



Benjamin Auer / Horst Rottmann

Statistik und Ökonometrie für Wirtschaftswissenschaftler

Eine anwendungsorientierte
Einführung



LEHRBUCH

Benjamin Auer / Horst Rottmann

Statistik und Ökonometrie für Wirtschaftswissenschaftler

Benjamin Auer / Horst Rottmann

Statistik und Ökonometrie

für Wirtschaftswissenschaftler

Eine anwendungsorientierte
Einführung



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.

Prof. Dr. Horst Rottmann ist Professor für Statistik und Volkswirtschaft an der Hochschule Amberg-Weiden. Des Weiteren ist er Forschungsprofessor am ifo Institut für Wirtschaftsforschung an der Universität München.

Dipl.-Betriebsw. (FH) Benjamin Auer ist Doktorand an der wirtschaftswissenschaftlichen Fakultät der Universität Leipzig sowie Buchautor in den Bereichen Mathematik, Statistik und Buchführung.

1. Auflage 2010

Alle Rechte vorbehalten

© Gabler Verlag | Springer Fachmedien Wiesbaden GmbH 2010

Lektorat: Jutta Hauser-Fahr | Renate Schilling

Gabler Verlag ist eine Marke von Springer Fachmedien.

Springer Fachmedien ist Teil der Fachverlagsgruppe Springer Science+Business Media.

www.gabler.de



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

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürfen.

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg

Druck und buchbinderische Verarbeitung: Ten Brink, Meppel

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Printed in the Netherlands

ISBN 978-3-8349-0323-5

Vorwort

Liebe Leserinnen und Leser,

mit diesem Werk liegt Ihnen das erste deutsche Lehrbuch vor, das die beiden für empirisch arbeitende Wirtschaftswissenschaftler relevanten Gebiete - die Statistik und die Ökonometrie - kombiniert und umfassend darstellt. Bei der Aufbereitung der Themenbereiche achten wir besonders auf Klarheit und die Verwendung einer Vielzahl theoretischer und empirischer Beispiele. Auf komplizierte Herleitungen und Beweise verzichten wir bewusst, wenn sie unseres Erachtens nur das Verständnis erschweren.

Inhalte und Beispiele mit Praxisbezug

Das vorliegende Lehrbuch behandelt die traditionellen Bereiche der deskriptiven und der induktiven Statistik sowie der Wahrscheinlichkeitsrechnung. Es unterscheidet sich allerdings von anderen einschlägigen Lehrbüchern in der speziellen Darstellungsweise. Jedes Kapitel enthält zahlreiche Beispiele und Übungen, die ausführlich erläutert werden und somit die abstrakten Zusammenhänge anschaulich vertiefen. Sie behandeln auch die Anwendung der statistischen Methoden für typische betriebswirtschaftliche und sozialwissenschaftliche Problemstellungen.

Umfangreiche Darstellung der Regressionsanalyse

Darüber hinaus stellt die Schwerpunktbildung dieses Buches ein entscheidendes Alleinstellungsmerkmal dar. So bietet es beispielsweise ein Kapitel zur angewandten Ökonometrie (veranschaulicht unter Verwendung der Software EViews 6.0), das weit über die in der Masse der Lehrbücher enthaltene deskriptive Behandlung der linearen Regression hinausgeht. Dies ist besonders wertvoll für all jene, die die Regressionsanalyse für empirische Untersuchungen, etwa für eine praxisorientierte Bachelor- oder Masterarbeit, verwenden wollen.

Konkrete Anwendungen der Methoden in der Praxis

Des Weiteren finden sich eine Vielzahl sehr konkreter wirtschafts- und insbesondere finanzwirtschaftlicher Anwendungsfälle (z.B. die Schätzung von Aktienvolatilitäten, die Analyse von Wettbewerbsprozessen und Einkommensverteilungen, Kaufkraftvergleiche, Aktienindizes, subjektive Wahrscheinlichkeiten und Wettquotienten, Portfoliorisiko und Diversifikation, Qualitätskontrolle, die Schätzung von Nachfrage-, Produktions- und Konsumfunktionen, die Untersuchung von Trends in der Goldpreisentwicklung, die Beurteilung des langfristigen Einflusses von Zentralbanken auf das Zinsniveau und der langfristigen Beziehung zwischen Geldmarktzinssätzen verschiedener Fristigkeiten, die Analyse der Performance von Investmentfonds und Zinsprognosen). Wir haben Kapitel, die derartige Anwendungen enthalten, im nachfolgenden Inhaltsverzeichnis mit dem Zeichen **A** kenntlich gemacht. Darüber hinaus wurde ein separates Anwendungsverzeichnis angelegt.

Dozenten- und Studentennutzen

Dieses Buch soll Ihnen als Lehrenden helfen, die heiklen Themen Statistik und Ökonometrie verständlich, studenten- und vor allem praxisgerecht für Ihren Unterricht aufzubereiten. Den Studierenden soll es dazu dienen, alle relevanten Themen für ihre Klausur noch einmal zu wiederholen und mit Hilfe zahlreicher Übungsaufgaben zu festigen. Bei der Lösung dieser Aufgaben lassen wir Sie nicht alleine. Anders als in einschlägigen Lehrbüchern wird Ihnen nicht einfach ein Ergebnis ohne Rechenweg und Erläuterungen präsentiert. Stattdessen erklären wir nachvollziehbar und anschaulich jeden einzelnen Rechenschritt und die allgemeine Vorgehensweise.

Internetseite mit umfassendem Material

Zur Unterstützung von Dozenten und Studierenden wurde eigens die Internetseite www.statistik-auer-rottmann.de eingerichtet. Hier finden Sie beispielsweise diverse **Excel-Tools**, mit denen Rechenbeispiele und Verfahren einfach nachvollzogen und praktisch umgesetzt werden können, sowie die **Datentabellen** ökonometrischer Schätzungen im Excel- und EViews-Format. Auf die Verfügbarkeit derartigen Materials wird im Text durch das Symbol  am Seitenrand hingewiesen. Weiterhin bieten wir Ihnen online zusätzliche **Übungsaufgaben** inklusive der ausführlichen Lösungen und eine auf das Buch abgestimmte **Formelsammlung**. Speziell für Dozenten stellen wir die Lösungen der Übungsaufgaben mit geraden Aufgabennummern, die im Buch nicht veröffentlicht sind, sowie alle Grafiken des Buches in bearbeitbarer Form zur Verfügung. *Benutzernamen und Passwort* für den Zugang zu diesem Onlineangebot können Sie über das Kontaktmenü auf der o.g. Seite anfordern.

Danksagung

Bei der Konzeption und Gestaltung des Buches haben uns viele Personen unterstützt. Wir danken Prof. Dr. Christopher Dietmaier, Prof. Dr. Gebhard Flaig, Prof. Dr. Norbert Janz, Prof. Dr. Thomas Jost, Prof. Dr. Erich Ruppert, Prof. Dr. Franz Seitz und Dr. Timo Wollmershäuser für wertvolle Hinweise und Verbesserungsvorschläge. Ihnen möchten wir für ihre hilfreichen kritischen Auseinandersetzungen mit unseren Ideen ganz besonders danken. Für das Gelingen dieser Arbeit war zudem der wissenschaftliche Freiraum hilfreich, den Herr Benjamin Auer am Lehrstuhl von Prof. Dr. Frank Schuhmacher erhalten hat. Besonderer Dank gilt außerdem Herrn Steffen Burkhardt für eine kritische Durchsicht des Manuskripts, Korrekturen und wertvolle formale Hinweise, sowie Herrn Mark Kirstein für ein abschließendes Korrekturlesen.

Allen Lesern wünschen wir auf diesem Weg, dass sie durch dieses Buch einen Zugang zu den interessanten Fragestellungen finden, die sich mit Hilfe der Statistik und Ökonometrie beantworten lassen. Bei Fragen, Anregungen und Kritik würden wir uns über eine E-Mail via www.statistik-auer-rottmann.de sehr freuen.

Weiden i. d. OPf. und Leipzig,
Juni 2010

Prof. Dr. Horst Rottmann
Dipl.-Bw. (FH) Benjamin R. Auer

Inhaltsverzeichnis

Vorwort	V
Abbildungsverzeichnis	XIX
Anwendungsverzeichnis	XXIII
I Deskriptive Statistik	1
1. Grundbegriffe.....	3
1.1 Der Statistikbegriff.....	3
1.2 Merkmalsträger, Grundgesamtheiten und Stichproben	4
1.3 Klassifikation von Merkmalen.....	6
1.3.1 Klassifikation nach dem Skalenniveau	6
1.3.2 Klassifikation in diskrete und stetige Merkmale.....	10
1.3.3 Klassifikation in qualitative und quantitative Merkmale	11
2. Eindimensionale Häufigkeitsverteilungen.....	13
2.1 Häufigkeitsverteilung	13
2.1.1 Häufigkeitsverteilung bei diskreten Merkmalen.....	13
2.1.2 Empirische Verteilungsfunktion bei diskreten Merkmalen	18
2.1.3 Klassierte Häufigkeitsverteilung bei stetigen Merkmalen	21
2.1.4 Typische Häufigkeitsverteilungen.....	26
2.1.5 Quantile	28
2.2 Maßzahlen	31
2.2.1 Lageparameter	31
2.2.1.1 Modus	32
2.2.1.2 Median	34
2.2.1.3 Arithmetisches Mittel	35
2.2.1.4 Geometrisches Mittel.....	38
2.2.1.5 Exkurs: Renditen und Renditedurchschnitte	40
2.2.1.6 Lageregeln	44
2.2.2 Streuungsparameter.....	45
2.2.2.1 Spannweite und Quartilsabstand	45
2.2.2.2 Mittlere absolute Abweichung	47
2.2.2.3 Varianz und Standardabweichung	49

2.2.2.4	Exkurs: Volatilität	56	A
2.2.2.5	Variationskoeffizient	59	
2.2.2.6	Box-Whisker-Plot.....	61	
2.2.3	Momente und Schiefemaße	62	
2.2.3.1	Empirische Momente	63	
2.2.3.2	Schiefemaße	63	
2.2.4	Konzentrationsmessung.....	65	
2.2.4.1	Maßzahlen der absoluten Konzentration	66	
2.2.4.2	Maßzahlen der relativen Konzentration	70	A
3.	Zweidimensionale Häufigkeitsverteilungen.....	81	
3.1	Grundlagen	81	
3.1.1	Kontingenztabelle	81	
3.1.2	Randhäufigkeiten und -verteilungen.....	85	
3.1.3	Bedingte Häufigkeiten und Verteilungen	86	
3.1.4	Statistische Unabhängigkeit	89	
3.2	Korrelationsanalyse.....	92	
3.2.1	Kovarianz und Bravais-Pearson-Korrelationskoeffizient.....	92	
3.2.2	Spearman-Rangkorrelationskoeffizient.....	98	
3.2.3	Kontingenzkoeffizient.....	102	
3.2.4	Lineartransformationen und Linearkombinationen.....	104	
3.2.5	Kritische Anmerkungen zur Korrelationsanalyse.....	106	
4.	Messzahlen und Indizes	109	
4.1	Messzahlen	109	
4.2	Indexzahlen	112	
4.2.1	Preisindizes	112	
4.2.1.1	Grundlegendes	112	
4.2.1.2	Preisindex nach Laspeyres.....	114	
4.2.1.3	Preisindex nach Paasche	116	
4.2.1.4	Weitere Preisindizes.....	116	
4.2.1.5	Preisindexreihen und Inflationsmessung	118	A
4.2.1.6	Preisbereinigung und reale Größen	119	A
4.2.1.7	Interregionale Kaufkraftvergleiche	121	A
4.2.1.8	Umbasierung und Verknüpfung.....	123	
4.2.2	Mengenindizes	125	
4.2.3	Wertindex.....	127	
4.2.4	Wichtige Indizes aus der Wirtschaftspraxis.....	128	

4.2.4.1	Verbraucherpreisindex (VPI)	128	A
4.2.4.2	Harmonisierter Verbraucherpreisindex (HVPI)	131	A
4.2.4.3	Deutscher Aktienindex (DAX)	132	A
5.	Aufgaben	135	
II Wahrscheinlichkeitsrechnung	145	
1.	Grundlagen der Wahrscheinlichkeitstheorie	147	
1.1	Grundbegriffe	147	
1.2	Ereignisse und ihre Darstellung.....	149	
1.3	Wahrscheinlichkeitsregeln und -definitionen.....	155	
1.3.1	Axiome der Wahrscheinlichkeitsrechnung.....	156	
1.3.2	Klassische Wahrscheinlichkeitsdefinition	159	
1.3.3	Statistische Wahrscheinlichkeitsdefinition	162	
1.3.4	Subjektive Wahrscheinlichkeitsdefinition	163	A
1.4	Zufallsauswahl und Kombinatorik.....	166	
1.4.1	Zufallsauswahl und Urnenmodell.....	167	
1.4.2	Kombinatorik.....	167	
1.4.2.1	n-Fakultät und Binomialkoeffizient	167	
1.4.2.2	Prinzipien der Kombinatorik.....	169	
1.4.2.3	Zusammenfassung und Vergleich.....	174	
1.5	Bedingte Wahrscheinlichkeiten	176	
1.5.1	Definition und Interpretation.....	176	
1.5.2	Multiplikationssatz.....	177	
1.5.3	Unabhängigkeit von Ereignissen	180	
1.5.4	Satz der totalen Wahrscheinlichkeit.....	183	
1.5.5	Formel von Bayes	185	A
2.	Zufallsvariablen	191	
2.1	Begriff der Zufallsvariable.....	191	
2.2	Diskrete Zufallsvariablen	194	
2.2.1	Wahrscheinlichkeitsfunktion	194	
2.2.2	Verteilungsfunktion	196	
2.2.3	Zusammenfassende Gegenüberstellung	198	
2.3	Stetige Zufallsvariablen	200	
2.3.1	Verteilungsfunktion	200	
2.3.2	Dichtefunktion.....	201	
2.3.3	Zusammenfassende Gegenüberstellung	204	

2.4 Kennzahlen von Wahrscheinlichkeitsverteilungen	205
2.4.1 Erwartungswert	205
2.4.1.1 Definition	205
2.4.1.2 Eigenschaften	207
2.4.2 Varianz und Standardabweichung.....	211
2.4.2.1 Definition	211
2.4.2.2 Eigenschaften	212
2.4.2.3 Standardisierung von Zufallsvariablen	214
2.4.3 Höhere Momente	216
2.4.4 Quantile.....	217
2.5 Ungleichung von Tschebyscheff	219
2.6 Anwendungsbeispiele.....	221
2.6.1 Renditen als Zufallsvariablen.....	221 A
2.6.2 Zufallsvariablen beim Roulette	222 A
2.7 Mehrdimensionale Zufallsvariablen.....	225
2.7.1 Begriff.....	225
2.7.2 Wahrscheinlichkeits- und Verteilungsfunktion	226
2.7.2.1 Gemeinsame Wahrscheinlichkeitsfunktion	226
2.7.2.2 Gemeinsame Verteilungsfunktion	228
2.7.2.3 Randverteilungen	228
2.7.2.4 Bedingte Verteilungen	229
2.7.3 Stochastische Unabhängigkeit	231
2.7.4 Kennzahlen zweidimensionaler Wahrscheinlichkeitsverteilungen....	232
2.7.4.1 Erwartungswert und Varianz	232
2.7.4.2 Kovarianz und Korrelationskoeffizient.....	234
2.7.5 Linearkombinationen von Zufallsvariablen.....	237
2.7.6 Formelzusammenstellung für stetige Zufallsvariablen.....	239
2.7.7 Anwendungsbeispiel: Portfoliotheorie	240 A
3. Theoretische Verteilungen	245
3.1 Diskrete Verteilungen	245
3.1.1 Binomialverteilung.....	245
3.1.1.1 Wahrscheinlichkeits- und Verteilungsfunktion	245
3.1.1.2 Eigenschaften	249
3.1.1.3 Praxisanwendung: Operationscharakteristiken.....	250 A
3.1.2 Hypergeometrische Verteilung	252
3.1.2.1 Wahrscheinlichkeits- und Verteilungsfunktion	252

3.1.2.2	Eigenschaften	255
3.1.2.3	Approximation durch die Binomialverteilung.....	256
3.1.3	Poissonverteilung	257
3.1.3.1	Wahrscheinlichkeits- und Verteilungsfunktion	257
3.1.3.2	Eigenschaften	259
3.1.3.3	Approximation	259 A
3.2	Stetige Verteilungen	261
3.2.1	Gleichverteilung	261
3.2.1.1	Dichte- und Verteilungsfunktion	261
3.2.1.2	Diskretes Gegenstück.....	262
3.2.2	Exponentialverteilung	264
3.2.2.1	Dichte- und Verteilungsfunktion	264
3.2.2.2	Diskretes Gegenstück.....	266
3.2.3	Normalverteilung.....	268
3.2.3.1	Dichte- und Verteilungsfunktion	268
3.2.3.2	Standardnormalverteilung	271
3.2.3.3	Reproduktionseigenschaft.....	275
3.2.4	Logarithmische Normalverteilung.....	276
3.3	Test-Verteilungen	278
3.3.1	Chi-Quadrat-Verteilung	278
3.3.2	t-Verteilung.....	280
3.3.3	F-Verteilung	281
3.4	Bedeutung der Normalverteilung	283
3.4.1	Zentraler Grenzwertsatz	283
3.4.2	Approximation diskreter Verteilungen	285
3.4.2.1	Binomialverteilung	285
3.4.2.2	Hypergeometrische Verteilung	286
3.4.2.3	Poissonverteilung	287
3.4.2.4	Überblick zur Approximation eindimensionaler Verteilungen....	289
3.4.2.5	Empirische Verteilungen	290
4.	Aufgaben	293
III	Induktive Statistik.....	307
1.	Punktschätzung.....	309
1.1	Stichproben	309
1.2	Schätzer und ihre Stichprobenverteilungen	310

1.2.1	Grundlagen der Punktschätzung.....	310
1.2.2	Verteilung des Stichprobenmittels.....	313
1.2.2.1	Ziehen mit Zurücklegen	313
1.2.2.2	Ziehen ohne Zurücklegen	316
1.2.3	Verteilung des Stichprobenanteilswerts	318
1.2.3.1	Ziehen mit Zurücklegen	318
1.2.3.2	Ziehen ohne Zurücklegen	319
1.2.4	Verteilung der Stichprobenvarianz	321
1.2.5	Verteilung weiterer Stichprobengrößen	322
1.2.5.1	Differenz zweier Stichprobenmittel.....	322
1.2.5.2	Differenz zweier Stichprobenanteilswerte	323
1.2.5.3	Quotient zweier Stichprobenvarianzen.....	324
1.3	Güte von Schätzern	326
1.3.1	Erwartungstreue	326
1.3.2	Asymptotische Erwartungstreue	327
1.3.3	Effizienz.....	328
1.3.4	Konsistenz	328
1.3.5	Mittlerer quadratischer Fehler.....	329
1.4	Konstruktion von Schätzern	330
1.4.1	Methode der kleinsten Quadrate.....	331
1.4.2	Maximum-Likelihood-Methode	331
2.	Intervallschätzung.....	335
2.1	Grundlagen	335
2.2	Konfidenzintervalle für das arithmetische Mittel	336
2.2.1	Normalverteilte Grundgesamtheit mit bekannter Varianz	338
2.2.2	Normalverteilte Grundgesamtheit mit unbekannter Varianz	340
2.2.3	Beliebig verteilte Grundgesamtheit	341
2.3	Konfidenzintervall für den Anteilswert	342
2.4	Konfidenzintervall für die Varianz	344
2.5	Überblick über die behandelten Konfidenzintervalle.....	345
2.6	Planung des Stichprobenumfangs	346 A
2.6.1	Konfidenzintervall für das arithmetische Mittel	346
2.6.2	Konfidenzintervall für den Anteilswert	347
2.6.3	Konfidenzintervall für die Varianz	347
3.	Testen von Hypothesen	349
3.1	Allgemeines Testschema.....	349

3.2 Testklassifizierung	353
3.3 Parametertests	354
3.3.1 Einstichprobentests	354
3.3.1.1 Einstichprobentest für den Anteilswert	354
3.3.1.2 Einstichprobentest für das arithmetische Mittel	361
3.3.1.3 Statistische Qualitätskontrolle	365
3.3.1.4 Einstichprobentest für die Varianz	366
3.3.2 Zweistichprobentests	368
3.3.2.1 Vergleich zweier arithmetischer Mittel	369
3.3.2.2 Vergleich zweier Anteilswerte	372
3.3.2.3 Vergleich zweier Varianzen	373
3.3.3 Parametertests bei verbundenen Stichproben	375
3.3.3.1 Differenzentest	376
3.3.3.2 Korrelationstest	378
3.3.4 Gütfunktionen von Parametertests	381
3.4 Verteilungstests	386
3.4.1 Chi-Quadrat-Anpassungstest	386
3.4.1.1 Anpassungstest bei diskret verteilter Grundgesamtheit	386
3.4.1.2 Anpassungstest bei stetig verteilter Grundgesamtheit	391
3.4.2 Chi-Quadrat-Unabhängigkeitstest	392
3.4.3 Chi-Quadrat-Homogenitätstest	397
3.5 Einfache Varianzanalyse	399
3.6 Überblick über die behandelten Testverfahren	403
4. Aufgaben	405
IV Einführung in die Ökonometrie	413
1. Grundlagen	415
1.1 Was ist Regressionsanalyse?	415
1.1.1 Ziele der Regressionsanalyse	415
1.1.2 Grundgedanken und Abgrenzungen	417
1.2 Das Prinzip der kleinsten Quadrate	418
1.2.1 OLS bei Modellen mit einer erklärenden Variablen	418
1.2.2 OLS und Linearität	424
1.2.3 OLS bei Modellen mit mehreren erklärenden Variablen	426
1.2.4 Güte einer geschätzten Regressionsgleichung	428
1.2.4.1 Das Bestimmtheitsmaß	428

1.2.4.2	Einfacher Korrelationskoeffizient	431
1.2.4.3	Angepasstes Bestimmtheitsmaß.....	432
2.	Das lineare Regressionsmodell und seine Annahmen	435
2.1	Das lineare Regressionsmodell.....	435
2.1.1	Die Regressionsfunktion der Grundgesamtheit	435
2.1.2	Die Regressionsfunktion der Stichprobe.....	440
2.2	Klassische Annahmen	443
2.2.1	Annahmenkatalog	443
2.2.2	Bedeutung deterministischer und stochastischer Regressoren.....	451
2.2.3	Duplikation der Annahmen des CLRM durch OLS	452
2.3	Statistische Eigenschaften der OLS-Schätzer	452
2.3.1	Verteilung der OLS-Schätzer.....	452
2.3.2	Gauß-Markov-Theorem	456
3.	Testen von Hypothesen und Konfidenzintervalle.....	459
3.1	Testen einzelner Regressionsparameter – t-Test.....	459
3.1.1	Hypothesen, t-Statistik und Entscheidungsregel.....	459 A
3.1.2	Der p-Wert	463
3.1.3	Beschränkungen des t-Tests	464
3.1.4	Konfidenzintervalle für Regressionsparameter.....	465 A
3.2	Simultanes Testen mehrerer Parameter – F-Test.....	467
3.2.1	Hypothesen, F-Statistik und Entscheidungsregel	467
3.2.2	F-Test für die Gesamt signifikanz	468
3.2.3	Weitere Anwendungen des F-Tests und der Chow-Test	470
3.3	Test der Normalverteilungsannahme	473
4.	Verletzungen der Annahmen des klassischen Regressionsmodells	475
4.1	Modellspezifikation I: Variablenwahl.....	475
4.1.1	Vernachlässigte Variablen.....	475
4.1.2	Überflüssige Variablen.....	478
4.1.3	Modellspezifikationskriterien und Spezifikationstests	480
4.1.4	Verzögerte erklärende Variablen.....	483
4.2	Modellspezifikation II: Funktionale Form	486
4.2.1	Bedeutung des konstanten Terms.....	486
4.2.2	Alternative funktionale Formen.....	488
4.2.2.1	Lineare Form	488
4.2.2.2	Doppel-Log-Form.....	489
4.2.2.3	Semi-Log-Form.....	490

4.2.2.4	Polynom-Form.....	492
4.2.2.5	Inverse Form	493
4.2.2.6	Zusammenfassender Überblick.....	494
4.2.3	Dummy-Variablen	495
4.2.3.1	Achsenabschnitts-Dummies	495
4.2.3.2	Steigungs-Dummies.....	499
4.2.4	Folgen der Wahl einer falschen funktionalen Form	501
4.3	Multikollinearität.....	503
4.3.1	Formen von Multikollinearität.....	503
4.3.2	Konsequenzen von Multikollinearität.....	505
4.3.3	Aufdecken von Multikollinearität.....	506 A
4.3.4	Vorgehensweise bei festgestellter Multikollinearität.....	510
4.4	Heteroskedastizität	516
4.4.1	Ursachen von Heteroskedastizität.....	516
4.4.2	Konsequenzen von Heteroskedastizität.....	518
4.4.3	Aufdecken von Heteroskedastizität	520 A
4.4.3.1	Grafische Methode	520
4.4.3.2	Breusch-Pagan LM-Test.....	523
4.4.3.3	White-Test	525
4.4.4	Vorgehensweise bei festgestellter Heteroskedastizität	528
4.4.4.1	Gewichtetes Prinzip der kleinsten Quadrate (WLS).....	528
4.4.4.2	White Standardfehler.....	531
4.4.4.3	Variablenredefinition.....	533
4.5	Autokorrelation	534
4.5.1	Formen von Autokorrelation	534
4.5.2	Konsequenzen von Autokorrelation	540
4.5.3	Aufdecken von Autokorrelation	542 A
4.5.3.1	Grafische Methode	542
4.5.3.2	Durbin-Watson d-Test.....	543
4.5.3.3	Breusch-Godfrey LM-Test	546
4.5.4	Vorgehensweise bei festgestellter Autokorrelation	548
4.5.4.1	Verallgemeinertes Prinzip der kleinsten Quadrate (GLS)	549
4.5.4.2	Newey-West Standardfehler.....	552
4.6	Korrelation zwischen erklärenden Variablen und stochastischem Störterm ..	554
4.6.1	Konsequenzen.....	554
4.6.2	Ursachen.....	555

4.6.2.1	Vernachlässigte Variablen.....	555
4.6.2.2	Messfehler	555
4.6.2.3	Verzögerte endogene Variable	556
4.6.2.4	Simultanität.....	557
4.6.3	Instrumentenvariablenschätzung.....	558
4.6.3.1	Instrumentenvariablen	558
4.6.3.2	Zweistufige Methode der kleinsten Quadrate (TSLS)	560 A
4.6.3.3	Hausman-Test und Verletzung von Annahme 2b.....	564
4.6.3.4	Sargan-Test und Güte von Instrumenten	567
4.7	Besonderheiten bei der Arbeit mit Zeitreihen	570
4.7.1	Dynamische Modelle	570
4.7.1.1	Grundlagen	570
4.7.1.2	Problem der Autokorrelation in ARDL-Modellen.....	571
4.7.2	Nichtstationäre Zeitreihen und Kointegration.....	572
4.7.2.1	Stationarität vs. Nicht-Stationarität.....	572
4.7.2.2	Random Walks und Unit Roots	573
4.7.2.3	Differenzstationarität vs. Trendstationarität.....	576
4.7.2.4	Scheinregression und ihre Bekämpfung	578
4.7.2.5	Prüfung auf Stationarität	582 A
4.7.2.6	Kointegration und Fehlerkorrekturmodell	591 A
4.7.2.7	Zusammenfassung	596
5.	Zusammenfassende Anwendungen aus dem Finanzbereich	597
5.1	Capital Asset Pricing Model.....	597 A
5.2	Investmentfondsperformance	600 A
6.	Prognose mit geschätzten Regressionsmodellen	605
6.1	Grundlagen der Prognose	605
6.2	Bedingte Prognosen.....	608
6.2.1	Prognosefehler bei bedingten Prognosen.....	608
6.2.2	Beurteilung der Güte von Prognosen	611 A
6.2.3	Prognose bei Vorliegen von Autokorrelation	615
6.2.4	Trendprognosen.....	618
6.3	Unbedingte Prognosen	620
6.4	Zusammenfassung	622
7.	Aufgaben.....	623

V Lösungen	637
1. Kapitel I – Deskriptive Statistik	639
2. Kapitel II – Wahrscheinlichkeitsrechnung	645
3. Kapitel III – Induktive Statistik	655
4. Kapitel IV – Ökonometrie	665
VI Anhang.....	679
1. Statistische Tafeln.....	681
1.1 Binomialkoeffizienten	681
1.2 Binomialverteilung - Verteilungsfunktion	682
1.3 Poissonverteilung - Verteilungsfunktion.....	689
1.4 Standardnormalverteilung - Verteilungsfunktion	692
1.5 Standardnormalverteilung - Wichtige Quantile	693
1.6 Chi-Quadrat-Verteilung - Quantile	694
1.7 t-Verteilung - Quantile	696
1.8 F-Verteilung - Quantile	697
2. Ökonometrische Tafeln	703
2.1 Kritische Werte der Durbin-Watson-Statistik	703
2.2 Kritische Dickey-Fuller t-Werte.....	705
Literaturverzeichnis	707
Stichwortverzeichnis	715

Abbildungsverzeichnis

Abbildung I 1: Deskriptive und induktive Statistik	3
Abbildung I 2: Grundgesamtheit oder Stichprobe und Merkmalsträger	4
Abbildung I 3: Statistische Maßskalen	6
Abbildung I 4: Stetige und diskrete Merkmale	10
Abbildung I 5: Häufigkeitstabelle	15
Abbildung I 6: Häufigkeitstabelle bei klassierten Daten	22
Abbildung I 7: Histogramme typischer Häufigkeitsverteilungen	27
Abbildung I 8: Quartile in der Urliste	28
Abbildung I 9: Statistische Maßzahlen	31
Abbildung I 10: Lageparameter und Skalierung	32
Abbildung I 11: Zusammenhang zwischen stetigen und diskreten Renditen	42
Abbildung I 12: Lageregeln	44
Abbildung I 13: Arbeitstabelle zur Varianzberechnung	51
Abbildung I 14: Einfacher Box-Whisker-Plot	61
Abbildung I 15: Spezieller Box-Whisker-Plot	62
Abbildung I 16: Konstruktion der Lorenz-Kurve	70
Abbildung I 17: Arbeitstabelle zur Berechnung des Gini-Koeffizienten	75
Abbildung I 18: Streudiagramm	82
Abbildung I 19: Kontingenztabelle	83
Abbildung I 20: Kontingenztabelle mit Randhäufigkeiten	85
Abbildung I 21: Korrelationsmaße und Skalierung	92
Abbildung I 22: Interpretation der Kovarianz	93
Abbildung I 24: Bravais-Pearson-Korrelationskoeffizient und Streudiagramm	98
Abbildung I 24: Spearman-Rangkorrelationskoeffizient und Streudiagramm	100
Abbildung I 25: Kausale versus statistische Zusammenhänge	106
Abbildung II 1: Teilereignis und äquivalente Ereignisse	151
Abbildung II 2: Vereinigung und Durchschnitt	152
Abbildung II 3: Disjunkte und komplementäre Ereignisse	153
Abbildung II 4: Differenz	154
Abbildung II 5: Durchschnitte bei der Zerlegung	159
Abbildung II 6: Urnenmodell	167
Abbildung II 7: Überblick Kombinatorik	176
Abbildung II 8: Vierfeldertafel	181

Abbildung II 9: Herleitung der totalen Wahrscheinlichkeit I	183
Abbildung II 10: Herleitung der totalen Wahrscheinlichkeit II	184
Abbildung II 11: Zufallsvariablen als Abbildung von Ω auf die reelle Achse	192
Abbildung II 12: Zusammenhang Wahrscheinlichkeits- und Verteilungsfunktion	199
Abbildung II 13: Verteilungsfunktion einer stetigen Zufallsvariable.....	200
Abbildung II 14: Dichtefunktion einer stetigen Zufallsvariable	202
Abbildung II 15: Zusammenhang Verteilungs- und Dichtefunktion.....	205
Abbildung II 16: Symmetrie und Erwartungswert (stetige ZV)	209
Abbildung II 17: Bestimmung des Median (stetige ZV)	218
Abbildung II 18: Ungleichung von Tschebyscheff	220
Abbildung II 19: Gemeinsame Wahrscheinlichkeitsfunktion	227
Abbildung II 20: Dichtefunktion der stetigen Gleichverteilung	261
Abbildung II 21: Verteilungsfunktion der stetigen Gleichverteilung	262
Abbildung II 22: Dichte- und Verteilungsfunktion der Normalverteilung	269
Abbildung II 23: Skizzierte NV-Dichtefunktionen für verschiedene σ	270
Abbildung II 24: Skizzierte NV-Dichtefunktionen für verschiedene μ	270
Abbildung II 25: Dichtefunktion der Standardnormalverteilung.....	271
Abbildung II 26: Dichtefunktion, Quantile und Intervallwahrscheinlichkeiten.....	275
Abbildung II 27: Diskrete Verteilungen im Überblick.....	282
Abbildung II 28: Stetige Verteilungen im Überblick.....	283
Abbildung II 29: Approximationsregeln ausgewählter Verteilungen.....	289
Abbildung III 1: Verteilung des Stichprobenmittels	317
Abbildung III 2: Wichtige Stichprobenverteilungen im Überblick.....	325
Abbildung III 3: Erwartungstreue und verzerrte Schätzer	326
Abbildung III 4: Dichten einer konsistenten Schätzfunktion für verschiedene n	329
Abbildung III 5: Vorteilhaftigkeit leicht verzerrter Schätzer	330
Abbildung III 6: Wahrscheinlichkeitsintervall.....	337
Abbildung III 7: Intervallschätzungen für den Mittelwert	338
Abbildung III 8: Konfidenzintervalle im Überblick	345
Abbildung III 9: Fehler 1. und 2. Art	351
Abbildung III 10: Interpretation eines Testergebnisses	353
Abbildung III 11: Dichtefunktion von V bei korrekter Nullhypothese a)	356
Abbildung III 12: Dichtefunktion von V bei korrekter Nullhypothese b)	357
Abbildung III 13: Dichtefunktion von V bei korrekter Nullhypothese c).....	357
Abbildung III 14: Einstichprobentests für das arithmetische Mittel	362
Abbildung III 15: Überwachungsgrafik zur statistischen Qualitätskontrolle.....	366
Abbildung III 16: Gütfunktion eines unverfälschten und eines idealen Tests.....	382
Abbildung III 17: Gütfunktion vs. OC-Kurve.....	383

Abbildung III 18: Gütefunktionen beim einfachen Gaußtest	385
Abbildung III 19: Hilfstabelle für den Chi-Quadrat-Anpassungstest	387
Abbildung III 20: Kontingenztabelle für den Chi-Quadrat-Unabhängigkeitstest	393
Abbildung III 21: Kontingenztabelle für den Chi-Quadrat-Homogenitätstest	398
Abbildung IV 1: Linearer Zusammenhang zweier Variablen.....	418
Abbildung IV 2: OLS-Schätzung	419
Abbildung IV 3: Zerlegung der Varianz von Y	429
Abbildung IV 4: Bestimmtheitsmaß.....	430
Abbildung IV 5: Regressionsfunktion der Grundgesamtheit I.....	436
Abbildung IV 6: Regressionsfunktion der Grundgesamtheit II.....	438
Abbildung IV 7: Wahre und geschätzte Regressionsgerade I.....	442
Abbildung IV 8: Wahre und geschätzte Regressionsgerade II	442
Abbildung IV 9: Bedingte Verteilung der ϵ_i	444
Abbildung IV 10: Homoskedastizität.....	446
Abbildung IV 11: Heteroskedastizität	447
Abbildung IV 12: Korrelationsmuster des Störterms	448
Abbildung IV 13: Einfluss von Ausreißern auf OLS-Schätzungen	449
Abbildung IV 14: Veranschaulichung der BLUE-Eigenschaft von OLS	456
Abbildung IV 15: Vernachlässigung des konstanten Terms	487
Abbildung IV 16: Doppel-Log-Form.....	490
Abbildung IV 17: Semi-Log-Formen	491
Abbildung IV 18: Inverse Form	494
Abbildung IV 19: Achsenabschnitts-Dummy.....	496
Abbildung IV 20: Steigungs-Dummies	500
Abbildung IV 21: Vollkommene vs. unvollkommene Multikollinearität	504
Abbildung IV 22: Homo- vs. Heteroskedastizität.....	517
Abbildung IV 23: Grafische Aufdeckung von Heteroskedastizität	521
Abbildung IV 24: Positive, negative und fehlende Autokorrelation.....	537
Abbildung IV 25: Autokorrelation des Störterms und funktionale Form	539
Abbildung IV 26: Durbin-Watson d-Test.....	545
Abbildung IV 27: Prognosearten (Zeitreihenmodelle).....	606
Abbildung IV 28: Prognosekonfidenzintervalle	610

Anwendungsverzeichnis

Renditen und Renditedurchschnitte	40
Schätzung von Aktienvolatilitäten	56
Analyse von Wettbewerbsprozessen und Einkommensverteilungen.....	70
Inflationsmessung	118
Inflationsbereinigung	119
Kaufkraftvergleiche	121
Indizes aus der Wirtschaftspraxis (VPI, HVPI, DAX)	128
Wettquotienten.....	165
Das Dreitorespiel	189
Roulette	222
Portfoliorisiko und Diversifikation	240
Statistische Qualitätskontrolle	250
Kreditrisiko	260
Planung von Stichprobenumfängen	346
Schätzung von Konsumfunktionen	460
Schätzung von Produktionsfunktionen	466
Erklärung von Importvolumina in Wirtschaftsräumen.....	509
Erklärung von Immobilienpreisen.....	522
Schätzung von Nachfragefunktionen	542
Erklärung von Löhnen	562
Untersuchung von Trends in der Goldpreisentwicklung	589
Beurteilung des langfristigen Zinseinflusses von Zentralbanken	592
Beurteilung der langfristigen Beziehung zwischen Geldmarktzinssätzen.....	595
Prüfung der empirischen Validität des Capital Asset Pricing Models.....	597
Analyse der Performance von Investmentfonds	600
Zinsprognose.....	613

I DESKRIPTIVE STATISTIK

Thema dieses ersten Kapitels ist die Aufbereitung von Zahlenmaterial mittels überschaubarer Tabellenstrukturen, grafischer Abbildungen und aussagekräftiger Kennzahlen. Nach einem Grundlagenabschnitt, in dem wir zunächst die wichtigsten Grundbegriffe der deskriptiven Statistik klären, befassen wir uns ausführlich mit sog. eindimensionalen Häufigkeitsverteilungen, die sich über Lage-, Streuungs-, Schiefe- und Konzentrationsparameter charakterisieren lassen. Im Anschluss daran erfolgt die detaillierte Behandlung wichtiger Kennzahlen mehrdimensionaler Häufigkeitsverteilungen.

Den Abschluss des Kapitels zur deskriptiven Statistik bildet die Behandlung von Preismesszahlen und -indizes. Neben verschiedenen Indextypen und -berechnungsmethoden stellen wir dabei insbesondere verschiedene in der Praxis wichtige Indizes zur Abbildung des gesamtwirtschaftlichen Preisniveaus und der Lage auf Finanzmärkten vor.

1.

Grundbegriffe

Bevor wir auf statistische Methoden zur Zahlenaufbereitung eingehen, werden wir in diesem einführenden Abschnitt zunächst den Begriff Statistik definieren (1.1) und seine Teilgebiete kurz erläutern. Darauf folgend klären wir die für die deskriptive Statistik bedeutenden Begriffe Merkmalsträger, Merkmal, Merkmalsausprägung, Grundgesamtheit und Stichprobe (1.2) und gehen näher auf Maßskalen und andere Kriterien ein, die sich zur Klassifizierung von Merkmalen verwenden lassen (1.3).

1.1 Der Statistikbegriff

Der Begriff **Statistik** hat im Allgemeinen keine eindeutig festgelegte Definition. Er ist mehrdeutig. Zum einen verstehen wir unter Statistik eine Zusammenstellung (tabellarisch oder grafisch) von Zahlen und Daten über bestimmte Sachverhalte. Typische Beispiele hierfür sind Bevölkerungs- und Umsatzstatistiken. Zum anderen ist Statistik die Gesamtheit aller Methoden zur Erhebung, Aufbereitung und Analyse empirischer Informationen. Gerade diese Methoden gehören mittlerweile zum unverzichtbaren Instrumentarium vieler Fachwissenschaften wie etwa der Physik, der Biologie, der Geografie und natürlich der Wirtschaftswissenschaften.

Die Statistik kann in die Teilgebiete deskriptive und induktive Statistik untergliedert werden (vgl. Abbildung I 1). Die **deskriptive Statistik** (beschreibende, explorative Statistik) befasst sich damit, die umfangreichen und unüberschaubaren Einzeldaten statistischer Erhebungen durch möglichst wenige, aussagekräftige Kennzahlen zu charakterisieren. Die Aufgabe der **induktiven Statistik** (schließende Statistik) ist der Rückschluss auf eine Grundgesamtheit anhand von Stichproben. Dies ist in der Praxis von entscheidender Wichtigkeit, da Vollerhebungen im Vergleich zu Teilerhebungen (vgl. Abschnitt 1.2) oft zu kostenintensiv oder einfach unmöglich sind.

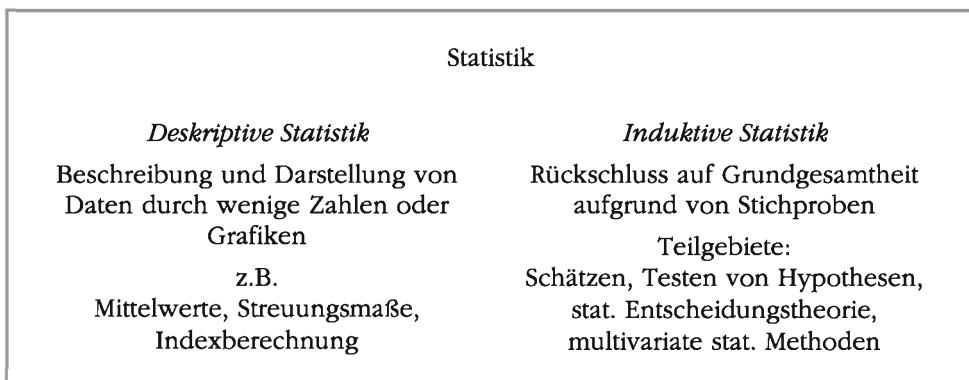


Abbildung I 1: Deskriptive und induktive Statistik

Als weiteres Teilgebiet könnte die **Wahrscheinlichkeitsrechnung** aufgeführt werden. Sie wird im Rahmen dieses Lehrbuches als eigenständiges Kapitel II behandelt. Da sie für das Verständnis der Verfahren der induktiven Statistik unerlässlich ist, wird sie häufig in einem Zug mit der induktiven Statistik genannt.

1.2 Merkmalsträger, Grundgesamtheiten und Stichproben

In der Statistik bezeichnen wir die Gegenstände einer Untersuchung als **Merkmalsträger** (statistische Einheiten). Die Gesamtheit der Merkmalsträger mit übereinstimmenden Identifikationskriterien (sachlich, räumlich, zeitlich) bzw. die für eine statistische Untersuchung relevanten Merkmalsträger bilden die sog. **Grundgesamtheit** (statistische Masse). Jeder Merkmalsträger besitzt gewisse Eigenschaften, die im Rahmen einer Untersuchung relevant bzw. zu analysieren sind. Diese nennen wir **Merkmale**. Jedes Merkmal besitzt zwei oder mehr **Merkmalsausprägungen** (Menge der Ergebnisse), die bei der Beobachtung auftreten können. Wird die Grundgesamtheit bei einer Untersuchung vollständig erfasst, so sprechen wir von einer **Voll- oder Totalerhebung**. Im Gegensatz dazu wird bei einer **Teil- oder Stichprobenerhebung** nur ein Bruchteil der Grundgesamtheit nach gewissen Zufallsgesetzen ausgewählt.

Diese Bezeichnungen sollen anhand von Abbildung I 2 und des konkreten Beispiels I 1-1 veranschaulicht werden.

Grundgesamtheit oder Stichprobe		
Merkmalsträger 1	Merkmal 1	Merkmalsausprägung
	Merkmal 2	Merkmalsausprägung
...	...	
	Merkmal m	Merkmalsausprägung
Merkmalsträger 2	Merkmal 1	Merkmalsausprägung
	Merkmal 2	Merkmalsausprägung
...	...	
	Merkmal m	Merkmalsausprägung
⋮		
Merkmalsträger n	Merkmal 1	Merkmalsausprägung
	Merkmal 2	Merkmalsausprägung
...	...	
	Merkmal m	Merkmalsausprägung

Abbildung I 2: Grundgesamtheit oder Stichprobe und Merkmalsträger

Beispiel I 1-1	Merkmalsbegriffe										
<p>Im Rahmen einer Vollerhebung soll das monatliche Durchschnittseinkommen von bayerischen Betriebswirtschaftsstudenten im Jahr 2010 untersucht werden. Dazu werden von den Hochschulverwaltungen Fragebögen an die Studentinnen und Studenten ausgegeben, worin Angaben über Alter, Geschlecht, Familienstand, Wohnort, Einkommenshöhe usw. zu machen sind. Die vorhergehend definierten Begrifflichkeiten können nun folgendermaßen auf dieses Beispiel übertragen werden:</p> <ul style="list-style-type: none"> ▪ Grundgesamtheit: Alle bayerischen (räumliche Identifikation) Betriebswirtschaftsstudenten (sachliche Identifikation) im Jahr 2010 (zeitliche Identifikation) ▪ Merkmalsträger: einzelner Student ▪ Merkmal: Mögliche Merkmalsausprägungen: <table> <tr> <td>Alter</td> <td>19, 20, 21, 22, 23, ...</td> </tr> <tr> <td>Geschlecht</td> <td>männlich, weiblich</td> </tr> <tr> <td>Familienstand</td> <td>ledig, verheiratet</td> </tr> <tr> <td>Wohnort</td> <td>München, Nürnberg, ...</td> </tr> <tr> <td>Einkommenshöhe</td> <td>300 €, 400 €, 450 €, ...</td> </tr> </table> 		Alter	19, 20, 21, 22, 23, ...	Geschlecht	männlich, weiblich	Familienstand	ledig, verheiratet	Wohnort	München, Nürnberg, ...	Einkommenshöhe	300 €, 400 €, 450 €, ...
Alter	19, 20, 21, 22, 23, ...										
Geschlecht	männlich, weiblich										
Familienstand	ledig, verheiratet										
Wohnort	München, Nürnberg, ...										
Einkommenshöhe	300 €, 400 €, 450 €, ...										

Neben **realen Grundgesamtheiten** (z.B. Erwerbstätige in Bayern zum 31.12.2008, herzinfarktbedingte Todesfälle im Jahr 2008, ...) existieren auch **hypothetische Grundgesamtheiten** (z.B. Menge der Ergebnisse eines fortlaufenden Würfelspiels, Geburten bis zum Menschheitsende, ...). Gerade bei den zuletztgenannten Grundgesamtheiten, welche auch als unendliche Grundgesamtheiten bezeichnet werden, ist ausschließlich eine Stichprobenbetrachtung sinnvoll oder möglich.

Nach der zeitlichen Abgrenzung der untersuchten Grundgesamtheit unterscheiden wir **Bestandsgrößen** (Bestandsmassen) und **Stromgrößen** (Bewegungsmassen). Bestandsgrößen beziehen sich auf einen bestimmten Zeitpunkt (z.B. Einwohner Bayerns zum 01.01.2008, Kassenbestand zum 31.12.2008), Stromgrößen sind in Bezug auf einen Zeitraum definiert (z.B. Geburten in Bayern in 2008, Insolvenzen in der BRD in 2007). In der statistischen Praxis werden Bestandsgrößen in der Regel aus Kostengründen durch geeignete Stromgrößen fortgeschrieben. Wir können nämlich den allgemeinen Zusammenhang

$$\frac{\text{Anfangsbestand}}{\text{Bestandsgröße}} + \frac{\text{Zugänge}}{\text{Stromgrößen}} - \frac{\text{Abgänge}}{\text{Bestandsgröße}} = \frac{\text{Schlussbestand}}{\text{Bestandsgröße}}$$

festhalten, den wir auch als **Bestandsgleichung** bezeichnen. In der Bevölkerungsstatistik wird beispielsweise die Wohnbevölkerung nur in größeren Abständen vollständig erfasst (Volkszählung) und in den Zwischenjahren durch die Beziehung Bevölkerung Ende Jahr x = Bevölkerung Ende des Vorjahres + Zugänge im Jahr x (Geburten, Zuwanderung) – Abgänge im Jahr x (Sterbefälle, Abwanderung) fortgeschrieben. Ein Beispiel für die Fortschreibung aus der betriebswirtschaftlichen Praxis ist etwa die Ermittlung von Lagerbeständen an einem Stichtag nach der Methode der laufenden Inventur.

1.3 Klassifikation von Merkmalen

1.3.1 Klassifikation nach dem Skalenniveau

Gerade das Beispiel I 1-1 aus dem vorhergehenden Abschnitt hat gezeigt, dass jedes Merkmal zwei oder mehr mögliche Ausprägungen besitzen kann. Je nach Art des betrachteten Merkmals können diese Merkmalsausprägungen anhand verschiedener **Skalen** gemessen werden. Abbildung I 3 zeigt die zu unterscheidenden Skalen im Überblick und gibt außerdem an, welche Rechenoperationen bei der Auswertung mit den Daten sinnvoll durchführbar und interpretierbar sind.



