvokabular des modernen, internationalen Geschäftslebens in einer besonders anwenderfreundlichen Weise an die Hand.

Der englisch-deutsche Teil umfasst die 11.000 wichtigsten Wörter und Begriffe des angloamerikanischen Sprachgebrauchs.

Der deutsch-englische Teil enthält entsprechend 14.000 aktuelle Fachbegriffe mit ihren Übersetzungen.

Im dritten Teil werden 3.000 Abkürzungen aus dem internationalen Wirtschaftsgeschehen mit ihren verschiedenen Bedeutungen aufgeführt.

Das Lexikon richtet sich an Studierende der Wirtschaftswissenschaften sowie alle Fach- und Führungskräfte, die Wirtschaftsenglisch für Ihren Beruf benötigen. Für ausländische Studenten bietet es einen Einstieg in das hiesige Wirtschaftsleben.



Prof. Dr. Axel Noack lehrt an der Fachhochschule Stralsund BWL, insbes. International Marketing.



Impulsgeber für die Wirtschaft



Bernd O. Weitz
Bedeutende Ökonomen
2008. VIII, 205 S., geb.
€ 19,80
ISBN 978-3-486-58222-2

Das Werk porträtiert herausragende Ökonomen vom 17. Jahrhundert bis heute. Die Autoren wollen neben dem wissenschaftlichen Vermächtnis der ausgewählten Wirtschaftswissenschaftler Eindrücke von deren historisch-sozialem Umfeld vermitteln, Querverbindungen zu anderen Ökonomen aufzeigen und verdeutlichen, welche Impulse für die weitere wirtschaftswissenschaftliche und gesellschaftliche Entwicklung erfolgten. Der Leser wird auf eine ökonomiehistorische Entdeckungsreise geschickt. In diesem Buch werden auch Werkauszüge, weitergehende Literaturanregungen sowie Hinweise auf vertiefende Quellen im Internet gegeben.

Behandelte Ökonomen: Adam Smith, Francois Quesnay, Johann Peter Becher, Jean-Baptiste Say, Johann Heinrich von Thünen, Thomas Robert Malthus, David Ricardo, Karl Marx, Leon Walras, Vilfredo Pareto, Max Weber, Joseph Alois Schumpeter, Walter Eucken, John Maynard Keynes, Friedrich von Hayek, Wassily Leontief, John Kenneth Galbraith, Ronald H. Coase, Milton Friedman, Ludwig Erhard, Alfred Müller-Armack.



Prof. Dr. Bernd O. Weitz lehrt an der Universität zu Köln Wirtschaftswissenschaft und ihre Didaktik.