Abbildung I 3: Statistische Maßskalen

Zur Verdeutlichung der in der Grafik dargestellten Sachverhalte betrachten wir nun die einzelnen Skalen näher und veranschaulichen sie durch geeignete Beispiele:

1. Nominalskala

Die Nominalskala findet bei Merkmalen Anwendung, bei denen die Ausprägungen lediglich eine Verschiedenartigkeit und keine Reihenfolge ausdrücken. Dies bedeutet, dass die Merkmalsausprägungen gleichberechtigt nebeneinander stehen. Es besteht zwar die Möglichkeit jeder Merkmalsausprägung eine Zahl zuzuordnen, doch dient eine solche Verschlüsselung der reinen Identifikation und nicht einer Rangfolgenbildung.

Beispiel I 1-2

Nominalskalierte Merkmale

Als typische Beispiele für nominalskalierte Merkmale sind etwa die Religionszugehörigkeit oder das Geschlecht eines Merkmalsträgers zu nennen. Betrachten wir das Merkmal Geschlecht näher, so zeigen sich die Ausprägungen "Mann" und "Frau". Es kann keine Rangfolge gebildet werden, da keine der Ausprägungen eine höhere Wertigkeit besitzt. Selbst wenn man die Verschlüsselung "1 = Mann" und "2 = Frau" vornimmt, dienen die Zahlen nicht der Rangfolgenbildung. Es kann mit und ohne Verschlüsselung nur Gleichheit oder Ungleichheit überprüft werden. Wir können also z.B. sagen, dass zwei Merkmalsträger unterschiedliches Geschlecht aufweisen. Ein "besseres Geschlecht" gibt es nicht.

Kann ein Merkmal nur zwei verschiedene Ausprägungen annehmen (z.B. "Mann" und "Frau" beim Merkmal "Geschlecht"), sprechen wir auch von einem *dichotomen* Merkmal.

2. Ordinalskala

Neben der Verschiedenartigkeit besitzen die Merkmalsausprägungen ordinalskalierten Merkmale eine natürliche Rangfolge. Das heißt, es kann eine "größer als"-Beziehung zwischen den Ausprägungen aufgestellt werden. Wichtig ist dabei aber, dass die Abstände zwischen den Merkmalsausprägungen nicht quantifizierbar sind.

Beispiel I 1-3

Ordinalskalierte Merkmale

Typische ordinalskalierte Merkmale sind z.B. der Intelligenzgrad, der Tabellenplatz in der Fußballbundesliga und Schulnoten. Für das Merkmal Schulnoten zeigen sich z.B. die Ausprägungen 1, 2, 3, 4, 5, 6. Die Noten erfüllen das Verschiedenartigkeitskriterium einer Nominalskala. Zusätzlich kann die Aussage gemacht werden, dass z.B. die Note 2 besser (höherwertiger) ist als die Note 4. Es kann aber nicht geschlossen werden, dass die 1 im Vergleich zur 2 genau so viel besser ist wie die 4 im Vergleich zur 5. Die Berechnung eines Notendurchschnitts ist daher streng genommen nur zulässig, wenn wir *unterstellen*, dass die Abstände zwischen den einzelnen Noten gleichartig sind. Nur dann kann eine Durchschnittsnote von 1,5 als genau zwischen 1 und 2 liegend interpretiert werden.

Im Zusammenhang mit Schulnoten und dem Tabellenplatz in der Bundesliga sprechen wir auch vom Vorliegen einer sog. **Rangskala**, da es sich um eine Skala handelt, die nur ganzzahlige Ordnungsziffern aufweist, die mit 1 beginnen und in ununterbrochener Reihenfolge hintereinander stehen.

Wird z.B. in einem Fragebogen die Qualität einer Lehrveranstaltung evaluiert, so könnten die Ausprägungen der Qualitätsvariablen folgende Gestalt aufweisen:

-2 = sehr niedrig -1 = niedrig 0 = mittelmäßig +1 = hoch +2 = sehr hoch

Solche Bewertungs- oder Einschätzungsskalen nennen wir auch **Rating-Skalen**.

3. Intervallskala

Zusätzlich zu den Eigenschaften einer Ordinalskala können bei intervallskalierten Merkmalen zwischen den Merkmalsausprägungen Entfernung oder Abstände angegeben und sinnvoll verglichen werden. Es existiert kein natürlicher Nullpunkt. Er ist willkürlich definiert. Zwischen den Merkmalsträgern können keine Produkte oder Quotienten gebildet werden.

Beispiel I 1-4

Intervallskalierte Merkmale

Die Kalenderzeitrechnung und Temperaturen in °C können als intervallskaliert betrachtet werden. Es kann nämlich z.B. für das Merkmal Temperatur (von Wasser) zwischen den Ausprägungen 20°C und 30°C eine Differenz von 10°C festgestellt werden. Dieser Unterschied wird als genau so groß erachtet wie der Unterschied zwischen 50°C und 60°C. Eine Aussage wie "20°C ist viermal so warm wie 5°C." ist hingegen nicht möglich.

4. Verhältnisskala

Zu den Eigenschaften der Intervallskala kommt bei der Verhältnisskala noch ein natürlicher Nullpunkt hinzu, wodurch eine Produkt- und Quotientenbildung möglich wird. Dies ist typischerweise bei Skalen zur Längen-, Mengen- und Gewichtsmessung der Fall.

Beispiel I 1-5	Verhältnisskalierte Merkmale
----------------	------------------------------

Alter, Körpergröße und Umsatz können zu den verhältnisskalierten Merkmalen gezählt werden. Es kann nämlich z.B. für die Ausprägungen 100 € und 200 € des Merkmals Umsatz ausgesagt werden, dass 200 € zweimal so viel ist wie 100 €.

Intervall- und Verhältnisskalen werden auch unter den Begriffen **Kardinalskalen** oder **metrische Skalen** zusammengefasst, da ihnen ein Maßsystem zugrunde liegt.

Mit jedem Skalentyp ist ein eindeutig festgelegtes Informationsniveau verbunden. In der Praxis kann es nun aber aus verschiedenen Gründen erforderlich sein, eine **Transformation von Merkmalsausprägungen** vorzunehmen. Eine solche ist natürlich nur zulässig, wenn dadurch die in den Skalenwerten ursprünglich enthaltenen Informationen nicht verändert werden (*skalenerhaltende Transformation*). Würden wir z.B. die Ausprägungen 1, 2, 3, 4, 5 und 6 des Merkmals Schulnote in 10, 20, 30, 40, 50 und 60 transformieren, so wären weiterhin alle Merkmale einer Ordinalskala erfüllt und es würden keine Informationen verlorengehen. Würden wir aber z.B. zu Körpergrößen eine gewisse Konstante addieren, so würde dies zu Interpretationsfehlern der neuen Größen führen. Es ist also deutlich zu erkennen, dass für jeden Skalentyp nur ganz bestimmte Transformationen möglich sind. Diese betrachten wir im Folgenden näher:

1. Nominalskala

Hier sind alle *Transformationen* zulässig, die die Klassenbezeichnungen ändern. So können beispielsweise in einem Produktsortiment Artikel Nr. 321 und Nr. 322 mit neuen Artikelnummern 2005 und 1998 versehen werden.

2. Ordinalskala

Zulässig sind hier nur sog. *streng monotone Transformationen*. Dies bedeutet, dass aus jedem Merkmalswert x_i mit einer Funktion $f(x_i)$ der neue Merkmalswert y_i gebildet wird. Die Rangfolge der Merkmalsträger darf sich dadurch nicht verändern.

$$y_i = f(x_i) \quad \text{wobei} \quad x_i < x_{i+1} \rightarrow y_i < y_{i+1} \quad \text{für alle } i \quad (\text{I1: 1})$$

Als Beispiel kann hier die bereits genannte Schulnotentransformation aufgeführt werden. Vor und nach der Transformation $y_i = 10x_i$ ist die Rangfolge identisch. Alternativ wären hier ebenso Transformationen wie $y_i = \ln x_i$ oder $y_i = \sqrt{x_i}$ denkbar. Auch bei ihnen bleibt nämlich die Rangfolge erhalten.

3. Intervallskala

Bei intervallskalierten Merkmalen dürfen ausschließlich *lineare Transformationen* durchgeführt werden. Dies bedeutet, dass sich die neuen Merkmalswerte y_i als lineare Funktion von x_i ergeben.

$$y_i = ax_i + b \quad \text{mit } a > 0 \quad (\text{I1: 2})$$

Nehmen wir etwa an, dass die Ausprägungen 30°C und 40°C des Merkmals Temperatur in ${}^\circ\text{F}$ (Fahrenheit) transformiert werden sollen, so erfolgt dies über

$$\underbrace{y_i}_{{}^\circ\text{F}} = 32 + \frac{9}{5} \underbrace{x_i}_{{}^\circ\text{C}},$$

was zu den neuen Werten 86°F und 104°F führt, welche weiterhin intervallskaliert sind und keine Informationen eingebüßt haben.

4. Verhältnisskala

Die wohl restriktivsten Transformationen sind bei Verhältnisskalen möglich. Es sind hier nämlich nur sog. *Ähnlichkeitstransformationen* (*proportionale Transformationen*), d.h. Multiplikationen mit positiven Konstanten zulässig.

$$y_i = ax_i \quad \text{mit } a > 0 \quad (\text{I1: 3})$$

Ähnlichkeitstransformationen finden etwa bei der Umrechnung von Körpergrößen Anwendung. Sollen z.B. die Größen 1,60 m und 1,70 m in cm-Werte umgerechnet werden, erfolgt dies über die Transformation $y_i = 100x_i$, welche die cm-Werte 160 und 170 liefert. Auch die Umrechnung von DM in Euro über die geläufige Transformation $y_i = 1/1,95583 \cdot x_i$ ist eine typische Ähnlichkeitstransformation.

In Ausnahmefällen kann es erforderlich sein *skalensenkende Transformationen* durchzuführen, d.h. solche, die mit einem Informationsverlust verbunden sind. Da die von uns behandelten Skalen in der Reihenfolge Verhältnis-, Intervall-, Ordinal- und Nominalskala dominant sind, bedeutet dies, dass wir durch Transformationen metrisch skalierte Merkmale in ordinalskalierte oder nominalskalierte Merkmale und ordinalskalierte in nominalskalierte Merkmale überführen können. Eine Umkehrung ist nicht möglich, d.h. es gibt keine skalenerhöhenden Transformationen.

Beispiel I 1-6

Skalensenkende Transformationen

Üblicherweise werden zur Messung von Körpergrößen Verhältnisskalen (z.B. Zentimeter-skala) eingesetzt. Eine Messung kann jedoch auch auf ordinalem oder nominalem Niveau erfolgen. So können wir z.B. metrische Größenklassen bilden und diese durch eine skalensenkende Transformation ordinal oder nominal abbilden.

Verhältnisskala (5 Größenklassen)	Ordinalskala (5 Größenränge)	Nominalskala (2 Größenklassen)
bis 140 cm	1 sehr klein	1 normgerechte Größe
über 140 bis 160 cm	2 klein	(150 cm bis 190 cm)
über 160 bis 180 cm	3 mittelgroß	2 nicht normgerechte Größe
über 180 bis 200 cm	4 groß	(unter 150 oder über 190 cm)
über 200 cm	5 sehr groß	

1.3.2 Klassifikation in diskrete und stetige Merkmale

Neben der Untergliederung bezüglich der Skalierung ist die Einteilung in diskrete und stetige Merkmale von Bedeutung. Wir bezeichnen ein Merkmal als **diskret**, wenn es nur endlich viele oder höchstens abzählbar unendlich viele unterschiedliche Ausprägungen annehmen kann. Allgemein können alle nominal- und ordinalskalierten Merkmale als diskret angesehen werden. Auch metrisch skalierte Merkmale können diskret sein, wenn sie nur isolierte Zahlenwerte annehmen. Dies ist typischerweise der Fall, wenn ihnen ein *Zählvorgang* zugrunde liegt.

Beispiel I 1-7

Diskrete Merkmale

Typische diskrete Merkmale sind die Anzahl der Studenten an einer Hochschule oder die Anzahl der Einwohner eines Landes. Betrachten wir das Merkmal Einwohnerzahl näher, so sind Ausprägungen von 1 bis n mit $n \in \mathbb{N}^+$ möglich. Es können also beispielsweise Werte wie 50 oder 51, aber keine beliebigen Zwischenwerte wie 50,5 o. ä. auftreten.

Im Gegensatz zu diskreten Merkmalen sprechen wir von **stetigen** Merkmalen, wenn in einem bestimmten Intervall jeder beliebige Wert angenommen werden kann, wobei die Intervallgrenzen auch offen sein können. Dies gilt in der Regel für Merkmale, denen ein beliebig genauer *Messvorgang* zugrunde liegt.

Beispiel I 1-8

Stetige Merkmale

Lebensdauern, Längen, Breiten und Höhen sowie Füllgewichte sind typische stetige Merkmale. Für das Beispiel Füllgewicht können so etwa 5 kg oder 6 kg, aber auch beliebige Zwischenwerte wie 5,1 kg oder 5,25 kg bei einer Messung festgestellt werden.

Häufig werden aus Zweckmässigkeitsgründen bei der Datenauswertung diskrete Merkmale mit sehr vielen Ausprägungen (z.B. Umsätze) als stetig betrachtet. Wir bezeichnen solche Merkmale als **quasistetig**. Umgekehrt können stetige Merkmale durch sog. **Klassierung** (Gruppierung) als diskrete Merkmale behandelt werden (vgl. Abbildung I 4). Dies ist z.B. üblich, wenn der Datenerhebung nur ein sehr grober Messvorgang zugrunde liegt. Anstelle des genauen Merkmalswertes wird nur noch das Intervall angegeben, in welches der Wert fällt. Hierauf werden wir im Rahmen des Abschnitts I 2 zu Häufigkeitsverteilungen noch genauer eingehen.

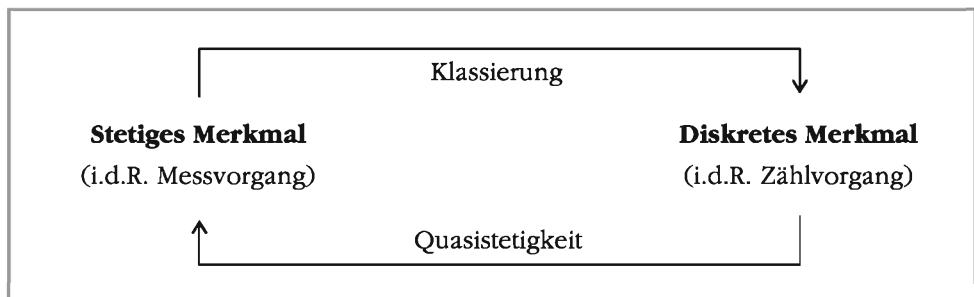


Abbildung I 4: Stetige und diskrete Merkmale

1.3.3 Klassifikation in qualitative und quantitative Merkmale

Neben den bisherigen Klassifizierungen können wir Merkmale auch in qualitative und quantitative Merkmale unterteilen. **Quantitative Merkmale** sind Merkmale, deren Ausprägungen Zahlen sind. **Qualitative Merkmale** hingegen besitzen in der Regel Ausprägungen, die durch einen verbalen Ausdruck charakterisiert sind. Dies bedeutet also, dass nominalskalierte Merkmale als qualitativ und metrisch skalierte Merkmale als quantitativ zu betrachten sind. Ordinalskalierte Merkmale hingegen können nicht genau zugeordnet werden.

Werden die Ausprägungen qualitativer Merkmale durch Zahlen ersetzt, wird dieser Vorgang als **Quantifizierung** bezeichnet. Aus dem qualitativen Merkmal wird also (nur äußerlich) ein quantitatives. Bei der statistischen Auswertung des Datenmaterials ist jedoch zwischen einem echt quantitativen Merkmal und einem quantifizierten Merkmal zu unterscheiden.

Beispiel I 1-9

Quantitative und quantifizierte Merkmale

Für das verhältnisskalierte und damit quantitative Merkmal Einkommen liegen nach Befragung zweier Personen die Ausprägungen 1.000 € und 3.000 € vor. Es kann also ein Durchschnittseinkommen von 2.000 € angegeben werden.

Für dieselben Personen haben sich die Religionszugehörigkeiten 1 und 3 (vorherige Quantifizierung: 1 = katholisch, 2 = evangelisch, 3 = sonstige) ergeben. Hier von einer "mittleren Konfession" 2 zu sprechen wäre unsinnig. Trotz Quantifizierung bleibt die Religionszugehörigkeit ein nominalskaliertes Merkmal, sodass lediglich Vergleiche der Merkmalsausprägungen zulässig sind.

2. Eindimensionale Häufigkeitsverteilungen

Thema dieses Abschnitts ist die Auswertung eindimensionalen (univariaten) Datenmaterials, d.h. Daten zu einem einzigen Merkmal einer Grundgesamtheit oder Stichprobe. Der erste Schritt besteht dabei im Auszählen von Häufigkeiten, mit der die Ausprägungen des interessierenden Merkmals einzeln oder in Klassen zusammengefasst auftreten. Dies liefert uns die sog. Häufigkeitsverteilung. In einem zweiten Schritt können wir Maßzahlen berechnen, die diese Häufigkeitsverteilung charakterisieren. Maßzahlen geben zwar einen kompakten Überblick über wesentliche Eigenschaften einer Häufigkeitsverteilung, doch lassen sich in grafischen Darstellungen oft Strukturen im Datenmaterial erkennen, die Maßzahlen nicht widerspiegeln. Gerade daher empfiehlt es sich, tabellarische und grafische Darstellungen von Häufigkeitsverteilungen mit ihren Maßzahlen zu kombinieren, um das vorliegende Datenmaterial möglichst gut zu beschreiben.

2.1 Häufigkeitsverteilung

Das Auszählen der Häufigkeiten von Ausprägungen diskreter Merkmale und derer stetiger (bzw. quasistetiger) Merkmale erfordert unterschiedliche Vorgehensweisen. Bei diskreten Merkmalen (mit nicht zu vielen unterschiedlichen möglichen Ausprägungen) kommen im auszuwertenden Datenmaterial die einzelnen Ausprägungen meist mehrmals, d.h. bei mehreren Merkmalsträgern vor. Bei stetigen (und quasi-stetigen) Merkmalen ist das mehrmalige Auftreten hingegen nicht die Regel.

2.1.1 Häufigkeitsverteilung bei diskreten Merkmalen

In statistischen Erhebungen werden von einer Grundgesamtheit mit insgesamt N Elementen (Merkmalsträgern) in der Regel genau n Elemente (Stichprobe) hinsichtlich eines einzigen Merkmals X untersucht. Die durch die Stichprobe gewonnenen Werte x_1, x_2, \dots, x_n bezeichnen wir als Beobachtungswerte. Durch Aneinanderreihung aller Beobachtungswerte erhalten wir die sog. Beobachtungsreihe oder **Urliste**. Anders ausgedrückt ist die Urliste das n -Tupel (x_1, x_2, \dots, x_n) aller n Beobachtungswerte.

Zur Aufbereitung dieses Datenbestandes bzw. der Urliste empfiehlt es sich, die Häufigkeitsverteilung des Merkmals X zu gewinnen. Dazu benötigen wir die absoluten und relativen Häufigkeiten, der in der Urliste vorkommenden individuellen Merkmalsausprägungen von X . Wir bezeichnen diese Merkmalsausprägungen mit a_1, a_2, \dots, a_k , d.h. wir gehen davon aus, dass in der Stichprobe genau k verschiedene Ausprägungen des Merkmals X auftreten und k nicht zu groß ist. Wir können damit für die gesuchten Häufigkeiten die folgenden Zusammenhänge festhalten:

1. Absolute Häufigkeit

Unter der absoluten Häufigkeit $h(a_j)$ verstehen wir die Anzahl von Elementen der Urliste, die auf eine bestimmte Merkmalsausprägung a_j entfallen.

Es gilt allgemein

$$0 \leq h(a_j) \leq n \quad (\text{I2: 1})$$

und

$$h(a_1) + h(a_2) + \dots + h(a_k) = \sum_{j=1}^k h(a_j) = n. \quad (\text{I2: 2})$$

Der Wert der absoluten Häufigkeit muss also nach (I2: 1) zwingend größer oder gleich Null und kleiner oder gleich der Gesamtzahl n der Elemente der Urliste sein. Zudem ergibt nach (I2: 2) die Summe aller absoluten Häufigkeiten stets n .

2. Relative Häufigkeit

Dividieren wir die absoluten Häufigkeiten $h(a_j)$ durch die Gesamtzahl der Elemente n , so erhalten wir die relative Häufigkeit

$$f(a_j) = \frac{h(a_j)}{n}, \quad (\text{I2: 3})$$

für die

$$0 \leq f(a_j) \leq 1 \quad (\text{I2: 4})$$

und

$$f(a_1) + f(a_2) + \dots + f(a_k) = \sum_{j=1}^k f(a_j) = 1 \quad (\text{I2: 5})$$

gilt. Da die relative Häufigkeit $f(a_j)$ nichts anderes angibt als den Anteil, mit dem a_j in der Stichprobe des Umfangs n vertreten ist, liegt $f(a_j)$ nach (I2: 4) logischerweise stets zwischen 0 und 1 oder stimmt mit diesen Werten überein. Zudem muss die Summe aller Anteilsraten zwingend 1 ergeben, was (I2: 5) zum Ausdruck bringt. Relative Häufigkeiten werden häufig auch in Prozent angegeben. Ihre Summe liegt dann bei 100 %.

Unter der **absoluten Häufigkeitsverteilung** verstehen wir nun nichts anderes als die Paare $(a_j, h(a_j))$ mit $j = 1, 2, \dots, k$ und unter der **relativen Häufigkeitsverteilung** die Paare $(a_j, f(a_j))$ mit $j = 1, 2, \dots, k$. Die einzelnen Merkmalsausprägungen a_j werden in einer Häufigkeitsverteilung meist zur besseren Übersicht in aufsteigender Reihenfolge sortiert.

Häufigkeitsverteilungen können wir übersichtlich in Form von *Häufigkeitstabellen* darstellen. Es empfiehlt sich dabei die Verwendung der in Abbildung I 5 skizzierten Struktur. Wir erkennen daran recht schön, dass die Eigenschaften (I2: 2) und (I2: 5) von absoluten und relativen Häufigkeiten hier als Spaltensummen überprüfbar sind. Für eine anschauliche Tabellengestaltung empfiehlt es sich, mit möglichst wenig Linierungen zu arbeiten, um die Tabelle optisch nicht zu überfrachten.

Laufende Nummer	Merkmals- ausprägung	absolute Häufigkeit	relative Häufigkeit
j	a_j	$h(a_j)$	$f(a_j)$
1	a_1	$h(a_1)$	$f(a_1)$
2	a_2	$h(a_2)$	$f(a_2)$
...
k	a_k	$h(a_k)$	$f(a_k)$
Σ	-	n	1

Abbildung I 5: Häufigkeitstabelle

Beispiel I 2-1**Tabellarische Darstellung einer Häufigkeitsverteilung**

Anlässlich einer Mitarbeiterstatistik wurde in einer Abteilung eines Unternehmens das Alter X von $n = 25$ Mitarbeitern erfasst. Dabei ergab sich die folgende Urliste:

	x_i in Jahren		x_i in Jahren		x_i in Jahren
1	23	10	23	19	35
2	25	11	25	20	25
3	23	12	29	21	23
4	29	13	25	22	35
5	35	14	29	23	25
6	25	15	25	24	29
7	23	16	23	25	25
8	25	17	35		
9	29	18	25		

Wir erkennen daran, dass in der Stichprobe lediglich die $k = 4$ Ausprägungen 23, 25, 29 und 35 des Merkmals X auftauchen. Durch Auszählung dieser Ausprägungen erhalten wir die absoluten Häufigkeiten. Diese liefern dividiert durch $n = 25$ schließlich die relativen Häufigkeiten. Wir können damit die Häufigkeitsverteilung tabellarisch wie folgt darstellen:

Laufende Nummer	Merkmals- ausprägung	Absolute Häufigkeit	Relative Häufigkeit	Prozentwert
j	a_j	$h(a_j)$	$f(a_j)$	$100 \% \cdot f(a_j)$
1	23	6	0,24	24 %
2	25	10	0,40	40 %
3	29	5	0,20	20 %
4	35	4	0,16	16 %
Σ	-	25	1	100 %

Alternativ zur Häufigkeitstabelle kann die Darstellung der absoluten und relativen Häufigkeitsverteilung auch *grafisch* in Form eines Piktogrammes, Stab- oder Säulendiagrammes, Häufigkeitspolygons oder Kreissektorendiagrammes erfolgen. Auch wenn sich in der Praxis überwiegend das Säulendiagramm durchgesetzt hat, wollen wir im Folgenden dennoch alle Möglichkeiten kurz betrachten:

1. Piktogramm

Wie die Diagrammbezeichnung schon verrät, erfolgt eine Darstellung der Häufigkeiten durch unterschiedliche Anzahl oder Größe von Bildsymbolen.

2. Stabdiagramm

Ein Stabdiagramm (höhenproportionale Darstellung) entsteht durch Eintragung von Punkten $P(a_i, h(a_i))$ oder $P(a_i, f(a_i))$ in ein Koordinatensystem, wobei jeder Punkt durch eine Senkrechte mit der x-Achse bzw. a_i -Achse verbunden wird. Die Längen dieser Senkrechten (Stäbe) entsprechen daher den absoluten bzw. relativen Häufigkeiten. Werden relative Häufigkeiten abgetragen, so muss die Gesamtlänge aller Stäbe zusammen gleich 1 sein. Wir können daher bei Stabdiagrammen für relative Häufigkeiten immer den gleichen Maßstab auf der y-Achse wählen. Dies bietet sich insbesondere beim Vergleich der Ergebnisse mehrerer Stichproben an.

3. Häufigkeitspolygon

Wie bei der Konstruktion eines Stabdiagrammes werden hier zunächst entweder die Punkte $P(a_i, h(a_i))$ oder $P(a_i, f(a_i))$ in ein Koordinatensystem eingetragen. Im Anschluss daran werden diese durch Geraden verbunden.

4. Säulendiagramm

Im Vergleich zum Stabdiagramm werden beim Säulendiagramm (flächenproportionale Darstellung) die Häufigkeiten durch Flächen (Rechtecke) und nicht durch Strecken abgebildet. Um eine Verwechslung mit einem Histogramm (vgl. Abschnitt I 2.1.3) zu vermeiden, ist zu beachten, dass die Rechtecke nicht aneinanderstoßen.

5. Kreissektorendiagramm

Hier wird zu jeder Merkmalsausprägung a_i ein Kreissektor gebildet, wobei die Flächen der Sektoren und damit auch die Innenwinkel proportional zu den Häufigkeiten gewählt werden. Dazu muss für den Innenwinkel α_i des j -ten Sektors

$$\alpha_i = 360^\circ \cdot f(a_i) \quad (\text{I2: 6})$$

gelten. Kreissektorendiagramme eignen sich besonders gut für nominalskalierte Merkmale, da hier die Merkmalsausprägungen nicht auf einem Zahlenstrahl abgetragen werden müssen und somit aufgrund ihrer Anordnung nicht fälschlicherweise eine Rangfolge vermutet wird.

Beispiel I 2-2

Grafische Darstellung einer Häufigkeitsverteilung



Eine Untersuchung über die Industriesektorzugehörigkeit von Unternehmen in einer bestimmten Region lieferte das Ergebnis, dass 25 % der Unternehmen zum primären (P), 25 % zum sekundären (S) und 50 % zum tertiären Sektor (T) zählen. Auch wenn bei einem derartig einfach überschaubaren Resultat eigentlich keine grafische Darstellung erforderlich ist, wollen wir dieses einfache Beispiel im Folgenden dazu nutzen, die vorhergehend beschriebenen Darstellungsformen zu veranschaulichen:

1. Piktogramm:

- ⊕ ⊕ ⊕ ⊕ ⊕
- ⊕ ⊕ ⊕ ⊕ ⊕
- ⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕ ⊕

Primärer Sektor

Sekundärer Sektor

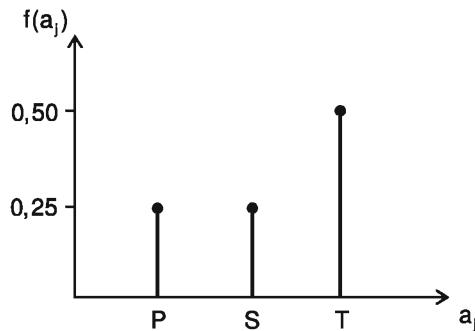
Tertiärer Sektor

Zur Konstruktion von Stab- und Säulendiagramm sowie des Häufigkeitspolygons im Koordinatensystem benötigen wir die Punkte

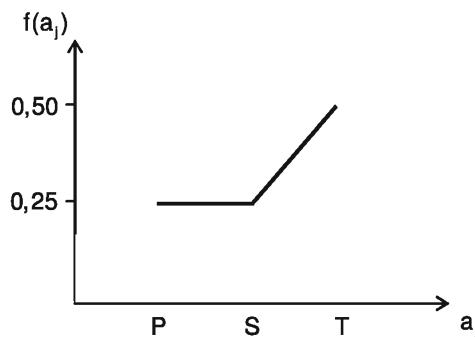
$$P_1(P; 0,25) \quad P_2(S; 0,25) \quad P_3(T; 0,50).$$

Wir erkennen daran, dass in derartigen grafischen Darstellungen nicht zwangsläufig Zahlenwerte auf der x-Achse abgetragen werden müssen.

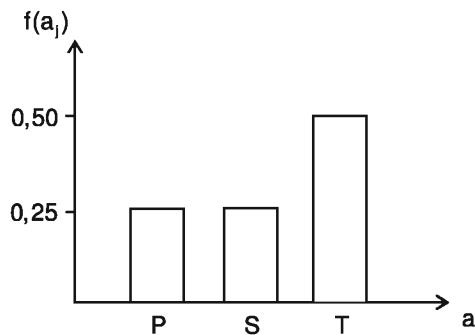
2. Stabdiagramm:



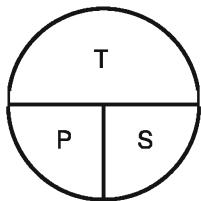
3. Häufigkeitspolygon:



4. Säulendiagramm:



5. Kreissektordiagramm:



Winkelberechnung:

$$\alpha_1 = 360^\circ \cdot f(a_1) = 360^\circ \cdot 0,25 = 90^\circ$$

$$\alpha_2 = 360^\circ \cdot f(a_2) = 360^\circ \cdot 0,25 = 90^\circ$$

$$\alpha_3 = 360^\circ \cdot f(a_3) = 360^\circ \cdot 0,50 = 180^\circ$$

2.1.2 Empirische Verteilungsfunktion bei diskreten Merkmalen

Ist ein Merkmal X *ordinal- oder metrisch skaliert*, dann unterliegen seine Merkmalsausprägungen a_j einer *Rangordnung*. Es ist hier (nur hier) sinnvoll, die absoluten und relativen Häufigkeiten in aufsteigender Reihenfolge der a_j zu summieren.

Durch fortlaufende Aufsummierung (Kumulierung) der absoluten Häufigkeiten einer Häufigkeitsverteilung mit $a_1 < a_2 < \dots < a_k$ lassen sich sog. *absolute Summenhäufigkeiten* oder *kumulierte absolute Häufigkeiten*

$$H(x) = H(X \leq x) = \sum_{a_j \leq x} h(a_j) \quad (\text{I2: 7})$$

ermitteln, wobei $H(x)$ die Anzahl jener Merkmalsträger angibt, welche einen Merkmalswert besitzen, der höchstens (kleiner oder gleich) x beträgt. Falls x genau einem der a_j entspricht, sind also lediglich die absoluten Häufigkeiten bis zu dieser Ausprägung a_j aufzusummieren, d.h. $H(a_j) = h(a_1) + h(a_2) + \dots + h(a_j)$. Die ungewöhnliche Summenschreibweise in (I2: 7) wird verwendet, da x auch Werte annehmen kann, für die kein a_j existiert. In einem solchen Fall werden die absoluten Häufigkeiten aufsummiert, die zu Merkmalswerten a_j gehören, die kleiner als x sind.

Analog lassen sich sog. *relative Summenhäufigkeiten* oder *kumulierte relative Häufigkeiten* ermitteln. Für diese gilt

$$F(x) = F(X \leq x) = \sum_{a_j \leq x} f(a_j), \quad (\text{I2: 8})$$

wobei $F(x)$ den Anteil der Merkmalsträger angibt, welche einen Merkmalswert aufweisen, der höchstens x beträgt. Alternativ kann die Bestimmung von $F(x)$ auch über den Zusammenhang

$$F(x) = \frac{H(x)}{n} \quad (\text{I2: 9})$$

erfolgen. Wir bezeichnen (I2: 8) auch als **empirische Verteilungsfunktion**. Da für die Häufigkeitsverteilung $a_1 < a_2 < \dots < a_k$ gilt, ist der Funktionswert für ein jedes x , welches kleiner als a_1 ausfällt, unweigerlich Null. Es gibt schließlich keine Merkmalsträger, die eine Merkmalsausprägung aufweisen, die kleiner als a_1 sind, da dies bereits der kleinste Wert im Datensatz ist. Der Funktionswert für ein jedes beliebiges x , das über dem größten Merkmalswert a_k liegt, ist gleich 1. Der Anteil der-

jeniger Merkmalsträger, die eine Merkmalsausprägung von weniger als x aufweisen, ist schließlich 100 %. Wir können also zusammenfassend festhalten, dass die empirische Verteilungsfunktion beschränkt ist, d.h. $0 \leq F(x) \leq 1$ gilt. Wegen $f(a_j) \geq 0$ gilt außerdem $F(a_1) \leq \dots \leq F(a_j) \leq \dots \leq F(a_n)$, d.h. die empirische Verteilungsfunktion ist monoton steigend. Wie wir im Beispiel I 2-3 sehen werden, hat sie außerdem grafisch die Gestalt einer **Treppenfunktion**, welche bei jedem a_j eine Sprungstelle mit einem Sprung in der Höhe von $f(a_j)$ aufweist.

Bei vorliegenden kumulierten absoluten oder relativen Häufigkeiten können wir natürlich unmittelbar wieder auf die Einzelhäufigkeiten schließen. Es gelten nämlich die Zusammenhänge

$$h(a_j) = H(a_j) - H(a_{j-1}) \quad (\text{I2: 10})$$

und

$$f(a_j) = F(a_j) - F(a_{j-1}). \quad (\text{I2: 11})$$

Beispiel I 2-3

Empirische Verteilungsfunktion

In 100 Filialen einer Buchhandlung wurden an einem bestimmten Tag entweder 0, 1, 2 oder 3 Exemplare (Ausprägungen des Merkmals Absatzzahl) eines bestimmten Buches verkauft. Die resultierende Häufigkeitstabelle, die uns die Häufigkeit (absolut und relativ) der Filialen angibt, in denen es zu den jeweiligen Verkaufszahlen kam, zeigt sich wie folgt:

j	a_j	$h(a_j)$	$f(a_j)$
1	0	10	0,10
2	1	40	0,40
3	2	20	0,20
4	3	30	0,30
-	-	100	1,00

Ermitteln wir nun zunächst die absoluten und relativen kumulierten Häufigkeiten:

j	a_j	$h(a_j)$	$H(a_j)$	← Berechnung
1	0	10	10	$h(0)$
2	1	40	50	$h(0) + h(1)$
3	2	20	70	$h(0) + h(1) + h(2)$
4	3	30	100	$h(0) + h(1) + h(2) + h(3)$
-	-	100	-	-

j	a_j	$f(a_j)$	$F(a_j)$	← Berechnung
1	0	0,10	0,10	$f(0)$
2	1	0,40	0,50	$f(0) + f(1)$
3	2	0,20	0,70	$f(0) + f(1) + f(2)$
4	3	0,30	1,00	$f(0) + f(1) + f(2) + f(3)$
-	-	1,00	-	-

Zur besseren Veranschaulichung der Berechnung einzelner kumulierter Häufigkeiten wollen wir diese nun noch einmal für den Wert $F(2)$ im Detail darlegen:

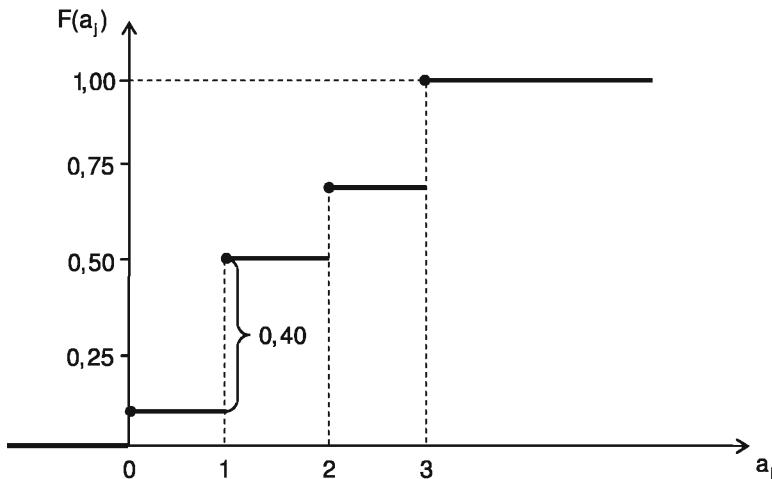
$$F(2) = f(0) + f(1) + f(2) = 0,1 + 0,4 + 0,2 = 0,7 \quad \text{oder alternativ} \quad F(2) = \frac{H(2)}{n} = \frac{70}{100} = 0,7$$



Soll nun beispielsweise der Wert von $f(1)$ aus den relativen kumulierten Häufigkeiten errechnet werden, kann dies wie folgt geschehen:

$$f(1) = F(1) - F(0) = 0,50 - 0,10 = 0,40$$

Wie wir gleich sehen werden, lässt sich dieser Wert auch aus der grafischen Darstellung der empirischen Verteilungsfunktion ablesen, da in dieser die Sprunghöhe zwischen einzelnen Treppenstufen an einer Stelle a_j jeweils den zugehörigen Wert $f(a_j)$ angibt. Für die grafische Konstruktion der empirischen Verteilungsfunktion tragen wir lediglich die Punkte $P(a_j, F(a_j))$ in ein Koordinatensystem ein und versehen diese mit Linienzügen:



Auch die konkrete Funktionsschreibweise der empirischen Verteilungsfunktion können wir angeben. Außerdem können wir das in absoluten kumulierten Häufigkeiten formulierte Gegenstück zur empirischen Verteilungsfunktion bestimmen:

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0,10 & " \quad 0 \leq x < 1 \\ 0,50 & " \quad 1 \leq x < 2 \\ 0,70 & " \quad 2 \leq x < 3 \\ 1,00 & " \quad x \geq 3 \end{cases} \quad H(x) = \begin{cases} 0 & \text{für } x < 0 \\ 10 & " \quad 0 \leq x < 1 \\ 50 & " \quad 1 \leq x < 2 \\ 70 & " \quad 2 \leq x < 3 \\ 100 & " \quad x \geq 3 \end{cases}$$

Beantworten wir nun mittels dieser beiden Funktionen folgende Fragen:

- a) In wie viel Prozent der Filialen wurden höchstens 2 Bücher verkauft?
Die Antwort auf diese Fragestellung kann direkt als $F(2) = 0,70$ bzw. als 70 % angegeben werden.
- b) In wie vielen Filialen wurde mindestens 1 Buch verkauft?
Hier können wir die Beziehung $n - H(0) = 100 - 10 = 90$ nutzen, d.h. wir subtrahieren von der Gesamtzahl der Filialen die Anzahl derjeniger Filialen, in denen kein Buch verkauft wurde.

2.1.3 Klassierte Häufigkeitsverteilung bei stetigen Merkmalen

Liegt ein *stetiges* (oder quasistetiges) *Merkmal X* vor, so taucht das Problem auf, dass die Elemente der Urliste häufig alle voneinander verschieden sind. Das Aufstellen einer "regulären" Häufigkeitstabelle des im Abschnitt I 2.1.1 vorgestellten Typs bietet hier keinen Informationsgewinn, da die meisten Merkmalsausprägungen nur einmal vorkommen. Es empfiehlt sich in einem solchen Fall, die Ausprägungen von X in der Stichprobe in k **Klassen** K_j ($j = 1, 2, \dots, k$) einzuteilen und die zugehörigen Klassenhäufigkeiten zu ermitteln. Die **absolute Klassenhäufigkeit** h_j ist dabei die Anzahl derjeniger Elemente, deren Merkmalswerte in die Klasse K_j fallen. Wir können also sagen, dass hier die Klassen K_j die Rolle der a_j aus der "regulären" Häufigkeitstabelle übernehmen. Die **relativen Klassenhäufigkeiten** f_j ergeben sich wieder über

$$f_j = \frac{h_j}{n} \quad (\text{I2: 12})$$

und weisen damit die gleichen Eigenschaften wie in (I2: 4) und (I2: 5) auf. Auch für die h_j gelten (I2: 1) und (I2: 2) analog. Zudem sei erwähnt, dass wir auch *kumulierte Klassenhäufigkeiten* gemäß der unter I 2.1.2 beschriebenen Vorgehensweise bestimmen können.

Wieviele Klassen sinnvollerweise gebildet werden sollten, kann aus DIN 55302 (Blatt 1) entnommen werden. Hier finden sich unter anderem die Empfehlungen bei Stichprobenumfängen n von 100, 1.000 und 10.000 eine Anzahl von 10, 13 bzw. 16 Klassen zu wählen. Natürlich kann aus Gründen der Übersichtlichkeit der Darstellung auch bereits bei geringerem n eine Klassierung vorgenommen werden. Außerdem kann aufgrund des Untersuchungsobjekts auch eine andere Klassenanzahl geeigneter erscheinen.

Um Probleme bei der Zuordnung einzelner Merkmalswerte zu den Klassen zu vermeiden, empfiehlt es sich die Klassen K_j in der Form $a_j^u < x \leq a_j^o$ bzw. $[a_j^u, a_j^o]$ zu definieren. Jede Klasse K_j ist damit gekennzeichnet durch eine **Klassenuntergrenze** a_j^u und eine **Klassenobergrenze** a_j^o . Da die einzelnen Klassen direkt ineinander übergehen, entspricht bei k Klassen die Untergrenze einer Klasse jeweils der Obergrenze der vorhergehenden Klasse.

$$a_j^u = a_{j-1}^o \quad (j = 2, 3, \dots, k) \quad (\text{I2: 13})$$

Die Untergrenze der ersten Klasse können wir anhand des kleinsten Merkmalswerts der Stichprobe wählen. Häufig wird sie auch auf Null gesetzt. Die letzte Klasse K_k besitzt häufig eine Form $a_k^u < x < \infty$ bzw. $[a_k^u, \infty]$ und wird als *offene Randklasse* bezeichnet. Dies bedeutet, dass in diese Klasse alle Merkmalsausprägungen fallen, die größer als a_k^u sind.

Die sog. **Klassenbreite** Δa_j der j-ten Klasse, die nach Möglichkeit durchgehend einheitlich (*konstant*) gestaltet werden sollte, ist (außer für offene Randklassen) als

$$\Delta a_j = a_j^o - a_j^u \quad (j = 1, 2, \dots, k) \quad (\text{I2: 14})$$

definiert. Bei der Wahl konstanter Klassenbreiten sind jedoch häufig manche Klassen sehr schwach besetzt. Dies trifft vor allem auf die Randklassen zu. Dann ist es evtl. sinnvoll, diese Klassen breiter zu wählen.

Besonders für die im Kapitel I 2.2 behandelnden Messzahlen ist es von entscheidender Wichtigkeit, jeder Klasse K_j einen repräsentativen Zahlenwert zuzuordnen. Dazu bietet sich die **Klassenmitte** a_j^m an, für welche (außer bei offenen Randklassen) folgendes gilt:

$$a_j^m = \frac{1}{2} \cdot (a_j^u + a_j^o) \quad (j=1,2,\dots,k) \quad (\text{I2: 15})$$

Die Kombination der Klassen K_j mit ihren Klassenhäufigkeiten h_j bzw. f_j bildet die **klassierte Häufigkeitsverteilung** des Merkmals X. Wieder haben wir die Möglichkeit einer tabellarischen Darstellung. Abbildung I 6 veranschaulicht eine typische **klassierte Häufigkeitstabelle**.

Klassennummer	Klasse K_j	Klassenbreite	Klassenmitte	absolute Klassenhäufigkeit	relative Klassenhäufigkeit
j	über a_j^u bis a_j^o	Δa_j	a_j^m	h_j	f_j
1	$[a_1^u, a_1^o]$	Δa_1	a_1^m	h_1	f_1
2	$[a_1^o, a_2^o]$	Δa_2	a_2^m	h_2	f_2
...
k	$[a_{k-1}^o, a_k^o]$	Δa_k	a_k^m	h_k	f_k
Σ	-	-	-	n	1

Abbildung I 6: Häufigkeitstabelle bei klassierten Daten

Beispiel I 2-4

Häufigkeitstabelle bei klassierten Daten



Nehmen wir an, eine Erhebung in 47 amerikanischen Staaten ergab folgende Anzahlen polizeilich gemeldeter Straftaten bezogen auf 100.000 Einwohner:

79,3 163,4 57,8 196,1 123,5 68,1 96,3 155,8 85,3 70,9 167,4 84,8 51,2 66,4 79,1
94,3 53,1 92,4 75,1 122,5 74,8 43,3 121,9 96,5 52,7 199,1 34,2 121,6 104,5 69,3
37,7 75,9 107,2 92,4 65,1 127,6 83,3 56,5 82,8 115,8 88,1 54,2 82,4 103,0 44,5
50,8 84,9

Eine erste Untersuchung des Datenmaterials zeigt, dass die kleinste Kriminalitätsrate bei 34,2 und die größte bei 199,1 liegt. Die Mehrzahl der Staaten verzeichnete Kriminalitätsraten im Bereich von 50 bis 120. Um sich einen genaueren Überblick zu verschaffen, bietet sich hier die Aufstellung einer klassierten Häufigkeitstabelle mit Klassenbreiten von 25 an. Da dabei die entstehende untere Randklasse jedoch nicht besetzt wäre, verdoppeln wir diese und erhalten damit folgende Tabelle (Achtung: Rundungsungenauigkeiten bei relativen Häufigkeiten!):

Nr.	Klasse K_j	Klassenhäufigkeit		Kumulierte Klassenhäufigkeit		Klassenbreite Δa_j	Klassenmitte a_j^m
		absolut	relativ	absolut	relativ		
j	über a_j^u bis a_j^o	h_j	f_j				
1	0 - 50	4	0,09	4	0,09	50	25
2	50 - 75	13	0,28	17	0,37	25	62,5
3	75 - 100	16	0,34	33	0,71	25	87,5
4	100 - 125	8	0,17	41	0,88	25	112,5
5	125 - 150	1	0,02	42	0,90	25	137,5
6	> 150	5	0,11	47	1,00	-	-
Σ	-	47	1,00	-	-	-	-

Die *grafische Darstellung* einer Häufigkeitsverteilung klassierter Daten erfolgt durch ein sog. **Histogramm**. Als Darstellungsmittel dienen hier Rechtecke, die über den jeweiligen x-Achsen-Abschnitten der Klassen gezeichnet werden. Die Grundlinien der Rechtecke entsprechen somit den Klassenbreiten. Anders als beim Säulendiagramm für unklassierte Daten ist es hier nicht sinnvoll, die Höhe der Rechtecke als Maß für die Häufigkeiten zu wählen. Bei ungleichen Klassenbreiten entstünde ein falscher Eindruck von der Häufigkeitsverteilung.

Der Betrachter eines Histogramms orientiert sich beim Vergleich der Größe zweier Rechtecke nicht an ihrer Höhe, sondern an ihrem Flächeninhalt. Die Höhe der Rechtecke ist deshalb so zu wählen, dass ihr jeweiliger Flächeninhalt proportional zur Klassenhäufigkeit ist. Dies wird erreicht, indem als Höhe der Rechtecke die sog. *normierten absoluten oder normierten relativen Klassenhäufigkeiten* h_j^* bzw. f_j^* verwendet werden. Diese ergeben sich zu

$$h_j^* = \frac{h_j}{\Delta a_j} \quad \text{bzw.} \quad f_j^* = \frac{f_j}{\Delta a_j} \quad (j = 1, 2, \dots, k). \quad (\text{I2: 16})$$

Liegen einem Histogramm normierte absolute Klassenhäufigkeiten zugrunde, ist die Gesamtfläche F aller Rechtecke gleich n, im Falle normierter relativer Klassenhäufigkeiten ist sie gleich 1. Es gilt nämlich

$$F = \sum_{j=1}^k h_j^* \cdot \Delta a_j = \sum_{j=1}^k \frac{h_j}{\Delta a_j} \cdot \Delta a_j = n \quad \text{bzw.} \quad F = \sum_{j=1}^k f_j^* \cdot \Delta a_j = \sum_{j=1}^k \frac{f_j}{\Delta a_j} \cdot \Delta a_j = 1.$$

Da im Histogramm relativer Häufigkeiten unabhängig vom Stichprobenumfang n die Fläche immer auf 1 normiert ist, eignet sich dieses insbesondere beim Vergleich verschiedener Datensätze mit unterschiedlichen Stichprobenumfängen. Wir bezeichnen die normierten relativen Klassenhäufigkeiten auch als *Häufigkeitsdichten*.

Beispiel I 2-5

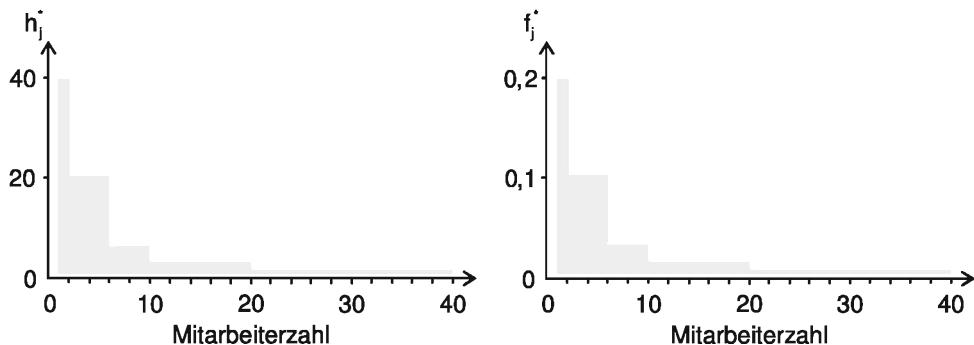
Histogramm

Für 200 Unternehmen einer Region wurden jeweils die Mitarbeiterzahlen erfasst, dafür Klassen unterschiedlicher Breiten gebildet und die jeweiligen Häufigkeiten ausgezählt. Die Ergebnisse sind in nachfolgender Tabelle zu sehen. Darin sind außerdem bereits die normierten absoluten und relativen Klassenhäufigkeiten berechnet, die wir zur Konstruktion der Histogramme benötigen. Die Berechnungen bereiten hier keine sonderlichen Schwie-

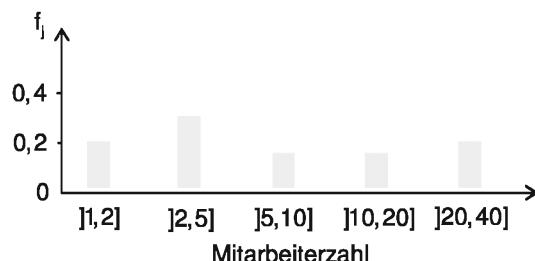
rigkeiten. Wäre jedoch eine offene Randklasse enthalten, so könnten wir keine Klassenbreite und damit auch keine normierten Klassenhäufigkeiten für unser Histogramm bestimmen. In einem solchen Fall behelfen wir uns in der Praxis meist damit, einen künstlichen Höchstwert für die Klasse festzulegen, der sich am größten Wert der Stichprobe orientieren kann.

Nr.	Klasse K_j	Klassenhäufigkeit		Klassenbreite Δa_j	Normierte Klassenhäufigkeit	
		absolut	relativ		absolut	relativ
j	über a_j^u bis a_j^o	h_j	f_j		h_j^*	f_j^*
1	1 - 2	40	0,20	1	40	0,200
2	2 - 5	60	0,30	3	20	0,100
3	5 - 10	30	0,15	5	6	0,030
4	10 - 20	30	0,15	10	3	0,015
5	20 - 40	60	0,20	20	2	0,100
Σ	-	200	1,00	-	-	-

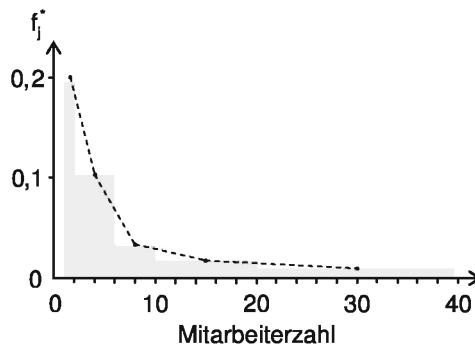
Wir erhalten aus diesem Datenmaterial folgende Histogramme:



Hätten wir nur ein Säulendiagramm konstruiert, welches als Säulenhöhen die nichtnormierten relativen Häufigkeiten verwendet, so würden wir ein völlig anderes Bild von der Häufigkeitsverteilung vermitteln:



In der Praxis werden Histogramme gelegentlich auf sog. *Häufigkeitspolygone* reduziert. Solche erhalten wir durch Verbinden der Mittelpunkte der oberen Säulenseiten. Sie kommen insbesondere dann zum Einsatz, wenn in einer Grafik verschiedene Häufigkeitsverteilungen miteinander verglichen werden sollen. Im hier vorliegenden Fall hätte das Häufigkeitspolygon folgende Gestalt:



Zur Interpretation eines Histogramms:

Im vorliegenden Beispiel ist eine recht einfache Interpretation des Histogramms möglich. Das zugrunde liegende Merkmal X (Mitarbeiteranzahl) ist *diskret* und kann nur *natürliche Zahlen* $1, 2, 3, \dots$ als Ausprägungen annehmen. Wählen wir auf der x -Achse des Histogramms einen beliebigen ganzzahligen Wert wie z.B. 15, kann auf der y -Achse eine dazu gehörige Häufigkeitsdichte abgelesen werden. Diese ist interpretierbar als diejenige Häufigkeit, die wir beim Auszählen der unklassierten Daten erhalten hätten, wenn die beobachteten Werte innerhalb der Klassen vollständig gleichmäßig verteilt wären. Eine solche Interpretation ist allerdings unzulässig, wenn ein Merkmal vorliegt, das auch nichtganzzahlige Werte annehmen kann. Dies gilt insbesondere für stetige Merkmale. Auch wenn wir ein diskretes Merkmal mit nur natürlichen Zahlen als Ausprägung vorliegen haben, werden wir in der Praxis von der hier geschilderten Interpretation Abstand halten, da in der Mehrzahl der Fälle nicht davon ausgegangen werden kann, dass eine Gleichverteilung innerhalb der Klassen gegeben ist.

Analog zu "regulären" Häufigkeitsverteilungen können wir auch für klassierte Daten eine *empirische Verteilungsfunktion* bestimmen und skizzieren. Es übernehmen ja, wie bereits angemerkt, die Klassen K_j die Rolle der a_i .

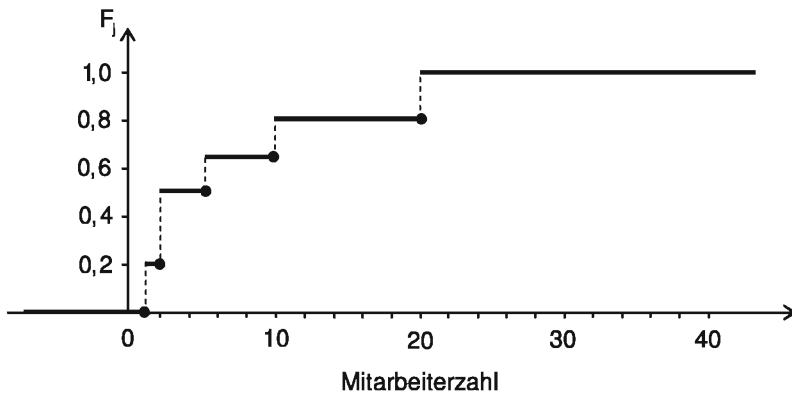
Beispiel I 2-6

Empirische Verteilungsfunktion bei klassierten Daten

Skizzieren wir für unser Beispiel I 2-5 die empirische Verteilungsfunktion. Wir benötigen dazu zunächst die kumulierten relativen Klassenhäufigkeiten F_j , die wir in der folgenden Tabelle bestimmen:

j	über a_j^u bis a_j^o	f_j	F_j
1	1 - 2	0,20	0,20
2	2 - 5	0,30	0,50
3	5 - 10	0,15	0,65
4	10 - 20	0,15	0,80
5	20 - 40	0,20	1,00
Σ	-	1,00	

Die empirische Verteilungsfunktion erhalten wir damit grafisch wie folgt. Die Sprunghöhen sind nun relative Klassenhäufigkeiten f_j .

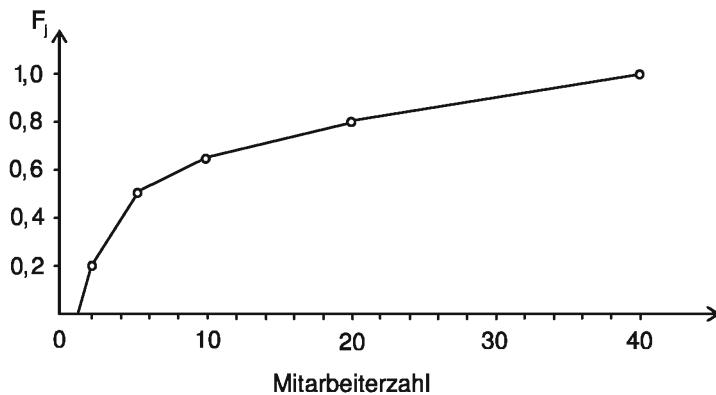


In der Literatur findet sich außerdem gelegentlich der Vorschlag die empirische Verteilungsfunktion klassierter Daten durch ein sog. *Summenpolygon* abzubilden. Dazu werden die Punkte $P(a_i^o, F_j)$, d.h. die Klassenobergrenzen mit zugehörigem kumulierten relativen Klassenhäufigkeitswert, in ein Koordinatensystem eingetragen. Unter der Annahme, dass die Elemente innerhalb einer Klasse gleichmäßig verteilt sind, dürfen diese Punkte linear miteinander verbunden werden. Der so entstehende stetige Graf heißt schließlich Summenpolygon.

Beispiel I 2-7

Summenpolygon

Das Summenpolygon zu Beispiel I 2-5 bzw. I 2-6 hat folgende Gestalt:



2.1.4 Typische Häufigkeitsverteilungen

Bei der Beschreibung von Datenmaterial mittels tabellarischer oder grafischer Hilfsmittel finden sich immer wieder typische Formen von Häufigkeitsverteilungen. Wichtige Charakteristika dieser Strukturen sind z.B. die Anzahl der Gipfel, die Symmetrie und die Streuung. Wir unterscheiden

- nach Anzahl der Gipfel:
 - Verteilungen ohne Gipfel (z.B. Gleichverteilung)
 - Verteilungen mit einem Gipfel (unimodale Verteilungen)
 - Verteilungen mit mehreren Gipfeln (multimodale Verteilungen)
- nach Symmetrie:
 - symmetrische Verteilungen (z.B. unimodal-symmetrisch, U-förmig)
 - unsymmetrische Verteilungen (z.B. linksschief, rechtsschief)
- nach Streuung:
 - geringe Streuung (Daten konzentrieren sich um ein Datenzentrum)
 - große Streuung (Daten streuen sehr stark um ein Zentrum)

Diese Charakteristika können wir grafisch gut mittels Stab- und Säulendiagrammen oder Histogrammen (vgl. Abbildung I 7) veranschaulichen. Sie können aber auch zahlenmäßig mit Hilfe von Maßzahlen beschrieben werden. Dies ist Thema von Abschnitt I 2.2.

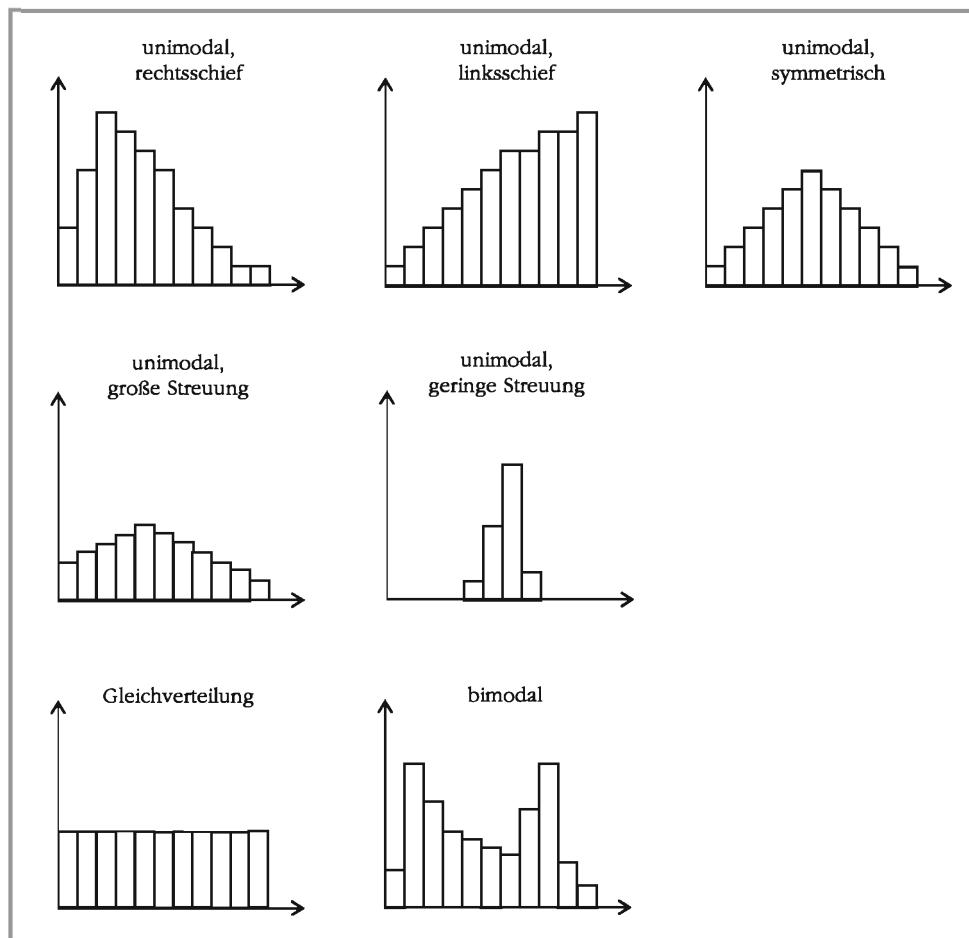


Abbildung I 7: Histogramme typischer Häufigkeitsverteilungen

2.1.5 Quantile

In diesem Abschnitt wollen wir kurz den Begriff des *Quantils* (auch als *Fraktil* bezeichnet) einführen. Wir werden im Abschnitt I 2.2 sehen, dass wir Quantile sowohl zur Beschreibung der Lage als auch der Streuung von Häufigkeitsverteilungen einsetzen können. Aufgrund ihrer hohen praktischen Bedeutung wollen wir ihnen jedoch bereits an dieser Stelle einen eigenen Abschnitt widmen.

Liegt eine der Größe nach aufsteigend sortierte *Urliste* ($x_1 < x_2 < \dots < x_n$) und ein vorgegebener Anteilssatz α mit $0 < \alpha < 1$ vor, so bezeichnen wir eine Zahl x_α als α -Quantil oder $(100 \cdot \alpha)\%$ -Quantil, wenn mindestens $(100 \cdot \alpha)\%$ der Werte der Urliste kleiner oder gleich und mindestens $100 \cdot (1 - \alpha)\%$ größer oder gleich x_α sind.

Ist das Produkt $\alpha \cdot n$ aus dem vorgegebenen Anteilssatz und der Anzahl der Elemente der Urliste *nicht ganzzahlig*, so ist das α -Quantil der Beobachtungswert, der auf den $\alpha \cdot n$ -ten folgt. Ist $\alpha \cdot n$ hingegen *ganzzahlig*, besteht in der Literatur Uneinigkeit darüber, ob der $\alpha \cdot n$ -te Beobachtungswert (*Variante A*) oder der Mittelwert aus der $\alpha \cdot n$ -ten und der $(\alpha \cdot n + 1)$ -ten Beobachtung der Urliste (*Variante B*) als α -Quantil gelten soll. Vorteil von Variante B ist, dass ein so ermittelter Zahlenwert die Urliste exakt im Verhältnis α zu $1 - \alpha$ teilt. Als nachteilig wird jedoch zumeist angesehen, dass ein nach Variante B ermittelter Wert in der Urliste gar nicht existiert. Soll ein echter Beobachtungswert als Quantil dienen, ist auf Variante A zurückzugreifen, wobei das genannte exakte Teilungsverhältnis hiermit aber nicht erreicht wird.

Je nach Ausprägung von α haben sich in der Literatur verschiedene Bezeichnungen und auch Symboliken für Quantile eingebürgert:

- **Quartile:** $Q_1 = x_{0,25}$, $Q_2 = x_{0,50}$, $Q_3 = x_{0,75}$
- **Dezile:** $x_{0,10}$, $x_{0,20}$, ...
- **Perzentile:** $x_{0,01}$, $x_{0,02}$, ...

Die bedeutendsten Quantile sind die *Quartile*. Wir bezeichnen Q_1 als unteres oder erstes, Q_2 als mittleres oder zweites und Q_3 als oberes oder drittes Quartil. Quartile teilen die Daten in vier etwa gleichgroße Teilgesamtheiten ein. Betrachten wir dazu das Beispiel in Abbildung I 8.

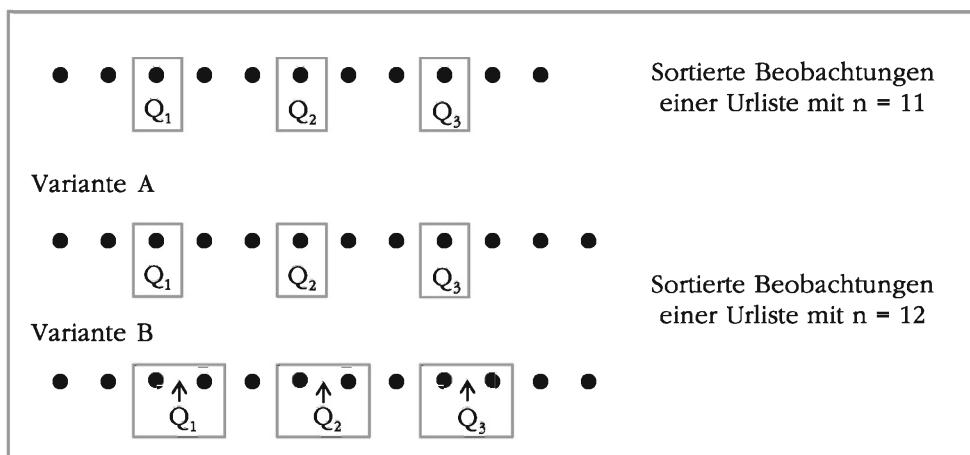


Abbildung I 8: Quartile in der Urliste

Im Falle der *ungeraden Urliste* mit $n = 11$ ist das Produkt aus n und $0,25, 0,50$ bzw. $0,75$ jeweils nicht ganzzahlig, sodass wir den dritten, sechsten und neunten Beobachtungswert als Q_1 , Q_2 und Q_3 erhalten. $3/11 = 27,27\%$ der Beobachtungen sind kleiner oder gleich Q_1 und $9/11 = 81,81\%$ sind größer oder gleich Q_3 . Analoges gilt für Q_2 und Q_3 . In der *geraden Urliste* mit $n = 12$ erhalten wir nach Variante A die Quantile Q_1 , Q_2 und Q_3 als den dritten, sechsten und neunten Beobachtungswert. $3/12 = 25\%$ der Beobachtungen sind kleiner oder gleich Q_1 und $10/12 = 83,33\%$ sind größer oder gleich Q_3 . Ähnliches gilt für Q_2 und Q_3 . Nach Variante B erhalten wir Q_1 hingegen als Mittelwert aus der dritten und vierten Beobachtung, sodass $3/12 = 25\%$ links und $9/12 = 75\%$ der Beobachtungen neben diesem gemittelten Wert liegen. Er existiert zwar nicht als Beobachtung in der Urliste, erlaubt jedoch eine exakte 25-75-Einteilung der Urliste. Ähnliche Aussagen können wir auch über die nach Variante B bestimmten Quartile Q_2 und Q_3 machen.

Beispiel I 2-8

Quartile aus Urlisten

Die Auszählung der Kassenbestände in 11 Filialen einer Supermarktkette ergab folgende gerundete und bereits der Größe nach sortierte Euro-Werte:

420 430 440 440 500 550 700 700 800 950 990

Bestimmen wir das 25 %-, 50 %- und 75 %-Quantil bzw. das untere (Q_1), mittlere (Q_2) und obere (Q_3) Quartil aus dieser Urliste. Da aufgrund von $n = 11$ das Produkt $\alpha \cdot n$ hier keine ganze Zahl ergeben kann, runden wir $\alpha \cdot n$ zur nächsten ganzen Zahl auf. Wir erhalten:

- Q_1 ist der $11 \cdot 0,25 \approx 3$ -te Beobachtungswert, d.h. $Q_1 = 440$ €.
- Q_2 ist der $11 \cdot 0,50 \approx 6$ -te Beobachtungswert, d.h. $Q_2 = 550$ €.
- Q_3 ist der $11 \cdot 0,75 \approx 9$ -te Beobachtungswert, d.h. $Q_3 = 800$ €.

Bei der Ermittlung von Quantilen aus *Häufigkeitstabellen* ist minimal anders vorzugehen, wie folgendes Beispiel zeigt. Hier ist entweder "im Hinterkopf" die Urliste aus der Häufigkeitstabelle zu rekonstruieren oder die empirische Verteilungsfunktion zur Quantilsbestimmung heranzuziehen.

Beispiel I 2-9

Quartile aus Häufigkeitstabellen

Eine Fertigungsabteilung hat über einen Zeitraum von 30 Tagen die Anzahl von Ausschussteilen je Fertigungstag ermittelt. Dazu ergab sich folgende Häufigkeitstabelle:

j	a _j	Ausschussteile	Anzahl der Tage mit a _j Ausschussteilen h(a _j)	Quartile aus Häufigkeitstabellen	
				f(a _j)	F(a _j)
1	0		6	0,20	0,20
2	1		6	0,20	0,40
3	2		9	0,30	0,70
4	3		3	0,10	0,80
5	10		6	0,20	1,00
-	-		30	1,00	-

Ermitteln wir das 25 %-, 50 %- und 75 %-Quantil. Da für Q_1 und Q_3 sowohl $0,25 \cdot 30 = 7,5$ als auch $0,75 \cdot 30 = 22,5$ nicht ganzzahlig sind, bestimmen wir diese nach der beschriebenen Aufrundungsmethodik. Bei Q_2 haben wir die Wahl zwischen den Varianten A und B.

- Q_1 ist der $0,25 \cdot 30 \approx 8$ -te Beobachtungswert, d.h. $Q_1 = 1$ Stück.
- Q_2 ist entweder der $0,50 \cdot 30 = 15$ -te Beobachtungswert (Variante A), sodass $Q_2 = 2$ Stück, oder der Durchschnitt aus 15-tem und 16-tem Beobachtungswert (Variante B), sodass $Q_2 = (2 + 2)/2 = 2$ Stück. Es spielt hier also keine Rolle, welches Verfahren wir heranziehen.
- Q_3 ist der $0,75 \cdot 30 \approx 23$ -te Beobachtungswert, d.h. $Q_3 = 3$ Stück.

Diese Quantile können wir auch mittels der empirischen Verteilungsfunktion bestimmen, deren wesentliche Werte (die kumulierten relativen Häufigkeiten) in unserer Ausgangstabelle enthalten sind. Q_1 ist diejenige Ausprägung von a_j , für die $F(x)$ den Wert 0,25 erreicht bzw. überschreitet. Für Q_2 und Q_3 gilt selbiges für die Werte 0,50 und 0,75.

Im Falle *klassierter Daten* bedienen wir uns ebenfalls der empirischen Verteilungsfunktion zur Quantilsfindung. Als Ergebnis erhalten wir hier jedoch eine Klasse. Bei der Verwendung eines konkreten Wertes aus der Klasse als Quantil besteht wieder das Problem, dass das so bestimmte Quantil als Beobachtung in der Urliste vielleicht nicht existiert.

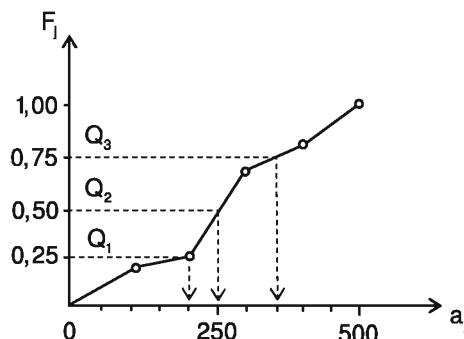
Beispiel I 2-10

Quartile aus klassierten Häufigkeitstabellen

Nehmen wir an, uns liegt der folgende Auszug aus einer klassierten Häufigkeitstabelle vor:

j	über a_j^u bis a_j^o	a_j^m	F_j
1	0 - 100	50	0,20
2	100 - 200	150	0,25
3	200 - 300	250	0,70
4	300 - 400	350	0,80
5	400 - 500	450	1,00

Da die kumulierten Klassenhäufigkeiten für Klasse 2 den Wert 0,25 erreichen, bestimmt diese das Quartil Q_1 . Wir können daher $Q_1 = 200$ angeben. Analog erhalten wir $Q_2 = 250$ und $Q_3 = 350$. Diese Quantilsfindung können wir anschaulich im Summenpolygon der klassierten Häufigkeitsverteilung skizzieren. Quartile werden darin dadurch gefunden, dass zunächst horizontale Linien bei 0,25, 0,50 und 0,75 eingezeichnet werden. Von den Schnittpunkten dieser Linien mit dem Summenpolygon werden dann Lote auf die x-Achse gefällt, um so das entsprechende Quartil ablesen zu können.



2.2 Maßzahlen

Neben Häufigkeitstabellen und Diagrammen können wir zur Beschreibung von Häufigkeitsverteilungen auf statistische Maßzahlen (Parameter) zurückgreifen. Sie bringen die wichtigsten Charakteristika einer Verteilung zum Ausdruck und lassen sich grob in Lage-, Streuungs-, Schiefe- und Konzentrationsparameter einteilen (vgl. Abbildung I 9). Da mit der Berechnung von Maßzahlen eine Verdichtung der Daten und damit auch ein Informationsverlust verbunden ist, erlauben einzelne Maßzahlen keinen Rückschluss auf den Typ der Verteilung. Sie werden erst kombiniert aussagefähig.

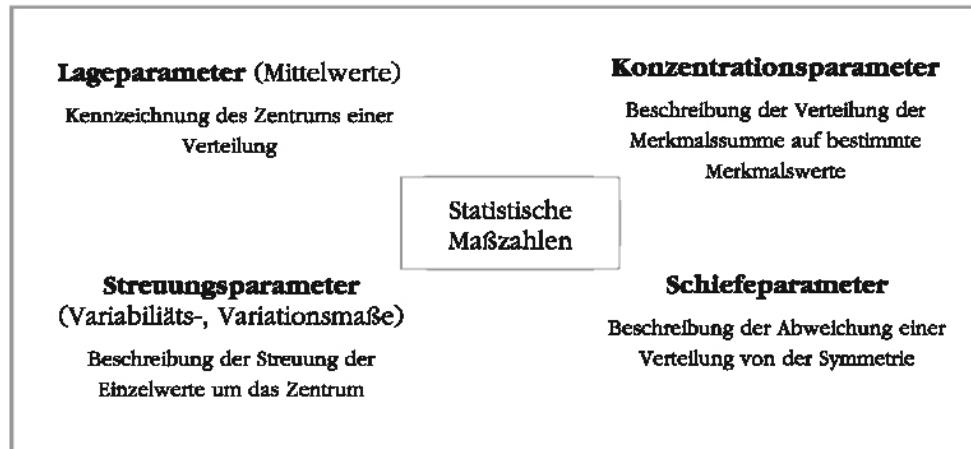


Abbildung I 9: Statistische Maßzahlen

Die Berechnung der Maßzahlen bzw. deren Berechnungsformeln unterscheiden sich je nach vorliegendem Datenmaterial (Urliste, Häufigkeitstabelle, Häufigkeitstabelle mit klassierten Daten). Für den Fall *klassierter Daten* sei bereits vorweg genommen, dass hier immer mit einem repräsentativen Wert für jede Klasse gerechnet wird. Dieser ist in der Regel die Klassenmitte. Da dies allerdings nur als Näherung angesehen werden kann, ist die Berechnung von Maßzahlen aus der Urliste oder der *unklassierten* Häufigkeitstabelle zu bevorzugen. Wir greifen auf eine derartige näherungsweise Berechnung in der Praxis in der Regel nur zurück, wenn uns *nur* eine klassierte Häufigkeitstabelle (z.B. in statistischen Veröffentlichungen) und weder Urliste noch unklassierte Häufigkeitstabelle vorliegen.

2.2.1 Lageparameter

Lageparameter dienen der Kennzeichnung des Zentrums einer Verteilung durch eine einzige Zahl. Gerade diese Komprimierung ermöglicht den Vergleich zweier oder mehrerer Gruppen bezüglich eines Merkmals (z.B. "In welchem von zwei Betrieben wird durchschnittlich besser verdient?").

Im Folgenden werden der Modus, der Median, das arithmetische Mittel und das geometrische Mittel näher betrachtet. Es ist zu beachten, dass die Berechnung die-



ser Lageparameter von der Skalierung der zugrunde liegenden Daten abhängt. Abbildung I 10 zeigt deutlich, dass gewisse Lageparameter nur bei bestimmten Skalierungen angewendet werden dürfen. So ist das geometrische Mittel beispielsweise nur bei verhältnisskalierten Daten sinnvoll.

Nominalskala	Modus
Ordinalskala	Median
Intervallskala	arithm. Mittel
Verhältnisskala	geom. Mittel *

* Es dürfen zusätzlich nur positive Werte vorliegen.

Abbildung I 10: Lageparameter und Skalierung

Außer mit der Berechnung o.g. Lagemaße, werden wir uns mit wichtigen Eigenschaften der Maße auseinandersetzen. Wir werden dabei insbesondere Datentransformationen und Ausreißer bzw. ihre Auswirkungen auf die Lagemaße erörtern.

Transformationen erleichtern uns die inhaltliche Interpretation und vor allem den Vergleich von Daten, die z.B. in unterschiedlichen Einheiten gemessen werden. So können wir z.B. das Zentrum zweier Umsatzverteilungen, die in verschiedenen Währungen erfasst wurden, einfacher vergleichen, wenn wir die Beobachtungen einer Verteilung in die Währung der anderen umrechnen.

Ausreißer stellen einzelne extreme Beobachtungswerte x dar, die außerhalb des Intervalls liegen, in dem sich der Großteil der anderen befindet, d.h. die weit vom Zentrum der Verteilung entfernt liegen. Das Auftreten derartiger Beobachtungen kann unterschiedliche Gründe haben. Ein Ausreißer kann ein untypischer Wert sein, der sich auf eine Sonderstellung des betroffenen Merkmalsträgers in der Grundgesamtheit zurückführen lässt. Ausreißer können aber auch aufgrund von Messfehlern entstehen. Besonders bei komplexen Messvorgängen sind Messfehler ein häufiges Phänomen.

2.2.1.1 Modus

Der auch als dichtester Wert, häufigster Wert oder Modalwert bezeichnete **Modus** x_{Mod} kann bei jedem Skalenniveau einer Merkmalsvariablen X ermittelt werden. Er ist diejenige Merkmalsausprägung in einer Beobachtungsreihe, die am häufigsten vorkommt, und ist für nominalskalierte Merkmale der einzige sinnvolle Lageparameter. Der Modus muss dabei nicht eindeutig sein. Gibt es zwei oder mehr Merkmalsausprägungen, die die gleiche maximale Häufigkeit aufweisen, dann ist jeder dieser Werte ein Modus. Wir sprechen dann von einer *multimodalen Verteilung*. Um die Aussagekraft des Modus zu erhöhen, sollte man ihn nur bei eingipfligen Verteilungen verwenden.

Modus aus der Urliste:

Ein einfaches Auszählen ist zur Ermittlung des Modus ausreichend. Der Modus ist jenes x_i , welches am häufigsten vorkommt.

Modus aus der Häufigkeitstabelle:

Der Modus ist hier jenes a_i , das am häufigsten vorkommt.

Modus aus der Häufigkeitstabelle mit klassierten Daten:

Bei klassierten Daten können wir lediglich eine sog. *Modalklasse* angeben. Dies ist diejenige Klasse K_i mit der größten Klassenhäufigkeit h_i bzw. f_i .¹ Als Modus könnte man nun die Mitte dieser Modalklasse verwenden. Problem kann dann natürlich sein, dass der so bestimmte Modus gar nicht als Beobachtung in der Urliste existiert oder nicht der am häufigsten vorkommende Wert sein muss.

Beispiel I 2-11**Basisbeispiel für die Berechnung von Lageparametern**

Ein Assessment-Center beurteilte im Rahmen einer Einstellungsmaßnahme eines Unternehmens mittels diverser Intelligenz- und Leistungstests den IQ von 30 Bewerbern. Für die IQs ergaben sich folgende sortierte Werte:

70 70 70 80 80 80 80 80 90 90 90 90 90 90 90 90 90 90 90 90 90 90 100 100
100 100 100 100 110 110 110

Da die Beobachtung 90 am häufigsten vorkommt, ergibt sich $x_{Mod} = 90$.

Aus der vorliegenden Urliste erhalten wir folgende Häufigkeitstabelle. Bei Betrachtung der Häufigkeiten wird darin deutlich, dass es sich bei der vorliegenden Verteilung um eine *symmetrische Verteilung* handelt. Ihr Zentrum liegt bei $a_3 = 90$.

Lfd. Nr. j	IQ a_j	abs. Häufigkeit $h(a_j)$	rel. Häufigkeit $f(a_j)$	rel. Summenhäufigkeit $F(a_j)$
1	70	3	0,10	0,10
2	80	6	0,20	0,30
3	90	12	0,40	0,70
4	100	6	0,20	0,90
5	110	3	0,10	1,00
-	-	30	1,00	-

Aus der Häufigkeitstabelle erhalten wir hier ebenfalls $x_{Mod} = 90$, da die absolute Häufigkeit der Beobachtung 90 mit 12 maximal ist.

Eigenschaften des Modus:

- Der Modus ist zwar einfach bestimmbar und auch anschaulich interpretierbar, jedoch besitzt er einen geringen Informationsgehalt.
- x_{Mod} sollte nur bei eingipfligen Verteilungen und nominal- oder ordinalskalierten Daten verwendet werden, da sonst der Informationsverlust zu groß ist.
- Der Modus ist unempfindlich gegenüber Ausreißern, da solche nichts an der Häufigkeit der Modusbeobachtung ändern.

¹ In der Literatur wird hier auch gelegentlich vorgeschlagen, die normierten Klassenhäufigkeiten h'_i bzw. f'_i nach (I2: 16) zu verwenden.

2.2.1.2 Median

Ein weiteres Lagemaß ist der **Median** oder Zentralwert x_{Med} . Er kann aus den n vorliegenden Merkmalswerten bestimmt werden, wenn diese der Größe nach geordnet sind. Für die Anwendung des Medians muss das untersuchte Merkmal X daher zumindest ordinalskaliert sein. Der Median ist dadurch charakterisiert, dass mindestens 50 % der Merkmalswerte kleiner oder gleich und mindestens 50 % größer oder gleich x_{Med} sind. Er stellt somit nichts anderes als einen Sonderfall der unter I 2.1.5 behandelten Quantile dar. Konkret gilt $x_{\text{Med}} = x_{0,50} = Q_2$, d.h. der Median ist das 50 %-Quantil bzw. das Quartil Q_2 . Vereinfacht ausgedrückt wird auch häufig formuliert, dass der Median derjenige Merkmalswert ist, der genau *in der Mitte* der sortierten Beobachtungsreihe liegt.

Median aus der Urliste:

Aus einer sortierten Urliste der Form $x_1 \leq x_2 \leq \dots \leq x_n$ ergibt sich der Median bei *ungerader Beobachtungszahl* n als

$$x_{\text{Med}} = x_{\frac{n+1}{2}}. \quad (\text{I2: 17})$$

Ist n allerdings *gerade*, erfüllt jeder Wert im Intervall $[x_{n/2}, x_{n/2+1}]$ die Bedingung (I2: 17). Es ist hier üblich, den Median als Mitte dieses Intervalls, d.h. als

$$x_{\text{Med}} = \frac{1}{2} \cdot (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) \quad (\text{I2: 18})$$

anzugeben. Dieser berechnete Median existiert jedoch in der Urliste nicht, was insbesondere bei ordinalskalierten Merkmalen problematisch sein kann.

Median aus der Häufigkeitstabelle:

Der Median ist hier jener Merkmalswert a_j , bei dem die empirische Verteilungsfunktion $F(x)$ den Wert 0,5 erreicht oder überschreitet.

Median aus der Häufigkeitstabelle mit klassierten Daten:

Im Falle klassierter Daten wird der Median als die Mitte derjenigen Klasse K_j definiert, bei der $F(x)$ den Wert 0,5 erreicht oder überschreitet. Es ist klar, dass hier Medianwerte resultieren können, die in der Urliste nicht existieren, und wir das Zentrum der Verteilung nur näherungsweise beschreiben können. Haben wir die klassierte Häufigkeitstabelle selbst erstellt, werden wir diese Näherungslösung meiden und den Median stattdessen aus Urliste oder Häufigkeitstabelle bestimmen. Nur bei klassierten Tabellen aus anderen Quellen nutzen wir diese Näherung.

Beispiel I 2-12

Berechnung des Median

Berechnen wir für unser IQ-Datenmaterial aus Beispiel I 2-11 den Median. Da die Beobachtungszahl $n = 30$ gerade ist, erhalten wir den Median aus der Urliste über (I2: 18) als Mittelwert aus den Merkmalswerten der 15-ten und 16-ten Beobachtung zu

$$x_{\text{Med}} = 0,5 \cdot (x_{\frac{30}{2}} + x_{\frac{30}{2}+1}) = 0,5 \cdot (x_{15} + x_{16}) = 0,5 \cdot (90 + 90) = 90.$$

Auch in der Häufigkeitstabelle ist $x_{\text{Med}} = 90$, da für $a_j = 90$ die kumulierte relative Häufigkeit den Wert 0,5 überschreitet.

Würden wir den IQ-Datensatz um eine Beobachtung mit dem Wert 110 ergänzen ($n = 31$), so wäre der Median der Merkmalswert der 16-ten Beobachtung, d.h. $x_{\text{Med}} = x_{\frac{31+1}{2}} = x_{16} = 90$.

Eigenschaften des Median:

- Liegen die Merkmalswerte als Zahlen vor, hat die Summe der absoluten Abweichungen der Merkmalswerte von einer beliebigen Zahl λ ein Minimum, wenn diese Zahl gleich x_{Med} ist.

$$\text{Min} \sum_{i=1}^n |x_i - \lambda| \quad \text{für } \lambda = x_{\text{Med}} \quad (\text{I2: 19})$$

Dies ist die sog. *Optimalitäts- oder Minimumseigenschaft* des Medians.

- Der Median ist unempfindlich gegenüber Ausreißern. Nehmen wir z.B. an, es ist eine geordnete Urliste von Werten 1, 2, 3, 4, 5 gegeben. Der Median liegt hier klar bei $x_{\text{Med}} = 3$. Tritt nun an die Stelle des Wertes 5 der Wert 100, so liegt ein Ausreißer vor. Dieser hat jedoch keinen Einfluss auf den Median, da der Wert 3 schließlich seine Position in der Mitte behält.
- Der Median ist das wichtigste Lagemaß für ordinalskalierte Daten.

2.2.1.3 Arithmetisches Mittel

Der wohl bekannteste und wichtigste Lageparameter ist das auch als Durchschnittswert bezeichnete **arithmetische Mittel** \bar{x} . Es findet bei zumindest intervallskalierten Merkmalen Anwendung.

Arithmetisches Mittel aus der Urliste:

Liegt eine Urliste x_1, x_2, \dots, x_n vor, so ermitteln wir das arithmetische Mittel \bar{x} als

$$\bar{x} = \frac{1}{n} \cdot (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \cdot \sum_{i=1}^n x_i. \quad (\text{I2: 20})$$

Es werden also lediglich alle Merkmalswerte aufsummiert und durch die Anzahl der Merkmalsträger n dividiert.

Arithmetisches Mittel aus der Häufigkeitstabelle:

Ist eine Häufigkeitsverteilung vorhanden, erleichtert sich die Berechnung erheblich. Zur Bestimmung des arithmetischen Mittels ist dann nämlich lediglich die Summe der Produkte der Merkmale a_j und ihrer absoluten Häufigkeiten $h(a_j)$ zu bestimmen und diese durch n zu dividieren. Es gilt nämlich

$$\begin{aligned} \bar{x} &= \frac{1}{n} \cdot (\underbrace{a_1 + \dots + a_1}_{h(a_1)-\text{mal}} + \underbrace{a_2 + \dots + a_2}_{h(a_2)-\text{mal}} + \dots + \underbrace{a_k + \dots + a_k}_{h(a_k)-\text{mal}}) \\ &= \frac{1}{n} \cdot (h(a_1) \cdot a_1 + h(a_2) \cdot a_2 + \dots + h(a_k) \cdot a_k), \end{aligned}$$

woraus

$$\bar{x} = \frac{1}{n} \cdot \sum_{j=1}^k h(a_j) \cdot a_j \quad (\text{I2: 21})$$

folgt. Bringen wir $1/n$ in den Summenausdruck und substituieren $h(a_j)/n = f(a_j)$, erhalten wir folgende alternative Berechnungsformel mit relativen Häufigkeiten:

$$\bar{x} = \sum_{j=1}^k f(a_j) \cdot a_j \quad (\text{I2: 22})$$

Wir erkennen an (I2: 22), dass das arithmetische Mittel ein Sonderfall des sog. *gewogenen arithmetischen Mittels* ist. Dieses ist definiert als

$$\bar{x} = \sum_{i=1}^n g_i \cdot x_i \quad \text{mit} \quad 0 \leq g_i \leq 1 \quad \text{und} \quad \sum_{i=1}^n g_i = 1. \quad (\text{I2: 23})$$

Die Beobachtungswerte x_i werden hier durch vorgegebene Zahlen g_i gewichtet und aufsummiert. Die Gewichte müssen nichtnegative reelle Zahlen sein, deren Summe sich auf Eins beläuft. Das arithmetische Mittel (I2: 22) erhalten wir, wenn wir in (I2: 23) die Beobachtungen a_j mit ihren relativen Häufigkeiten $f(a_j)$ gewichten.

Arithmetisches Mittel aus der Häufigkeitstabelle mit klassierten Daten:

Da wir in der Praxis (z.B. in statistischen Veröffentlichungen) die einzelnen Beobachtungen nicht kennen, die einer klassierten Häufigkeitstabelle zugrunde liegen, sind wir nur in der Lage, aus dieser ein approximatives arithmetisches Mittel zu bestimmen. Wir ersetzen dazu in den Formeln (I2: 21) und (I2: 22) die a_j durch die Klassenmitten a_j^m als repräsentative Werte für die einzelnen Klassen und verwenden korrespondierende Klassenhäufigkeiten. Wir erhalten damit die Näherung

$$\bar{x} \approx \frac{1}{n} \cdot \sum_{j=1}^k h_j \cdot a_j^m \quad \text{bzw.} \quad \bar{x} \approx \sum_{j=1}^k f_j \cdot a_j^m. \quad (\text{I2: 24})$$

Haben wir die klassierte Häufigkeitstabelle selbst erstellt, werden wir natürlich nicht auf (I2: 24), sondern auf die Urlistenformel (I2: 20) zurückgreifen.

Beispiel I 2-13

Berechnung des arithmetischen Mittels

Für unseren IQ-Datensatz aus Beispiel I 2-11 wollen wir nun das arithmetische Mittel aus den verschieden aufbereiteten Daten berechnen:

Urliste:

$$\bar{x} = \frac{1}{30} \cdot (70 + 70 + 70 + 80 + 80 + 80 + 80 + 80 + 80 + \dots + 110 + 110 + 110) = 90$$

Häufigkeitstabelle:

$$\bar{x} = \frac{1}{30} \cdot (3 \cdot 70 + 6 \cdot 80 + 12 \cdot 90 + 6 \cdot 100 + 3 \cdot 110) = 90$$

Vergleichen wir Modus, Median und arithmetisches Mittel für unseren IQ-Datensatz, so stellen wir fest, dass alle drei Maße den Wert 90 annehmen. Wir werden unter I 2.2.1.6 sehen, dass für den Fall einer symmetrischen unimodalen Verteilung Modus, Median und arithmetisches Mittel stets identisch sind.

Häufigkeitstabelle mit klassierten Daten:

Zwar kennen wir hier die Urliste und sollten daher das beschriebene Näherungsverfahren nicht anwenden, doch wollen wir dies zur Veranschaulichung des entstehenden Rechenfehlers trotzdem tun. Wir führen dazu die folgende Klassierung ein:

j	über a_j^u bis a_j^o	a_j^m	h_j	f_j
1	60 - 80	70	9	0,30
2	80 - 90	85	12	0,40
3	90 - 110	100	9	0,30
-			30	1,00

Wir erhalten daraus nicht den Wert 90, sondern

$$\bar{x} \approx \frac{1}{30} \cdot (9 \cdot 70 + 12 \cdot 85 + 9 \cdot 100) = 85.$$

Eigenschaften des arithmetischen Mittels:

- Die Summe der Abweichungen der Merkmalswerte vom arithmetischen Mittel ist stets Null. Wir sprechen hier auch von der *Zentraleigenschaft* von \bar{x} .

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n \cdot \bar{x} = n \cdot \bar{x} - n \cdot \bar{x} = 0 \quad (\text{I2: 25})$$

Berechnen wir beispielsweise das arithmetische Mittel von 2, 4, 6 und 8, so erhalten wir 5. Die Abweichungen der Einzelwerte von 5 sind $-3, -1, 1$ und 3 , was in der Summe genau Null ergibt

- Die Summe der quadrierten Abweichungen der Merkmalswerte von einer beliebigen Zahl λ hat ein Minimum, wenn diese Zahl gleich \bar{x} ist. Wir sprechen hier auch von der *Optimalitäts- oder Minimumseigenschaft* von \bar{x} .

$$\text{Min} \sum_{i=1}^n (x_i - \lambda)^2 \quad \text{für } \lambda = \bar{x} \quad (\text{I2: 26})$$

Diese Eigenschaft können wir beweisen, indem wir zunächst die erste Ableitung der Summe nach λ bilden und gleich Null setzen (notwendige Bedingung für einen Extremwert):

$$\sum_{i=1}^n 2 \cdot (x_i - \lambda) \cdot (-1) = -2 \sum_{i=1}^n (x_i - \lambda) = 0$$

Division durch -2 und Auflösen nach λ liefert

$$\sum_{i=1}^n x_i - n\lambda = 0 \quad \Leftrightarrow \quad \sum_{i=1}^n x_i = n\lambda \quad \Leftrightarrow \quad \lambda = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Hinreichende Bedingung für ein Minimum ist eine positive zweite Ableitung an der Stelle $\lambda = \bar{x}$, was gegeben ist, da die zweite Ableitung $2n$ lautet.

- Werden die Merkmalswerte x_i in der Form $y_i = a + b \cdot x_i$ mit $a, b \in \mathbb{R}$ und $b \neq 0$ linear transformiert (vgl. Beispiel I 2-14), so kann das neue arithmetische Mittel \bar{y} durch dieselbe Transformationsvorschrift aus \bar{x} ermittelt werden. Es gilt nämlich

$$\sum_{i=1}^n y_i = \sum_{i=1}^n (a + b \cdot x_i) = \sum_{i=1}^n a + \sum_{i=1}^n b \cdot x_i = \sum_{i=1}^n a + b \cdot \sum_{i=1}^n x_i = n \cdot a + b \cdot n \cdot \bar{x},$$

woraus wir nach Division durch n folgende Beziehung erhalten:

$$\bar{y} = a + b \cdot \bar{x} \quad (\text{I2: 27})$$

- Ist eine Gesamtheit in zwei Teilgesamtheiten zerlegt ($n = n_1 + n_2$), so ermittelt sich das arithmetische Mittel folgendermaßen (vgl. Beispiel I 2-15):

$$\bar{x}_{\text{ges}} = \frac{1}{n_1 + n_2} \cdot \left(\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} x_i \right) = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n}$$

(I2: 28)

Liegen mehrere Teilgesamtheiten ($n = n_1 + n_2 + \dots + n_m$) vor, gilt

$$\bar{x}_{\text{ges}} = \frac{1}{n} \sum_{k=1}^m n_k \bar{x}_k = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_m \bar{x}_m}{n}. \quad (\text{I2: 29})$$

Eine nähere Betrachtung von (I2: 29) bzw. eine Umformung

$$\bar{x}_{\text{ges}} = \sum_{k=1}^m \frac{n_k}{n} \bar{x}_k$$

zeigt, dass es sich hierbei um nichts anderes als ein gewogenes arithmetisches Mittel der Mittelwerte der Teilgesamtheiten handelt. Die jeweiligen Gewichte ergeben sich zu n_k/n .

- Das arithmetische Mittel reagiert empfindlich auf Ausreißer, da in \bar{x} anders als beim Modus und Median alle Beobachtungswerte der Verteilung eingehen und so ein einzelner extremer Wert Einfluss auf das Maß nehmen kann (vgl. Beispiel I 2-16).

Beispiel I 2-14

Transformation und arithmetisches Mittel

Ein Vertriebsmitarbeiter erhält ein fixes Grundgehalt von 1.500 €. Als Provision stehen ihm zusätzlich 5 % des von ihm erwirtschafteten Umsatzes zu. Ferner sei bekannt, dass dieser Mitarbeiter im Durchschnitt einen Umsatz von 10.000 € je Monat erzielt. Bestimmen wir auf Basis dieser Daten sein monatliches Durchschnittseinkommen:

$$\bar{y} = a + b \cdot \bar{x} = 1.500 \text{ €} + 0,05 \cdot 10.000 \text{ €} = 2.000 \text{ €}$$

Beispiel I 2-15

Arithmetisches Mittel aus Teilgesamtheiten

In einer Stadt mit 10.000 Einwohnern (6.000 Männer, 4.000 Frauen) liegt das Durchschnittseinkommen der Männer bei 2.500 € und der Frauen bei 2.000 €. Wie hoch ist das durchschnittliche Einkommen aller Einwohner?

$$\bar{x}_{\text{ges}} = \frac{6.000 \cdot 2.500 \text{ €} + 4.000 \cdot 2.000 \text{ €}}{10.000} = 2.300 \text{ €}$$

Beispiel I 2-16

Arithmetisches Mittel und Ausreißer

Haben 10 Personen ein Jahreseinkommen von 20.000 €, liegt ihr Durchschnittseinkommen bei 20.000 €. Besitzt nun jedoch eine der Personen ein Jahreseinkommen von 500.000 €, so erhalten wir ein Durchschnittseinkommen in Höhe von 63.636,36 €, was ein "falsches" Bild von der Einkommensverteilung liefert. Gerade hieran erkennen wir, dass Lageparameter allein nicht ausreichen, um eine Häufigkeitsverteilung zutreffend zu beschreiben.

2.2.1.4 Geometrisches Mittel

Ein weiteres in der Praxis bedeutendes Mittel ist das geometrische Mittel \bar{x}_{geo} . Es kann grundsätzlich nur für verhältnisskalierte Merkmale bestimmt werden, deren Ausprägungen x_i ausschließlich positive Werte aufweisen.

Geometrisches Mittel aus der Urliste:

Liegt eine Urliste n positiver Beobachtungswerte x_i vor, so ergibt sich das geometrische Mittel als n -te Wurzel des Produkts aller Beobachtungen. Es wird also über

$$\bar{x}_{\text{geo}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (\text{I2: 30})$$

berechnet. Häufig sind diese x_i im Zeitverlauf aufgezeichnete relative Änderungen (Wachstumsfaktoren) eines Merkmals, sodass \bar{x}_{geo} Aussagen über das durchschnittliche Wachstum des Merkmals ermöglicht.

Beispiel I 2-17**Geometrisches Mittel aus der Urliste**

Über einen Zeitraum von 4 Jahren wurde die Mitarbeiterzahl z_t eines Unternehmens jeweils zum Jahresende aufgezeichnet und in der nachfolgenden Tabelle niedergeschrieben. Von Interesse ist nun das durchschnittliche Jahreswachstum der Belegschaft dieses Unternehmens. Dazu werden zunächst zwischen den Perioden die *Wachstumsraten*

$$w_t = \frac{z_t - z_{t-1}}{z_{t-1}}$$

bestimmt, aus denen *Wachstumsfaktoren* $x_t = 1 + w_t$ gewonnen werden. Wie zu erkennen ist, wird im Zeitraumkontext das Subskript i meist durch t ersetzt. Auf die Gültigkeit der behandelten Formeln hat dies aber keinen Einfluss.

t	Jahr	Mitarbeiterzahl z_t	Wachstumsrate w_t	Wachstumsfaktor x_t
0	2002	2.300	-	-
1	2003	3.000	0,3043	1,3043
2	2004	2.700	-0,1000	0,9000
3	2005	3.500	0,2963	1,2963

Setzen wir die x_t in (I2: 30) ein, so ergibt sich damit das geometrische Mittel zu

$$\bar{x}_{\text{geo}} = \sqrt[3]{1,3043 \cdot 0,9000 \cdot 1,2963} = 1,1502 ,$$

d.h. das durchschnittliche Jahreswachstum liegt bei $(1,1502 - 1) \cdot 100 \% = 15,02 \% .$

Um das Prinzip zu verdeutlichen, welches sich in Formel (I2: 30) verbirgt, empfiehlt sich die Betrachtung des nachfolgenden Beispiels, welches eine "intuitive Herleitung" des geometrischen Mittels im Wachstumsfaktorenkontext liefert.

Beispiel I 2-18**Prinzip des geometrischen Mittels**

Nehmen wir an, der Umsatz U_0 eines Unternehmens steigt in den Folgequartalen um 2 %, 3 %, 4 % und 5 %. Die Umsatzsteigerungen beziehen sich dabei jeweils auf den Umsatz des Vorquartals. Die Umsatzentwicklung zeigt sich daher wie folgt:

$$U_1 = U_0 \cdot (1+0,02)$$

$$U_2 = U_1 \cdot (1+0,03) = U_0 \cdot (1+0,02) \cdot (1+0,03)$$

$$U_3 = U_2 \cdot (1+0,04) = U_0 \cdot (1+0,02) \cdot (1+0,03) \cdot (1+0,04)$$

$$U_4 = U_3 \cdot (1+0,05) = U_0 \cdot (1+0,02) \cdot (1+0,03) \cdot (1+0,04) \cdot (1+0,05)$$

Eine durchschnittliche Zuwachsrate w der Quartalsumsätze ist nun eine für alle Quartale identische Zuwachsrate, die zum gleichen Jahresumsatz U_4 führt. Setzen wir also obigen Ausdruck mit $U_4 = U_0 \cdot (1 + w)^4$ gleich, so erhalten wir folgendes Ergebnis:

$$U_0 \cdot (1 + w)^4 = U_0 \cdot (1 + 0,02) \cdot (1 + 0,03) \cdot (1 + 0,04) \cdot (1 + 0,05)$$

$$(1 + w)^4 = (1 + 0,02) \cdot (1 + 0,03) \cdot (1 + 0,04) \cdot (1 + 0,05)$$

$$1 + w = \sqrt[4]{1,1473}$$

$$w = 0,0349 = 3,49\%$$

Die Zeilen 2 und 3 lassen hier deutlich das geometrische Mittel und die in selbiges eingehenden Wachstumsfaktoren erkennen.

Geometrisches Mittel aus der Häufigkeitstabelle:

Liegen die positiven Merkmalswerte a_1, a_2, \dots, a_k und die dazugehörigen Häufigkeiten $h(a_1), h(a_2), \dots, h(a_k)$ bzw. $f(a_1), f(a_2), \dots, f(a_k)$ vor, so berechnen wir das geometrische Mittel wie folgt:

$$\bar{x}_{geo} = \sqrt[n]{a_1^{h(a_1)} \cdot a_2^{h(a_2)} \cdot \dots \cdot a_k^{h(a_k)}} \quad \text{mit} \quad n = \sum_{j=1}^k h(a_j) \quad (\text{I2: 31})$$

$$\bar{x}_{geo} = a_1^{f(a_1)} \cdot a_2^{f(a_2)} \cdot \dots \cdot a_k^{f(a_k)} \quad (\text{I2: 32})$$

Durch Logarithmierung von (I2: 30), (I2: 31) und (I2: 32) erhalten wir eine interessante Eigenschaft des geometrischen Mittels zu

$$\ln \bar{x}_{geo} = \frac{1}{n} \sum_{i=1}^n \ln x_i = \frac{1}{n} \sum_{j=1}^k h(a_j) \cdot \ln a_j = \sum_{j=1}^k f(a_j) \cdot \ln a_j. \quad (\text{I2: 33})$$

Der Logarithmus des geometrischen Mittels ist also gleich dem arithmetischen Mittel der logarithmierten Daten.

Das geometrische Mittel findet, wie bereits angesprochen, vor allem bei der Berechnung durchschnittlicher Wachstumsraten Anwendung. Einen besonderen Anwendungsbereich stellt daher die Berechnung von durchschnittlichen Zinssätzen oder Renditen dar, da diese nichts anderes als Wachstumsraten von Kapital sind.

2.2.1.5 Exkurs: Renditen und Renditedurchschnitte

Legen wir einen Geldbetrag K_{t-1} zu Beginn einer Periode $t-1$ zu einem Zins von r_t^d bis zum Ende dieser Periode an, so ergibt sich am Ende der Periode $t-1$ bzw. dem Beginn der Periode t ein Endkapital K_t als

$$K_t = (1 + r_t^d) \cdot K_{t-1}. \quad (\text{I2: 34})$$

Wir bezeichnen die Größe r_t^d dabei auch als sog. **diskrete Rendite**. Sie ist nichts anderes als die Wachstumsrate des Kapitals zwischen den Zeitpunkten $t-1$ und t und kann durch folgende allgemeine Formel bestimmt werden, die sich durch Umformung von (I2: 34) herleiten lässt:

$$r_t^d = \frac{K_t - K_{t-1}}{K_{t-1}} = \frac{K_t}{K_{t-1}} - 1 \quad (\text{I2: 35})$$

Die sog. **geometrische Durchschnitts- oder zeitgewichtete Rendite** \bar{r}_{geo} stellt die für jede Periode identische Rendite dar, die bei einer Kapitalanlage in Höhe K_0 über n Perioden zum gleichen Endkapital K_n führt, wie die n individuellen Periodenrenditen. Für die Kapitalstände K_0 und K_n kann deshalb entsprechend der Vorgehensweise aus Beispiel I 2-18 der Zusammenhang

$$\underbrace{(1+r_1^d) \cdot (1+r_2^d) \cdots (1+r_n^d)}_{K_n} \cdot K_0 = \underbrace{(1+\bar{r}_{\text{geo}})^n}_{K_n} \cdot K_0 \quad (\text{I2: 36})$$

festgehalten werden, aus dem durch Umstellung

$$\bar{r}_{\text{geo}} = \sqrt[n]{(1+r_1^d) \cdot (1+r_2^d) \cdots (1+r_n^d)} - 1 \quad (\text{I2: 37})$$

resultiert. Diesen Berechnungen unterliegt die Annahme, dass *Zinserträge* auf dem Anlagekonto verbleiben und daher in den Folgeperioden zusätzlich zum Anfangskapital *mitverzinst* werden.

Von der sog. **arithmetischen Durchschnittsrendite** sprechen wir hingegen bei

$$\bar{r}^d = \frac{1}{n} \cdot \sum_{t=1}^n r_t^d . \quad (\text{I2: 38})$$

Hierbei wird von der Annahme ausgegangen, dass jede Periode mit dem *gleichen Kapitaleinsatz* startet, entstandene Zinserträge also nicht reinvestiert werden.

Beispiel I 2-19

Diskrete Durchschnittsrenditen

Eine Anlage in Höhe von 1.000 € ergab bei Jahreszinssätzen von 2 %, 3 % und 4 % und Wiederanlage der Zinserträge ein Endkapital von $1.000 \cdot (1 + 0,02) \cdot (1 + 0,03) \cdot (1 + 0,04) = 1.092,62$ €. Die geometrische Durchschnittsrendite ergibt sich daraus zu

$$\bar{r}_{\text{geo}} = \sqrt[3]{(1+0,02) \cdot (1+0,03) \cdot (1+0,04)} - 1 = 0,0299 = 2,99 \%$$

Dies bedeutet, dass wir auch zu einem Jahreszins von 2,99 % hätten anlegen können, um (bei Wiederanlage der Zinserträge) den Endbetrag 1.092,62 € zu erhalten.

Werden die Zinserträge hingegen jedes Jahr entnommen, so erhalten wir lediglich ein Endkapital von $1.000 \cdot (1 + 0,02 + 0,03 + 0,04) = 1.090$ €. Die arithmetische Durchschnittsrendite ergibt sich daraus zu

$$\bar{r}^d = \frac{1}{3} \cdot (0,02 + 0,03 + 0,04) = 0,03 = 3 \% .$$

Dies bedeutet, dass wir durch Anlage zu einem Zins von 3 % pro Jahr (bei Entnahme der Zinserträge) ebenso ein Endkapital von 1.090 € erreichen können.

Unterstellen wir eine stetige Verzinsung des Kapitals K_{t-1} mit dem Satz r_t^s , so ergibt sich das Endkapital K_t nach einer Periode als

$$K_t = e^{r_t^s} \cdot K_{t-1}, \quad (\text{I2: 39})$$

wobei e für die Eulersche Zahl mit $e = 2,718281\dots$ steht. Durch Logarithmierung beider Seiten der Gleichung (I2: 39)

$$\ln K_t = r_t^s + \ln K_{t-1}$$

und anschließender Umformung erhalten wir die sog. **stetige Rendite**

$$r_t^s = \ln K_t - \ln K_{t-1} = \ln \left(\frac{K_t}{K_{t-1}} \right). \quad (\text{I2: 40})$$

Die Zusammenhänge, die zwischen stetigen und diskreten Renditen bestehen, sind in Abbildung I 11 veranschaulicht.

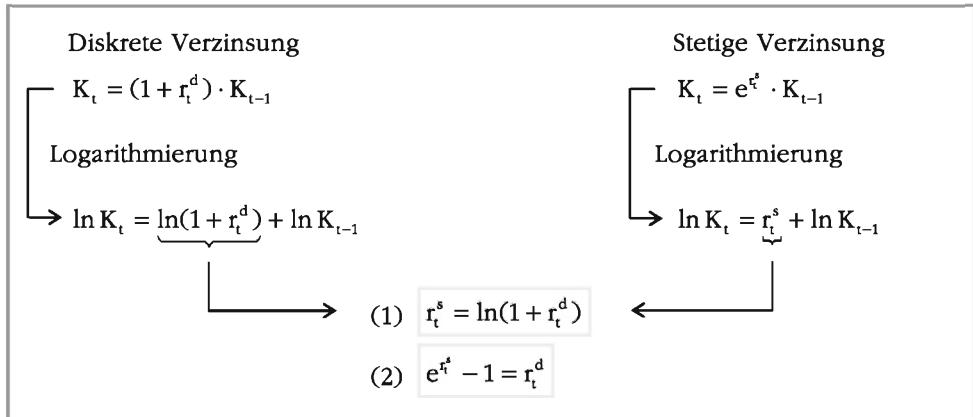


Abbildung I 11: Zusammenhang zwischen stetigen und diskreten Renditen

Betrachten wir nun insgesamt n Perioden, so gilt zunächst unter Verwendung diskreter Renditen nach (I2: 36) $K_n = (1 + r_1^d) \cdot (1 + r_2^d) \cdot \dots \cdot (1 + r_n^d) \cdot K_0$. Logarithmieren wir diese Gleichung, erhalten wir unter Nutzung der Beziehung (1) aus Abbildung I 11 den Zusammenhang $\ln K_n = r_1^s + r_2^s + \dots + r_n^s + \ln K_0$, woraus wir

$$\ln K_n - \ln K_0 = \sum_{i=1}^n r_i^s \quad (\text{I2: 41})$$

ableiten können. Wir können also sagen, dass sich die *stetige Rendite einer Gesamtperiode* als *Summe der stetigen Renditen der dazugehörigen Teilperioden* ermitteln lässt. Für diskrete Renditen gilt ein solcher Zusammenhang nicht. Hier sind die um Eins erhöhten diskreten Teilperiodenrenditen miteinander zu multiplizieren und Eins zu subtrahieren, um die Gesamtperiodenrendite zu erhalten.

Beispiel I 2-20

Mikro- und Makroperiodenrenditen

Eine Investition wirft innerhalb eines Jahres die stetigen Quartalsrenditen 1 %, 5 %, -2 % und 3 % ab. Die stetige Jahresrendite ergibt sich daraus als

$$r_{\text{Jahr}}^s = \sum_{t=1}^4 r_{\text{Quartal } t}^s = 0,01 + 0,05 - 0,02 + 0,03 = 0,07 = 7 \text{ %.}$$

Nehmen wir an, es handelt sich bei den gegebenen Renditen um diskrete Quartalsrenditen, so erhalten wir die diskrete Jahresrendite zu

$$r_{\text{Jahr}}^d = \prod_{t=1}^4 (1 + r_{\text{Quartal } t}^s) - 1 = (1 + 0,01) \cdot (1 + 0,05) \cdot (1 - 0,02) \cdot (1 + 0,03) - 1 = 0,0704 = 7,04 \text{ %.}$$

Spezieller Exkurs Finance:

Es soll nicht unerwähnt bleiben, warum gerade stetige Renditen in vielen statistischen Verfahren und Modellen Anwendung finden. Sie besitzen nämlich eher als diskrete Renditen die Eigenschaften der Symmetrie und der Normalverteilung (vgl. Abschnitt II 3.2.3), welche Voraussetzung für diverse Verfahren der Finanzmarktstatistik und -ökonomie sind.

Beispiel I 2-21

Symmetrie stetiger Renditen

Gegeben sei die Kapitalentwicklung $K_1 = 2 \cdot K_0$ und $K_2 = 0,5 \cdot K_1$, woraus sich die folgenden stetigen und diskreten Renditen ergeben:

- Stetig: $r_1^s = \ln 2$, $r_2^s = \ln 0,5$
- Diskret: $r_1^d = 1$, $r_2^d = -0,5$

Bei den stetigen Renditen ist die Symmetriebedingung $r_1^s = |r_2^s|$ erfüllt, wohingegen bei den diskreten Renditen $r_1^d \neq |r_2^d|$ gilt.

Es ist außerdem interessant, dass die *geometrische Durchschnittsrendite in stetiger Schreibweise gleich dem arithmetischen Durchschnitt der stetigen Einzelrenditen* ist. Um diesen Zusammenhang zu erklären, betrachten wir zunächst die Formel (I2: 36)

$$(1 + \bar{r}_{\text{geo}})^n = \underbrace{(1 + r_1^d) \cdot (1 + r_2^d) \cdot \dots \cdot (1 + r_n^d)}_{\text{Wachsumsmultiplikator WM}}.$$

Durch Logarithmierung erhalten wir daraus

$$n \cdot \ln(1 + \bar{r}_{\text{geo}}) = r_1^s + r_2^s + \dots + r_n^s,$$

was schließlich mittels Division durch n zum eingangs beschriebenen Zusammenhang führt:

$$\ln(1 + \bar{r}_{\text{geo}}) = \frac{r_1^s + r_2^s + \dots + r_n^s}{n} = \bar{r}^s$$

(I2: 42)

Wir stellen fest, dass dieses Resultat (I2: 33) entspricht.

Beispiel I 2-22

Stetige Durchschnittsrendite

Eine Aktienanlage wird vier Jahre betrachtet. Die diskreten Renditen der einzelnen Jahre sind +30 %, -10 %, +50 % und -20 %. Das Vermögen ist gesamt von 100 € auf 140,40 € gestiegen. In diesem Beispiel lässt sich die stetige Durchschnittsrendite nun auf zwei verschiedene Arten berechnen:

1. Berechnung der stetigen Einzelrenditen, Aufsummierung und Division durch n :

Die gegebenen diskreten Renditen werden zunächst mit der Formel $r_i^s = \ln(1 + r_i^d)$ in stetige Renditen umgerechnet. Es ergeben sich +26,2 %, -10,5 %, +40,5 % und -22,3 %. Addiert und durch 4 (Jahre) dividiert, ergibt sich so eine stetige Durchschnittsrendite von 8,48 %.

2. Geometrische Ermittlung der diskreten Durchschnittsrendite und Darstellung in stetiger Schreibweise:

$$\bar{r}_{\text{geo}} = \sqrt[4]{(1+0,3) \cdot (1-0,1) \cdot (1+0,5) \cdot (1-0,2)} - 1 = \sqrt[4]{1,404} - 1 = 8,85 \%$$

Der sich hier ergebende Wachstumsmultiplikator von 1,404 kann schneller direkt als Quotient aus Vermögensendbestand und -anfangsbestand ermittelt werden.

Die stetige Durchschnittsrendite ist damit $\bar{r}^s = \ln(1 + \bar{r}_{geo}) = \ln(1 + 0,0885) = 8,48\%$.

2.2.1.6 Lageregeln

Für jede Art von Verteilung (symmetrisch, rechtsschief, linksschief) gilt zwischen dem geometrischen und arithmetischen Mittel die Beziehung $\bar{x}_{geo} \leq \bar{x}$. Wie Abbildung I 12 skizziert, sind des Weiteren für *symmetrische Verteilungen* arithmetisches Mittel, Median und Modus immer identisch, d.h. es gilt

$$\bar{x} = x_{Med} = x_{Mod}$$

Dies haben wir bereits in Beispiel I 2-13 gesehen. Für *linksschiefe Verteilungen* (auch rechtssteil genannt) gilt stets

$$\bar{x} < x_{Med} < x_{Mod},$$

d.h. das arithmetische Mittel ist stets kleiner als Median und Modus. Für *rechtschiefe Verteilungen* (auch linkssteil genannt) gilt dies umgekehrt, d.h. es liegt stets

$$x_{Mod} < x_{Med} < \bar{x}$$

vor. Das arithmetische Mittel ist also hier stets größer als Median und Modus. Diese "Größenverschiebung" des arithmetischen Mittels ist nicht überraschend, da linksschiefe Verteilungen einige besonders niedrige Merkmalsausprägungen und rechtschiefe Verteilungen einige besonders hohe Merkmalsausprägungen aufweisen.

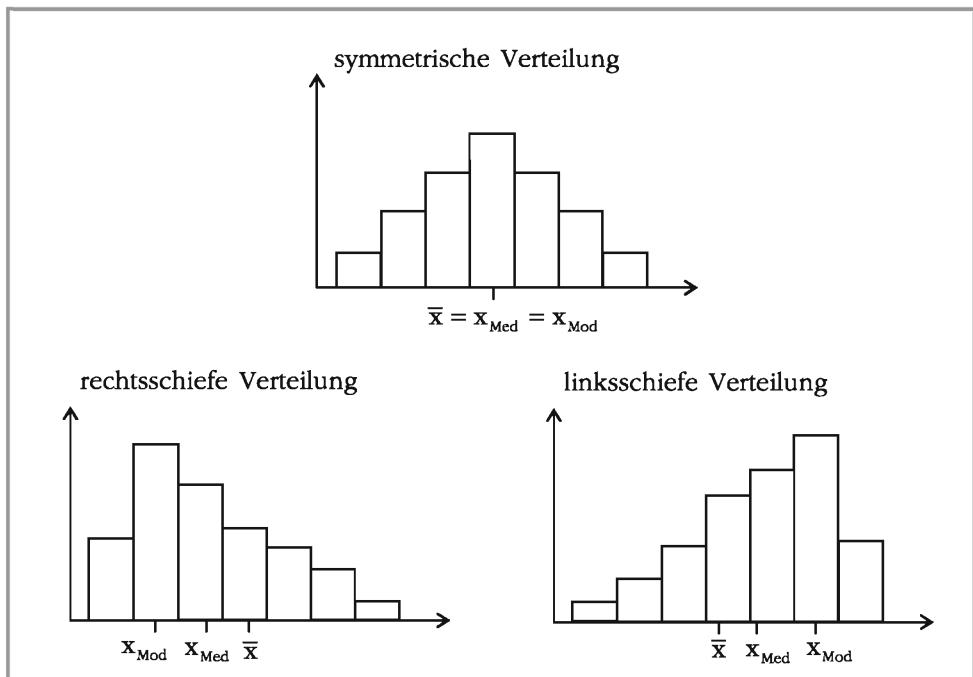


Abbildung I 12: Lageregeln

Wir erkennen, dass außer beim Auftreten von Ausreißern auch durch die Asymmetrie oder Schiefe einer Verteilung die Interpretation von Lagemaßen erschwert werden kann. Bei ausgeprägt schiefen Verteilungen ist die Frage nach dem geeigneten Lageparameter besonders schwer zu beantworten, da sich die Maße z.T. deutlich unterscheiden. Hier empfiehlt sich zur Beschreibung der Verteilung die Verwendung weiterer Parameter (siehe Abschnitt I 2.2.2 bis I 2.2.4).



2.2.2 Streuungsparameter

Die bisher behandelten Lageparameter dienten lediglich der Kennzeichnung des Zentrums einer Verteilung. Oftmals ist jedoch von Interesse, wie stark die Einzelwerte vom Zentrum abweichen, d.h. wie eng oder weit sie um das Zentrum der Verteilung streuen. Zur Gewinnung dieser Informationen sind sog. **Streuungsparameter** geeignet, die wir im Folgenden näher betrachten. Sie setzen stets eine **Kardinalskala** voraus.

Die Streuungsparameter, die wir im Folgenden behandeln, lassen sich grob in zwei Kategorien einteilen. Zur ersten zählen die Spannweite und der Quartilsabstand. Bei diesen Maßen dienen die *Abstände zwischen speziellen Beobachtungen* der Häufigkeitsverteilung als Maß für die Streuung. Zur zweiten Kategorie zählen die mittlere absolute Abweichung, die Varianz und die Standardabweichung. Hier sind die *Abweichungen aller Daten von einem Lagemaß* entscheidendes Kriterium für die Beurteilung der Streuung. Je kleiner diese Kennzahlen sind, desto stärker sind die Daten um den jeweiligen Lageparameter konzentriert.

2.2.2.1 Spannweite und Quartilsabstand

Die **Spannweite SP** gehört zu den wohl einfachsten Streuungsmaßen. Unter ihr verstehen wir lediglich die Differenz zwischen dem größten und kleinsten Merkmalswert im vorliegenden Datenmaterial. Sie gibt uns allgemein Informationen über den Ausdehnungsbereich des Datenbestandes.

Aus einer der Größe nach sortierten *Umliste* errechnet sich die Spannweite als Differenz zwischen größtem (x_n) und kleinstem (x_1) Beobachtungswert.

$$SP = x_n - x_1 \quad (\text{I2: 43})$$

Aus der *Häufigkeitsverteilung* ermitteln wir die Spannweite ebenfalls als Differenz zwischen größter (a_n) und kleinster (a_1) Merkmalsausprägung.

$$SP = a_n - a_1 \quad (\text{I2: 44})$$

Beispiel I 2-23

Basisbeispiel für die Berechnung von Streuungsmaßen

An 10 Zeitpunkten wird der Kurs des A-Dollars und des B-Dollars notiert. Der A-Dollar hatte folgende Notierungen: 1,10, 1,15, 1,10, 1,20, 1,25, 1,30, 1,30, 1,25, 1,20 und 1,15 B\$. Der B-Dollar lag dreimal bei 1,60, zweimal bei 1,80, viermal bei 2,30 und einmal bei 2,40 A\$.

Wir wollen diese beiden Verteilungen nutzen, um die Berechnung der von uns vorgestellten Streuungsmaße zu veranschaulichen. Bei Einführung eines neuen Maßes werden wir jeweils wieder auf dieses Beispiel zurückgreifen, weshalb die im Folgenden aufgestellten Tabellen bereits Berechnungen enthalten, die wir erst nach und nach benötigen werden.

Für den A-Dollar zeigt sich die folgende Urliste:

i	x_i	$ x_i - x_{Med} $	x_i^2
1	1,10	0,10	1,21
2	1,10	0,10	1,21
3	1,15	0,05	1,32
4	1,15	0,05	1,32
5	1,20	0,00	1,44
6	1,20	0,00	1,44
7	1,25	0,05	1,56
8	1,25	0,05	1,56
9	1,30	0,10	1,69
10	1,30	0,10	1,69
-	12,00	0,60	14,44

Wir erhalten daraus die Spannweite

$$SP = x_n - x_1 = x_{10} - x_1 = 1,30 - 1,10 = 0,20 \text{ A\$}.$$

Für den B-Dollar ergibt sich folgende Häufigkeitstabelle:

j	a_j	h_j	$h_j \cdot a_j$	$h_j \cdot a_j - x_{Med} $	a_j^2	$h_j \cdot a_j^2$
1	1,60	3	4,80	1,35	2,56	7,68
2	1,80	2	3,60	0,50	3,24	6,48
3	2,30	4	9,20	1,00	5,29	21,16
4	2,40	1	2,40	0,35	5,76	5,76
-	-	10	20,00	3,20	-	41,08

Dies liefert eine Spannweite von

$$SP = a_k - a_1 = a_4 - a_1 = 2,40 - 1,60 = 0,80 \text{ B\$}.$$

Aufgrund der Tatsache, dass die Spannweite gerade durch ihr Wesen äußerst empfindlich auf Ausreißer reagiert und aufgrund der sehr geringen Ausnutzung des Datenmaterials nur einen geringen Informationsgehalt besitzt, wird sie in der Praxis kaum als Streuungsmaß eingesetzt. Vielmehr dient sie dazu, Hinweise auf Ausreißer zu geben. Ein anderes, der Spannweite sehr ähnliches Maß, das hingegen robust gegen Ausreißer ist, findet häufiger Anwendung. Es handelt sich dabei um den sog. **Quartilsabstand QA**, der als die Differenz zwischen oberem und unterem Quartil definiert ist und nichts anderes als die Spannweite der 50 % mittleren Beobachtungswerte darstellt.

$$QA = Q_3 - Q_1 \quad (\text{I2: 45})$$

Gelegentlich wird auch der sog. **mittlere Quartilsabstand**

$$\overline{QA} = \frac{1}{2}(Q_3 - Q_1). \quad (\text{I2: 46})$$

als Streuungsmaß genutzt. Auch er ist robust gegen Ausreißer und kann wegen

$$\overline{QA} = \frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}([Q_3 - Q_2] - [Q_1 - Q_2])$$

als mittlerer Abstand der Quartile vom Median interpretiert werden.

2.2.2.2 Mittlere absolute Abweichung

Definieren wir den Abstand eines Messwertes x_i der *Urliste* vom Median x_{Med} als $|x_i - x_{Med}|$, so ist die **mittlere absolute Abweichung von x_{Med}** , MA, das arithmetische Mittel dieser Abweichungen, d.h. es gilt

$$MA = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - x_{Med}| . \quad (\text{I2: 47})$$

In Abschnitt I 2.2.1.2 haben wir gesehen, dass die Summe der absoluten Abweichungen vom Median minimal ist. Die mittlere absolute Abweichung ist daher immer dann ein sinnvolles Streuungsmaß, wenn der Median als Lagemaß verwendet wird. Das Zentrum einer Verteilung wird umso besser durch den Median beschrieben, je kleiner die mittlere absolute Abweichung ist. Gelegentlich wird bei der Berechnung von MA das arithmetische Mittel \bar{x} herangezogen. Dadurch geht jedoch die Optimalitätseigenschaft verloren. Wird das arithmetische Mittel als Lagemaß verwendet, wird die Streuung meist durch die sog. Standardabweichung (vgl. Abschnitt I 2.2.2.3) gemessen.

MA aus der Häufigkeitsverteilung:

Bei Vorliegen einer Häufigkeitsverteilung können wir MA unter Verwendung von absoluten und relativen Häufigkeiten bestimmen. Es können nämlich bestimmte absolute Abweichungen mehrfach auftreten, sodass wir

$$MA = \frac{1}{n} \cdot \sum_{j=1}^k h(a_j) \cdot |a_j - x_{Med}| = \sum_{j=1}^k f(a_j) \cdot |a_j - x_{Med}| \quad (\text{I2: 48})$$

festhalten können.

MA aus der Häufigkeitsverteilung klassierter Daten:

Liegt das Datenmaterial in klassierter Form vor, werden an Stelle von a_j in Formel (I2: 48) die Klassenmitten a_j^m und die jeweiligen absoluten bzw. relativen Klassenhäufigkeiten herangezogen. Mit einer derartigen Vorgehensweise erhalten wir jedoch wieder nur eine näherungsweise Lösung.

Beispiel I 2-24

Berechnung der mittleren absoluten Abweichung

Berechnen wir für unsere verschiedenen Währungsverteilungen aus Beispiel I 2-23 die mittlere absolute Abweichung bezüglich des Median. Dazu benötigen wir zunächst die Mediane der Verteilungen. Für den A-Dollar ergibt sich $x_{Med} = 1,20$ A\$ und für den B-Dollar $x_{Med} = 2,05$ B\$. Wir erhalten damit den bereits in den Tabellen von Beispiel I 2-23 enthaltenen Nebenrechnungen folgende Ergebnisse:

$$\text{A-Dollar: } MA = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - x_{Med}| = \frac{1}{10} \cdot 0,60 = 0,06 \text{ A\$}$$

$$\text{B-Dollar: } MA = \frac{1}{n} \cdot \sum_{j=1}^k h_j \cdot |a_j - x_{Med}| = \frac{1}{10} \cdot 3,20 = 0,32 \text{ B\$}$$

Eigenschaften der mittleren absoluten Abweichung:

- Die mittlere absolute Abweichung nimmt ihr Minimum an, wenn sie mittels x_{Med} berechnet wird.
- Im Vergleich zur Varianz, die wir unter I 2.2.2.3 vorstellen, weist sie nur wenige befriedigende mathematische Eigenschaften auf. Deswegen wird sie in der Praxis selten verwendet.
- Die mittlere absolute Abweichung reagiert im Vergleich zur Varianz weniger stark auf Ausreißer.
- Werden die Merkmalswerte x_i linear transformiert ($y_i = a + b \cdot x_i$, mit $a, b \in \mathbb{R}$ und $b \neq 0$), so kann die neue mittlere absolute Abweichung $MA_{y_{\text{Med}}}$ aus der mittleren absoluten Abweichung $MA_{x_{\text{Med}}}$ der Ausgangsdaten hergeleitet werden, ohne dass eine Neuberechnung notwendig wird. Dies geschieht über

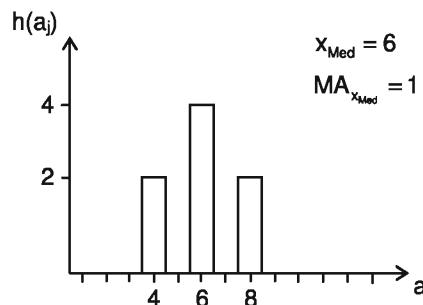
$$MA_{y_{\text{Med}}} = |b| \cdot MA_{x_{\text{Med}}} . \quad (\text{I2: 49})$$

Wir erkennen daran, dass eine Transformation der Form $y_i = a + x_i$ ($b = 1$) die Streuung der Daten nicht verändert. Die Transformation bewirkt lediglich eine *Lageverschiebung* der Daten. Erst die Multiplikation der Ausgangsdaten mit einer Konstanten $b \neq 1$ führt zu einer Änderung der Streuung.

Beispiel I 2-25

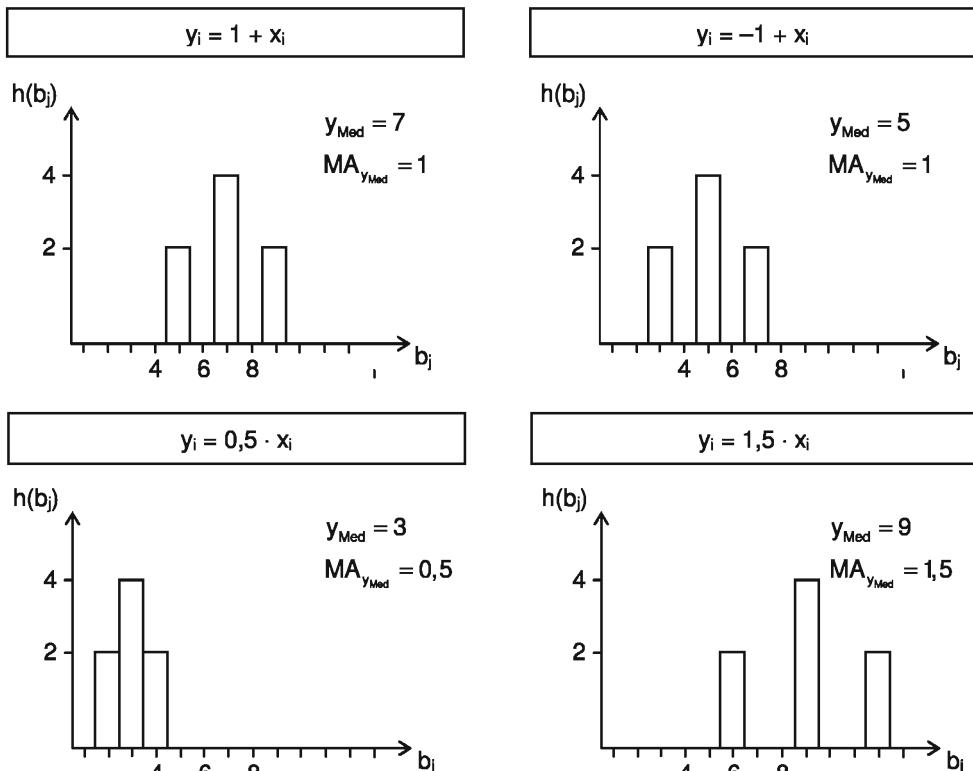
Eigenschaften der mittleren absoluten Abweichung

Betrachten wir die Auswirkungen verschiedener Transformationen einer einfachen Ausgangsverteilung auf die neue mittlere absolute Abweichung.



Addieren wir zu den Beobachtungen des Merkmals X die Konstante $a = 1$, so bewirkt dies eine Rechtsverschiebung der Verteilung, sodass sich ihre Lageparameter bzw. hier der Median ändert. Die Streuung wird davon nicht beeinflusst. Bei $a = -1$ kommt es zu einer Linksverschiebung bei unveränderter Streuung. Die Grafiken auf der Folgeseite veranschaulichen dies.

Multiplizieren wir mit einem Faktor zwischen 0 und 1 bzw. hier 0,5, so ändert sich sowohl Lage als auch Streuung der Verteilung. Wir erkennen eine Linksverschiebung bei gleichzeitiger Verringerung der Streuung. Für einen Faktor größer als 1 bzw. hier 1,5 beobachten eine Rechtsverschiebung bei Erhöhung der Streuung.



2.2.2.3 Varianz und Standardabweichung

Die Varianz s^2 bzw. die sich aus ihr ergebende Standardabweichung s ist das in der Praxis am häufigsten verwendete Streuungsmaß. s^2 ist definiert als arithmetisches Mittel der quadrierten Abweichungen der einzelnen Merkmalswerte vom arithmetischen Mittel \bar{x} . Gerade daher wird die Varianz auch häufig als **mittlere quadratische Abweichung** bezeichnet. s ist die positive Wurzel aus der Varianz.

Die Varianz und die Standardabweichung werden stets in Bezug auf das arithmetische Mittel berechnet, da für \bar{x} die Summe der quadrierten Abweichungen nach der Optimalitätseigenschaft des arithmetischen Mittels minimal ist. Das Zentrum einer Verteilung wird also umso besser durch das arithmetische Mittel beschrieben, je kleiner die Varianz bzw. die Standardabweichung ist.

Varianz aus der Urliste:

Die Varianz lässt sich in der Urliste über

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{mit} \quad \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (\text{I2: 50})$$

berechnen. Die darin zur Anwendung kommende Quadrierung der Abweichungen dient auf der einen Seite dazu, *großen* im Vergleich zu kleinen *Abweichungen mehr Gewicht* zu verleihen. Dies unterscheidet die Varianz von der mittleren ab-

soluten Abweichung. Andererseits wird durch die Quadrierung verhindert, dass sich positive und negative Abweichungen gegenseitig aufheben.

Zur Rechenvereinfachung lässt sich die Varianzformel umformen. Es handelt sich bei dieser Umformung um einen *Spezialfall* des sog. *Verschiebungssatzes*:

$$\text{Version A: } \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i - c)^2 - n \cdot (\bar{x} - c)^2 \quad (\text{I2: 51a})$$

$$\text{Version B: } \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - c)^2 - (\bar{x} - c)^2 \quad (\text{I2: 51b})$$

Für Interessierte führen wir die Herleitung des Verschiebungssatzes im Folgenden kurz auf. Sie kann jedoch auch übersprungen werden, da sie für das Verständnis des Folgetextes nicht relevant ist. Version A erhalten wir wie folgt:

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i - c - (\bar{x} - c))^2 \\ &= \sum_{i=1}^n (x_i - c)^2 - 2 \cdot \sum_{i=1}^n (x_i - c) \cdot (\bar{x} - c) + \sum_{i=1}^n (\bar{x} - c)^2 \\ &= \sum_{i=1}^n (x_i - c)^2 - 2 \cdot (\bar{x} - c) \cdot \underbrace{\sum_{i=1}^n (x_i - c)}_{n \cdot \bar{x} - n \cdot c} + n \cdot (\bar{x} - c)^2 \\ &= \sum_{i=1}^n (x_i - c)^2 - 2 \cdot n \cdot (\bar{x} - c)^2 + n \cdot (\bar{x} - c)^2 \\ &= \sum_{i=1}^n (x_i - c)^2 - n \cdot (\bar{x} - c)^2 \end{aligned}$$

Dividieren wir beide Seiten von Version A durch n, erhalten wir Version B.

Setzen wir nun in Version B des Verschiebungssatzes $c = 0$, so erhalten wir die folgende alternative Berechnungsformel für die Varianz aus der Urliste:

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (\text{I2: 52})$$

Da diese Formel unter Statistikneulingen oft zu Verwirrung führt, wollen wir klar darauf hinweisen, dass bei der Berechnung zunächst die Summe der quadrierten x_i durch n dividiert wird und erst im Anschluss daran das quadrierte arithmetische Mittel abgezogen wird. Analog gilt dies auch für die Formeln in (I2: 55).

Varianz aus der Häufigkeitsverteilung:

Liegt eine Häufigkeitsverteilung mit k Merkmalswerten a_1, a_2, \dots, a_k mit den zugehörigen absoluten Häufigkeiten $h(a_j)$ bzw. relativen Häufigkeiten $f(a_j)$ vor, so erfolgt die Berechnung der Varianz nach den folgenden Formeln:

$$s^2 = \frac{1}{n} \cdot \sum_{j=1}^k h(a_j) \cdot (a_j - \bar{x})^2 \quad \text{mit} \quad \bar{x} = \frac{1}{n} \cdot \sum_{j=1}^k h(a_j) \cdot a_j \quad (\text{I2: 53})$$

$$s^2 = \sum_{j=1}^k f(a_j) \cdot (a_j - \bar{x})^2 \quad \text{mit} \quad \bar{x} = \sum_{j=1}^k f(a_j) \cdot a_j \quad (\text{I2: 54})$$

Unsere Rechenvereinfachung (I2: 52) liefert daraus

$$s^2 = \frac{1}{n} \cdot \sum_{j=1}^k h(a_j) \cdot a_j^2 - \bar{x}^2 \quad \text{bzw.} \quad s^2 = \sum_{j=1}^k f(a_j) \cdot a_j^2 - \bar{x}^2. \quad (\text{I2: 55})$$

Zur Erleichterung einer von Hand durchgeführten Varianzberechnung empfiehlt es sich, die *Arbeitstabelle* aus Abbildung I 13 heranzuziehen. Wir erhalten daraus die Varianz, indem wir vom Mittelwert (absolute Häufigkeiten) bzw. der Summe (relative Häufigkeiten) der letzten Spalte \bar{x}^2 subtrahieren. \bar{x} können wir dabei aus der jeweiligen vorletzten Spalte ermitteln.

j	a_j	$h(a_j)$	$h(a_j) \cdot a_j$	$h(a_j) \cdot a_j^2$	$f(a_j)$	$f(a_j) \cdot a_j$	$f(a_j) \cdot a_j^2$
1
...
k
Σ	-	n	1,00

Abbildung I 13: Arbeitstabelle zur Varianzberechnung

Varianz aus der Häufigkeitsverteilung klassierter Daten:

Im Falle klassierter Daten werden zur näherungsweisen Berechnung in obigen Formeln wieder lediglich die Werte a_j durch die Klassenmitten a_j^m und die Häufigkeiten $h(a_j)$ und $f(a_j)$ durch die Klassenhäufigkeiten h_j und f_j ersetzt.

Beispiel I 2-26

Berechnung von Varianz und Standardabweichung

Berechnen wir für unser Währungsbeispiel I 2-23 die Varianzen und Standardabweichungen. Wir erhalten unter Verwendung der bereits in Beispiel I 2-23 enthaltenen Arbeitstabellen und der arithmetischen Mittel

$$\text{A-Dollar: } \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{10} \cdot 12,00 = 1,20 \text{ A\$}$$

$$\text{B-Dollar: } \bar{x} = \frac{1}{n} \cdot \sum_{j=1}^k h_j \cdot a_j = \frac{1}{10} \cdot 20,00 = 2,00 \text{ B\$}$$

folgende Ergebnisse:

$$\text{A-Dollar: } s^2 = \frac{1}{n} \cdot \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{14,44}{10} - 1,20^2 = 0,0040 \text{ A\2$

$$s = \sqrt{0,0040} = 0,0632 \text{ A\$}$$

$$\text{B-Dollar: } s^2 = \frac{1}{n} \cdot \sum_{j=1}^k h_j \cdot a_j^2 - \bar{x}^2 = \frac{41,08}{10} - 2,00^2 = 0,1080 \text{ B\2$

$$s = \sqrt{0,1080} = 0,3286 \text{ B\$}$$

Eigenschaften der Varianz (und der Standardabweichung):

- Die Varianz s^2 (und damit auch die Standardabweichung $s = \sqrt{s^2}$) ist stets größer oder gleich Null. Nimmt s^2 den Wert Null an, so liegt überhaupt keine Streuung vor, d.h. alle Merkmalswerte sind identisch.
- Die Dimension der Varianz ist das Quadrat der Dimension der einzelnen Beobachtungswerte. Die Dimension der Standardabweichung stimmt mit der der betrachteten Merkmalswerte überein. Besitzt also z.B. ein Merkmal die Einheit Euro, so ist die Dimension der Varianz €^2 und die der Standardabweichung € .
- Aufgrund der gleichen Dimension ist die Standardabweichung im Gegensatz zur Varianz ein mit der mittleren absoluten Abweichung vergleichbares Maß. Dabei gilt stets die Ungleichung

$$s \geq \text{MA}_{\bar{x}} \geq \text{MA}_{x_{\text{Med}}} . \quad (\text{I2: 56})$$

- Verteilungen können bei gleichen arithmetischen Mitteln trotzdem unterschiedliche Varianzen und damit auch Standardabweichungen aufweisen.

Beispiel I 2-27

Arithmetisches Mittel und Varianz

Für zwei Unternehmen wurde von einem Marktforschungsinstitut über 4 Jahre eine Kundenzufriedenheitsanalyse durchgeführt. Der Anteil zufriedener Kunden ist in folgender Tabelle angegeben.

Jahr	2002	2003	2004	2005	2006
Unternehmen A	0,70	0,95	0,50	0,55	0,80
Unternehmen B	0,30	0,70	1,00	0,60	0,90

Für beide Unternehmen ergibt sich eine durchschnittliche Kundenzufriedenheit von $\bar{x} = 0,7$ bzw. 70 %. Bei Unternehmen B ($s = 0,060$) ist allerdings eine stärkere Fluktuation der Kundenzufriedenheit festzustellen als bei Unternehmen A ($s = 0,027$).

- Das sog. Tschebyscheff-Theorem (vgl. Abschnitt II 2.9) besagt, dass für beliebige Verteilungen mindestens $1 - 1/k^2(\%)$ der Beobachtungen in einem Intervall $[\bar{x} - k \cdot s; \bar{x} + k \cdot s]$ liegen.
- Praxisregel: Für eingipflige Verteilungen gilt,² dass im Intervall
 - $[\bar{x} - s; \bar{x} + s]$ zumeist zwischen 60 % und 80 %, für fast symmetrische Verteilungen ca. 70 % und für sehr schiefe Verteilungen fast 90 % der Beobachtungswerte liegen.
 - $[\bar{x} - 2 \cdot s; \bar{x} + 2 \cdot s]$ für symmetrische Verteilungen ca. 95 % und für sehr schiefe Verteilungen bis fast 100 % der Beobachtungswerte liegen.
 - $[\bar{x} - 3 \cdot s; \bar{x} + 3 \cdot s]$ knapp 100 % der Beobachtungen liegen.
- Wir hatten bereits angesprochen, dass der Grund dafür, dass die Varianz bezüglich \bar{x} bestimmt wird, darin liegt, dass die Summe der quadratischen Abweichungen der einzelnen Merkmalsausprägungen x_i von einem beliebigen

² Vgl. Bamberg, G., Baur, F. (2004), Kapitel 3.

Wert λ dann ein Minimum hat, wenn $\lambda = \bar{x}$ ist. Dies hatten wir bereits als Optimalitätseigenschaft des arithmetischen Mittels in (I2: 26) festgehalten. Ist die Summe der quadratischen Abweichungen minimal, ist es natürlich auch ihr Mittelwert und damit die Varianz.

- **Lineare Transformationen:**

Transformieren wir die Einzelwerte x_i in der linearen Form $y_i = a + b \cdot x_i$, mit $a, b \in \mathbb{R}$ und $b \neq 0$, so wissen wir bereits, dass nach (I2: 27) für das arithmetische Mittel $\bar{y} = a + b \cdot \bar{x}$ gilt. Für die neue Varianz s_y^2 und die neue Standardabweichung s_y gilt bei einer derartigen Transformation

$$\begin{aligned}s_y^2 &= \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \cdot \sum_{i=1}^n (a + b \cdot x_i - a - b \cdot \bar{x})^2 \\ &= \frac{b^2}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = b^2 \cdot s_x^2,\end{aligned}$$

sodass wir insgesamt

$$s_y^2 = b^2 \cdot s_x^2 \quad \text{und} \quad s_y = |b| \cdot s_x \quad (\text{I2: 57})$$

festhalten können. Dies zeigt, dass eine Transformation der Form $y_i = a + x_i$, d.h. mit $b = 1$, wie wir auch bereits bei der mittleren absoluten Abweichung gesehen haben, keinerlei Einfluss auf Streuung und damit die Varianz und die Standardabweichung des Datenmaterials hat. Es gilt nämlich dann $s_y^2 = s_x^2$ bzw. $s_y = s_x$.

Beispiel I 2-28

Lineare Transformationen und Varianz

Innerhalb einer Woche fielen vor Euroumstellung in der Filiale einer Bank folgende Kosten (in DM) für den Versand von Geschäftsbriefen an:

Mo	Di	Mi	Do	Fr
5,50	9,00	3,00	12,00	1,50

Es ergaben sich damit durchschnittliche Versandkosten von

$$\bar{x} = \frac{1}{5} \cdot (5,50 + 9,00 + 3,00 + 12,00 + 1,50) = 6,20 \text{ DM}$$

bei einer Varianz und einer Standardabweichung von

$$\begin{aligned}s_x^2 &= \frac{1}{5} \cdot (5,50^2 + 9,00^2 + 3,00^2 + 12,00^2 + 1,50^2) - 6,20^2 = 14,86 \text{ DM}^2 \\ s_x &= \sqrt{14,86} = 3,85 \text{ DM}.\end{aligned}$$

Im Zuge der Euroumstellung werden nun die DM- in Euro-Werte umgerechnet. Die dazugehörige Transformationsformel lautet

$$y_i = \frac{1}{1,95583} \cdot x_i \quad \text{mit} \quad a = 0 \quad \text{und} \quad b = \frac{1}{1,95583}.$$

Ohne auf die neuen Euro-Werte zugreifen zu müssen, können wir neben dem neuen Mittelwert $\bar{y} = \frac{1}{1,95583} \cdot \bar{x} = 3,17 \text{ €}$ die neue Varianz und Standardabweichung bestimmen:

$$s_y^2 = \left(\frac{1}{1,95583}\right)^2 \cdot 14,86 = 3,88 \text{ €}^2 \quad \text{und} \quad s_y = \left|\frac{1}{1,95583}\right| \cdot 3,85 = 1,97 \text{ €}.$$

■ **Zentrierung, Standardisierung:**

Liegt eine Urliste x_1, x_2, \dots, x_n mit dem arithmetischen Mittel \bar{x} und der Standardabweichung s_x vor, so heißt eine Lineartransformation

$$y_i = x_i - \bar{x} \Leftrightarrow y_i = a + b \cdot x_i \quad \text{mit } a = -\bar{x} \text{ und } b = 1 \quad (\text{I2: 58a})$$

Zentrierung. Das Resultat einer solchen Zentrierung ist

$$\bar{y} = 0 \text{ und } s_y = s_x. \quad (\text{I2: 58b})$$

Wir erhalten also ein neues arithmetisches Mittel von Null und die gleiche Standardabweichung (Streuung) wie die Ausgangsdaten.

Von besonderer praktischer Bedeutung ist die sog. *Standardisierung*. Darunter verstehen wir eine Lineartransformation der Form

$$z_i = \frac{x_i - \bar{x}}{s_x} \Leftrightarrow z_i = a + b \cdot x_i \quad \text{mit } a = -\frac{\bar{x}}{s_x} \text{ und } b = \frac{1}{s_x}. \quad (\text{I2: 59a})$$

Standardisierte Daten besitzen die besonderen Eigenschaften

$$\bar{z} = 0 \text{ und } s_z = 1, \quad (\text{I2: 59b})$$

was wir anhand von

$$\bar{z} = a + b \cdot \bar{x} = -\frac{\bar{x}}{s_x} + \frac{1}{s_x} \cdot \bar{x} = 0$$

$$s_z^2 = b^2 \cdot s_x^2 = \left(\frac{1}{s_x} \right)^2 \cdot s_x^2 = 1 \rightarrow s_z = 1$$

anschaulich zeigen können. Die Tatsache, dass durch die Standardisierung der Merkmalswerte ein arithmetisches Mittel von Null entsteht, bietet den Vorteil, dass nun sofort erkennbar ist, ob die jeweiligen Merkmalswerte größer oder kleiner als das arithmetische Mittel sind. Alle neuen positiven Merkmalswerte liegen über und alle neuen negativen Merkmalswerte unter dem Durchschnitt. Wir sind außerdem in der Lage zu erkennen, um wie viele Standardabweichungen der jeweilige Wert vom arithmetischen Mittel abweicht. Dies wird im nachfolgenden Beispiel klar.

Beispiel I 2-29

Standardisierung

Für ein bestimmtes Jahr sind die Zahlen der Übernachtungen pro Monat in einem Hotel bekannt:

Jan	Feb	Mär	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez
80	10	30	80	120	160	200	200	200	120	10	10

Wir führen nun eine Standardisierung dieser Werte mittels der Formel

$$z_i = \frac{x_i - \bar{x}}{s} = \frac{x_i - 101,67}{73,24}$$

durch, welche folgende Resultate liefert:

Jan	Feb	Mär	Apr	Mai	Jun
-0,30	-1,25	-0,98	-0,30	0,25	0,80
Jul	Aug	Sep	Okt	Nov	Dez
1,34	1,34	1,34	0,25	-1,25	-1,25

Wie wir erkennen, sind alle Merkmalswerte, die ehemals größer als das arithmetische Mittel waren, nun positiv, alle anderen hingegen negativ. Außerdem sehen wir z.B., dass die Anzahl der Übernachtungen im Juli um 1,34 Standardabweichungen über dem arithmetischen Mittel liegt, was wir über die Rechnung $101,67 + 73,24 \cdot 1,34 \approx 200$ zeigen können.

- Für die Varianz einer Gesamtheit n , die sich aus zwei Teilgesamtheiten n_1 und n_2 mit den Merkmalswerten x_{1i} und x_{2i} zusammensetzt, gilt allgemein

$$s_{\text{ges}}^2 = \frac{1}{n_1 + n_2} \cdot \left(\sum_{i=1}^{n_1} (x_{1i} - \bar{x})^2 + \sum_{i=1}^{n_2} (x_{2i} - \bar{x})^2 \right), \quad (\text{I2: 60})$$

wobei sich darin \bar{x} nach (I2: 28) ergibt. Bezeichnen wir mit \bar{x}_1 bzw. \bar{x}_2 und s_1^2 bzw. s_2^2 die arithmetischen Mittel und Varianzen der beiden Teilgesamtheiten, so ergibt sich die Gesamtvarianz ebenfalls mit

$$s_{\text{ges}}^2 = \frac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2} + \frac{n_1 \cdot (\bar{x}_1 - \bar{x})^2 + n_2 \cdot (\bar{x}_2 - \bar{x})^2}{n_1 + n_2}. \quad (\text{I2: 61})$$

Beispiel I 2-30

Gesamtvarianz aus Teilgesamtheiten

Die Leistungsbeurteilung (Bewertung 1 bis 7) in zwei Unterabteilungen A und B mit 12 und 14 Mitarbeitern der Controlling-Abteilung eines Unternehmens ergab folgendes Ergebnis:

Abteilung A: 1 2 2 2 3 3 3 3 4 4 4 5

Abteilung B: 2 3 3 4 4 4 5 5 5 5 6 6 7

Wir erhalten daraus

$$\bar{x}_A = 3,00 \quad s_A^2 = 1,17$$

$$\bar{x}_B = 4,50 \quad s_B^2 = 1,68.$$

Um nun die Varianz der Gesamtabteilung Controlling zu bestimmen, könnten wir die Beurteilungsergebnisse beider Abteilungen zusammenlegen und aus der so entstehenden Gesamtheit die Varianz berechnen. Schneller erhalten wir das Ergebnis, wenn wir zunächst nach (I2: 28)

$$\bar{x}_{\text{ges}} = \frac{12 \cdot 3,00 + 14 \cdot 4,50}{12 + 14} = 3,81$$

bestimmen und dieses Ergebnis in (I2: 61) nutzen, um

$$s_{\text{ges}}^2 = \frac{12 \cdot 1,17 + 14 \cdot 1,68}{12 + 14} + \frac{12 \cdot (3,00 - 3,81)^2 + 14 \cdot (4,50 - 3,81)^2}{12 + 14} = 2,00$$

zu ermitteln.

Für m Teilgesamtheiten, deren jeweilige arithmetische Mittel $\bar{x}_1, \dots, \bar{x}_m$ und Varianzen s_1^2, \dots, s_m^2 sowie das Gesamtmittel \bar{x}_{ges} nach (I2: 29) bekannt sind, berechnen wir die Varianz für die Gesamtheit $n = n_1 + \dots + n_m$ nach

$$\underbrace{s_{\text{ges}}^2}_{\text{Gesamte Varianz}} = \frac{1}{n} \cdot \sum_{k=1}^m n_k \cdot s_k^2 + \frac{1}{n} \cdot \sum_{k=1}^m n_k \cdot (\bar{x}_k - \bar{x}_{\text{ges}})^2, \quad (\text{I2: 62})$$

Varianz innerhalb der Teilgesamtheiten

Varianz zwischen den Teilgesamtheiten

wobei der erste Summand häufig als *interne Varianz* (Varianz innerhalb der Teilgesamtheiten) oder gewogenes Mittel der Teilgesamtheitsvarianzen (Gewichte n_k/n) und der zweite als *externe Varianz* (Varianz zwischen den Teilgesamtheiten) bezeichnet wird. Die angegebene Berechnung wird deshalb auch als *Formel der Streuungszerlegung* bezeichnet. An dieser Darstellung lässt sich gut erkennen, ob sich die Varianz der Daten vor allem durch die Variation innerhalb der Gruppen oder durch Unterschiede zwischen den Gruppen ergibt.

- Praxishinweis: Auf Finanzmärkten wird häufig die Standardabweichung der Renditen (nicht der Kurse) als Maß für die Beschreibung des Risikos verschiedener Finanzanlagen verwendet. Wir sprechen in diesem Zusammenhang von der *Volatilität* (vgl. Abschnitt I 2.2.2.4).

2.2.2.4 Exkurs: Volatilität

Die Einschätzung des Risikos verschiedener Anlageformen (Aktien, Anleihen, etc.) ist in der finanzwirtschaftlichen Praxis von erheblicher Bedeutung, da das mit einer Anlage verbundene Risiko ein wesentliches Kriterium dafür ist, ob eine Investition erfolgt oder unterbleibt. Unter *Risiko* ist dabei allgemein die Möglichkeit eines unerwünschten Ausgangs der Investitionsaktivität zu verstehen. Da ein unerwünschter Ausgang von den Präferenzen und Zielsetzungen des Investors abhängig ist, kann er unterschiedliche Formen, wie etwa den Verlust von Kapital,³ die Verfehlung einer Mindestrendite oder lediglich die Renditeschwankung im Zeitablauf annehmen.

Wie im vorhergehenden Abschnitt erwähnt wurde, wird zur Beurteilung des Risikos von Finanzanlagen in der Praxis meist die Standardabweichung der Renditen r_i herangezogen.⁴ Es wird dabei jedoch nicht auf die Quadratwurzel aus (I2: 50) zurückgegriffen, sondern die Volatilität σ einer Finanzanlage über

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (r_i - \bar{r})^2} \quad (\text{I2: 63})$$

geschätzt.⁵ Wie die nachfolgenden Beispiele zeigen, hat dieses Maß vor allem beim Vergleich der Vorteilhaftigkeit von Wertpapieren und der Beurteilung der Fragilität der Lage auf Aktien- und Anleihenmärkten große Bedeutung.

³ Verfolgen Investoren das Ziel des Kapitalerhalts, so müssten sie unter Inflationsberücksichtigung keine Mindestrendite von 0 %, sondern eine in Höhe der Inflationsrate anstreben.

⁴ Zur Renditedefinition vgl. Abschnitt I 2.2.1.5 bzw. die Formeln (I2: 35) und (I2: 40).

⁵ Die genaue Bedeutung der Division durch $n-1$ anstatt durch n werden wir im Rahmen der induktiven Statistik in Kapitel III behandeln.

Beispiel I 2-31

Volatilität

Die Jahresrenditen einer Aktie A lagen in den Jahren 2002 bis 2006 bei -15% , 40% , 30% , -40% und 35% . Anhand dieser Daten und einer anderen Aktie B (mittlere Rendite von 6% und Standardabweichung der Rendite von 30% für den gleichen Zeitraum) wollen wir nun die Vorteilhaftigkeit des Papiers A beurteilen:

$$\bar{r} = \frac{1}{5} \cdot (-0,15 + 0,40 + 0,30 - 0,40 + 0,35) = 0,10 = 10\%$$

$$\hat{\sigma} = \sqrt{\frac{1}{5-1} \cdot [(-0,15 - 0,10)^2 + \dots + (0,35 - 0,10)^2]} = \sqrt{0,1263} = 35,54\%$$

Da das Papier B bei niedrigerer Durchschnittsrendite eine niedrigere Standardabweichung aufweist, hängt die Vorteilhaftigkeit des Wertpapiers A von den Präferenzen des Investors ab. Ist der Investor bereit, für höhere Ertragschancen zusätzliche Risiken einzugehen, so wird er sich für Papier A entscheiden.

Würden beide Papiere unter sonst unveränderten Bedingungen die gleiche Durchschnittsrendite aufweisen, so wäre Papier B als vorteilhafter einzustufen, da bei geringerem Risiko die gleiche Durchschnittsrendite erzielbar ist wie bei Papier A.

Da die "normale" Standardabweichung aus (12: 63) auch positive Abweichungen vom Durchschnitt berücksichtigt, die ja für den Investor nicht unbedingt ein Risiko darstellen müssen, verwendet man zum Teil sog. *Downside-Risikomaße*, welche nur negative Abweichungen aufnehmen.⁶ Zu diesen zählt etwa die *Semi-Volatilität*, welche wir aus historischen Daten mittels

$$\hat{\sigma}_{\text{semi}} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (r_i - \bar{r})^2}$$

schätzen können, wobei \tilde{n} für die Anzahl negativer Abweichungen von \bar{r} steht.

Beispiel II 2-32

Semi-Volatilität

Über einen Zeitraum von $n = 8$ Werktagen ergaben sich für eine Aktie folgende Renditen (mit $\bar{r} = 0,0025$):

r_i	-0,02	0,02	-0,03	0,01	0,00	-0,01	0,01	0,04
$r_i - \bar{r}$	-0,0225	0,0175	-0,0325	0,0075	-0,0025	-0,0125	0,0075	0,0375

Mit $\tilde{n} = 4$ negativen Abweichungen ergibt sich daraus die Semi-Volatilität zu

$$\hat{\sigma}_{\text{semi}} = \sqrt{\frac{1}{8} \cdot [(-0,0225)^2 + (-0,0325)^2 + (-0,0025)^2 + (-0,0125)^2]} = 0,0147.$$

Diese Semi-Volatilität fällt hier natürlich niedriger aus als die Volatilität, die auch positive Abweichungen berücksichtigt ($\hat{\sigma} = 0,0225$), da erstere nur 4 quadrierte Abweichungen aufnimmt.

⁶ Einen umfassenden Überblick über Risikomaße und deren Implementierung geben z.B. Auer, B.R., Seitz, F. (2008). Praktische Anwendungen finden sich z.B. bei Rottmann, H., Franz, T. (2007, 2008).

Die Volatilitäten aus den Beispielen II 2-31 und II 2-32 sind nicht direkt miteinander vergleichbar, da (abgesehen von der Tatsache, dass verschiedene Risikomaße verwendet werden) in Beispiel II 2-31 Jahresdaten und in Beispiel II 2-32 Tagesdaten verwendet wurden. In der Praxis wird daher eine sog. *Annualisierung* mit folgenden Formeln vorgenommen:

$$\begin{aligned} \text{bei Tagesdaten: } \hat{\sigma}_{\text{annualisiert}} &= \hat{\sigma}_{\text{Tagesdaten}} \cdot \sqrt{252} \\ \text{bei Monatsdaten: } \hat{\sigma}_{\text{annualisiert}} &= \hat{\sigma}_{\text{Monatsdaten}} \cdot \sqrt{12} \\ \text{bei Quartalsdaten: } \hat{\sigma}_{\text{annualisiert}} &= \hat{\sigma}_{\text{Quartalsdaten}} \cdot \sqrt{4} \end{aligned}$$

In Beispiel II 2-32 ergibt sich somit $\hat{\sigma}_{\text{annualisiert}} = 0,0225 \cdot \sqrt{252} = 0,3572$. Wie wir unschwer erkennen, ist die Zahl unter der Wurzel stets die Anzahl der zu berücksichtigenden Subperioden. Bei Tagesdaten ist dabei zu beachten, dass Handelstage und nicht tatsächliche Kalendertage zu verwenden sind.

Beispiel I 2-33

Volatilitäten auf Aktien- und Anleihenmärkten

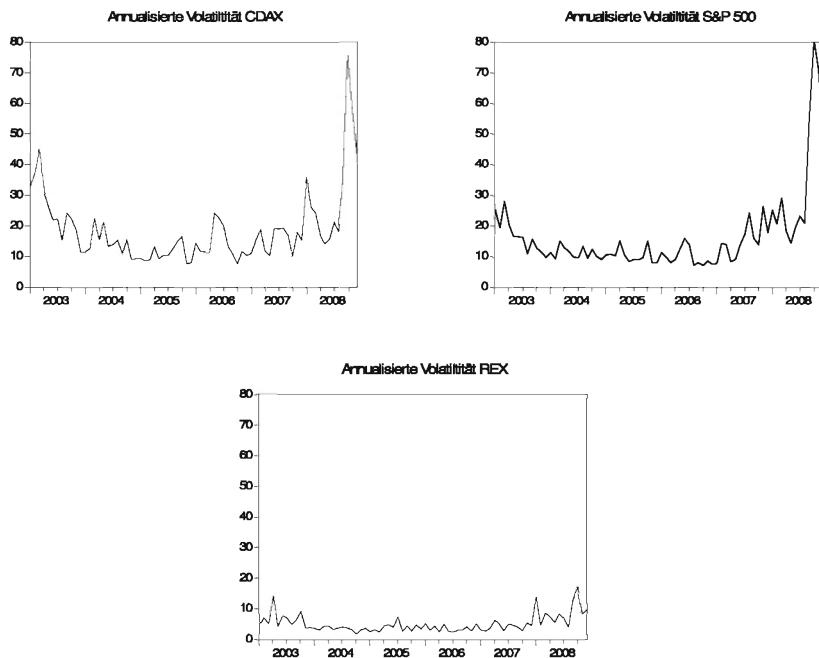


In der Praxis geben uns Volatilitäten vor allem Auskunft über die Fragilität der Lage auf den Finanzmärkten. Betrachten wir dazu folgende Tabelle, die die annualisierten Renditestandardabweichungen der Aktienindizes CDAX (Deutschland) und S&P 500 (USA) sowie des REX (Index für deutsche Staatsanleihen) für die Jahre 2003 bis 2008 zeigt. Die annualisierten Standardabweichungen wurden aus den Tagesrenditen der jeweiligen Jahre ermittelt, d.h. ihnen liegt die oben behandelte Formel $\hat{\sigma}_{\text{annualisiert}} = \hat{\sigma}_{\text{Tagesdaten}} \cdot \sqrt{252}$ zugrunde.

	CDAX	S&P 500	REX
2003	27,0746	17,0676	3,7965
2004	14,5774	11,0935	2,7339
2005	11,1793	10,2835	2,9039
2006	14,9402	10,0253	2,5242
2007	15,6750	15,9856	2,9950
2008	36,3609	40,9733	7,0120

Wie deutlich zu erkennen ist, kommt es im Zuge der im Jahr 2008 verstärkt zu Tage tretenden Finanzmarktkrise auf allen Märkten (Aktien- und Anleihenmarkt) zu stark gestiegenen Volatilitäten. Insbesondere auf dem amerikanischen Aktienmarkt ist eine deutliche Steigerung der Renditeschwankungen zu verzeichnen. Selbst die Anleihenmärkte, die gewöhnlich durch niedrigere Volatilitäten gekennzeichnet sind, werden durch die Krise in Mitleidenschaft gezogen.

Zur genaueren Analyse (bzw. besseren Verdeutlichung) der sich verändernden Volatilität auf den Märkten empfiehlt sich auch eine Betrachtung auf Monatsbasis. Wir berechnen dazu aus den Tagesrenditen für jeden Monat des Betrachtungszeitraums die geschätzte Volatilität und annualisieren diese. Die daraus resultierende Volatilitätsentwicklung ist in den nachfolgenden Grafiken abgebildet. Auch die monatliche Betrachtung macht deutlich, dass es krisenbedingt zu einem Volatitätsanstieg kam, der auf den Aktienmärkten stärker ausfiel als auf den Anleihenmärkten.



2.2.2.5 Variationskoeffizient

Soll die Streuung zweier Verteilungen mit *stark verschiedenen Mittelwerten* verglichen werden, sind die *absoluten Streuungsmaße* Varianz und Standardabweichung oftmals ungeeignet, da sie nicht das Niveau der Daten berücksichtigen. Nehmen wir etwa an, wir stellen für zwei Verteilungen A und B die Mittelwerte 10.000 und 1 und jeweils eine Varianz von 10 fest. Für Verteilung A würden wir die Varianz als gering bewerten, bei Verteilung B hingegen als eher groß. Ohne die gleichzeitige Betrachtung der Mittelwerte hätten wir eine solche Schlussfolgerung nicht treffen können. Zudem ist es in der Regel so, dass die absoluten Streuungsmaße bei zahlenmäßig größeren Merkmalswerten größer ausfallen als bei kleineren. So fällt z.B. die Standardabweichung bei Pkw-Preisen höher aus als jene bei Kaugummi-Preisen, was auch hier einen direkten Vergleich erschwert. Um dennoch aussagekräftige Vergleiche durchführen zu können, empfiehlt sich die Verwendung sog. *relativer Streuungsmaße*, die das Niveau der Daten berücksichtigen. Sie werden konstruiert, indem ein Streuungsmaß zu dem entsprechenden Lagemaß ins Verhältnis gesetzt wird. Relative Streuungsmaße sind dimensionslose Größen, d.h. sie besitzen keine Einheit, und sind daher besonders bei Vergleichen von Verteilungen mit unterschiedlichen Maßeinheiten wertvoll. Das gebräuchlichste relative Streuungsmaß ist der **Variationskoeffizient** VC, der auf der Standardabweichung und dem arithmetischen Mittel basiert:

$$VC = \frac{s}{\bar{x}} \quad (\text{I2: 64})$$

Der Variationskoeffizient wird in der Praxis *gelegentlich* als Maß für die *Streuung der Wechsel- oder Wertpapierkurse* verwendet. Generell hat sich jedoch die Standardabweichung der Renditen als Volatitätsmaß durchgesetzt.

Beispiel I 2-34
Variationskoeffizient und Aktienvolatilität

Der Kurs der X-Aktie weist in einem Zeitraum von 200 Handelstagen bei einem Mittelwert von $\bar{x} = 49,30 \text{ €}$ eine Standardabweichung von $s_x = 35,14 \text{ €}$ auf. Für den identischen Zeitraum ergab sich für die Y-Aktie ein Mittelwert von $\bar{y} = 385,17 \text{ €}$ und eine Standardabweichung von $s_y = 179,55 \text{ €}$.

Ein Vergleich der Variationskoeffizienten

$$VC_x = \frac{s_x}{|\bar{x}|} = \frac{35,14}{49,30} = 0,71$$

$$VC_y = \frac{s_y}{|\bar{y}|} = \frac{179,55}{385,17} = 0,47$$

zeigt, dass trotz geringerer Standardabweichung die X-Aktie stärker streute.

Beispiel I 2-35
Variationskoeffizienten im Basisbeispiel

Für unser Wechselkursbeispiel I 2-23 erhalten wir folgende Variationskoeffizienten:

$$\text{A-Dollar: } VC = \frac{s}{|\bar{x}|} = \frac{0,0632}{1,20} = 0,0527$$

$$\text{B-Dollar: } VC = \frac{s}{|\bar{x}|} = \frac{0,3286}{2,00} = 0,1643$$

Um nun einen umfassenden Variationsvergleich durchführen zu können, wollen wir die Ergebnisse der bisher berechneten Streuungsmaße (unter Zusatzangabe der dabei verwendeten Lagemaße) in folgender Tabelle zusammenfassen:

	A-Dollar	B-Dollar
x_{Med}	1,20 A\$	2,05 B\$
\bar{x}	1,20 A\$	2,00 B\$
SP	0,20 A\$	0,80 B\$
MA	0,06 A\$	0,32 B\$
s^2	0,0040 A\$ ²	0,1080 B\$ ²
s	0,0632 A\$	0,3286 B\$
VC	0,0527	0,1643

SP, MA, s^2 und s sind als absolute Maße der Streuung beim B-Dollar am größten. Auch der Variationskoeffizient als relatives Streuungsmaß erlaubt die Schlussfolgerung, dass der B-Dollar stärker streut als der A-Dollar. In diesem speziellen Fall macht es also keinen Unterschied, ob wir absolute oder relative Maße für den Streuungsvergleich heranziehen. Dies lässt sich damit begründen, dass hier anders als im Beispiel I 2-34 der Unterschied zwischen den Mittelwerten nicht sonderlich groß ist. Wir könnten also genau so gut auch auf die Verwendung relativer Maße verzichten.

Weitere relative Streuungsmaße lassen sich aus den Quartilen einer Verteilung bestimmen. So wird gelegentlich der sog. *Quartilskoeffizient* (oder Quartilsdispersionskoeffizient)

$$QK = \frac{Q_3 - Q_1}{|Q_2|} \quad (I2: 65)$$

berechnet. Unter Verwendung anderer Quantile können des Weiteren noch andere *Quantilsverhältnisse*, wie etwa $x_{0,9} / x_{0,1}$ oder $x_{0,9} / x_{0,5}$ zur Beschreibung der Streuung gebildet werden.

2.2.2.6 Box-Whisker-Plot

Neben den bisher behandelten Methoden werden in der Praxis zum Vergleich von Verteilungen gelegentlich sog. **Box-Whisker-Plots** eingesetzt. Sie fassen die in einem Datenbestand enthaltene Information mit Hilfe von 5 Zahlen, x_1 , Q_1 , x_{Med} , Q_3 , x_n , zusammen und stellen damit Lage, Streuung, Schiefe und Ausreißerwerte der Verteilung grafisch dar. Wie Abbildung I 14 zeigt, setzt sich ein einfacher Box-Whisker-Plot aus folgenden Bestandteilen zusammen:

- Skala der Merkmalswerte x_i
- Rechteck (Box) vom unteren Quartil Q_1 bis zum oberen Quartil Q_3 ,
- Senkrechte Striche zur Markierung von Median x_{Med} und der beiden Extremwerte x_1 und x_n ,
- Waagerechte Striche von der Box zu den beiden senkrechten Extremwertstrichen

Das Zentrum von mit Box-Whisker Plots visualisierten Verteilungen wird durch den Median x_{Med} lokalisiert. Die beiden Extremwerte x_1 und x_n informieren über den Datenausdehnungsbereich. Die Box zeigt den zentralen 50 %-Anteil der Daten. Die linke und rechte Seite der Box stellt jeweils 25 % der Daten dar. Die waagerechten Striche links und rechts von der Box symbolisieren jeweils 25 % der Randdaten.

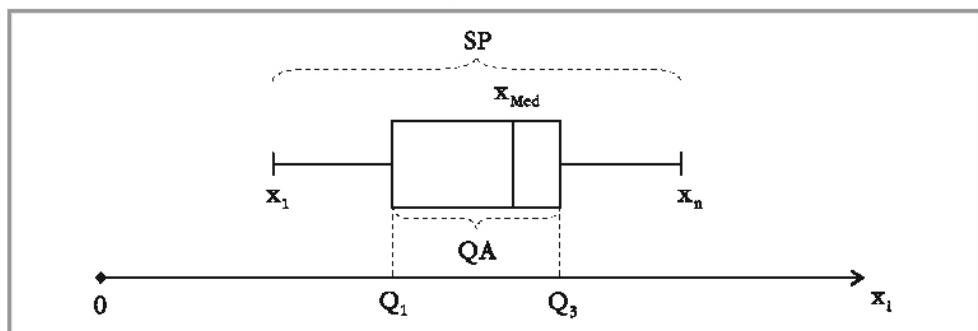


Abbildung I 14: Einfacher Box-Whisker-Plot

Hinsichtlich der *Schiefe* der Verteilung lässt sich mit einem Box-Whisker-Plot sagen, dass ein rechts (links) von der Mitte der Box liegender Median eine linksschiefe (rechtsschiefe) Verteilung aufzeigt. Befindet sich der Median in der Mitte der Box, liegt eine symmetrische Verteilung vor. Bei einer solchen Vorgehensweise berücksichtigen wir allerdings nur die mittleren 50 % der Beobachtungen zur Beurteilung der Schiefe.

Bei der Interpretation von Box-Whisker-Plots ist zu berücksichtigen, dass die Länge der waagerechten Striche von der Box zu den beiden Extrema durch wenige Ausreißerwerte stark beeinflusst werden kann. Deswegen werden extreme Werte meist gesondert behandelt. Dazu werden zunächst sog. *Ausreißerzäune* definiert, die zur Identifikation von Ausreißern dienen sollen. Diesen liegt der Gedanke zugrunde, dass Werte, die weit außerhalb der Box liegen, als potenzielle Ausreißer betrachtet werden können. Wir berechnen den sog. *inneren Zaun* als

$$[Q_1 - 1,5 \cdot QA; Q_3 + 1,5 \cdot QA]$$

und den sog. *äußeren Zaun* als

$$[Q_1 - 3 \cdot QA; Q_3 + 3 \cdot QA].$$

Mit diesen Zäunen können wir nun z.B. sagen, dass ein Wert, der über den inneren Zaun hinausgeht als potenzieller Ausreißer betrachtet werden kann. Ein solcher Ausreißer ist als weniger krass einzustufen als einer, der den äußeren Zaun überschreitet. In der Praxis werden für Werte innerhalb des inneren Zauns Box-Whisker-Plots nach der eingangs beschriebenen Methodik gezeichnet. Treten Werte außerhalb des inneren Zauns auf, werden diese als Zahlenwerte im Plot angegeben, um eine zu starke Streckung der Grafik über die Spannweite SP zu vermeiden. Die senkrechten Striche dienen in einem solchen Plot für den kleinsten Wert x_u , der größer oder gleich $Q_1 - 1,5 \cdot QA$ ist, und den größten Wert x_o , der kleiner oder gleich $Q_3 + 1,5 \cdot QA$ ist. Abbildung I 15 skizziert dies.

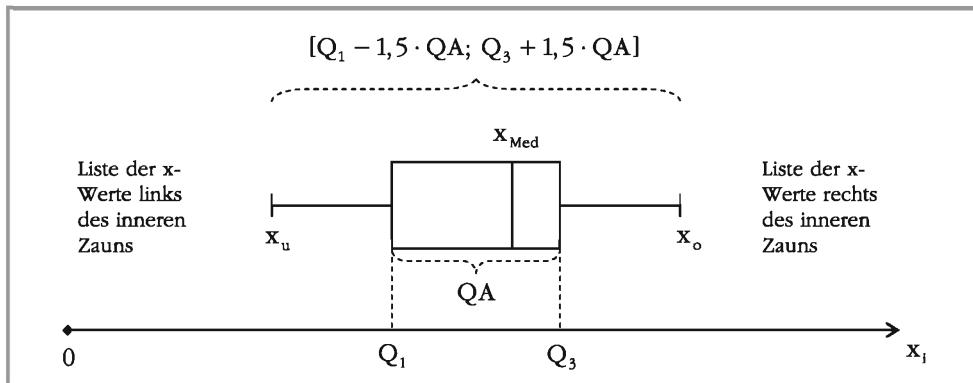


Abbildung I 15: Spezieller Box-Whisker-Plot

2.2.3 Momente und Schiefemaße

Nach Lage und Streuung ist die Schiefe ein drittes wichtiges Einzelcharakteristikum von Häufigkeitsverteilungen. Wie wir bereits gesehen haben, führen die Lageparameter x_{Mod} , x_{Med} und \bar{x} nur bei symmetrischen Verteilungen zu einheitlichen Ergebnissen. Bei asymmetrischen oder schiefen Verteilungen fallen die Ergebnisse auseinander. Die Auswahl und Interpretation von Lagemaßen wird hier zum Problem. Um das Ausmaß der Schiefe einer Verteilung abschätzen zu können, ist es sinnvoll, die Schiefe anhand einer Kennzahl zu messen. Hier bietet sich der **Schiefekoeffizient** an, der aus dem **Konzept der Momente** entsteht.

2.2.3.1 Empirische Momente

Beim Konzept der Momente handelt es sich um ein einheitliches System von Kennzahlen für metrische Merkmale, mit dem wir in der Lage sind verschiedene Eigenschaften von Häufigkeitsverteilungen (insbesondere auch Lage, Streuung und Schiefe) zu beschreiben.

Ist eine Urliste x_1, x_2, \dots, x_n gegeben und ist δ eine natürliche Zahl, so erhalten wir das sog. *δ -te Moment um Null* als

$$m_\delta(0) = \frac{1}{n} \sum_{i=1}^n x_i^\delta \quad (\text{I2: 66})$$

und das *zentrale Moment der Ordnung δ* als

$$m_\delta(\bar{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^\delta. \quad (\text{I2: 67})$$

Einige dieser Momente sind uns bereits begegnet. Das erste Moment um Null entspricht nämlich dem arithmetischen Mittel

$$m_1(0) = \frac{1}{n} \sum_{i=1}^n x_i^1 = \bar{x}$$

und das zentrale Moment zweiter Ordnung der Varianz

$$m_2(\bar{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2.$$

Aufgrund der Zentraleigenschaft des arithmetischen Mittels ist das zentrale Moment erster Ordnung immer Null:

$$m_1(\bar{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^1 = \frac{1}{n} \sum_{i=1}^n x_i - \bar{x} = 0$$

2.2.3.2 Schiefemaße

Für symmetrische Verteilungen nehmen die zentralen Momente $m_\delta(\bar{x})$ ungerader Ordnung $\delta = 3, 5, 7, \dots$ den Wert Null an. Aus $m_\delta(\bar{x}) \neq 0$ für ein $\delta \in \{3, 5, 7, \dots\}$ können wir daher auf eine schiefe Verteilung schließen. Jedes dieser zentralen Momente ist daher theoretisch zur Messung der Schiefe geeignet. In der Praxis hat sich jedoch ein dimensionsloses Schiefemaß durchgesetzt, welches auf den zentralen Momenten $m_2(\bar{x}) = s^2$ und $m_3(\bar{x})$ basiert und als **Schiefekoeffizient** bezeichnet wird.

Den Schiefekoeffizient g_m berechnen wir in der *Urliste* als

$$g_m = \frac{m_3(\bar{x})}{s^3} = \frac{m_3(\bar{x})}{(s^2)^{1.5}} = \frac{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1.5}}. \quad (\text{I2: 68})$$

Für die *Häufigkeitstabelle* gilt

$$g_M = \frac{\frac{1}{n} \cdot \sum_{j=1}^k h(a_j) \cdot (a_j - \bar{x})^3}{\left(\frac{1}{n} \cdot \sum_{j=1}^k h(a_j) \cdot (a_j - \bar{x})^2 \right)^{1,5}} = \frac{\sum_{j=1}^k f(a_j) \cdot (a_j - \bar{x})^3}{\left(\sum_{j=1}^k f(a_j) \cdot (a_j - \bar{x})^2 \right)^{1,5}}. \quad (\text{I2: 69})$$

Liegt eine *Häufigkeitstabelle mit klassierten Daten* vor, so werden zur näherungsweisen Berechnung in (I2: 69) lediglich die a_j durch die a_j^m und die Häufigkeiten durch Klassenhäufigkeiten ersetzt.

Eigenschaften des Schiefekoeffizienten:

- Die Division durch s^3 generiert ein dimensionsloses Schiefemaß, welches den Vergleich zwischen verschiedenen Verteilungen ermöglicht.
- Die Wahl von 3 an Stelle von 2 als Exponent bei der Abweichungsberechnung im Zähler bewirkt, dass große Abweichungen stärker ins Gewicht fallen und die Vorzeichen der Abweichungen erhalten bleiben.
- Das Vorzeichen des Schiefekoeffizienten gibt Auskunft über die Schiefe der Verteilung. Es gilt konkret

$g_M > 0$	\rightarrow	rechtsschief	.
$g_M < 0$	\rightarrow	linksschief	
$g_M = 0$	\rightarrow	symmetrisch	

(I2: 70)

Je größer der Betrag $|g_M|$ der Kennzahl ist, desto schiefer ist die Verteilung.

Beispiel I 2-36

Schiefekoeffizient

In einem Konzern mit 50.000 Mitarbeitern wurde von der Personalabteilung aufgrund bevorstehender Entlassungen eine Erhebung zur Anzahl der Kinder je Mitarbeiter (0 bis 3) durchgeführt. Da kinderlose Mitarbeiter aufgrund der sozialen Entlassungspolitik zuerst ausgestellt werden, sind die Ergebnisse von besonderer Bedeutung.

Es liegt nun dazu folgende Häufigkeitstabelle (mit $\bar{x} = 0,77$) vor. In diesem einfachen, übersichtlichen Fall kann bereits aus der Häufigkeitstabelle eine Aussage über die Schiefe der Verteilung gemacht werden. Aufgrund der kleiner werdenden Werte von $f(a_j)$, ist nämlich eine rechtsschiefe Verteilung zu vermuten. Bei komplexeren Verteilungen sind konkrete Aussagen jedoch nur noch über den Schiefekoeffizienten möglich.

j	a_j	$h(a_j)$	$f(a_j)$	$a_j - \bar{x}$	$f(a_j) \cdot (a_j - \bar{x})^2$	$f(a_j) \cdot (a_j - \bar{x})^3$
1	0	24.510	0,49	-0,77	0,29	-0,22
2	1	15.430	0,31	0,23	0,02	0,00
3	2	7.050	0,14	1,23	0,21	0,26
4	3	3.010	0,06	2,23	0,30	0,67
-	-	50.000	1,00	-	0,82	0,71

Es ergibt sich daraus ein positiver Schiefekoeffizient von

$$g_m = \frac{0,71}{0,82^{1,5}} = 0,96,$$

der anzeigen, dass die vorliegende Verteilung rechtsschief ist.

Neben g_m existiert in der Literatur eine Vielzahl weiterer Schiefemaße. Da er im Vergleich zu g_m unempfindlich gegenüber Ausreißern ist, ist davon vor allem der *Schiefequartilkoeffizient von Bowley* interessant. Er ist definiert als

$$g_Q = \frac{(Q_3 - x_{\text{Med}}) - (x_{\text{Med}} - Q_1)}{QA}. \quad (\text{I2: 71})$$

Auch für ihn gilt (I2: 70) analog. Die Division des Zählers durch den Quartilsabstand bewirkt eine Normierung der Kennzahl auf das Intervall $-1 \leq g_Q \leq 1$. Bei ausgeprägter Rechtsschiefe gilt $Q_1 = x_{\text{Med}}$, woraus $g_Q = 1$ folgt. Analog gilt bei ausgeprägter Linksschiefe $x_{\text{Med}} = Q_3$, woraus $g_Q = -1$ folgt.

2.2.4 Konzentrationsmessung

In den Wirtschaftswissenschaften bezeichnen wir eine Ballung der Verfügungsgewalt über Produktionsfaktoren (natürliche Ressourcen, Beschäftigte, usw.) oder eine Ballung wirtschaftlicher Erträge (Umsätze, Gewinne, Einkommen, usw.) in den Händen weniger Wirtschaftssubjekte als *Konzentration*. Konzentration können wir mittels spezieller Maßzahlen quantifizieren. Diese lassen sich grob in zwei Klassen einteilen: Maßzahlen der **absoluten Konzentration** untersuchen, ob der Großteil des gesamten Merkmalsbetrages (z.B. Umsatz) auf eine kleine *Anzahl* von Merkmalsträgern (z.B. Unternehmen) entfällt. Jene der **relativen Konzentration** messen, ob der Großteil des gesamten Merkmalsbetrages auf einen kleinen *Anteil* von Merkmalsträgern entfällt.

Ein in der Praxis bedeutendes Einsatzgebiet der Konzentrationsmessung ist die Abschätzung von Ausmaß und Entwicklung der Ballung wirtschaftlicher Macht auf einzelne oder wenige Unternehmen (*Unternehmenskonzentration*). Die Beherrschung eines Marktes durch wenige große Unternehmen beeinträchtigt nämlich den Wettbewerb unter den Unternehmen. Gerade daher hat das Statistische Bundesamt ein Berichtssystem zur kontinuierlichen Beobachtung von Konzentrationsprozessen geschaffen. Zudem existieren mit dem Bundeskartellamt und der Monopolkommission Instanzen, die die Unternehmenskonzentration begutachten und überwachen. Ein weiteres Einsatzgebiet der Konzentrationsmessung ist die Untersuchung der Einkommens- oder Vermögensverteilung von Wirtschaftssubjekten in einer Volkswirtschaft (*Einkommens- und Vermögenskonzentration*). Diese ist von Bedeutung, da sie die Entscheidungsgrundlage für eine staatliche Verteilungspolitik liefert. Auch einige betriebswirtschaftliche Fragestellungen lassen sich mit der Konzentrationsmessung beantworten. So können Konzentrationsmaße z.B. zur Analyse der Kostenstruktur von Beschaffungsgütern im Rahmen der sog. *ABC-Analyse* eingesetzt werden. In der Regel konzentriert sich nämlich ein hoher Anteil der Beschaffungskosten auf wenige Güter (sog. A-Güter), die es zu identifizieren gilt, damit ihre Beschaffungspolitik optimiert werden kann.

2.2.4.1 Maßzahlen der absoluten Konzentration

Voraussetzung für viele Verfahren der Konzentrationsmessung ist zunächst, dass die n Merkmalswerte *der Größe nach geordnet* sind, also $x_1 \leq x_2 \leq \dots \leq x_n$ gilt. Die sog. Merkmalsumme MS ergibt sich daraus als

$$MS = \sum_{i=1}^n x_i. \quad (I2: 72)$$

Wir untersuchen ausschließlich metrisch skalierte Merkmale X , die nur **nichtnegative Zahlen** als Werte annehmen können (z.B. Unternehmensumsätze, Haushaltseinkommen, Kosten von Beschaffungsgütern, usw.) und nicht alle x_i gleichzeitig Null sind. Nur dann können wir nämlich die Anteile der einzelnen Merkmalswerte an der Merkmalsumme

$$p_i = \frac{x_i}{MS} \quad (I2: 73)$$

bestimmen, für welche aufgrund der Sortierung der einzelnen Merkmalswerte $p_1 \leq p_2 \leq \dots \leq p_n$ gilt. Aus diesen Anteilswerten lassen sich die Konzentrationsrate, der Herfindahl-Index und der Exponentialindex berechnen:

1. Konzentrationsrate

Unter der **Konzentrationsrate** (-koeffizient) CR (engl. concentration ratio) verstehen wir den Anteil an der Merkmalsumme MS, der auf die m größten Merkmalsträger entfällt. Er ist definiert als

$$CR_m = \frac{\sum_{i=n-m+1}^n x_i}{\sum_{i=1}^n x_i} = \sum_{i=n-m+1}^n p_i \quad (I2: 74)$$

und ergibt sich also durch einfache Addition der Anteilswerte der m größten Merkmalswerte. Das "+1" in der Formel ergibt sich dadurch, dass wenn wir z.B. die 3 größten Werte aus einer sortierten Liste von 7 Werten betrachten wollen, wir uns konkret für die Werte Nr. 5, 6 und 7 interessieren. Um die Nr. des Startwerts (hier also 5) zu erhalten, können wir also nicht einfach 3 von 7 subtrahieren.

Eigenschaften der Konzentrationsrate:

- Die Anwendung von CR ist in der Praxis sehr beliebt, da sich dieses Maß durch seine einfache Berechenbarkeit auszeichnet.
- Durch Beschränkung auf ein einziges, oft willkürlich festgelegtes m bleibt die gesamte sonstige, in der Verteilung enthaltene Information unausgeschöpft. Bei der Beurteilung von Konzentrationsunterschieden in verschiedenen Verteilungen können wir leicht zu Fehlschlüssen gelangen, wenn der Vergleich ausschließlich für einen einzigen Wert m durchgeführt wird. Durch geschickte Wahl von m kann das Ergebnis eines Vergleichs sogar beeinflusst werden.
- Bei Gleichverteilung der Merkmalsumme MS auf die n Merkmalsträger gilt

$$p_1 = p_2 = \dots = p_n = \frac{1}{n} \rightarrow CR_m = \sum_{i=n-m+1}^n \frac{1}{n} = \frac{m}{n}. \quad (I2: 75)$$

Beispiel I 2-37**Basisbeispiel zur Konzentrationsmessung**

Grundlage für die Beispiele, die wir in den nachfolgenden Abschnitten verwenden, sind folgende Tabellen. Sie zeigen 8 Verteilungen (A-H) von Unternehmensumsätzen. Betrachten wir z.B. Verteilung D, so können wir sagen, dass Firma 4 genau $x_4 = 280$ Mio. Euro Umsatz bzw. $p_4 = 0,14 = 14\%$ des Gesamtumsatzes (bzw. der Merkmalssumme MS) aller 5 Unternehmen dieser Verteilung aufweist.

Firma	Verteilungen (Umsatz in Mio.)							
	A		B		C		D	
	x_i	p_i	x_i	p_i	x_i	p_i	x_i	p_i
1	0	0,00	400	0,20	120	0,06	220	0,11
2	0	0,00	400	0,20	160	0,08	240	0,12
3	0	0,00	400	0,20	400	0,20	260	0,13
4	0	0,00	400	0,20	600	0,30	280	0,14
5	2.000	1,00	400	0,20	720	0,36	1.000	0,50
MS	2.000	1,00	2.000	1,00	2.000	1,00	2.000	1,00

Firma	Verteilungen (Umsatz in Mio.)							
	E		F		G		H	
	x_i	p_i	x_i	p_i	x_i	p_i	x_i	p_i
1	440	0,11	60	0,03	200	0,10	1	0,0005
2	480	0,12	60	0,03	200	0,10	1	0,0005
3	520	0,13	80	0,04	200	0,10	1	0,0005
4	560	0,14	80	0,04	200	0,10	1	0,0005
5	2.000	0,50	200	0,10	200	0,10	1	0,0005
6			200	0,10	200	0,10	399	0,1995
7			300	0,15	200	0,10	399	0,1995
8			300	0,15	200	0,10	399	0,1995
9			360	0,18	200	0,10	399	0,1995
10			360	0,18	200	0,10	399	0,1995
MS	4.000	1,00	2.000	1,00	2.000	1,00	2.000	1,00

Vergleichen wir nun zunächst die Konzentration in den Verteilungen A, B und C mittels der Konzentrationsrate für $m = 3$, liegt der Anteil der drei größten Unternehmen am Gesamtumsatz in den Verteilungen bei

$$CR_3^A = p_5 + p_4 + p_3 = 1,00 + 0,00 + 0,00 = 1,00$$

$$CR_3^B = p_5 + p_4 + p_3 = 0,20 + 0,20 + 0,20 = 0,60$$

$$CR_3^C = p_5 + p_4 + p_3 = 0,36 + 0,30 + 0,20 = 0,86 ,$$

was uns zeigt, dass die Verteilung A die höchste absolute Konzentration aufweist. Dies ist auch nicht verwunderlich, da in dieser der gesamte Umsatz von nur einem Unternehmen erwirtschaftet wird.

Zur einfachen Veranschaulichung der Manipulierbarkeit des Konzentrationsverhältnisses bestimmen wir den Anteil der $m = 1$, $m = 2$ und $m = 3$ größten Unternehmen am Gesamtumsatz in den Verteilungen C und D. Wir erhalten damit folgende Werte:

$$CR_1^C = p_5 = 0,36$$

$$CR_1^D = 0,50$$

$$CR_2^C = p_5 + p_4 = 0,36 + 0,30 = 0,66$$

$$CR_2^D = 0,50 + 0,14 = 0,64$$

$$CR_3^C = p_5 + p_4 + p_3 = 0,36 + 0,30 + 0,20 = 0,86$$

$$CR_3^D = 0,50 + 0,14 + 0,13 = 0,77$$

Wählen wir also ein m größer als 1, so ist die Konzentration in Verteilung C höher. Entscheiden wir uns hingegen für $m = 1$, ist die Konzentration bei D höher. Der Leser statistischer Veröffentlichungen sollte sich einer derartigen Problematik bewusst sein.

2. Herfindahl-Index

Der *Herfindahl-Index* H ist die Summe der quadrierten Anteile p_i der Merkmalsträger an der Merkmalssumme MS. Da alle Anteile p_i in die Berechnung eingehen, müssen diese hier *nicht der Größe nach geordnet* sein.

$$H = \sum_{i=1}^n p_i^2 = \frac{\sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i\right)^2} \quad (\text{I2: 76})$$

Eigenschaften des Herfindahl-Index:

- Der Herfindahl-Index lässt sich sehr einfach aus nicht sortierten Merkmalsausprägungen berechnen und schöpft die gesamten in der Verteilung enthaltenen Informationen aus. Anders als bei der Konzentrationsrate CR kommt es bei ihm also zu keiner Informationsvernachlässigung.
- Haben alle Merkmalswerte den gleichen Anteil an der Merkmalssumme, so liegt *minimale Konzentration* vor. Der Herfindahl-Index ist in diesem Fall gleich $1/n$. Es gilt nämlich

$$p_1 = p_2 = \dots = p_n = \frac{1}{n} \quad \rightarrow \quad H = n \cdot \frac{1}{n^2} = \frac{1}{n}. \quad (\text{I2: 77})$$

Im Falle *maximaler Konzentration* entfällt die gesamte Merkmalssumme auf einen Merkmalsträger, wobei alle anderen Merkmalsträger einen Anteil von Null an der Merkmalssumme haben. Der Herfindahl-Index ist dann 1, da

$$p_1 = p_2 = \dots = p_{n-1} = 0, p_n = 1 \quad \rightarrow \quad H = 1 \quad (\text{I2: 78})$$

gilt. Aufgrund dieser Eigenschaft bewegt sich H im Intervall

$$\frac{1}{n} \leq H \leq 1. \quad (\text{I2: 79})$$

- Zwischen dem Herfindahl-Index und dem Variationskoeffizienten VC bzw. der Varianz s^2 lässt sich folgender Zusammenhang feststellen:

$$H = \frac{1}{n} \cdot (VC^2 + 1) = \frac{1}{n} \cdot \left(\frac{s^2}{\bar{x}^2} + 1 \right) \quad (\text{I2: 80})$$

In dieser Darstellung formulieren wir den Variationskoeffizienten nicht wie in (I2: 64) mit Absolutstrichen, da wir ja unsere Betrachtungen nur auf nicht-negative x_i beschränken.

Beispiel I 2-38

Herfindahl-Index

Für die Verteilungen A, B und C aus Beispiel I 2-37 erhalten wir die folgenden Herfindahl-Indizes:

$$H_A = 0,00^2 + 0,00^2 + 0,00^2 + 0,00^2 + 1,00^2 = 1,00$$

$$H_B = 0,20^2 + 0,20^2 + 0,20^2 + 0,20^2 + 0,20^2 = 0,20$$

$$H_C = 0,06^2 + 0,08^2 + 0,20^2 + 0,30^2 + 0,36^2 = 0,27$$

Für Verteilung A nimmt H den Wert 1 an, was maximale Konzentration belegt. Für Verteilung B erhalten wir $0,20 = 1/5 = 1/n$, sodass hier minimale Konzentration bestätigt wird.

Betrachten wir die Verteilungen D und E, so ist festzustellen, dass eine proportionale Veränderung (z.B. Verdoppelung) der Merkmalswerte keinerlei Auswirkung auf den Herfindahl-Index hat. Es gilt $H_D = H_E = 0,31$.

Vergleichen wir die Verteilungen B und H, zeigt sich, dass das Hinzufügen einer Anzahl von sehr kleinen Merkmalsträgern kaum Einfluss auf den Wert des Herfindahl-Index hat. Es gilt nämlich $H_B = 0,200$ und $H_H = 0,199$.

Der Herfindahl-Index der Verteilung B mit $H_B = 0,20$ ist doppelt so hoch wie der der Verteilung G mit $H_G = 0,10$. Es macht also einen Unterschied, ob sich die Merkmalssumme von 2.000 gleichmäßig über 5 oder 10 Merkmalsträger verteilt. Ähnliches gilt auch für die Verteilungen C ($H_C = 0,2696$) und F ($H_F = 0,1348$).

3. Exponentialindex

Potenzieren wir alle Anteilswerte p_i mit sich selbst und multiplizieren diese Potenzen miteinander, erhalten wir den *Exponentialindex* E als

$$E = p_1^{p_1} \cdot p_2^{p_2} \cdots p_n^{p_n}. \quad (\text{I2: 81})$$

Eigenschaften des Exponentialindex:

- Bei Auftreten von Anteilswerten $p_i = 0$ ist $p_i^{p_i} = 1$.
- E kann wie der Herfindahl-Index Werte zwischen $1/n$ und 1 annehmen.

Beispiel I 2-39

Exponentialindex

Für die Verteilungen A, B und C aus Beispiel I 2-37 erhalten wir folgende Werte für den Exponentialindex:

$$E_A = 0^0 \cdot 0^0 \cdot 0^0 \cdot 0^0 \cdot 1^1 = 1$$

$$E_B = 0,20^{0,20} \cdot 0,20^{0,20} \cdot 0,20^{0,20} \cdot 0,20^{0,20} \cdot 0,20^{0,20} = 0,20$$

$$E_C = 0,06^{0,06} + 0,08^{0,08} + 0,20^{0,20} + 0,30^{0,30} + 0,36^{0,36} = 0,24$$

Wie beim Herfindahl-Index wird auch hier bei Verteilung A maximale Konzentration ($E = 1$) und bei B minimale Konzentration ($E = 0,20 = 1/5 = 1/n$) bestätigt.

2.2.4.2 Maßzahlen der relativen Konzentration



Das wohl wichtigste grafische Hilfsmittel zur Bestimmung von Konzentrationstendenzen ist die sog. Lorenz-Kurve. Zu ihrer Konstruktion werden n Punkte (u_i, v_i) einer in *Urlistenform* vorliegenden Verteilung bestimmt, zusammen mit einer Diagonalen und einem Punkt $(0; 0)$ in ein quadratisches Schaubild (vgl. Abbildung I 16) eingezeichnet und miteinander verbunden. Der entstehende Streckenzug heißt **Lorenz-Kurve der Konzentration**.

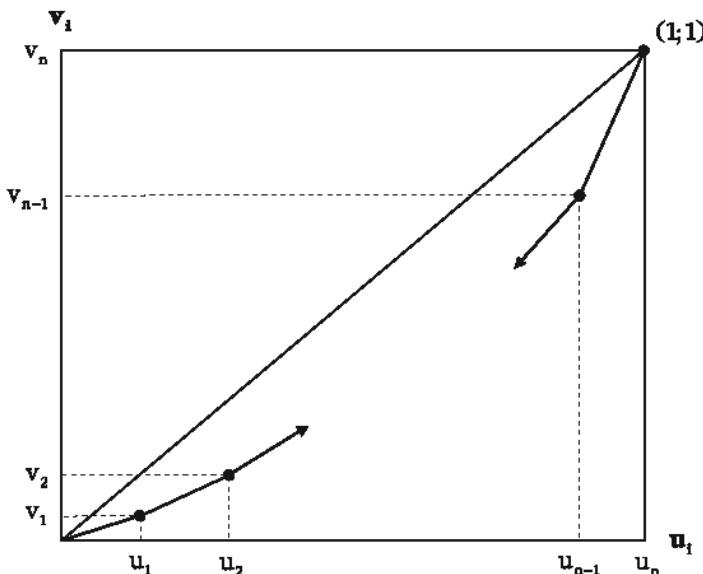


Abbildung I 16: Konstruktion der Lorenz-Kurve

Die n Punkte (u_i, v_i) werden folgendermaßen ermittelt:

- Grundvoraussetzung zur Berechnung der genannten Punkte sind wieder eine der Größe nach sortierte Urliste $x_1 \leq x_2 \leq \dots \leq x_n$ und nichtnegative x_i , die eine positive Merkmalssumme MS garantieren.
- Die y-Koordinaten v_i ($i = 1, \dots, n$) geben die relativen Anteile der Merkmalssumme MS an, den die i kleinsten Merkmalsträger auf sich vereinigen und werden daher durch Summierung der Anteile p_i der i kleinsten Merkmalsträger an der Merkmalssumme bestimmt:

$$v_i = \frac{\sum_{j=1}^i x_j}{MS} = \sum_{j=1}^i p_j \quad (\text{I2: 82})$$

- Die x-Koordinaten u_i ($i = 1, \dots, n$) entsprechen den relativen Anteilen der i kleinsten Merkmalsträger an der Gesamtzahl n der Merkmalsträger:

$$u_i = \frac{i}{n} \quad (\text{I2: 83})$$

Häufig werden die Anteilssätze u_i und v_i in Prozent ausgedrückt. Die Lorenz-Kurve verläuft dann durch den Ursprung (0; 0) sowie die n Punkte ($u_i \cdot 100\%$; $v_i \cdot 100\%$).

Formal gesehen ist die Lorenz-Kurve eine auf dem Intervall [0; 1] definierte monoton wachsende, konvexe Funktion $L(u_i)$ von u_i , deren Werte die 45°-Diagonale (Winkelhalbierende) nicht übersteigen. Ein Wert $L(u_i) \cdot 100 = v_i \cdot 100$ gibt an, welcher prozentuale Anteil an der Merkmalssumme auf $u_i \cdot 100$ Prozent der kleinsten Merkmalsträger entfällt. Entsprechend gibt $(1 - L(u_i)) \cdot 100 = (1 - v_i) \cdot 100$ an, welcher prozentuale Anteil der Merkmalssumme auf $(1 - u_i) \cdot 100$ Prozent der größten Merkmalsträger entfällt.

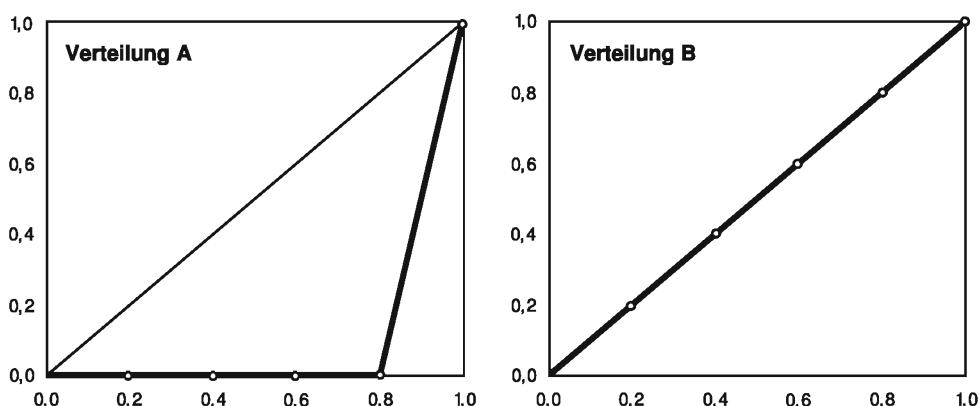
Beispiel I 2-40

Lorenz-Kurve

Für die Umsatzverteilungen A und B aus Beispiel I 2-37 ergeben sich die im Folgenden dargestellten Koordinaten (u_i, v_i) und die daraus resultierenden Lorenz-Kurven.

Verteilung A					MS = 2.000 n = 5
i	x_i	p_i	u_i	v_i	
1	0	0	$1 / 5 = 0,20$	$0 / 2.000 = 0,00$	
2	0	0	$2 / 5 = 0,40$	$0 / 2.000 = 0,00$	
3	0	0	$3 / 5 = 0,60$	$0 / 2.000 = 0,00$	
4	0	0	$4 / 5 = 0,80$	$0 / 2.000 = 0,00$	
5	2.000	2.000	$5 / 5 = 1,00$	$2.000 / 2.000 = 1,00$	
Σ	2.000	-	-	1,00	
				○	

Verteilung B					MS = 2.000 n = 5
i	x_i	p_j	u_i	v_i	
1	400	400	$1 / 5 = 0,20$	$400 / 2.000 = 0,20$	
2	400	800	$2 / 5 = 0,40$	$800 / 2.000 = 0,40$	
3	400	1.200	$3 / 5 = 0,60$	$1.200 / 2.000 = 0,60$	
4	400	1.600	$4 / 5 = 0,80$	$1.600 / 2.000 = 0,80$	
5	400	2.000	$5 / 5 = 1,00$	$2.000 / 2.000 = 1,00$	
Σ	2.000	-	-	3,00	



Die Verteilungen A und B zeigen zwei Extremfälle von Konzentration. In Verteilung A liegt **maximale Konzentration** vor. Ein Merkmalsträger vereint die gesamte Merkmalssumme auf sich. Die Lorenz-Kurve verläuft bis zum Punkt $((n-1)/n, 0)$ auf der x-Achse und steigt dann geradlinig zum Punkt $(1; 1)$ an. Die Funktion ist äußerst stark durchgebogen und die Fläche zwischen der 45° -Diagonale und der Lorenz-Kurve umfasst nahezu die gesamte Quadrathälfte unterhalb der Diagonalen.

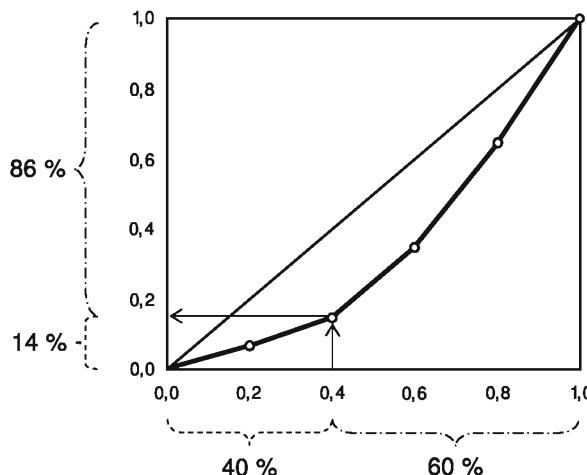
Verteilung B zeigt **minimale Konzentration**. Jeder Merkmalsträger weist den gleichen Anteil an der Merkmalssumme auf. Es entfallen auf $(i/n) \cdot 100\%$ der kleinsten Merkmalsträger immer $(i/n) \cdot 100\%$ der Merkmalssumme. Die Lorenz-Kurve fällt mit der 45° -Diagonale zusammen und die Fläche zwischen Lorenz-Kurve und 45° -Diagonale ist logischerweise Null.

Wir können daraus schließen, dass je stärker die Konzentration ist,

- desto größer ist das Flächenstück zwischen Diagonale und Lorenz-Kurve,
- desto stärker ist die Lorenz-Kurve "gebaucht" bzw. desto stärker "hängt sie durch".

Betrachten wir auch Verteilung C, stellen wir fest, dass weder minimale noch maximale Konzentration vorliegt. Die Lorenz-Kurve zeigt hier einen typischen in der Praxis beobachteten Verlauf. Wir wollen nun diese Verteilung nutzen, um die typische Interpretation der Werte der Lorenz-Kurve zu veranschaulichen. Wir können nämlich hier z.B. sagen, dass auf die $u_2 \cdot 100\% = 40\%$ der umsatzschwächsten Unternehmen genau $v_2 \cdot 100\% = 14\%$ des Gesamtumsatzes entfällt. Die $(1 - u_2) \cdot 100\% = 60\%$ der umsatzstärksten Unternehmen realisieren $(1 - v_2) \cdot 100\% = 86\%$ des Marktumsatzes.

Verteilung C					$MS = 2.000$
i	x_i	p_i	u_i	v_i	$n = 5$
1	120	120	$1 / 5 = 0,20$	$120 / 2.000 = 0,06$	
2	160	280	$2 / 5 = 0,40$	$280 / 2.000 = 0,14$	
3	400	680	$3 / 5 = 0,60$	$680 / 2.000 = 0,34$	
4	600	1.280	$4 / 5 = 0,80$	$1.280 / 2.000 = 0,64$	
5	720	2.000	$5 / 5 = 1,00$	$2.000 / 2.000 = 1,00$	
Σ	2.000	-	-		2,18



Kommt es beim Vergleich der Konzentration zweier Verteilungen anhand ihrer Lorenz-Kurven zu sich *schneidenden Lorenz-Kurven*, so kann nicht immer anhand reiner optischer Analyse ausgesagt werden, welche die stärkere Konzentration aufweist. Hilfreich wäre also eine aussagekräftige Kennzahl zur Quantifizierung *relativer Konzentration*. Wie wir bereits in Beispiel I 2-37 festgestellt haben, ist mit steigender Konzentration auch eine Zunahme der Fläche zwischen 45°-Diagonale und Lorenz-Kurve verbunden. Diese Fläche, die auch als **Konzentrationsfläche F** bezeichnet wird, käme also theoretisch als Konzentrationsmaß in Frage. Wie sich zeigen lässt, können wir die Konzentrationsfläche über

$$F = \frac{1}{2} - \frac{V}{n} \quad \text{mit} \quad V = \sum_{i=1}^n v_i - \frac{1}{2} \quad (\text{I2: 84})$$

bestimmen.⁷ Dividieren wir die Konzentrationsfläche F durch die Fläche zwischen der 45°-Diagonalen und der x-Achse (Dreiecksfläche mit Flächeninhalt 0,5), so erhalten wir den sog. **Gini-Koeffizienten** als Maß der relativen Konzentration zu

$$G = \frac{F}{0,5} = 2F = 1 - \frac{2V}{n} \quad (\text{I2: 85})$$

Wir wissen bereits, dass die Konzentrationsfläche im Falle minimaler Konzentration den Wert Null annimmt, da dann 45°-Diagonale und Lorenz-Kurve genau aufeinanderfallen. Auch G nimmt im Falle minimaler Konzentration den Wert Null an, da sich G schließlich direkt aus F ergibt. Bei maximaler Konzentration nimmt F aufgrund der Tatsache, dass die Summe der v_i dann gleich 1 ist (vgl. Beispiel I 2-37), den Wert $(n-1)/2n$ und damit G den Wert $(n-1)/n$ an. Wir können damit festhalten, dass sich G im Intervall

$$0 \leq G \leq \frac{n-1}{n} \quad (\text{I2: 86})$$

bewegt. Es zeigt sich also, dass selbst im Fall maximaler Konzentration anders als z.B. beim Herfindahl-Index oder Exponentialindex das Maß G nicht den Wert 1 annimmt, was jedoch eine wünschenswerte Eigenschaft für ein Konzentrationsmaß ist. Da außerdem das Maximum des Gini-Koeffizienten von n abhängt, kann dies bei Vergleichen von Datensätzen unterschiedlicher Größe problematisch sein. Zur Korrektur dieser Schwächen des Gini-Koeffizienten wird der Koeffizient daher häufig mit dem Kehrwert von $(n-1)/n$ multipliziert. Wir bezeichnen das dadurch entstehende relative Konzentrationsmaß auch als **normierten Gini-Koeffizienten**:

$$G_{\text{norm}} = \frac{n}{n-1} \cdot G \quad (\text{I2: 87})$$

G_{norm} weist damit im Falle maximaler Konzentration den Wert Eins auf. Es gilt also insgesamt

$$0 \leq G_{\text{norm}} \leq 1. \quad (\text{I2: 88})$$

⁷ Eine detaillierte Herleitung dieser Formel über die Berechnung diverser Dreiecksflächen unter der Lorenz-Kurve findet sich z.B. in Bleymüller, J., Gehlert, G., Gülicher, H. (2004), Kapitel 26.

Es lässt sich G und damit über (I2: 87) zudem G_{nom} auch direkt durch die Daten der Urliste ausdrücken. So gilt unter Verwendung der Merkmalswerte x_i

$$G = \frac{2 \sum_{i=1}^n i \cdot x_i - (n+1) \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i} \quad (\text{I2: 89})$$

und bei Verwendung der Anteilswerte p_i

$$G = \frac{2 \sum_{i=1}^n i \cdot p_i - (n+1)}{n}. \quad (\text{I2: 90})$$

Zusammengefasst können wir für F , G und G_{nom} festhalten, dass alle drei Maße im Falle minimaler Konzentration den Wert Null annehmen und hohe Konzentration mit hohen Werten der Maße einhergeht. Bei maximaler Konzentration ist die Konzentrationsfläche fast 0,5 bzw. $(n-1)/2n$, der einfache Gini-Koeffizient $(n-1)/n$ und der normierte Gini-Koeffizient 1.

Beispiel I 2-41

Gini-Koeffizient aus Urliste

Für die Verteilungen A, B und C, deren Lorenz-Kurven wir in Beispiel I 2-40 skizziert haben, ergeben sich die in folgender Tabelle dargestellten relativen Konzentrationsmaße. Wir können darin sehr schön erkennen, dass die Formel (I2: 94) zur Bestimmung des normierten Gini-Koeffizienten aus dem einfachen Gini-Koeffizienten exakt funktioniert, da im Falle der Verteilung A genau ein Wert von 1 (maximale Konzentration) resultiert.

	Verteilung A ($V = 1,00 - 0,50 = 0,50$)	Verteilung B ($V = 3,00 - 0,50 = 2,50$)	Verteilung C ($V = 2,18 - 0,50 = 1,68$)
F	$\frac{1}{2} - \frac{0,50}{5} = 0,40$	$\frac{1}{2} - \frac{2,5}{5} = 0,00$	$\frac{1}{2} - \frac{1,68}{5} = 0,16$
G	$\frac{0,40}{0,50} = 0,80$	$\frac{0,00}{0,50} = 0,00$	$\frac{0,16}{0,50} = 0,32$
G_{norm}	$0,80 \cdot \frac{5}{5-1} = 1,00$	$0,00 \cdot \frac{5}{5-1} = 0,00$	$0,32 \cdot \frac{5}{5-1} = 0,40$

Die bisher behandelten Formeln haben sich auf die Konstruktion der Lorenz-Kurve und die Bestimmung des Gini-Koeffizienten aus der Urliste bezogen. Liegt nun aber keine Urliste, sondern eine *Häufigkeitstabelle* vor, so gilt für die Ermittlung der nun k Punkte (u_i, v_i) der Lorenz-Kurve zunächst

$$v_i = \sum_{j=1}^i p_j = \frac{\sum_{j=1}^i h(a_j) \cdot a_j}{MS} \quad \text{mit} \quad MS = \sum_{j=1}^k h(a_j) \cdot a_j \quad (\text{I2: 91})$$

Die v_i werden also mit den auf die einzelnen Ausprägungen entfallenden Merkmalssummen $h(a_j) \cdot a_j$ gebildet. Die u_i erhalten wir durch Summierung der relativen Häufigkeiten $f(a_j)$ der relevanten Merkmalsausprägungen:

$$u_i = \frac{\sum_{j=1}^i h(a_j)}{n} = \sum_{j=1}^i f(a_j). \quad (\text{I2: 92})$$

Unter Verwendung dieser Werte erhalten wir die Konzentrationsfläche F zu

$$F = \frac{1}{2} - \frac{1}{2} \sum_{i=1}^k f(a_i) \cdot (v_i + v_{i-1}) \quad (\text{I2: 93})$$

und damit den Gini-Koeffizienten

$$G = 1 - \sum_{i=1}^k f(a_i) \cdot (v_i + v_{i-1}). \quad (\text{I2: 94})$$

Formel (I2: 94) ist natürlich auch auf die Urliste anwendbar, wenn wir die relativen Häufigkeiten durch 1/n ersetzen und die v_i nach (I2: 82) bestimmen.

Zur Berechnung von G aus einer Häufigkeitstabelle empfiehlt sich die Verwendung der Arbeitstabelle aus Abbildung I 17. Den normierten Gini-Koeffizienten erhalten wir durch Einsetzen von (I2: 94) in (I2: 87).

i	a_i	$h(a_i)$	$f(a_i)$	$h(a_i) \cdot a_i$	$\sum_{j=1}^i h(a_j) \cdot a_j$	u_i	v_i	$v_i + v_{i-1}$	$f(a_i) \cdot (v_i + v_{i-1})$
1
...
k	1	1
Σ	-	n	1	MS	-	-	-	-	...

Abbildung I 17: Arbeitstabelle zur Berechnung des Gini-Koeffizienten

Beispiel I 2-42

Gini-Koeffizient aus Häufigkeitstabelle

Eine statistische Untersuchung ergab, dass sich in einem Land 61 Firmen einen 300 Mio. Euro Markt für Schweinehälften wie folgt teilen:

a_j	Umsatz in Mio. Euro	Anzahl der Firmen $h(a_j)$
1		40
2		10
3		5
5		5
200		1
300		61

Berechnen wir nun daraus den normierten Gini-Koeffizienten und versuchen, darauf basierend, Aussagen über die Konzentration auf diesem Markt zu machen:

i	a_i	$h(a_i)$	$f(a_i)$	$h(a_i) \cdot a_i$	$\sum_{j=1}^i h(a_j) \cdot a_j$	u_i	v_i	$v_i + v_{i-1}$	$f(a_i) \cdot (v_i + v_{i-1})$
1	1	40	0,66	40	40	0,66	0,13	0,13	0,09
2	2	10	0,16	20	60	0,82	0,20	0,33	0,05
3	3	5	0,08	15	75	0,90	0,25	0,45	0,04
4	5	5	0,08	25	100	0,98	0,33	0,58	0,05
5	200	1	0,02	200	300	1,00	1,00	1,33	0,03
Σ	-	61	1,00	300	-	-	-	-	0,26

Es ergibt sich daraus

$$G = 1 - \sum_{i=1}^k f(a_i) \cdot (v_i + v_{i-1}) = 1 - 0,26 = 0,74$$

$$G_{\text{norm}} = \frac{n}{n-1} \cdot G = \frac{61}{61-1} \cdot 0,74 = 0,75$$

Die Konzentration auf diesem Markt kann also als sehr stark angesehen werden, was nicht verwunderlich ist, da ein Unternehmen 2/3 des Umsatzes auf sich vereint.

Bei Vorliegen *klassierter Daten* sind für die Konstruktion der Lorenz-Kurve folgende Besonderheiten zu beachten:

Fall 1: Sind für jede der k Klassen die Klassenhäufigkeiten h_j und die auf die Klassen entfallenden Merkmalssummen m_j bekannt, so gehen wir wie soeben beschrieben vor, d.h. wir verwenden die Klassenmerkmalssummen in (I2: 91). Sind die arithmetischen Mittel der Klassen gegeben, so ist entsprechend $m_j = h_j \cdot \bar{x}_j$ für die jeweilige Merkmalssumme der Klasse j zu verwenden. Die Berechnung der u_i erfolgt unter Verwendung der h_i bzw. f_i analog zu (I2: 92).

Fall 2: Liegen nur Klassenhäufigkeiten und -grenzen vor, behelfen wir uns damit, für das arithmetische Mittel der Klassen die jeweiligen Klassenmitten $(a_j^u + a_j^o) / 2$ heranzuziehen.

Die Berechnung des Gini-Koeffizienten kann im Falle klassierter Daten analog zu (I2: 94) unter Verwendung relativer Klassenhäufigkeiten f_i und der Arbeitstabelle aus Abbildung I 17 durchgeführt werden.

Beispiel I 2-43

Gini-Koeffizient aus klassierter Häufigkeitstabelle

Betrachten wir folgende fiktive Einkommensverteilung. Die für 7,6 Mio. Bürger erhobenen Einkommen sind in 4 Einkommensklassen eingeteilt. Für jede der Klassen liegen die Anzahl h_j (in Mio.) und der Anteil f_j der Bürger vor, die in diese Klasse fallen. Außerdem kennen wir die Einkommenssumme m_j (in Mrd. Euro), die den Bürgern einer Klasse zufließt, sowie den Anteil p_j dieser Summe am Gesamteinkommen aller 7,6 Mio. Bürger.

j	Einkommensklasse		Einkommensbezieher		Einkommen pro Klasse	
	über a_j^u bis a_j^o	h_j	f_j	m_j	p_j	
1	5.000	2,0	0,26	5	0,10	
2	5.000 - 9.000	2,0	0,26	11	0,21	
3	9.000 - 13.000	1,8	0,24	11	0,21	
4	13.000 - 35.000	1,8	0,24	25	0,48	
Σ		7,6	1,00	52	1,00	

Hier erleichtert sich die Berechnung der Werte u_i und v_i erheblich, da die zur Kumulation benötigten Daten f_j und p_j bereits vorliegen. Durch eine Kontrollrechnung wie $p_1 = m_1 / MS = 5 / 52 = 0,10$ erkennen wir jedoch, dass sich dahinter genau die geschilderte Vorgehensweise (Fall 1) verbirgt.

Zur Berechnung des Gini-Koeffizienten stellen wir die folgende Arbeitstabelle auf:

i	f_i	p_i	u_i	v_i	$v_i + v_{i-1}$	$f_i \cdot (v_i + v_{i-1})$
1	0,26	0,10	0,26	0,10	0,10	0,03
2	0,26	0,21	0,52	0,31	0,41	0,11
3	0,24	0,21	0,76	0,52	0,83	0,20
4	0,24	0,48	1,00	1,00	1,52	0,36
Σ	1,00	1,00	-	-	-	0,70

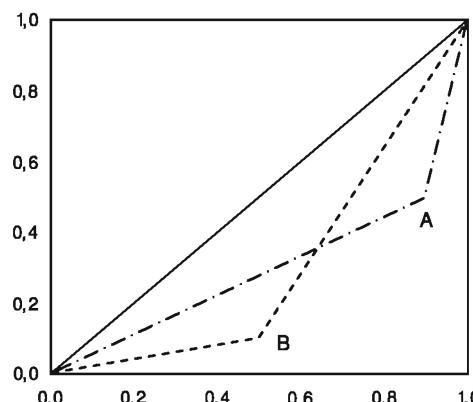
Daraus ergibt sich ein Gini-Koeffizient von $G = 1 - 0,70 = 0,30$. Es liegt also eine relativ schwach ausgeprägte Konzentration vor. Eine Normierung des Gini-Koeffizienten ist hier nicht erforderlich, da der Normierungsfaktor aufgrund des hohen Wertes für n (7.600.000) ziemlich genau bei 1 liegt.

Wir haben bisher gesehen, dass wir relative Konzentrationsmaße für komparative Konzentrationsuntersuchungen verschiedener Verteilungen einsetzen können. Kritisch ist in diesem Zusammenhang jedoch anzumerken, dass der Lorenz-Kurve und dem Gini-Koeffizienten gewisse Grenzen gesetzt sind. Die Ansätze führen zu eindeutigen Ergebnissen, wenn für zwei Verteilungen A und B die Lorenzkurve der Verteilung A stets unterhalb der von B liegt. Es gilt dann offensichtlich $G_A > G_B$. Bei sich schneidenden Kurven können nun jedoch Kompensationseffekte auftreten, so dass $G_A \approx G_B$ gilt. Die Ergebnisse sind in einem solchen Fall schwer interpretierbar, wie das folgende Beispiel verdeutlicht. Eine Konzentrationsmessung sollte daher nicht allein durch den Gini-Koeffizienten erfolgen, sondern durch die Lorenz-Kurve unterstützt werden.

Beispiel I 2-44

Grenzen des Gini-Koeffizienten

Zwei Märkte A und B werden durch 10 Unternehmen beliefert. Auf Markt A besitzen 9 Unternehmen einen 50/9 % Marktanteil. Der Marktanteil des 10. Unternehmens beträgt 50 %. Auf Markt B weisen 5 Unternehmen jeweils einen 2 % und 5 Unternehmen jeweils einen 18 % Marktanteil auf. Die nachfolgende Grafik zeigt die zugehörigen Lorenzkurven.



Rein optisch erkennen wir sofort, dass die Konzentrationsflächen F_A und F_B identisch sind. Wir können zudem leicht berechnen, dass $G_A = G_B = 0,4$ gilt. Obwohl also auf Markt A die deutlich stärkere Konzentration vorliegt, würden wir durch einen Vergleich der Gini-Koeffizienten zu dem Schluss gleicher Konzentration kommen. Die Gini-Koeffizienten allein reichen also zum Vergleich nicht aus.

Beschließen wollen wir diesen Abschnitt mit einer kurzen Diskussion der unterschiedlichen *Reaktion absoluter und relativer Konzentrationsmaße* auf Änderungen der zugrunde liegenden Verteilungen:

Für die Verteilungen E und D aus Beispiel I 2-37 ergibt sich derselbe Wert für den normierten Gini-Koeffizienten. Dieser ist somit wie auch der Herfindahl-Index invariant gegenüber einer proportionalen Veränderung des auf die einzelnen Merkmalsträger entfallenden Merkmalsbetrages.

Beim Vergleich der Herfindahl-Indizes der Verteilung B und G ergab sich $H_B = 0,20$ und $H_G = 0,10$, d.h. ein doppelt so hoher Indexwert (vgl. Beispiel I 2-38). Es macht also einen Unterschied, ob sich die Merkmalssumme von 2.000 gleichmäßig über 5 oder 10 Merkmalsträger verteilt. Vergleichen wir die normierten Gini-Koeffizienten beider Verteilungen, so erhalten wir in beiden Fällen den Wert Null.

Ein Vergleich von Verteilung B und H hatte beim Herfindahl-Index nahezu denselben Wert ergeben. Der normierte Gini-Koeffizient reagiert jedoch außerordentlich empfindlich auf die Hinzufügung von fünf Firmen mit ganz geringem Anteil am gesamten Merkmalsbetrag. Während sich für die Verteilung B ein Wert von $G_{\text{nom}} = 0$ ergab, erhalten wir für Verteilung H den Wert $G_{\text{nom}} = 0,55$. Diese Eigenschaft des normierten Gini-Koeffizienten ist jedoch nicht als Schönheitsfehler anzusehen. Sie zeigt auf, dass in der praktischen Anwendung der Untersuchungszweck über die Wahl absoluter oder relativer Konzentrationsmaße entscheiden sollte:

Wird die Stellung von Unternehmen auf einem Markt analysiert, so wird der Markteintritt umsatzschwacher Unternehmen den normierten Gini-Koeffizienten stark erhöhen, wohingegen der Herfindahl-Index nahezu unverändert bleibt. Ist also die *Analyse von Wettbewerbsprozessen* Untersuchungsgegenstand, so ist die Verwendung *absoluter Konzentrationsmaße* zu empfehlen. Die Marktmacht der etablierten Unternehmen wird durch die Markteinritte umsatzschwacher Unternehmen trotz intensiveren Wettbewerbs zunächst nicht empfindlich beeinträchtigt werden. Deshalb ist ein Konzentrationsmaß wählen, das sich kaum verändert. Ein relatives Konzentrationsmaß wie der Gini-Koeffizient würde hingegen in einer solchen Situation stark ansteigen und eine zunehmende Konzentration und damit schwächeren Wettbewerb anzeigen.

Bei der *Darstellung von Einkommensverteilungen* ist die letztgenannte Eigenschaft relativer Konzentrationsmaße eventuell problematisch. Die dabei verwendeten Daten werden üblicherweise der Einkommensteuerstatistik entnommen. Wird der Mindestbetrag angehoben, ab dem die Steuerpflicht besteht, fallen am unteren Ende der Verteilung ein Teil der bisher Einkommensteuerpflichtigen heraus, d.h. sie werden bei der Untersuchung nicht berücksichtigt. Damit verschiebt sich die Lorenz-Kurve nach links, obwohl sich an der Verteilung der Einkommen nichts geändert hat. Vor allem bei internationalen Vergleichen ist daher größte Vorsicht geboten. Höhere oder niedrigere Konzentration der Einkommen hängt vielfach vornehmlich davon ab, welche Untergrenze in den einzelnen Ländern besteht. Da auf-

grund der Größenunterschiede zwischen den Ländern und aus inhaltlichen Erwägungen hier *relative Konzentrationsmaße* zu empfehlen sind, ist nun der Gebrauch der Lorenz-Kurve und des normierten Gini-Koeffizienten nicht von vornherein abzulehnen. Vielmehr ist bei der Interpretation der Ergebnisse auf solche Probleme zu achten.

3.

Zweidimensionale Häufigkeitsverteilungen

Im vorhergehenden Abschnitt haben wir gesehen, dass durch die Erfassung nur eines Merkmals von einer Grundgesamtheit von Merkmalsträgern eindimensionale Häufigkeitsverteilungen entstehen, die sich durch eine Reihe von Kennzahlen charakterisieren lassen. Werden mehrere Merkmale erfasst, entstehen mehrdimensionale Häufigkeitsverteilungen. Auch diese können durch gewisse Maßzahlen komprimiert dargestellt werden. Diese Maßzahlen und Grundlagen derartiger Verteilungen sind Thema dieses Abschnitts.

3.1 Grundlagen

In vielen statistischen Untersuchungen werden bei statistischen Einheiten gleichzeitig mehrere Merkmale erfasst. So kann etwa bei einer personenbezogenen Erhebung sowohl Einkommen, Geschlecht, Alter, Bildung, Berufserfahrung, u.ä. aufgezeichnet werden. Ähnliche Beispiele finden sich auch bei der Analyse von Unternehmen (Umsatz, Gewinn, Investitionen, Beschäftigtenzahl, ...), der Ertragsfähigkeit landwirtschaftlicher Nutzungsf lächen (Ertrag je Hektar, Düngemittel Einsatz, ...) oder dem Bremsverhalten von Fahrzeugen (Geschwindigkeit, Bremsweg, ...). Für jeden der n Merkmalsträger wird also ein m -Tupel von Merkmalswerten betrachtet. Die Urliste muss daher $n \cdot m$ Einzeldaten enthalten.

Bei solchem Datenmaterial ist häufig von Interesse, ob es Zusammenhänge zwischen den auftretenden Merkmalen gibt und wie stark diese sind. Wir stellen uns also die Frage, ob es eine wechselseitige Abhängigkeit zwischen den Merkmalen gibt. Auf die Beantwortung solcher Fragestellungen werden wir im Abschnitt I 3.2 (Korrelationsanalyse) näher eingehen. Im Folgenden müssen aber zunächst grundlegende Begriffe wie Kontingenztabelle, Randhäufigkeit, bedingte Häufigkeit und Streudiagramm erklärt werden. Wir beschränken uns dabei auf **zweidimensionales Datenmaterial ($m = 2$)**.

3.1.1 Kontingenztabelle

Werden bei n statistischen Einheiten die Merkmale X und Y mit den Ausprägungen x_1, x_2, \dots, x_n und y_1, y_2, \dots, y_n erhoben, so erhalten wir für die i -te statistische Einheit einen Merkmalswert x_i der Variablen X und einen Merkmalswert y_i der Variablen Y . Zur i -ten statistischen Einheit gehört somit ein Wertepaar (x_i, y_i) . Die **Urliste** besteht daher aus allen n zweidimensionalen Wertepaaren (x_i, y_i) für $i = 1, 2, \dots, n$.

Aus Gründen der Übersichtlichkeit kann die Urliste allgemein in zwei Formen dargestellt werden. Erstens kann ein sog. **Streudiagramm** erstellt werden, in welchem alle n Wertepaare (x_i, y_i) als Punkte in ein Koordinatensystem eingetragen werden (vgl. Abbildung I 18). Diese Vorgehensweise empfiehlt sich allerdings nur, wenn sich alle oder fast alle Wertepaare voneinander unterscheiden.



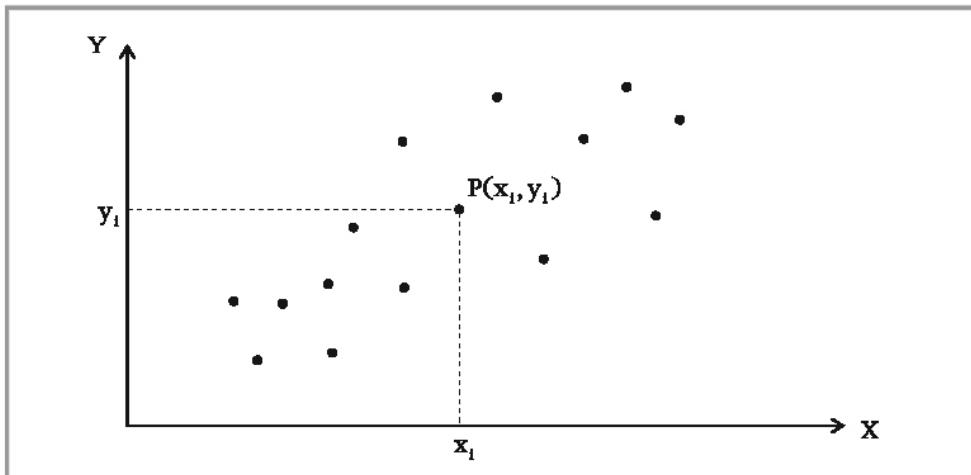


Abbildung I 18: Streudiagramm

Zweitens kann eine sog. **Kontingenztabelle** aufgestellt werden, wobei es sich um eine tabellarische Darstellung der Häufigkeitsverteilung handelt. Sie sollte herangezogen werden, wenn viele Wertepaare identisch sind. Da sie außerdem Grundlage für weitere Analysen des erhobenen Datenmaterials ist, wollen wir uns im Folgenden näher mit ihr beschäftigen. Da es sich bei einer Kontingenztabelle um eine Häufigkeitstabelle handelt, bezeichnen wir in dieser die verschiedenen Ausprägungen des Merkmals X mit a_i ($i = 1, 2, \dots, k$) und jene des Merkmals Y mit b_j ($j = 1, 2, \dots, r$). Zur Erstellung der Kontingenztabelle ist auszuzählen, bei wie vielen Beobachtungswerten der Urliste die Ausprägungskombination (a_i, b_j) auftritt. Der resultierende Zahlenwert wird als **gemeinsame absolute Häufigkeit** h_{ij} oder $h_{xy}(a_i, b_j)$ der Kombination (a_i, b_j) bezeichnet, die wir im Folgenden zumeist mit h_{ij} abkürzen. Für diese gilt wie bei eindimensionalen Häufigkeitsverteilungen, dass sie zwischen Null und n liegen und ihre Summe n ergeben muss:

$$0 \leq h_{ij} \leq n \quad \sum_{i=1}^k \sum_{j=1}^r h_{ij} = n \quad (\text{I3: 1})$$

Die **gemeinsame relative Häufigkeit** f_{ij} oder $f_{xy}(a_i, b_j)$ bzw. abgekürzt f_{ij} ist der Anteil der Beobachtungswerte in der Urliste, bei denen die Kombination (a_i, b_j) auftritt. Die Regeln eindimensionaler Verteilungen, also im Detail (I2: 3), (I2: 4) und (I2: 5), können daher analog übertragen werden:

$$f_{ij} = \frac{h_{ij}}{n} \quad 0 \leq f_{ij} \leq 1 \quad \sum_{i=1}^k \sum_{j=1}^r f_{ij} = 1 \quad (\text{I3: 2})$$

Die tabellarische Darstellung dieser $k \cdot r$ Häufigkeiten (absolut oder relativ) mittels Abbildung I 19 nennt man **Kontingenztabelle** oder Kreuztabelle. Die Menge aller Tripel (a_i, b_j, h_{ij}) bzw. (a_i, b_j, f_{ij}) für $i = 1, \dots, k$ und $j = 1, \dots, r$ bildet die sog. **zweidimensionale absolute bzw. relative Häufigkeitsverteilung** von X und Y. Gelegentlich wird auf dreidimensionale Stabdiagramme zur grafischen Darstellung zweidimensionaler Häufigkeitsverteilungen zurückgegriffen (vgl. Beispiel I 3-1). Im Falle klassierter zweidimensionaler Häufigkeitsverteilungen kommen dabei auch dreidimensionale Histogramme zum Einsatz, deren Praxisrelevanz jedoch eher gering ist.

		Ausprägungen von Y			
		b ₁	b ₂	...	b _r
Ausprägungen von X	a ₁	h ₁₁	h ₁₂	...	h _{1r}
	a ₂	h ₂₁	h ₂₂	...	h _{2r}

	a _k	h _{k1}	h _{k2}	...	h _{kr}

Abbildung I 19: Kontingenztabelle

Im Falle ordinal- oder metrisch skaliertener Daten ist außerdem die Definition der **absoluten kumulierten gemeinsamen Häufigkeit**

$$H(x, y) = H(X \leq x, Y \leq y) = \sum_{a_i \leq x} \sum_{b_j \leq y} h(a_i, b_j) \quad (\text{I3: 3})$$

(kurz H_p) sinnvoll, für welche

$$0 \leq H(x, y) \leq n \quad \text{und} \quad H(X \leq a_k, Y \leq b_r) = n \quad (\text{I3: 4})$$

gilt. Sie gibt die Anzahl der Merkmalsträger an, die bei Merkmal X höchstens die Ausprägung x und gleichzeitig bei Merkmal Y höchstens die Ausprägung y aufweisen (zur ungewöhnlichen Summenschreibweise vgl. Abschnitt I 2.1.2).

Analog ergibt sich die **relative kumulierte gemeinsame Häufigkeit** $F(x, y)$ (kurz F_p). Sie gibt den Anteilssatz der Merkmalsträger an, deren Merkmale X und Y beliebig vorgegebene Zahlenpaare (x, y) nicht übersteigen:

$$F(x, y) = F(X \leq x, Y \leq y) = \sum_{a_i \leq x} \sum_{b_j \leq y} f(a_i, b_j) \quad (\text{I3: 5})$$

Wir sprechen bei $F(x, y)$ auch von der **gemeinsamen empirischen Verteilungsfunktion** der Merkmale X und Y. Sie nimmt stets Werte im Intervall [0; 1] an.

Beispiel I 3-1

Kontingenztabelle

Von 80 Studienanfängern wurden die Abiturnoten in den Fächern Mathematik (X) und Englisch (Y) erfasst. Aus der Erhebung ergab sich die folgende bereits geordnete Urliste:

(1;1), (1;1), (1;1), (1;1), (1;2), (1;2), (1;3), (1;3), (1;3)

(2;1), (2;1), (2;1), (2;2), (2;2), (2;2), (2;2), (2;3), (2;3), (2;3), (2;3), (2;3), (2;3), (2;3), (2;3), (2;4), (2;4), (2;5), (2;5)

(3;1), (3;1), (3;2), (3;2), (3;2), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;3), (3;4), (3;4), (3;4), (3;4), (3;4), (3;4), (3;4), (3;4), (3;4), (3;4), (3;4), (3;4), (3;4), (3;5), (3;5)

(4;1), (4;1), (4;2), (4;2), (4;2), (4;3), (4;3), (4;3), (4;4), (4;4), (4;4), (4;4), (4;4), (4;5)

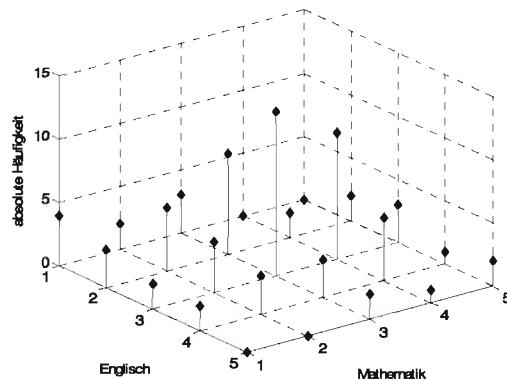
(5;3), (5;3), (5;4), (5;5), (5;5)

Eine Aufbereitung dieser Urliste liefert die folgende *Kontingenztabelle*:

		Englischnote Y				
		1	2	3	4	5
Mathe- matik- note X	1	4	2	3	-	-
	2	3	5	8	2	2
	3	2	4	13	10	3
	4	2	3	3	5	1
	5	-	-	2	1	2

Die Anzahl der Studenten, die sich z.B. sowohl mit einer Eins in Mathematik als auch in Englisch auszeichneten, lag somit bei 4.

Da hier Wertepaare häufiger auftreten, ist ein Streudiagramm nicht zur Visualisierung der Häufigkeitsverteilung geeignet. Ein dreidimensionales Stabdiagramm (komfortabel erzeugbar mit dem Befehl "stem3" in der Software Matlab 2009a) ergäbe im vorliegenden Fall folgende Gestalt:



Interessieren wir uns dafür, wie viele Studenten im Fach Mathematik nicht schlechter als Drei und in Englisch nicht schlechter als Zwei waren, können wir diese Häufigkeit aus einer Tabelle ablesen, die die absoluten gemeinsamen Häufigkeiten kumuliert:

		Englischnote Y				
		1	2	3	4	5
Mathe- matik- note X	1	4	6	9	9	9
	2	7	14	25	27	29
	3	9	20	44	56	61
	4	11	25	52	69	75
	5	11	25	54	72	80

Als Summe der Häufigkeiten im grau schattierten Bereich der Kontingenztabelle erhalten wir die kumulierte gemeinsame Häufigkeit 20, d.h. genau 20 Studenten sind in Mathematik nicht schlechter als Drei und gleichzeitig in Englisch nicht schlechter als Zwei. Obige Tabelle zeigt außerdem auf, dass $H(a_k, b_r) = n = 80$ gilt. Dividieren wir die Werte der Tabelle durch 80, erhalten wir die relativen Häufigkeiten der gemeinsamen empirischen Verteilungsfunktion.



3.1.2 Randhäufigkeiten und -verteilungen

Die Bestimmung sog. **Randhäufigkeiten** in einer Kontingenztafel erfolgt durch Bildung der Zeilen- und Spaltensummen der gemeinsamen Häufigkeiten. Durch dieses Summieren der gemeinsamen Häufigkeiten wird jeweils ein Merkmal "eliminiert", sodass innerhalb der zweidimensionalen Häufigkeitsverteilung die eindimensionalen Häufigkeitsverteilungen von X bzw. Y mit abgebildet werden.

Im Folgenden bezeichnen wir die Anzahl derjenigen Merkmalsträger, die bzgl. des Merkmals X die Ausprägung a_i aufweisen (gleichgültig welches die jeweilige Ausprägung des Merkmals Y ist), mit $h(a_i)$ und entsprechend die Anzahl derjeniger Merkmalsträger, die bzgl. des Merkmals Y die Ausprägung b_j aufweisen, mit $h(b_j)$. Für diese absoluten Randhäufigkeiten (relative Randhäufigkeiten analog) gilt damit formal

$$h(a_i) = \sum_{j=1}^r h_{ij} \quad \text{für } i = 1, 2, \dots, k \quad (\text{I3: 6})$$

und

$$h(b_j) = \sum_{i=1}^k h_{ij} \quad \text{für } j = 1, 2, \dots, r. \quad (\text{I3: 7})$$

Die Zeilensummen (häufig auch bezeichnet mit $h_i(a)$ oder $h_{\cdot i}$) liefern also die Häufigkeiten der eindimensionalen Verteilung des Merkmals X und die Spaltensummen (häufig auch bezeichnet mit $h_j(b)$ oder $h_{i \cdot j}$) die Häufigkeiten der eindimensionalen Verteilung des Merkmals Y. Diese bilden zusammen mit den jeweiligen Merkmalsausprägungen die sog. **Randverteilungen** von X und Y und sind in Abbildung I 20 dargestellt. Da sie eindimensionale Verteilungen sind, können auf sie die bereits behandelten Auswertungsmöglichkeiten angewendet werden.

		Y						Zeilensummen (Randverteilung X)
		b_1	b_2	...	b_j	...	b_r	
X	a_1	h_{11}	h_{12}	...	h_{1j}	...	h_{1r}	$h(a_1)$
	a_2	h_{21}	h_{22}	...	h_{2j}	...	h_{2r}	$h(a_2)$
	:	:	:		:		:	:
	a_i	h_{i1}	h_{i2}	...	h_{ij}	...	h_{ir}	$h(a_i)$
	:	:	:		:		:	:
	a_k	h_{k1}	h_{k2}	...	h_{kj}	...	h_{kr}	$h(a_k)$
Spaltensummen (Randverteilung Y)		$h(b_1)$	$h(b_2)$...	$h(b_j)$...	$h(b_r)$	n

Abbildung I 20: Kontingenztafel mit Randhäufigkeiten

Die relativen Randhäufigkeiten $f(a_i)$ und $f(b_j)$ bestimmen wir wie bisher über

$$f(a_i) = \frac{h(a_i)}{n} \quad \text{und} \quad f(b_j) = \frac{h(b_j)}{n}. \quad (\text{I3: 8})$$

Eine wichtige Eigenschaft der absoluten Randhäufigkeiten ist aus Abbildung I 20 abzulesen. So ist die Summe der Zeilensummen stets gleich der Summe der Spaltensummen. Dies gilt auch für die relativen Randhäufigkeiten.

$$\sum_{i=1}^k h(a_i) = \sum_{j=1}^r h(b_j) = \sum_{i=1}^k \sum_{j=1}^r h_{ij} = n \quad \sum_{i=1}^k f(a_i) = \sum_{j=1}^r f(b_j) = \sum_{i=1}^k \sum_{j=1}^r f_{ij} = 1 \quad (\text{I3: 9})$$

Summe der Zeilensummen Summe der Spaltensummen Summe der gemeinsamen Häufigkeiten

Summe der Zeilensummen Summe der Spaltensummen Summe der gemeinsamen Häufigkeiten

Beispiel I 3-2

Kontingenztabelle und Randverteilungen

Für die Kontingenztabelle der Schulnoten aus Beispiel I 3-1 ergibt sich:

		Englischnote Y					$h(a_i)$
		1	2	3	4	5	
Mathe- matik- note	1	4	2	3	-	-	9
	2	3	5	8	2	2	20
	3	2	4	13	10	3	32
	X	2	3	3	5	1	14
	5	-	-	2	1	2	5
$h(b_j)$		11	14	29	18	8	80

Daraus lassen sich nun die eindimensionalen Häufigkeitsverteilungen der Englisch- und Mathematiknoten ableiten, für die wieder die bekannten Kennzahlen berechenbar sind:

Englischnote	1	2	3	4	5
Häufigkeit $h(b_j)$	11	14	29	18	8
Mathematiknote	1	2	3	4	5
Häufigkeit $h(a_i)$	9	20	32	14	5

3.1.3 Bedingte Häufigkeiten und Verteilungen

Bei der Untersuchung des Zusammenhangs zweier Merkmale X und Y spielen die sog. **bedingten Häufigkeiten** eine besondere Rolle. Sie geben an, mit welcher relativen Häufigkeit eine Ausprägung eines Merkmals auftritt, wenn (unter der Bedingung, dass) das andere Merkmal auf einem bestimmten Wert festgehalten wird.



Die relative Häufigkeit der Ausprägung a_i des Merkmals X unter der Bedingung, dass das zweite Merkmal Y die konstante Ausprägung b_j besitzt, bezeichnen wir im Folgenden mit $f_x(a_i | b_j)$ für $i = 1, 2, \dots, k$. Anders ausgedrückt, betrachten wir also die relative Häufigkeit, mit der die Ausprägung a_i bei denjenigen Merkmalsträgern auftritt, die bzgl. des Merkmals Y die Ausprägung b_j aufweisen. Analog verwenden wir $f_y(b_j | a_i)$ für $j = 1, 2, \dots, r$. Diese bedingten Häufigkeiten können leicht aus einer vorliegenden Kontingenztabelle errechnet werden. Sie entstehen durch Division der gemeinsamen Häufigkeiten durch die entsprechenden Zeilen- oder Spalten-

summenhäufigkeiten (Randhäufigkeiten der Bedingungen). Es spielt dabei keine Rolle, ob absolute oder relative Häufigkeiten zur Berechnung der bedingten Häufigkeiten herangezogen werden.

$$f_x(a_i | b_j) = \frac{h_{ij}}{h(b_j)} = \frac{\frac{h_{ij}}{n}}{\frac{h(b_j)}{n}} = \frac{f_{ij}}{f(b_j)} \quad (\text{I3: 10})$$

$$f_y(b_j | a_i) = \frac{h_{ij}}{h(a_i)} = \frac{\frac{h_{ij}}{n}}{\frac{h(a_i)}{n}} = \frac{f_{ij}}{f(a_i)} \quad (\text{I3: 11})$$

Wie im unbedingten Fall, sind die bedingten relativen Häufigkeiten auf Eins normiert, d.h. es gilt

$$\sum_{i=1}^k f_x(a_i | b_j) = \sum_{j=1}^r f_y(b_j | a_i) = 1. \quad (\text{I3: 12})$$

Beispiel I 3-3

Bedingte Häufigkeiten

Eine Befragung von 1.000 berufstätigen Personen (Freiberufler, Angestellte, Beamte,...) ergab unter anderem, dass 250 Freiberufler nie beim Einzelhändler A einkaufen. Insgesamt meiden 500 der Befragten den besagten Einzelhändler. Die Anzahl der befragten Freiberufler lag bei 400.

Wir erhalten aus diesen Angaben folgende (unvollständige) Kontingenztabelle:

Einkaufsverhalten	Nie bei Einzelhändler A (= b_1)	...	Randhäufigkeiten
Berufsgruppe			
Freiberufler (= a_1)	$h_{11} = 250$...	$h(a_1) = 400$
...
Randhäufigkeiten	$h(b_1) = 500$...	$n = 1.000$

In diesem Beispiel lässt sich berechnen, dass die relative Häufigkeit der Freiberufler unter den Personen, die niemals beim Einzelhändler A einkaufen, bei 50 % liegt:

$$f_x(\text{Freiberufler} | \text{Niemals Kauf bei Einzelhändler A}) = f_x(a_1 | b_1) = \frac{h_{11}}{h(b_1)} = \frac{250}{500} = 0,50$$

Die relative Häufigkeit der Personen, die nie beim Einzelhändler A einkaufen, liegt unter den Freiberuflern bei 62,50 %:

$$f_y(\text{Niemals Kauf bei Einzelhändler A} | \text{Freiberufler}) = f_y(b_1 | a_1) = \frac{h_{11}}{h(a_1)} = \frac{250}{400} = 0,625$$

Aus (I3: 10) und (I3: 11) lassen sich sog. **bedingte Verteilungen** ableiten. Die bedingte Verteilung des Merkmals X bei gegebener Ausprägung b_j des Merkmals Y ergibt sich dabei aus den Ausprägungen a_1, a_2, \dots, a_k und den zugehörigen bedingten Häufigkeiten $f_x(a_1 | b_j), f_x(a_2 | b_j), \dots, f_x(a_k | b_j)$. Es existieren also insgesamt r

($j = 1 \dots r$) bedingte Verteilungen von X. Die bedingte Verteilung des Merkmals Y bei gegebener Ausprägung a_i des Merkmals X ergibt sich analog aus b_1, b_2, \dots, b_r mit $f_y(b_1 | a_i), f_y(b_2 | a_i), \dots, f_y(b_r | a_i)$, sodass insgesamt k ($i = 1 \dots k$) bedingte Verteilungen von Y aufstellbar sind.

Beispiel I 3-4

Bedingte Verteilungen

Zum bekannten Schulnotenbeispiel I 3-2 wollen wir nun die bedingte Verteilung der Mathematiknoten für die gegebene Englischnote 2 aufstellen. Dazu benötigen wir folgende (grau schattierte) Daten:

	Englischnote Y					
	1	2	3	4	5	$h(a_i)$
Mathe- matik- note X	1	4	2	3	-	9
	2	3	5	8	2	20
	3	2	4	13	10	32
	4	2	3	3	5	14
	5	-	-	2	1	5
$h(b_j)$	11	14	29	18	8	80

Die gesuchte bedingte Verteilung erhalten wir daraus durch Division der gemeinsamen Häufigkeiten durch die zugehörige Spaltensumme. Dies erklärt auch, warum die bedingten Häufigkeiten in der Summe Eins ergeben müssen.

a_i	h_{i2}	$f_x(a_i Y=2) = \frac{h_{i2}}{h(Y=2)}$
1	2	$2/14 = 0,14$
2	5	$5/14 = 0,36$
3	4	$4/14 = 0,29$
4	3	$3/14 = 0,21$
5	-	$0/14 = 0,00$
Σ	14	1,00

Wie die Randverteilungen sind auch die bedingten Verteilungen eindimensionale Häufigkeitsverteilungen, sodass wir hier ebenfalls die in Kapitel I 1 vorgestellten Maßzahlen zur verdichteten Beschreibung verwenden können. Wir bezeichnen sie in diesem Kontext als **bedingte Maßzahlen**. Ihre Berechnung erfolgt unter Nutzung der bedingten relativen Häufigkeiten.

Liegen zwei Merkmale X und Y mit den Realisationsmöglichkeiten a_i ($i = 1, 2, \dots, k$) bzw. b_j ($j = 1, 2, \dots, r$) sowie die zugehörigen bedingten Häufigkeiten $f(a_i | b_j)$ für $i = 1, 2, \dots, k$ bzw. $f(b_j | a_i)$ für $j = 1, 2, \dots, r$ vor, dann ist z.B. der bedingte Mittelwert von X gegeben Y = b_j gleich

$$\bar{x} | b_j = \sum_{i=1}^k a_i \cdot f_x(a_i | b_j) \quad (\text{I3: 13})$$

und der bedingte Mittelwert von Y gegeben X = a_i, gleich

$$\bar{y} | a_i = \sum_{j=1}^r b_j \cdot f_y(b_j | a_i) . \quad (\text{I3: 14})$$

Die bedingte Varianz von X gegeben Y = b_j ist

$$s_x^2 | b_j = \sum_{i=1}^k a_i^2 \cdot f_x(a_i | b_j) - (\bar{x} | b_j)^2 \quad (\text{I3: 15})$$

und die bedingte Varianz von Y gegeben X = a_i

$$s_y^2 | a_i = \sum_{j=1}^r b_j^2 \cdot f_y(b_j | a_i) - (\bar{y} | a_i)^2 . \quad (\text{I3: 16})$$

In entsprechender Weise können bedingte Mediane, bedingte Modi und andere bedingte Maßzahlen definiert werden.

Beispiel I 3-5

Bedingte Maßzahlen

Bestimmen wir für die in Beispiel I 3-4 aufgestellte bedingte Verteilung der Schulnoten bzw. die Häufigkeitsverteilung der Mathematiknoten unter den Studenten mit der Englischnote 2 den bedingten Mittelwert und die bedingte Varianz. Wir erhalten

$$\bar{x} | (Y=2) = 1 \cdot 0,14 + 2 \cdot 0,36 + \dots + 5 \cdot 0,00 = 2,57,$$

d.h. der Mathematiknotendurchschnitt unter den Studenten mit der Englischnote 2 liegt bei 2,57 bei einer Varianz von

$$s_x^2 | (Y=2) = 1^2 \cdot 0,14 + 2^2 \cdot 0,36 + \dots + 5^2 \cdot 0,00 - 2,57^2 = 0,95 .$$

3.1.4 Statistische Unabhängigkeit

Von statistischer Unabhängigkeit sprechen wir, wenn sich zwei Merkmale X und Y einer Häufigkeitsverteilung nicht beeinflussen, d.h. das Eintreten der verschiedenen Merkmalsausprägungen beim einen Merkmal unabhängig davon ist, welchen Wert das andere Merkmal annimmt. Ob Abhängigkeit oder Unabhängigkeit vorliegt, kann anhand der bedingten Verteilungen geprüft werden. **Statistische Unabhängigkeit** liegt vor, wenn die bedingten Verteilungen von X alle identisch sind und mit der jeweiligen Randverteilung übereinstimmen. Analoges muss für Y gelten.

$$\underbrace{f_x(a_i | b_1) = f_x(a_i | b_2) = \dots = f_x(a_i | b_r)}_{\text{bedingte Verteilungen}} = \underbrace{f(a_i)}_{\text{Randverteilung}} \quad \text{für } i = 1, \dots, k \quad (\text{I3: 17})$$

$$\underbrace{f_y(b_j | a_1) = f_y(b_j | a_2) = \dots = f_y(b_j | a_k)}_{\text{bedingte Verteilungen}} = \underbrace{f(b_j)}_{\text{Randverteilung}} \quad \text{für } j = 1, \dots, r \quad (\text{I3: 18})$$

Bedingungen bezüglich Y haben also keinen Einfluss auf die Verteilung von X und Bedingungen bezüglich X auch keinen Einfluss auf die Verteilung von Y. X und Y sind statistisch unabhängig.

Beispiel I 3-6**Statistische Unabhängigkeit**

Gegeben sei folgende zweidimensionale Verteilung der Merkmale X (Ausprägungen a_i) und Y (Ausprägungen b_j), die in Form absoluter und relativer Häufigkeiten vorliegt:

	b_1	b_2	b_3	$h(a_i)$		b_1	b_2	b_3	$f(a_i)$	
a_1	2	5	3	10		a_1	0,03	0,08	0,05	0,17
a_2	6	15	9	30		a_2	0,10	0,25	0,15	0,50
a_3	4	10	6	20		a_3	0,07	0,17	0,10	0,33
$h(b_j)$	12	30	18	60		$f(b_j)$	0,20	0,50	0,30	1,00

Um zu prüfen, ob die Merkmale X und Y unabhängig voneinander sind, benötigen wir die bedingten Verteilungen von X und Y. Die bedingten Verteilungen von X gegen Y sind in folgender Tabelle enthalten:

	b_1	b_2	b_3	* Rechenbeispiel:
a_1	0,17	0,17	0,17	$f_x(a_2 b_1) = \frac{h_{21}}{h(b_1)} = \frac{6}{12} = 0,50$
a_2	0,50	0,50	0,50	
a_3	0,33	0,33	0,33	

Wir erkennen, dass die bedingten Verteilungen (Spalten) alle identisch sind und mit der in der Ausgangstabelle schattierten Randverteilung von X übereinstimmen.

Die bedingten Verteilungen von Y gegen X haben folgende Gestalt:

	b_1	b_2	b_3	* Rechenbeispiel:
a_1	0,20	0,50	0,30	$f_y(b_1 a_2) = \frac{h_{21}}{h(a_2)} = \frac{6}{30} = 0,20$
a_2	0,20	0,50	0,30	
a_3	0,20	0,50	0,30	

Auch hier ist zu sehen, dass die bedingten Verteilungen (Zeilen) alle übereinstimmen und der schattierten Randverteilung von Y aus der Ausgangstabelle entsprechen.

Die beiden Merkmale können damit als statistisch unabhängig bezeichnet werden. Würden die Werte der bedingten Verteilungen von der beschriebenen Vorschrift abweichen, wären die Merkmale X und Y, wenn auch ggf. nur geringfügig, statistisch abhängig.

Aus (I3: 10) und dem Fakt, dass bei statistischer Unabhängigkeit $f_x(a_1 | b_1) = f(a_1)$ gilt, kann

$$f_{xy}(a_i, b_j) = f(a_i) \cdot f(b_j) \quad (\text{I3: 19})$$

gefolgert werden. Dies bedeutet, dass bei statistischer Unabhängigkeit die gemeinsamen relativen Häufigkeiten gleich dem Produkt der jeweiligen relativen Randhäufigkeiten sind. Die gemeinsamen absoluten Häufigkeiten erhalten wir, indem wir (I3: 19) wie folgt umformen:

$$n \cdot f_{xy}(a_i, b_j) = \frac{n \cdot f(a_i) \cdot n \cdot f(b_j)}{n} \leftrightarrow h_{xy}(a_i, b_j) = \frac{h(a_i) \cdot h(b_j)}{n} \quad (\text{I3: 20})$$

Die gesamte Information über die gemeinsame Verteilung ist also im Falle von Unabhängigkeit bereits in den beiden eindimensionalen Verteilungen von X und Y enthalten. Ist dies nicht der Fall, liegt *statistische Abhängigkeit* vor. Zwischen den Merkmalen bestehen sog. *Interdependenzen*.

Beispiel I 3-7		Statistische Abhängigkeit
-----------------------	--	----------------------------------

Eine Untersuchung unter 1.000 Personen zu den Merkmalen Raucher (X) mit den Ausprägungen "Raucher" (R) und "Nichtraucher" (NR) und Lungenkrebs (Y) mit den Ausprägungen "Lungenkrebs" (L) und "kein Lungenkrebs" (NL) ergab folgende Kontingenztabelle:

		Lungenkrebs Y		
		nein [NL]	ja [L]	$h(a_i) (f(a_i))$
Raucher X	nein [NR]	650 (0,65)	50 (0,05)	700 (0,70)
	ja [R]	250 (0,25)	50 (0,05)	300 (0,30)
$h(b_j) (f(b_j))$		900 (0,90)	100 (0,10)	1.000 (1,0)

Anhand des zuletzt behandelten Zusammenhangs (I3: 19) soll nun untersucht werden, ob die Merkmale Raucher und Lungenkrebs statistisch unabhängig sind. Dazu wird eine sog. *hypothetische Kontingenztafel* aufgestellt. Die gemeinsamen relativen Häufigkeiten werden darin durch Multiplikation der relativen Randhäufigkeiten bestimmt. Stimmen diese mit denen der ursprünglich vorhandenen Tabelle überein, liegt statistische Unabhängigkeit vor. Kommt es zu Abweichungen, so kann diese Aussage nicht mehr gemacht werden.

Hypothetische Kontingenztafel		Lungenkrebs Y		
		nein [NL]	ja [L]	$f_x(a_i)$
Raucher X	nein [NR]	0,63	0,07	0,70
	ja [R]	0,27	0,03	0,30
$f_y(b_j)$		0,90	0,10	1,0

Die gemeinsamen relativen Häufigkeiten f_{ij} weichen hier alle von denen der Ursprungstabelle ab. Die Merkmale X und Y sind also voneinander statistisch abhängig.

Würde anhand der Formel (I3: 18) auf Unabhängigkeit geprüft, so käme man zum gleichen Schluss. Es gilt nämlich (in vereinfachter Schreibweise):

$$f_y(NL|NR) = \frac{0,65}{0,70} = 0,93 \quad \neq \quad f_y(NL|R) = \frac{0,25}{0,30} = 0,83$$

$$f_y(L|NR) = \frac{0,05}{0,70} = 0,07 \quad \neq \quad f_y(L|R) = \frac{0,05}{0,30} = 0,17$$

Die Bedingungen haben also einen Einfluss auf die Verteilung von Y.

Von besonderer Bedeutung im Kontext des Begriffs der Unabhängigkeit ist, dass die von uns in diesem Abschnitt behandelten Unabhängigkeitsaussagen im Grunde nur für eine *vorliegende Grundgesamtheit* gelten. In einer Stichprobe wird es nämlich fast immer zufallsbedingt Abweichungen von der Unabhängigkeit geben, selbst wenn die beschriebenen Unabhängigkeitsbedingungen in der Grundgesamtheit erfüllt sind. Das Problem der Unabhängigkeit zweier Variablen wird deshalb in der **induktiven Statistik** differenziert behandelt (vgl. Kapitel III).

3.2 Korrelationsanalyse

Im Rahmen der Korrelationsanalyse werden diverse Maßzahlen zur Quantifizierung des Ausmaßes des statistischen Zusammenhangs zwischen verschiedenen Merkmalen berechnet. Im Folgenden beschränken wir uns dabei wieder auf nur zwei Merkmale X und Y und werden für diese die Kovarianz, den Bravais-Pearson-Korrelationskoeffizienten, den Spearman-Rangkorrelationskoeffizienten sowie den Kontingenzkoeffizienten näher behandeln. Die Anwendbarkeit der Maßzahlen hängt entscheidend vom Skalenniveau der Variablen X und Y ab. Abbildung I 21 zeigt, für welche Skalierungen welche Korrelationsparameter heranzuziehen sind. Zudem gibt sie einen ersten Überblick darüber, welche Informationen durch die jeweilige Messzahl gewonnen werden können.

X Y	kardinal	ordinal	nominal
kardinal	Bravais-Pearson-Korrelationskoeffizient ¹	↑	↑
ordinal	←	Spearman-Rangkorrelationskoeffizient ²	↑
nominal	←	←	Kontingenzkoeffizient ³

¹: Misst Stärke und Richtung eines linearen Zusammenhangs

²: Misst Stärke und Richtung eines monotonen Zusammenhangs

³: Misst nur die Stärke eines Zusammenhangs

Abbildung I 21: Korrelationsmaße und Skalierung

3.2.1 Kovarianz und Bravais-Pearson-Korrelationskoeffizient

Für zwei *metrisch skalierte* Merkmale X und Y liege eine *Urliste* vor, in der weder alle x_i -Werte noch alle y_j -Werte identisch sind. Zudem seien \bar{x} und \bar{y} die jeweiligen arithmetischen Mittel. Da zu jedem x_i nur gerade ein y_j gehört, können wir auf den Index j verzichten, d.h. $j = i$. Damit ist die sog. **Kovarianz** definiert als

$$\text{Cov}_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) . \quad (\text{I3: 21})$$

Im Falle einer vorliegenden *Häufigkeitstabelle* erhalten wir sie als

$$\text{Cov}_{XY} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^r (a_i - \bar{x}) \cdot (b_j - \bar{y}) \cdot h_{ij} = \sum_{i=1}^k \sum_{j=1}^r (a_i - \bar{x}) \cdot (b_j - \bar{y}) \cdot f_{ij} . \quad (\text{I3: 22})$$

Eine erste Betrachtung der Formel der Kovarianz zeigt, dass die Kovarianz eines Merkmals X mit sich selbst gleich seiner Varianz ist, d.h. $\text{Cov}_{xx} = s_x^2$ gilt.

Zur Interpretation des Werts der Kovarianz betrachten wir Abbildung I 22. Es enthält drei typische Streudiagramme zweidimensionaler Häufigkeitsverteilungen. Jedes der Streudiagramme wurde durch ein Fadenkreuz ergänzt, dessen Mittelpunkt der Schwerpunkt (\bar{x}, \bar{y}) des jeweiligen Datensatzes ist. Es entstehen so in jedem Diagramm 4 Quadranten I, II, III und IV.

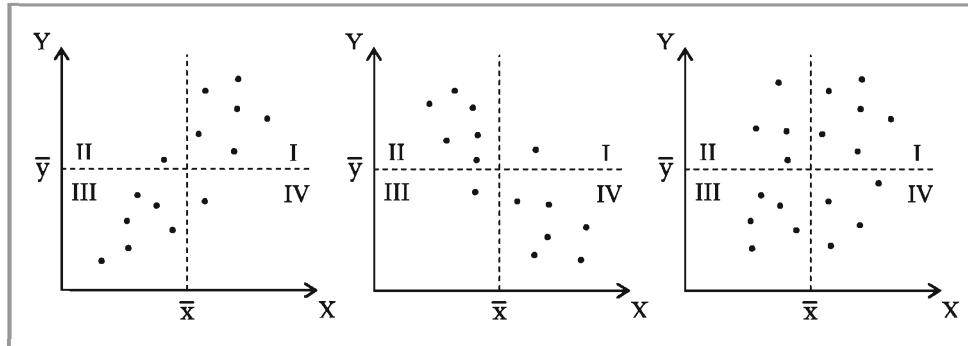


Abbildung I 22: Interpretation der Kovarianz

In den ersten beiden Streudiagrammen ist jeweils eine gemeinsame Tendenz der Beobachtungen von X und der von Y zu erkennen. Im ersten Diagramm liegen die Beobachtungspaare (x_i, y_i) überwiegend in den Quadranten I und III. Große X-Werte gehen hauptsächlich mit großen Y-Werten und kleine X-Werte mit kleinen Y-Werten einher. Es liegt ein sog. *positiver Zusammenhang* vor. Im zweiten Diagramm liegt die Mehrzahl der Beobachtungen in den Quadranten II und IV. Große X-Werte korrespondieren mit kleinen Y-Werten und umgekehrt. Es liegt ein sog. *negativer Zusammenhang* vor. Das dritte Diagramm lässt *keinen Zusammenhang* erkennen, da korrespondierend zu kleinen Werten der einen Variablen ähnlich viele kleine wie große Werte der anderen Variablen zu beobachten sind. Die Begriffe "klein" und "groß" sind hier relativ zum jeweiligen Mittelwert zu sehen.

Den Schlüssel zum Verständnis der Kovarianz bilden die Abweichungsprodukte $(x_i - \bar{x})(y_i - \bar{y})$. Liegt ein Beobachtungspaar (x_i, y_i) in

- | | | |
|--------------|------|--------------------------------------------------------------------------------|
| Quadrant I | gilt | $x_i > \bar{x}, y_i > \bar{y} \rightarrow (x_i - \bar{x})(y_i - \bar{y}) > 0,$ |
| Quadrant III | gilt | $x_i < \bar{x}, y_i < \bar{y} \rightarrow (x_i - \bar{x})(y_i - \bar{y}) > 0,$ |
| Quadrant II | gilt | $x_i < \bar{x}, y_i > \bar{y} \rightarrow (x_i - \bar{x})(y_i - \bar{y}) < 0,$ |
| Quadrant IV | gilt | $x_i > \bar{x}, y_i < \bar{y} \rightarrow (x_i - \bar{x})(y_i - \bar{y}) < 0.$ |

Da die Kovarianz das arithmetische Mittel der Abweichungsprodukte darstellt, gilt:

- Liegen die Datenpunkte hauptsächlich in den Quadranten I und III, so besteht ein positiver Zusammenhang. Cov_{xy} ist positiv.
- Liegen die Datenpunkte hauptsächlich in den Quadranten II und IV, so besteht ein negativer Zusammenhang. Cov_{xy} ist negativ.
- Sind die Punkte gleichmäßig auf die Quadranten verteilt, besteht kein Zusammenhang. Positive und negative Abweichungsprodukte heben sich bei der Mittelung weitgehend auf und Cov_{xy} liegt nahe bei Null.

Ebenso wie für die Varianz gilt auch für die Kovarianz ein *Verschiebungssatz*. Dieser liefert den Ausdruck

$$\text{Cov}_{XY} = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \bar{x} \cdot \bar{y} .^1 \quad (\text{I3: 23})$$

Im Falle einer Häufigkeitsverteilung hat (I3: 23) die Form

$$\text{Cov}_{XY} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^r a_i \cdot b_j \cdot h_{ij} - \bar{x} \cdot \bar{y} = \sum_{i=1}^k \sum_{j=1}^r a_i \cdot b_j \cdot f_{ij} - \bar{x} \cdot \bar{y} . \quad (\text{I3: 24})$$

Die Kovarianz ist in der Lage den empirischen Zusammenhang zweier Variablen X und Y anzuzeigen. $\text{Cov}_{XY} > 0$ zeigt einen positiven, $\text{Cov}_{XY} < 0$ einen negativen Zusammenhang. Die Kovarianz ist also in der Lage die *Richtung*, nicht jedoch die *Stärke* des Zusammenhangs aufzuzeigen. Ein sehr großer positiver Wert bedeutet beispielsweise nicht zwangsläufig, dass ein sehr starker positiver Zusammenhang vorliegt. Die Kovarianz ist nämlich eine dimensionsbehaftete Größe, die allein durch die Änderung der Maßeinheit größer oder kleiner werden kann. Um diesem Problem zu begegnen, wird eine *normierte Kovarianz* als Kenngröße verwendet:

Dividieren wir die Kovarianz Cov_{XY} durch das Produkt der Standardabweichungen von X und Y, also s_x und s_y , erhalten wir den **Bravais-Pearson-Korrelationskoeffizienten** r_{xy} , d.h.

$$r_{xy} = \frac{\text{Cov}_{XY}}{s_x \cdot s_y} = \frac{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} , \quad (\text{I3: 25})$$

bzw. vereinfacht

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} .$$

In (I3: 25) können entsprechend auch die Formeln verwendet werden, die sich des Verschiebungssatzes bedienen und/oder auf der Häufigkeitstabelle basieren. Durch die Division durch das Standardabweichungsprodukt (Normierung) ist r_{xy} automatisch *maßstabslos* und verläuft im Intervall

$$-1 \leq r_{xy} \leq +1 \quad \text{bzw. } |r_{xy}| \leq 1. \quad (\text{I3: 26})$$

$r_{xy} = \pm 1$ gilt genau dann, wenn zwischen X und Y eine *exakte lineare Beziehung* besteht. Existiert nämlich die lineare Beziehung $y_i = a + bx_i$, dann gilt

$$\bar{y} = a + b\bar{x} \quad \text{und} \quad s_y^2 = b^2 s_x^2 .$$

¹ Den rechnerischen Beweis der Gültigkeit dieses Zusammenhangs überlassen wir in Aufgabe MH6 dem Leser.

Ferner ist

$$\begin{aligned}\text{Cov}_{XY} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(a + bx_i - [a + b\bar{x}]) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(b[x_i - \bar{x}]) \\ &= \frac{b}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = bs_x^2.\end{aligned}$$

Daraus folgt schließlich

$$r_{XY} = \frac{\text{Cov}_{XY}}{s_X \cdot s_Y} = \frac{bs_x^2}{s_X \cdot |b| s_X} = \frac{b}{|b|} = \pm 1.$$

Aufgrund der genannten Eigenschaften des Korrelationskoeffizienten können wir nun folgern, dass der Korrelationskoeffizient r_{XY} (und damit natürlich auch die Kovarianz Cov_{XY}) ein Maß für den *linearen Zusammenhang* zweier Variablen X und Y ist. Gilt $|r_{XY}| = 1$, besteht eine *exakte lineare Beziehung*. Der lineare Zusammenhang ist umso schwächer, je kleiner der Absolutbetrag $|r_{XY}|$ des Koeffizienten ist. Gilt $r_{XY} = 0$, besteht keine *lineare Beziehung*. Im Fall $r_{XY} = 0$ heißen die Merkmale X und Y *unkorreliert*, im Fall $|r_{XY}| = 1$ *perfekt korreliert*. Bei $r_{XY} > 0$ bzw. $r_{XY} < 0$ bezeichnen wir sie als *positiv korreliert* bzw. *negativ korreliert*. Bei $|r_{XY}| \geq 0,8$ ist von starker, bei $|r_{XY}| \leq 0,5$ von schwacher Korrelation auszugehen. Abbildung I 24, die auf Seite 98 zu finden ist, illustriert diese Aussagen.

Beispiel I 3-8

Bravais-Pearson-Korrelationskoeffizient

Bei n Filialen einer Firma wurden die Variablen X = Verkaufsfläche und Y = Umsatz erhoben. Es soll geprüft werden, ob zwischen diesen beiden Merkmalen ein linearer Zusammenhang besteht. Wir verwenden dazu folgende Arbeitstabelle:

i	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	100	800	10.000	640.000	80.000
2	150	1.000	22.500	1.000.000	150.000
3	110	800	12.100	640.000	88.000
4	140	900	19.600	810.000	126.000
5	160	1.200	25.600	1.440.000	192.000
6	200	1.600	40.000	2.560.000	320.000
7	110	600	12.100	360.000	66.000
8	130	1.000	16.900	1.000.000	130.000
9	180	1.500	32.400	2.250.000	270.000
10	190	1.300	36.100	1.690.000	247.000
Σ	1.470	10.700	227.300	12.390.000	1.669.000

Es ergeben sich daraus zunächst die folgenden arithmetischen Mittel:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{10} \cdot 1.470 = 147 \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i = \frac{1}{10} \cdot 10.700 = 1.070$$

Anhand der mittels Verschiebungssatz umgestellten Formel ergibt sich r_{XY} zu

$$r_{XY} = \frac{\frac{1.669.000}{10} - 147 \cdot 1.070}{\sqrt{\frac{227.300}{10} - 147^2} \cdot \sqrt{\frac{12.390.000}{10} - 1.070^2}} = 0,936.$$

Der Wert des Korrelationskoeffizienten ist positiv und liegt nahe 1. Es liegt also eine starke positive Korrelation vor. Dies kann nun bedeuten, dass die Filialen durch eine Vergrößerung der Verkaufsfläche tendenziell eine Umsatzsteigerung erreichen oder auch, dass Filialen mit hohem Umsatz eher in der Lage sind größere Verkaufsflächen zu finanzieren. Die Korrelationsanalyse kann also keine Aussagen über Ursache und Wirkung (Kausalität) machen (vgl. Abschnitt I 3.2.5).

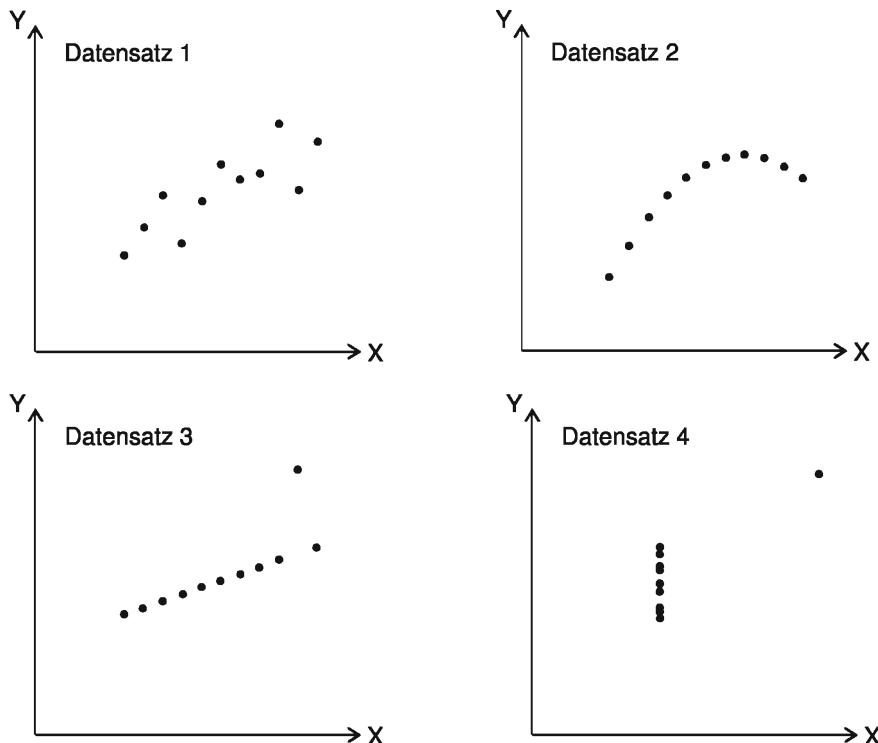
Beispiel I 3-9

Anscombe's Korrelationsdaten

Die folgende Tabelle enthält vier von Anscombe, F.J. (1973) konstruierte zweidimensionale Datensätze, die alle eine starke positive Korrelation von $r_{XY} = 0,82$ aufweisen. Die Datensätze 1 bis 3 haben alle dieselben X-Werte.

Datensatz Merkmal	1 bis 3 X	1 Y	2 Y	3 Y	4 X	4 Y
	10	8,04	9,14	7,46	8	6,58
	8	6,95	8,14	6,77	8	5,76
	13	7,58	8,74	12,74	8	7,71
	9	8,81	8,77	7,11	8	8,84
	11	8,33	9,26	7,81	8	8,47
	14	9,96	8,10	8,84	8	7,04
	6	7,24	6,13	6,08	8	5,25
	4	4,26	3,10	5,39	8	5,56
	12	10,84	9,13	8,15	8	7,91
	7	4,82	7,26	6,42	8	6,89
	5	5,68	4,74	5,73	19	12,50

Die dazugehörigen Streudiagramme auf der Folgeseite zeigen jedoch vollständig unterschiedliche Datenstrukturen. Betrachten wir zunächst die Datensätze 3 und 4, dann wird aufgrund der Streudiagramme ersichtlich, dass Cov_{XY} und r_{XY} keine robusten Maßzahlen sind. Sie sind empfindlich gegenüber Ausreißern. Einzelne extreme Beobachtungswerte können dazu führen, dass die Maßzahlen einen starken linearen Zusammenhang unterbewerten (Datensatz 3) oder fälschlicherweise eine stark lineare Beziehung ausweisen (Datensatz 4). Für die Datensätze 1 und 2 nimmt der Korrelationskoeffizient identische Werte an, obwohl, im Gegensatz zu Datensatz 1, zwischen den Variablen in Datensatz 2 eine perfekte Beziehung besteht. Hier muss berücksichtigt werden, dass Cov_{XY} und r_{XY} nicht Maße für statistische Zusammenhänge schlechthin sind. Sie beurteilen lediglich die lineare Beziehung zwischen zwei Variablen. In Datensatz 2 liegt aber eine nichtlineare Beziehung vor.



Gerade dieses Beispiel zeigt, dass eine Korrelationsrechnung immer durch eine grafische Beschreibung der Daten mittels eines Streudiagrammes begleitet werden sollte.

Beispiel I 3-10

Nichtlinearer Zusammenhang

Betrachten wir folgenden zweidimensionalen Datensatz:

x_i	-3	-2	-1	0	1	2	3
y_i	9	4	1	0	1	4	9

Wir erkennen sofort, dass für die Kovarianz und den Korrelationskoeffizienten gilt:

$$\text{Cov}_{XY} = \frac{1}{7} \sum_{i=1}^7 x_i y_i - \bar{x}\bar{y} = 0 - 0 = 0 \quad \rightarrow \quad r_{XY} = 0$$

Die Variablen weisen keine lineare Beziehung auf. Es gibt jedoch eine exakte Beziehung zwischen ihnen, da $y_i = x_i^2$ gilt. Der Zusammenhang ist quadratisch, also nichtlinear, und kann somit nicht vom Bravais-Pearson-Korrelationskoeffizienten aufgedeckt werden.

Im Falle *statistisch unabhängiger* Merkmale X und Y nimmt r_{xy} den Wert 0 an. Sind X und Y *unkorreliert*, dann sind sie *nicht* unbedingt auch *statistisch unabhängig*. Es besteht dann lediglich kein *linearer* Zusammenhang. Eine z.B. quadratische Abhängigkeit könnte durchaus bestehen.

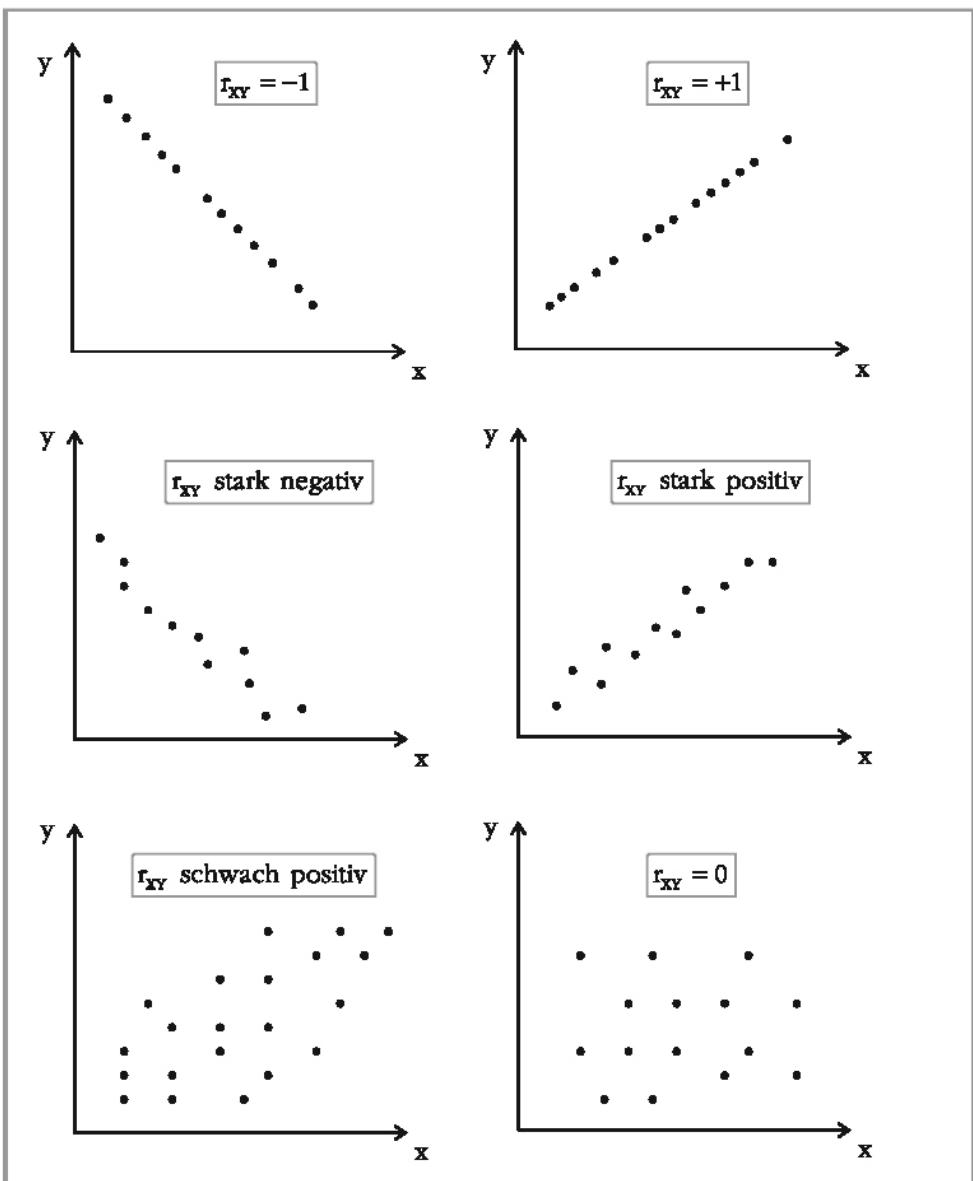


Abbildung I 23: Bravais-Pearson-Korrelationskoeffizient und Streudiagramm

3.2.2 Spearman-Rangkorrelationskoeffizient

Der sog. Spearman-Rangkorrelationskoeffizient ist eine *robuste* Variante des Bravais-Pearson-Korrelationskoeffizienten, die zudem auch bei *ordinalen* Merkmalen einsetzbar ist. Er basiert nicht direkt auf den Merkmalsausprägungen x_i und y_i der Merkmale X und Y, sondern auf ihnen zugeordneten Rangnummern R_i^x und R_i^y .



Im Zuge der Berechnung des Rangkorrelationskoeffizienten wird der größten Ausprägung von X der Rang 1, der zweitgrößten Ausprägung von X der Rang 2, usw. zugeordnet. Tritt ein Wert des Merkmals X mehrmals auf, d.h. liegen sog. *Bindungen* vor, dann wird diesen das arithmetische Mittel der entsprechenden Rangzahlen zugeordnet. Entsprechend werden auch den Ausprägungen des Merkmals Y Rangzahlen zugeordnet.

Beispiel I 3-11**Zuordnung von Rangzahlen**

Gegeben sei folgendes Datenmaterial:

i	1	2	3	4	5
x_i	8	3	9	10	7
y_i	2	1	1	2	2
R_i^x	3	5	2	1	4
R_i^y	2	4,5	4,5	2	2

Die Beobachtungen von X weisen keine Bindungen auf. Wir können daher dem Wert 10 den Rang 1, dem Wert 9 den Rang 2, usw. zuordnen. Die Beobachtungen von Y weisen Bindungen auf. Der größte Wert 2 tritt dreimal auf. Mit diesen drei Werten wären die Ränge 1, 2 und 3 verbunden, die nun gemittelt werden und so den Werten 2 einheitlich der Rang $(1 + 2 + 3)/3 = 2$ zugeordnet wird. Der Wert 1 weist erneut eine Bindung auf, sodass bei den Werten 1 der Rang $(4 + 5)/2 = 4,5$ zugeordnet wird.

Wenden wir den Bravais-Pearson-Korrelationskoeffizienten auf die ermittelten Rangziffern an, so erhalten wir den **Spearman-Rangkorrelationskoeffizient**

$$r_{xy}^{sp} = \frac{\text{Cov}_{R^x R^y}}{s_{R^x} \cdot s_{R^y}} = \frac{\frac{1}{n} \cdot \sum_{i=1}^n (R_i^x - \bar{R}^x) \cdot (R_i^y - \bar{R}^y)}{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (R_i^x - \bar{R}^x)^2} \cdot \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (R_i^y - \bar{R}^y)^2}}, \quad (\text{I3: 27})$$

wobei \bar{R}^x und \bar{R}^y die Mittelwerte der jeweiligen Rangziffern darstellen. Gibt es *keine Bindungen*, können wir r_{xy}^{sp} alternativ auch über

$$r_{xy}^{sp} = 1 - \frac{6 \cdot \sum_{i=1}^n (R_i^x - R_i^y)^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot \sum_{i=1}^n D_i^2}{n \cdot (n^2 - 1)} \quad \text{mit } D_i = R_i^x - R_i^y \quad (\text{I3: 28})$$

berechnen. Die D_i werden dabei als Rangdifferenzen bezeichnet. Sind im Datensatz relativ zum Stichprobenumfang n wenige Bindungen vorhanden, liefert (I3: 28) noch einen guten Näherungswert für (I3: 27).

Unabhängig von der Art der Berechnung kann r_{xy}^{sp} wie auch r_{xy} Werte im Intervall

$$-1 \leq r_{xy}^{sp} \leq +1 \quad (\text{I3: 29})$$

annehmen, da der Spearman-Rangkorrelationskoeffizient eine normierte, dimensionslose Größe ist. $r_{xy}^{sp} = \pm 1$ gilt genau dann, wenn eine beliebige monoton wachsende bzw. eine beliebige monoton fallende Funktion $f(x)$ existiert, sodass $y_i = f(x_i)$ erfüllt ist. Wir sagen dann auch, dass zwischen X und Y ein *monotoner Zusam-*

menhang besteht. Für den Fall strikter *Gleichhäufigkeit der Rangzahlen* ($R_i^x = R_i^y$ für $i = 1, 2, \dots, n$) nimmt der Rangkorrelationskoeffizient den Extremwert +1 an. Dies wäre z.B. der Fall, wenn in einer Untersuchung des Zusammenhangs zwischen Größe und Gewicht von Personen die größte Person zugleich die schwerste, die zweitgrößte Person zugleich die zweitschwerste, ... und die kleinste Person zugleich die leichteste ist. Der andere Extremwert -1 wird bei strikter *Gegenhäufigkeit der Rangnummern* (bei nicht vorliegenden Bindungen $R_i^x = n + 1 - R_i^y$ für $i = 1, 2, \dots, n$) erreicht. Dies wäre z.B. der Fall, wenn die größte Person zugleich die leichtestste, ... und die kleinste Person zugleich die schwerste ist. Abbildung I 24 veranschaulicht diesen und andere Fälle.

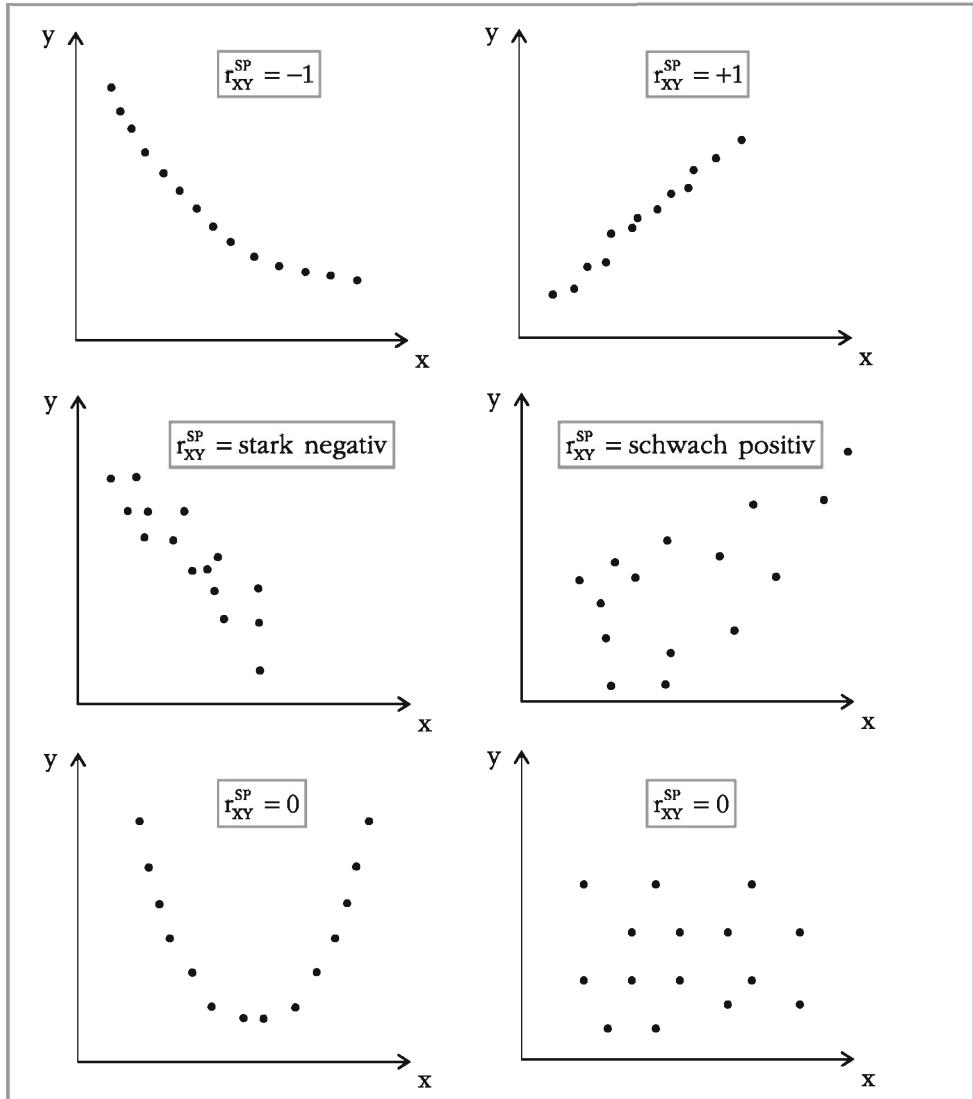


Abbildung I 24: Spearman-Rangkorrelationskoeffizient und Streudiagramm

Beispiel I 3-12**Spearman-Rangkorrelationskoeffizient**

10 Betriebswirtschaftsstudenten erreichten in ihren Drittversuchen in den Klausuren Statistik und Wirtschaftsmathematik folgende Gesamtpunktzahlen:

Student	1	2	3	4	5	6	7	8	9	10
Statistik	33	41	45	55	33	27	20	15	13	18
Wirtschaftsmathematik	37	38	32	49	44	21	11	11	11	28

Untersuchen wir nun hierfür, ob sich ein Zusammenhang zwischen den Leistungen in den beiden Fächern feststellen lässt. Zur Berechnung des Rangkorrelationskoeffizienten verwenden wir die folgende Arbeitstabelle. Diese ist nahezu selbsterklärend. Lediglich zwei Wertegruppen verdienen nähere Aufmerksamkeit. Da x_1 und x_5 identisch sind, teilen wir die Ränge 4 und 5 gleichmäßig auf diese beiden Beobachtungswerte auf, sodass wir für beide den Rang $(4 + 5)/2 = 4,5$ erhalten. Ein ähnliches Problem liegt bei y_7 , y_8 und y_9 vor. Diesen teilen wir daher den Rang $(8 + 9 + 10)/3 = 9$ zu.

i	x_i	y_i	R_i^x	R_i^y	$(R_i^x)^2$	$(R_i^y)^2$	$R_i^x \cdot R_i^y$	D_i	D_i^2
1	33	37	4,5	4	20,25	16	18	0,5	0,25
2	41	38	3	3	9	9	9	0	0
3	45	32	2	5	4	25	10	-3	9
4	55	49	1	1	1	1	1	0	0
5	33	44	4,5	2	20,25	4	9	2,5	6,25
6	27	21	6	7	36	49	42	-1	1
7	20	11	7	9	49	81	63	-2	4
8	15	11	9	9	81	81	81	0	0
9	13	11	10	9	100	81	90	1	1
10	18	28	8	6	64	36	48	2	4
Σ	-	-	55	55	384,5	383	371	-	25,5

Aus dieser Tabelle ergibt sich nun der Rangkorrelationskoeffizient nach (I3: 27) unter Berücksichtigung des Verschiebungssatzes für die Kovarianz und die Varianz zu

$$r_{xy}^{sp} = \frac{\frac{1}{10} \cdot 371 - \frac{55}{10} \cdot \frac{55}{10}}{\sqrt{\frac{1}{10} \cdot 384,5 - \left(\frac{55}{10}\right)^2} \cdot \sqrt{\frac{1}{10} \cdot 383 - \left(\frac{55}{10}\right)^2}} = \frac{6,85}{\sqrt{8,2} \cdot \sqrt{8,05}} = 0,8431.$$

Es liegt also ein stark positiver Zusammenhang zwischen den Leistungen in Statistik und Wirtschaftsmathematik vor. Dies bedeutet, dass Studenten, die im Drittversuch gut in Statistik sind auch tendenziell besser in Wirtschaftsmathematik abschneiden oder umgekehrt.

Eine Anwendung von (I3: 28) trotz vorhandener Bindungen ergäbe

$$r_{xy}^{sp} = 1 - \frac{6 \cdot \sum_{i=1}^n D_i^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 25,5}{10 \cdot (100 - 1)} = 0,8455.$$

Trotz der kleinen Stichprobe und der vorhandenen Bindungen erhalten wir also einen sehr guten Näherungswert.

Beispiel I 3-13**Anscombe's Korrelationsdaten**

Berechnen wir für die vier Datensätze aus Beispiel I 3-9 den Spearman-Rangkorrelationskoeffizienten, so erhalten wir folgende Ergebnisse:

Datensatz	1	2	3	4
r_{XY}^{Sp}	0,8182	0,6909	0,9909	0,5

Interessant ist hierbei der nahe 1 liegende Koeffizient für Datensatz 3. Dieser verdeutlicht, dass der Rangkorrelationskoeffizient hier kaum vom vorliegenden Ausreißer beeinflusst wird. Auch der im Vergleich zum Datensatz 1 geringere Rangkorrelationskoeffizient bei Datensatz 2 ist aufgrund der drei Beobachtungen mit den größten x-Werten einleuchtend. Hinsichtlich Datensatz 4 ist zu erwähnen, dass bei derartig vielen Bindungen i.d.R. auf eine Berechnung des Spearman-Rangkorrelationskoeffizienten verzichtet wird.

3.2.3 Kontingenzkoeffizient

Als letztes Korrelationsmaß wollen wir im Folgenden den sog. Kontingenzkoeffizienten vorstellen, welcher auf *alle Skalenniveaus* anwendbar ist. Die Berechnung dieser Maßzahl erfolgt in 4 Schritten:



Schritt 1:

In Abschnitt I 3.1.4 haben wir gesehen, dass im Falle der *statistischen Unabhängigkeit* zweier Merkmale X und Y der Zusammenhang (I3: 20) gilt, d.h. die gesamte Information bezüglich der gemeinsamen Häufigkeitsverteilung in den eindimensionalen Randverteilungen enthalten ist. Dies können wir zur Bildung einer **hypothetischen Kontingenztabelle** verwenden. Die in dieser Tabelle entstehenden hypothetischen bzw. bei statistischer Unabhängigkeit erwarteten gemeinsamen absoluten Häufigkeiten $\tilde{h}_{ij}(a_i, b_j)$, die wir im Folgenden mit \tilde{h}_{ij} abkürzen, sind

$$\tilde{h}_{ij} = \frac{h(a_i) \cdot h(b_j)}{n} = n \cdot f(a_i) \cdot f(b_j). \quad (\text{I3: 30})$$

Je stärker nun die tatsächlichen Häufigkeiten h_{ij} von den hypothetischen Häufigkeiten \tilde{h}_{ij} abweichen, desto größer ist die wechselseitige Abhängigkeit der Merkmale X und Y.

Schritt 2:

Ein Maß, welches auf den Differenzen $h_{ij} - \tilde{h}_{ij}$ basiert, ist der sog. **Chi-Quadrat-Koeffizient**

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^r \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}} \quad \text{mit} \quad 0 \leq \chi^2 < \infty. \quad (\text{I3: 31})$$

Er summiert im Zähler die quadrierten Abweichung der tatsächlich ermittelten Häufigkeiten h_{ij} von den erwarteten Häufigkeiten \tilde{h}_{ij} . Durch das Quadrieren wird verhindert, dass sich positive und negative Differenzen gegenseitig aufheben. Es gilt genau dann $\chi^2 = 0$, wenn alle Differenzen gleich Null sind. Im Falle statistischer

Unabhängigkeit nimmt (I3: 31) also den Wert 0 an. Je größer der Wert des Chi-Quadrat-Koeffizienten ist, umso stärker ist die statistische Abhängigkeit der Variablen X und Y. Ein kleiner Schönheitsfehler verhindert jedoch, dass die Größe selbst als Zusammenhangsmaß verwendet wird. Verdoppeln wir nämlich z.B. n und alle h_{ij} , so verdoppelt sich auch χ^2 . χ^2 kann also unbegrenzt große Werte annehmen.

Schritt 3:

Der **Kontingenzkoeffizient** K_{xy} , welcher aus χ^2 berechnet wird, beseitigt den Schönheitsfehler von χ^2 . K_{xy} weist im Falle statistischer Unabhängigkeit ebenfalls den Wert 0 auf und nähert sich bei zunehmender Stärke des Zusammenhangs dem Wert 1 an, erreicht diesen aber nicht. Konkret gilt

$$K_{xy} = \sqrt{\frac{\chi^2}{n + \chi^2}} \quad \text{mit} \quad 0 \leq K_{xy} < 1. \quad (\text{I3: 32})$$

Schritt 4:

Um eine bessere Vergleichbarkeit zu gewährleisten, empfiehlt es sich, den Kontingenzkoeffizienten zu normieren, sodass er höchstens den Wert 1 annehmen kann. Zu diesem Zweck dividieren wir K_{xy} durch seinen maximal möglichen Wert

$$K_{xy}^{\max} = \sqrt{\frac{M - 1}{M}} \quad \text{mit} \quad M = \min(k, r), \quad (\text{I3: 33})$$

wobei M das Minimum aus Zeilen- und Spaltenzahl der vorliegenden Häufigkeitstabelle darstellt. Für den **normierten Kontingenzkoeffizienten** folgt somit

$$K_{xy}^{\text{norm}} = \frac{K_{xy}}{K_{xy}^{\max}} \quad \text{mit} \quad 0 \leq K_{xy}^{\text{norm}} \leq 1. \quad (\text{I3: 34})$$

Wie der Wertebereich des normierten Kontingenzkoeffizienten zeigt, können wir lediglich Aussagen über die *Stärke* des Zusammenhangs der Variablen X und Y machen. Je näher der Koeffizient bei 1 liegt, desto stärker ist der Zusammenhang. Aussagen über die Richtung sind nicht möglich, da keine negativen Werte im Wertebereich vorkommen. Dies ist auch unmittelbar einleuchtend, da das Maß unter anderem für nominale Merkmale konzipiert ist, die keine Ordnungsstruktur besitzen, und daher nicht zwischen positivem und negativem Zusammenhang der Merkmale unterschieden werden kann.

Beispiel I 3-14

Kontingenzkoeffizient

Eine Untersuchung unter 500 Haushalten über den Zusammenhang von Wohnfläche und Haushaltsnettoeinkommen ergab die folgende Kontingenttafel (absolute Häufigkeiten):

		Einkommen Y		
		bis 1.500 €	über 1.500 €	
Wohnfläche X	bis 70 m ²	140	90	230
	über 70 m ²	100	170	270
		240	260	500

Schritt 1: Hypothetische Kontingenztabelle

		Einkommen Y		
		bis 1.500 €	über 1.500 €	
Wohnfläche X	bis 70 m ²	110,4	119,6	230
	über 70 m ²	129,6	140,4	270
		240	260	500

Schritt 2: Chi-Quadrat-Größe

$$\chi^2 = \sum_{l=1}^k \sum_{j=1}^r \frac{(h_{lj} - \tilde{h}_{lj})^2}{\tilde{h}_{lj}} = \frac{(140 - 110,4)^2}{110,4} + \frac{(100 - 129,6)^2}{129,6} + \frac{(90 - 119,6)^2}{119,6} + \frac{(170 - 140,4)^2}{140,4} \\ = 7,94 + 6,76 + 7,33 + 6,24 = 28,18$$

Schritt 3: Kontingenzkoeffizient

$$K_{XY} = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{28,18}{500 + 28,18}} = 0,23$$

Schritt 4: Normierter Kontingenzkoeffizient

$$M = \min(k, r) = \min(2, 2) = 2 \quad K_{XY}^{\max} = \sqrt{\frac{M-1}{M}} = \sqrt{\frac{2-1}{2}} = 0,71$$

$$K_{XY}^{\text{norm}} = \frac{K_{XY}}{K_{XY}^{\max}} = \frac{0,23}{0,71} = 0,33$$

Daraus lässt sich folgern, dass zwischen Einkommen und Wohnfläche ein schwach ausgeprägter Zusammenhang besteht.

Liegen zwei dichotome Merkmale X und Y vor, führt die tabellarische Darstellung ihrer gemeinsamen Häufigkeitsverteilung immer zu einer Vierfeldertafel. *Maximale Abhängigkeit* liegt in einer solchen Tafel immer dann vor, wenn *beide Diagonaleinträge* oder alternativ *beide Nicht-Diagonaleinträge Null* sind.

3.2.4 Lineartransformationen und Linearkombinationen

Betrachten wir zwei metrisch skalierte Merkmale X und Y deren Realisationen x_i und y_i in der Form

$$v_i = a + b \cdot x_i, \quad w_i = c + d \cdot y_i \quad \text{mit} \quad a, b, c, d \in \mathbb{R}, \quad b \neq 0, \quad d \neq 0 \quad (\text{I3: 35})$$

linear transformiert werden. Zwischen der Kovarianz Cov_{XY} der Ausgangsdaten und der Kovarianz Cov_{vw} der transformierten Daten besteht der Zusammenhang

$$\text{Cov}_{vw} = b \cdot d \cdot \text{Cov}_{XY}, \quad (\text{I3: 36})$$

den wir wie folgt beweisen können:

$$\text{Cov}_{vw} = \frac{1}{n} \sum_{i=1}^n (v_i - \bar{v})(w_i - \bar{w}) = \frac{1}{n} \sum_{i=1}^n (a + bx_i - a - b\bar{x})(c + dy_i - c - d\bar{y})$$

$$\begin{aligned}
 &= \frac{1}{n} \sum_{i=1}^n (b(x_i - \bar{x}))(d(y_i - \bar{y})) = \frac{b \cdot d}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\
 &= b \cdot d \cdot \text{Cov}_{XY}
 \end{aligned}$$

Für den Bravais-Pearson-Korrelationskoeffizienten gilt bei Lineartransformationen der Form (I3: 35)

$$r_{vw} = \frac{b \cdot d}{|b| \cdot |d|} \cdot r_{xy} \quad \text{bzw.} \quad |r_{vw}| = |r_{xy}|. \quad (\text{I3: 37})$$

Wir sagen, der Bravais-Pearson-Korrelationskoeffizient ist *invariant gegenüber Lineartransformationen*. Wir können dies wie folgt beweisen:

$$r_{vw} = \frac{\text{Cov}_{vw}}{s_v \cdot s_w} = \frac{b \cdot d \cdot \text{Cov}_{xy}}{|b| \cdot s_x \cdot |d| \cdot s_y} = \frac{b \cdot d}{|b| \cdot |d|} \cdot r_{xy}$$

(I3: 37) besagt, dass Lineartransformationen nicht die absolute *Stärke* eines korrelativen Zusammenhangs verändern können (Invarianzegenschaft). Allerdings kann sich ggf. die Richtung des Zusammenhangs umkehren. Es gilt

$$r_{vw} = \begin{cases} +r_{xy} & \text{falls } bd > 0 \\ -r_{xy} & \text{falls } bd < 0. \end{cases}$$

Die Aussage von (I3: 37) ist direkt auf den *Spearman-Rangkorrelationskoeffizienten* übertragbar. Dieser ist darüber hinaus *invariant gegenüber allen streng monotonen Transformationen*.

Werden die Beobachtungen x_i und y_i zweier metrisch skalierten Merkmale X und Y in der Form

$$w_i = a + b \cdot x_i + c \cdot y_i \quad \text{mit} \quad a, b, c \in \mathbb{R}, b \neq 0, c \neq 0 \quad (\text{I3: 38})$$

linear kombiniert, gilt für das arithmetische Mittel und die Varianz der Werte w_i

$$\bar{w} = a + b\bar{x} + c\bar{y}, \quad (\text{I3: 39})$$

$$s_w^2 = b^2 s_x^2 + c^2 s_y^2 + 2bc \text{Cov}_{xy}. \quad (\text{I3: 40})$$

Beweisen können wir den Zusammenhang (I3: 39) über

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i = \frac{1}{n} \sum_{i=1}^n (a + bx_i + cy_i) = \frac{1}{n} \left[n \cdot a + b \sum_{i=1}^n x_i + c \sum_{i=1}^n y_i \right] = a + b\bar{x} + c\bar{y}$$

und (I3: 40) über

$$\begin{aligned}
 s_w^2 &= \frac{1}{n} \sum_{i=1}^n (w_i - \bar{w})^2 = \frac{1}{n} \sum_{i=1}^n (a + bx_i + cy_i - [a + b\bar{x} + c\bar{y}])^2 \\
 &= \frac{1}{n} \sum_{i=1}^n (b(x_i - \bar{x}) + c(y_i - \bar{y}))^2 \\
 &= \frac{b^2}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{c^2}{n} \sum_{i=1}^n (y_i - \bar{y})^2 + \frac{2bc}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\
 &= b^2 s_x^2 + c^2 s_y^2 + 2bc \text{Cov}_{xy}.
 \end{aligned}$$

3.2.5 Kritische Anmerkungen zur Korrelationsanalyse

Oftmals werden durch die Korrelationsanalyse Zusammenhänge zwischen Merkmalen festgestellt, obwohl solche in der Wirklichkeit gar nicht existieren. Aus diesem Grund wollen wir uns im Folgenden kritisch mit der Korrelationsanalyse auseinandersetzen. Dazu müssen wir zunächst zwischen zwei Arten von Zusammenhängen unterscheiden. Von statistischen Zusammenhängen spricht man, wenn sich empirisch (mittels der bisher behandelten Verfahren) eine gemeinsame Variation verschiedener Variablen beobachten lässt. Von einem kausalen Zusammenhang spricht man hingegen, wenn außerdem eine Ursache-Wirkungs-Beziehung besteht, d.h. sich z.B. die Variable Y aufgrund einer Variation der Variable X verändert.

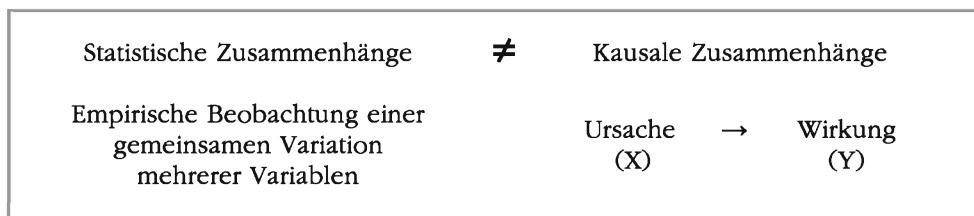


Abbildung I 25: Kausale versus statistische Zusammenhänge

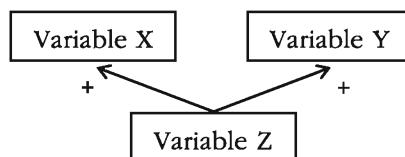
Ein festgestellter statistischer Zusammenhang darf nun nicht automatisch als kausaler Zusammenhang interpretiert werden. Dies beruht auf folgenden Tatsachen:

1. Richtung des Abhangigkeitsverhaltnisses

Bei kausalen Zusammenhängen kann die Abhängigkeit in die eine ($x_i \rightarrow y_j$) oder andere Richtung ($y_i \rightarrow x_j$) oder auch in beide Richtungen ($x_i \leftrightarrow y_j$) laufen. Die Korrelationsanalyse kann aber lediglich aussagen, ob ein positiver oder negativer Zusammenhang vorliegt und wie stark dieser ausgeprägt ist. Sie kann nicht feststellen, was Ursache und was Wirkung ist.

2. Dritte intervenierende Variable

Besteht kein kausaler Zusammenhang zwischen zwei Variablen (Merkmale) X und Y, so kann es vorkommen, dass trotzdem ein statistischer Zusammenhang festgestellt wird, da der Einfluss einer dritten Variablen bei der bivariaten Korrelationsanalyse nicht berücksichtigt wird. Nehmen wir beispielsweise an, dass die Variablen X und Y jeweils von einer Variable Z positiv beeinflusst werden, zwischen X und Y jedoch kein kausaler Zusammenhang besteht.



Im Rahmen einer statistischen Untersuchung werden nur die Variablen X und Y betrachtet. Steigt nun z.B. Z, steigen auch X und Y. Damit würde die Korrelationsanalyse einen Zusammenhang zwischen X und Y aufzeigen, obwohl keiner besteht.

Solche Fehler, die sog. **Scheinkorrelation**, können wir vermeiden, wenn wir wirklich alle möglichen Einflussvariablen berücksichtigen ($y = f(x, z, \dots)$). Dies ist typisch bei Verfahren wie der multiplen Regression (vgl. Kapitel IV), der Varianzanalyse und der Berechnung partieller Korrelationskoeffizienten.

3. Trendvariablen

Ein in der Praxis häufig gemachter Fehler tritt bei der Korrelationsuntersuchung von Trendvariablen (tendenziell steigender oder fallender Verlauf der Variablen) auf. Untersucht man nämlich z.B. die Variablen "nominales Bruttoinlandsprodukt" und "kumulierte Regenmenge" auf Korrelation, so wird mit Sicherheit ein statistischer Zusammenhang festgestellt. Der gesunde Menschenverstand besagt aber, dass es keinen Zusammenhang geben kann.

Zur Vermeidung von Scheinkorrelationen werden bei trendbehafteten Variablen die Korrelationsmesszahlen in der Praxis häufig nicht aus den originären Merkmalsausprägungen, sondern entweder aus den Veränderungen der Variablen von Periode zu Periode

$$\Delta x_t = x_t - x_{t-1} \quad \Delta y_t = y_t - y_{t-1} \quad (\text{I3: 41})$$

oder aus Wachstumsraten

$$w_t^x = \frac{x_t - x_{t-1}}{x_{t-1}} \quad w_t^y = \frac{y_t - y_{t-1}}{y_{t-1}} \quad (\text{I3: 42})$$

berechnet. Auf diese Weise kann der Trend in den Daten "eliminiert" werden.

4.

Messzahlen und Indizes

Mess- und Indexzahlen erlauben uns die Beschreibung der zeitlichen Entwicklung oder der räumlichen Unterschiede metrisch skalierter Merkmale. Während wir mit Messzahlen jeweils nur die Veränderung eines Merkmals beschreiben können, erfassen wir mit Indexzahlen die Entwicklung einer Gruppe von gleichartigen Merkmalen. Nach einem einführenden Abschnitt zu Messzahlen legen wir den Schwerpunkt dieses Kapitels auf Indexzahlen und dabei insbesondere auf Preisindizes. Als besondere Vertreter werden wir dabei den Verbraucherpreisindex (VPI) und den harmonisierten Verbraucherpreisindex (HVPI) sowie den Deutschen Aktienindex (DAX) kennenlernen.

4.1 Messzahlen

Messzahlen vergleichen zeitlich unterschiedliche Merkmalsausprägungen durch Bezugsbildung auf einen gemeinsamen Basiswert. Ausgangspunkt ist dabei die chronologisch geordnete Sequenz von im Zeitablauf gemessenen Beobachtungswerten x_t des Merkmals X, d.h. die sog. *Zeitreihe von Beobachtungswerten*

$$x_{-m}, \dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots, x_n, \quad (\text{I4: 1})$$

die die Entwicklung des Merkmals X vom Zeitpunkt $t = -m$ bis $t = n$ abbildet. Dividieren wir nun jedes Element dieser Zeireihe durch dasselbe vorher definierte Element, den sog. *Basiswert* x_0 , so erhalten wir eine *Zeitreihe von Messzahlen* zur *Basiszeit* $t = 0$, die sich wie folgt darstellt:

$$\frac{x_{-m}}{x_0}, \dots, \frac{x_{-2}}{x_0}, \frac{x_{-1}}{x_0}, 1, \frac{x_1}{x_0}, \frac{x_2}{x_0}, \dots, \frac{x_n}{x_0} \quad (\text{I4: 2})$$

Bezeichnen wir die **Messzahl** der *Berichtszeit* t zur Basiszeit $t = 0$ mit $r_{0,t}$, so können wir damit die allgemeine Formel

$$r_{0,t} = \frac{x_t}{x_0} \quad (\text{I4: 3})$$

zur Messzahloberechnung festhalten, in der x_t wie bisher für den Merkmalswert zur Berichtszeit t und x_0 für den Merkmalswert zur Basiszeit $t = 0$ stehen.

Multiplizieren wir die Messzahl (I4: 3) mit dem Faktor 100, so ermöglicht uns dies, prozentuale Abweichungen der Merkmalsausprägung der Berichtsperiode von der Merkmalsausprägung der Basisperiode unmittelbar abzulesen. Beispiel I 4-1 verdeutlicht dies.

Beispiel I 4-1

Messzahlen im Vergleich



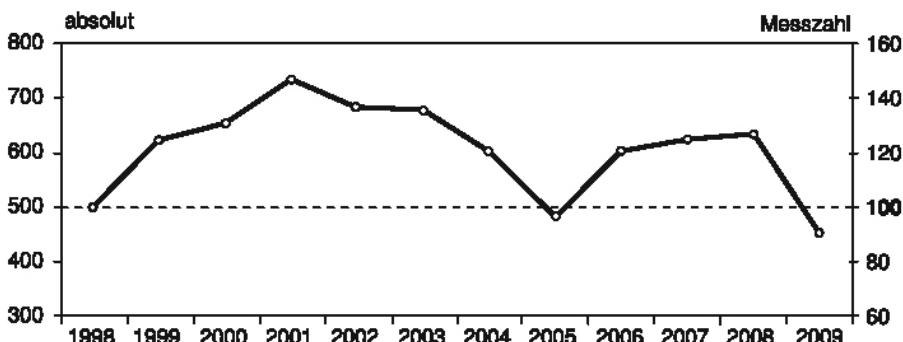
Über 12 Jahre wurden die Umsatzzahlen zweier Unternehmen A und B ermittelt. Um die Entwicklung beider Unternehmen besser vergleichen zu können, werden in beiden Fällen (bereits mit 100 multiplizierte) Messzahlen im Bezug auf den Umsatz in 1998 berechnet. Die Resultate sind in folgender Tabelle zu sehen:

Jahr	Periode	Umsatz in Tsd. €		Messzahlen (Basis 1995)	
		A	B	A	B
1998	0	500	1200	100	100
1999	1	620	1320	124	110
2000	2	650	1440	130	120
2001	3	730	1500	146	125
2002	4	680	1560	136	130
2003	5	670	1620	134	135
2004	6	600	1500	120	125
2005	7	480	1380	96	115
2006	8	600	1320	120	110
2007	9	620	1440	124	120
2008	10	630	1560	126	130
2009	11	450	1080	90	90

Betrachten wir nun die Messzahlen des Jahres 2003, so ist zu erkennen, dass der Umsatz des Unternehmens A im Vergleich zur Basisperiode um $(134 - 100) \% = 34 \% \text{ gestiegen ist.}$ Bei Unternehmen B liegt das entsprechende Umsatzwachstum bei 35 %. Wir können auch sagen, dass der Umsatz im Jahr 2003 genau 134 % bzw. 135 % des Umsatzes des Basisjahrs ausmacht. Vergleichen wir die Messzahlen des Unternehmens A für die Jahre 2007 und 2008, so ist zu erkennen, dass der Umsatz um 2 Prozentpunkte gestiegen ist. Dies entspricht einem Prozentsatz (Wachstumsrate) von

$$\frac{126 - 124}{124} \cdot 100 \% = 1,61 \%.$$

Stellen wir die Entwicklung der Messzahlen grafisch dar, so können wir daraus die Abweichungen von der Basis (100) sehr gut veranschaulichen. Für Unternehmen A ergibt sich so die folgende Grafik, in der neben den Messzahlen auch die absoluten Merkmalswerte an den Achsen abgetragen sind. Wie zu erkennen ist, wird der Basiswert 100 nur zweimal unterschritten. Dies bedeutet, dass es nur in den Jahren 2005 und 2009 zu einem Umsatzrückgang unter das Niveau von 1998 kam.



Sind wir gezwungen die Basis einer Messzahl bzw. einer Reihe von Messzahlen im Nachhinein zu ändern, bezeichnen wir diesen Vorgang auch als **Umbasierung**. Im Prinzip steckt dahinter nichts anderes als die Anwendung der Formel (I4: 3) zur Messzahlenberechnung auf die Zeitreihe der Beobachtungswerte (I4: 1) unter Nutzung einer neuen Basis x_s :

$$r_{s,t} = \frac{x_t}{x_s} \quad (\text{I4: 4})$$

Angenommen, die ursprünglichen Merkmalsausprägungen (I4: 1) liegen nicht mehr vor, sondern nur noch die Messzahlenreihe zur Basiszeit $t = 0$, können auch daraus die neuen Messzahlen zur Basiszeit $t = s$ berechnet werden:

$$r_{s,t} = \frac{\frac{x_t}{x_0}}{\frac{x_s}{x_0}} = \frac{r_{0,t}}{r_{0,s}} \quad (\text{I4: 5})$$

Beispiel I 4-2

Umbasierung von Messzahlen

Die folgende Tabelle enthält für den Zeitraum 1998 bis 2007 die Absatzzahlen eines Verlagshauses (in Mio. €) und die dazugehörigen (bereits mit 100 multiplizierten) Messzahlen zur Basis 1998. Durch Umbasierung auf die Basis 2003 ergeben sich folgende Werte:

Jahr	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
x_t	50	56	60	62	72	80	84	88	90	96
Messzahl (1998)	100	112	120	124	144	160	168	176	180	192
Messzahl (2003)	62,5	70	75	77,5	90	100	105	110	112,5	120

Wie die neuen Messzahlen zur Basis 2003 entstehen, wollen wir im Folgenden exemplarisch anhand der Messzahl für das Jahr 2004 darstellen. Die Berechnung anhand der Merkmalsausprägungen liefert

$$r_{2003,2004} = \frac{x_{2004}}{x_{2003}} \cdot 100 = \frac{84}{80} \cdot 100 = 105,$$

was natürlich mit der Berechnung anhand der Messzahlen zur Basis 1998

$$r_{2003,2004} = \frac{r_{1998,2004}}{r_{1998,2003}} \cdot 100 = \frac{168}{160} \cdot 100 = 105$$

übereinstimmt.

4.2 Indexzahlen

In diesem Abschnitt wollen wir nun die zeitliche Entwicklung mehrerer verwandter Größen beschreiben. Maßzahlen, die zeitlichen Vergleichen dienen und zudem dem Aspekt der Aggregation mehrerer Größen Rechnung tragen, nennen wir *Indexzahlen* oder kurz *Indizes*. Indizes fassen typischerweise Messzahlen der Einzelgrößen in geeigneter Weise zusammen. Damit wird klar, dass genau wie bei Kennzahlen zur Charakterisierung von Häufigkeitsverteilungen durch die Berechnung eines Index die zugrunde liegenden Einzelinformationen verloren gehen. Dies muss aber nicht unbedingt einen Nachteil darstellen. Sinn und Zweck eines Index ist es ja gerade die durchschnittliche Veränderung einer Vielzahl von gleichartigen Tatbeständen in einer einzigen Zahl auszudrücken. Durch diese Komprimierung der Ausgangsdaten sind Vergleiche und Analysen möglich, die unter Zugrundelegung der oft sehr umfangreichen anfänglichen Einzeldaten nur schwer umsetzbar wären.

Zur Erhöhung der Anschaulichkeit stehen im Folgenden spezielle Anwendungen von Indexzahlen im Vordergrund. Wir diskutieren die Messung der Preisentwicklung, der Mengenentwicklung und der Wertentwicklung von Güterbündeln über Preis-, Mengen- und Wertindizes. Unser Hauptaugenmerk liegt auf Preisindizes. Wir werden diese daher nutzen um wesentliche Indexoperationen zu veranschaulichen, die natürlich auf Mengen- und Wertindizes analog übertragbar sind. Zudem werden wir die praktische Bedeutung von Preisindizes herausarbeiten. Wir beschließen das Kapitel schließlich mit drei bedeutsamen Preisindizes aus der Wirtschaftspraxis.

4.2.1 Preisindizes

4.2.1.1 Grundlegendes

Eine Marktwirtschaft ist in der Regel gekennzeichnet durch sich im Zeitablauf ändernde Preise von Gütern und Dienstleistungen. Zur Quantifizierung dieser Preisänderungen nutzen wir in der Praxis sog. Preisindizes. Sie geben Auskunft über Kaufkraftänderungen des Einkommens und damit über die Stabilität oder Instabilität des Geldwertes. Rein definitorisch beschreibt ein Preisindex die *durchschnittliche Preisentwicklung einer vordefinierten Gruppe von Gütern*.

Ausgangsdatenmaterial für die rechnerische Bestimmung eines Preisindex sind die Preise $p_0^{(1)}, p_0^{(2)}, \dots, p_0^{(n)}$ von n Gütern in der Basisperiode 0 und $p_t^{(1)}, p_t^{(2)}, \dots, p_t^{(n)}$ in der Berichtsperiode t sowie die dazugehörigen Gütermengen $q_0^{(1)}, q_0^{(2)}, \dots, q_0^{(n)}$ und $q_t^{(1)}, q_t^{(2)}, \dots, q_t^{(n)}$. Bei den Preisen handelt es sich je nach Sachlage um Ein- oder Verkaufspreise, Import- oder Exportpreise, usw. Die Mengen sind je nach Sachlage konsumierte, produzierte, exportierte, importierte oder verkaufte Mengen der jeweiligen Güter. Wir sprechen dabei allgemein von *transferierten Mengen*.

Die n Güter und die dazugehörigen Mengen bilden die sog. **Warenkörbe**. Der Wert W dieser Warenkörbe bzw. die *tatsächlichen Ausgaben* eines Konsumenten, Produzenten, etc. (im Folgenden Individuen) für die Gütermengen kann wie folgt berechnet werden:

- a) Wert des Warenkorbs der Basisperiode in der Basisperiode

$$W_{0,0} = \sum_{i=1}^n p_0^{(i)} \cdot q_0^{(i)} \quad (\text{I4: 6})$$

- b) Wert des Warenkorbs der Berichtsperiode in der Berichtsperiode

$$W_{t,t} = \sum_{i=1}^n p_t^{(i)} \cdot q_t^{(i)} \quad (\text{I4: 7})$$

Neben diesen tatsächlichen Ausgaben lassen sich *hypothetische Ausgaben* bestimmen, die angeben, welche Zahlungen die Individuen für die Warenkörbe hätten leisten müssen, wenn die Preise der jeweils anderen Periode vorgeherrscht hätten:

- a) Was hätte das Individuum in der Berichtsperiode für den Warenkorb bzw. die Mengen der Basisperiode ausgeben müssen?

$$W_{t,0} = \sum_{i=1}^n p_t^{(i)} \cdot q_0^{(i)} \quad (\text{I4: 8})$$

- b) Was hätte das Individuum in der Basisperiode für den Warenkorb bzw. die Mengen der Berichtsperiode ausgeben müssen?

$$W_{0,t} = \sum_{i=1}^n p_0^{(i)} \cdot q_t^{(i)} \quad (\text{I4: 9})$$

Die Werte (I4: 6) bis (I4: 9) werden uns bei der Definition der Preis-, Mengen- und Wertindizes wertvolle Dienste leisten und uns helfen diese zu interpretieren.

Beispiel I 4-3

Warenkörbe

Ein Arbeitnehmerhaushalt kaufte im Basis- und Berichtsjahr folgende Güter 1-3:

Gut Nr. i	Basisjahr 2000		Berichtsjahr 2006	
	$p_{2000}^{(i)}$	$q_{2000}^{(i)}$	$p_{2006}^{(i)}$	$q_{2006}^{(i)}$
1	1,50 €/Stück	60 Stück	1,50 €/Stück	40 Stück
2	2,00 €/kg	20 kg	4,00 €/kg	10 kg
3	9,00 €/Liter	50 Liter	9,50 €/Liter	50 Liter

Betrachten wir dazu nun folgende Fragestellungen:

- a) Welchen Wert hat der Warenkorb der Basisperiode in der Basisperiode?

$$W_{2000,2000} = 1,50 \cdot 60 + 2,00 \cdot 20 + 9,00 \cdot 50 = 580 \text{ €}$$

- b) Welchen Wert hat der Warenkorb der Berichtsperiode in der Berichtsperiode?

$$W_{2006,2006} = 1,50 \cdot 40 + 4,00 \cdot 10 + 9,50 \cdot 50 = 575 \text{ €}$$

- c) Wieviel Euro hätte der Haushalt in der Berichtsperiode für den Warenkorb der Basisperiode ausgeben müssen?

$$W_{2006,2000} = 1,50 \cdot 60 + 4,00 \cdot 20 + 9,50 \cdot 50 = 645 \text{ €}$$

- d) Wieviel Euro hätte der Haushalt in der Basisperiode für den Warenkorb der Berichtsperiode ausgeben müssen?

$$W_{2000,2006} = 1,50 \cdot 40 + 2,00 \cdot 10 + 9,00 \cdot 50 = 530 \text{ €}$$

Die Entwicklung der Preise einzelner Güter aus dem Warenkorb können wir durch Preismesszahlen beschreiben. Für das i-te Gut erhalten wir so zur Berichtszeit t die Messzahl

$$\frac{p_t^{(i)}}{p_0^{(i)}}.$$

Eine denkbar einfache simultane Beschreibung der Preisentwicklung aller n Güter des Warenkorbs liefert das arithmetische Mittel der Preismesszahlen, also

$$P_{0,t} = \frac{1}{n} \sum_{i=1}^n \frac{p_t^{(i)}}{p_0^{(i)}}. \quad (\text{I4: 10})$$

Ein so definierter Preisindex ist allerdings nur dann ökonomisch sinnvoll interpretierbar, wenn allen n Gütern die *gleiche ökonomische Bedeutung* zukommt. In praktischen Anwendungen ist dies meist nicht der Fall. Es wird dann notwendig, die Güter bzw. die Preismesszahlen geeignet zu gewichten. Dafür gibt es eine Reihe verschiedener Lösungen, die im Folgenden vorgestellten Preisindizes liefern.

4.2.1.2 Preisindex nach Laspeyres

Der *Preisindex nach Laspeyres* $P_{0,t}^L$ ist ein gewogenes arithmetisches Mittel der n Preismesszahlen $p_t^{(i)} / p_0^{(i)}$ der im Warenkorb enthaltenen Güter. Er ist allgemein definiert als

$$P_{0,t}^L = \sum_{i=1}^n \frac{p_t^{(i)}}{p_0^{(i)}} \cdot g_0^{(i)} \quad \text{mit} \quad g_0^{(i)} = \frac{p_0^{(i)} \cdot q_0^{(i)}}{\sum_{i=1}^n p_0^{(i)} \cdot q_0^{(i)}} = \frac{p_0^{(i)} \cdot q_0^{(i)}}{W_{0,0}}. \quad (\text{I4: 11})$$

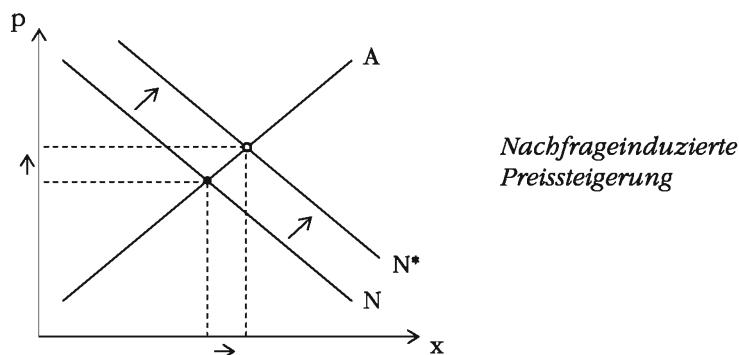
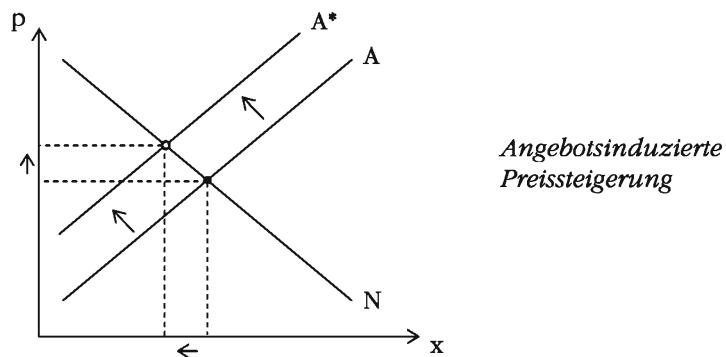
Wie wir sehen, entsprechen die Gewichte $g_0^{(i)}$ für jedes Gut i genau dem Anteil der Ausgaben für das Gut am Gesamtwert des Warenkorbs der Basisperiode. Die *Gewichte* (in ihrer Gesamtheit auch *Wägungsschema des Index* genannt) mit denen die Preismesszahlen in den Index eingehen, werden also ausschließlich mit Datenmaterial aus der *Basisperiode* bestimmt.

Zur genaueren Analyse und auch einfacheren Berechnung des Index erfolgt häufig eine Umformung in die sog. *Aggregatform*

$$P_{0,t}^L = \sum_{i=1}^n \frac{p_t^{(i)}}{p_0^{(i)}} \cdot \frac{p_0^{(i)} \cdot q_0^{(i)}}{W_{0,0}} = \frac{\sum_{i=1}^n p_t^{(i)} \cdot q_0^{(i)}}{W_{0,0}} = \frac{W_{t,0}}{W_{0,0}}. \quad (\text{I4: 12})$$

Diese zeigt, dass der Preisindex nach Laspeyres nichts anderes als das Verhältnis eines fiktiven Ausgabenwertes (wie viel der Warenkorb der Basisperiode zu Preisen der Berichtsperiode gekostet hätte) und den tatsächlichen Ausgaben für den Warenkorb der Basisperiode ist. (I4: 12) gibt damit an, wie sich das Preisniveau geändert hätte, wenn das in der Basisperiode geltende Mengenschema auch in der Berichtsperiode Gültigkeit hätte. Vorteil dieses Index ist es also, dass er von Mengenänderungen in der Berichtsperiode nicht beeinflusst wird und so die *reine Preisentwicklung* widerspiegelt. Während damit eine Vergleichbarkeit verschiedener Indexwerte gewährleistet wird, kann dies jedoch gerade bei langen Zeitreihen

einen erheblichen Nachteil darstellen. Güternachfragende Wirtschaftssubjekte reagieren nämlich im Allgemeinen auf Preisänderungen, und zwar durch den Kauf geringerer Mengen relativ teurer gewordener Güter und durch den Kauf größerer Mengen relativ billiger gewordener Güter. Bei Änderung ihrer Verbrauchsgewohnheiten werden sie außerdem "neue" Güter verstärkt und "alte" Güter weniger nachfragen. Aufgrund der unveränderten Gewichtung der Güter über- oder unterschätzt der Index bei solchen Gegebenheiten die Preissteigerung (sog. *Substitutions- oder Laspeyres-Effekt*). In der Literatur wird häufig ausschließlich von einer Überschätzung der Preissteigerung gesprochen. Es wird dabei vernachlässigt, dass zu unterscheiden ist, ob eine festgestellte Preissteigerung auf eine Angebots- (z.B. Lieferengpass einer Unternehmung und damit Angebotsverknappung) oder Nachfrageänderung (z.B. veränderte Verbrauchsgewohnheiten aufgrund neu auf den Markt gekommener Güter) zurückzuführen ist. Die Betrachtung der beiden folgenden Angebots-Nachfrage-Diagramme macht klar, dass eine angebotsinduzierte Preissteigerung mit einem Rückgang, eine nachfrageinduzierte Preissteigerung mit einer Zunahme der nachgefragten Menge verbunden ist. Bei unveränderter Gewichtung kommt es daher durch den Preisindex nach Laspeyres bei einem Angebotsschock zu einer Überschätzung der Preissteigerung und bei einem Nachfrageschock zu einer Unterschätzung der Preissteigerung.



Gerade bei der Problematik einer unveränderten Mengenstruktur setzt der Preisindex nach Paasche an, der die aktuelle Nachfragestruktur durch einen aktuellen Warenkorb berücksichtigt. Auch er ist jedoch nicht unproblematisch.

4.2.1.3 Preisindex nach Paasche

Auch der *Preisindex nach Paasche* $p_{0,t}^P$ ist als gewogenes arithmetisches Mittel von Preismesszahlen bzw. konkret als

$$p_{0,t}^P = \sum_{i=1}^n \frac{p_i^{(i)}}{p_0^{(i)}} \cdot g_t^{(i)} \quad \text{mit} \quad g_t^{(i)} = \frac{p_0^{(i)} \cdot q_t^{(i)}}{\sum_{i=1}^n p_0^{(i)} \cdot q_t^{(i)}} = \frac{p_0^{(i)} \cdot q_t^{(i)}}{W_{0,t}} \quad (\text{I4: 13})$$

definiert. Der Unterschied zum Preisindex nach Laspeyres besteht allein in der Gewichtung. Die Gewichte beim Preisindex nach Paasche verändern sich von Berichtsperiode zu Berichtsperiode, da sie die Anteile der Werte der in den Berichtsperioden transferierten Mengen zu Preisen der Basisperiode an der zugehörigen Gesamtwertesumme sind. Das *Gewichtungsschema* orientiert sich also an den Mengen der *Berichtsperiode* und ist damit *variabel*. Im Gegensatz dazu orientierte sich der *Preisindex nach Laspeyres* bei der Gewichtung an den Mengen der Basisperiode, sodass bei Berechnung des Preisindex für mehrere Berichtszeiten das *Gewichtungsschema* über die gesamte Indexreihe *konstant* beibehalten wird.

Auch $p_{0,t}^P$ lässt sich in *Aggregatform* darstellen und zwar als

$$p_{0,t}^P = \frac{\sum_{i=1}^n p_t^{(i)} \cdot q_t^{(i)}}{\sum_{i=1}^n p_0^{(i)} \cdot q_t^{(i)}} = \frac{W_{t,t}}{W_{0,t}}. \quad (\text{I4: 14})$$

Der Index ist also lediglich das Verhältnis der tatsächlichen Ausgaben für den Warenkorb der Berichtsperiode und einem fiktiven Ausgabenwert, der angibt, wie viel der Warenkorb der Berichtsperiode zu Preisen der Basisperiode gekostet hätte. Er gibt somit an, wie sich das Preisniveau geändert hätte, wenn das in der Berichtsperiode gültige Verbrauchsschema unverändert in der Basisperiode Gültigkeit hätte.

Zusammenfassend wird der Preisindex nach Paasche also anders als der Index nach Laspeyres nicht nur von Preis- sondern auch von Mengenänderungen beeinflusst. Trotz der größeren Aktualität der Gewichtungen neigt auch er bei Substitutionsprozessen oder veränderten Verbrauchsgewohnheiten zu verzerrten Messungen. Da ihm die Fiktion zugrunde liegt, dass die Wirtschaftssubjekte die nach dem Angebots- oder Nachfrageschock verbrauchten Güter auch in der Basiszeit (also vor dem Angebots- oder Nachfrageschock) verbraucht haben, kann der Index die Preissteigerung ebenso unter- oder überschätzen. Zudem ist nachteilig, dass die Verbrauchsgewohnheiten in jedem Berichtsjahr neu erhoben werden müssen, womit Kosten verbunden sind. Gerade daher sind die meisten Preisindizes in der Praxis vom Laspeyres-Typ.

4.2.1.4 Weitere Preisindizes

Wie wir gesehen haben, kann es bei Substitutionsprozessen oder veränderten Verbrauchsgewohnheiten bei den Preisindizes nach Laspeyres und Paasche zur Unter- oder Überschätzung der Preissteigerungen kommen. Dies hat zur Entwicklung weiterer Indizes geführt, von denen jedoch einzig der Preisindex nach Drobisch und der Fischersche Idealindex in der Praxis Beachtung gefunden haben.

Der **Preisindex nach Drobisch** ist lediglich das arithmetische Mittel aus den Preisindizes nach Laspeyres und Paasche.

$$p_{0,t}^{Dr} = \frac{p_{0,t}^L + p_{0,t}^P}{2} \quad (I4: 15)$$

Der **Fishersche Idealindex** ist das geometrische Mittel der Preisindizes nach Laspeyres und nach Paasche.

$$p_{0,t}^{Fl} = \sqrt{p_{0,t}^L \cdot p_{0,t}^P} \quad (I4: 16)$$

Darüber hinaus existiert eine Vielzahl weiterer Indexkonstrukte, die jedoch nahezu keine praktische Bedeutung besitzen. Zu diesen zählt z.B. der *Marshall-Edgeworth-Index*, der transferierte Mengen benutzt, die sich als Durchschnitt (arithmetisch oder geometrisch) der Verbrauchsmengen der Basis- und der Berichtsperiode ergeben. Der *Preisindex nach Lowe* verwendet transferierte Mengen, die sich als arithmetisches Mittel aller beobachteten Mengen zwischen Basis- und Berichtsperiode ergeben.²

In der Praxis multiplizieren wir Indexwerte häufig mit dem Faktor 100. Dies erlaubt uns eine ähnliche Interpretation, wie wir sie bei einer Preismesszahl vornehmen können. Der Wert eines Index in der Basisperiode ist nämlich wie der einer Messzahl in der Basisperiode immer 1 bzw. 100. Hier ist es jedoch die Preisentwicklung einer Gütergruppe, die wir untersuchen.

Beispiel I 4-4

Preisindizes im Überblick

Gegeben sei folgendes Datenmaterial:

Gut i	2003		2004		2005		2006	
	p	q	p	q	p	q	p	q
1	3	5	2	10	2	10	1	10
2	4	7	5	12	6	11	13	9
3	5	9	8	11	8	15	17	16

Berechnen wir nun daraus die bisher behandelten Preisindizes, wobei wir 2003 als Basis- und 2006 als Berichtsjahr verwenden:

$$p_{2003,2006}^L = \frac{1 \cdot 5 + 13 \cdot 7 + 17 \cdot 9}{3 \cdot 5 + 4 \cdot 7 + 5 \cdot 9} = \frac{249}{88} = 2,8295$$

$$p_{2003,2006}^P = \frac{1 \cdot 10 + 13 \cdot 9 + 17 \cdot 16}{3 \cdot 10 + 4 \cdot 9 + 5 \cdot 16} = \frac{399}{146} = 2,7329$$

$$p_{2003,2006}^{Dr} = 0,5 \cdot (2,8295 + 2,7329) = 2,7812$$

$$p_{2003,2006}^{Fl} = \sqrt{2,8295 \cdot 2,7329} = 2,7808$$

Betrachten wir den Preisindex nach Laspeyres, so erhalten wir aus diesen nach Multiplikation mit 100 den Wert 282,95. Wir können damit sagen, dass im gesamten Zeitraum 2003

² Schira, J. (2009) gibt einen detaillierten Überblick über diese und andere Preisindizes.

bis 2006 die Preise der drei Güter durchschnittlich um $(282,95 - 100) \% = 182,95 \%$ gestiegen sind. Interessieren wir uns für die durchschnittliche Preisänderung zwischen 2005 und 2006, so können wir diese nach Berechnung des Index $p_{2005,2006}^L = 1,4091$ bzw. des Wertes 140,91 wie folgt angeben:

$$\frac{282,95 - 140,91}{140,91} \cdot 100 \% = 100,80 \%$$

4.2.1.5 Preisindexreihen und Inflationsmessung

In der Praxis ermitteln wir Preisindizes in der Regel nicht nur für eine, sondern für mehrere aufeinanderfolgende Berichtsperioden $t = 1, 2, 3, \dots$. Die Folge von dabei entstehenden Indexwerten bezeichnen wir als *Indexreihe*. Sie hat die Gestalt

$$(p_{0,0} = 1), p_{0,1}, p_{0,2}, \dots, p_{0,t}, \dots, p_{0,T} . \quad (\text{I4: 17})$$

Aus derartigen Indexreihen lassen sich, wie wir bereits in Beispiel I 4-4 gesehen haben, durchschnittliche Preissteigerungsraten berechnen. Wir bezeichnen diese in der Praxis auch als *Inflationsraten*. Unter **Inflation** verstehen wir allgemein einen *kontinuierlichen Anstieg des Preisniveaus* bzw. einen anhaltenden Rückgang der Kaufkraft. Die Inflationsrate π_t bzw. die prozentuale Preisänderung für den gesamten Zeitraum *von der Basisperiode 0 bis zur Berichtsperiode t* kann allgemein als

$$\pi_t = (p_{0,t} - 1) \cdot 100 \% \quad (\text{I4: 18})$$

bestimmt werden. Wurden die Indexformeln um ein "mal 100" ergänzt, so ergibt sich die Inflationsrate als $\pi_t = (p_{0,t} - 100) \%$.

Die Inflationsrate zwischen zwei unterschiedlichen *Perioden t und t'* mit $t' < t$ bestimmt sich als

$$\pi_t = \frac{p_{0,t} - p_{0,t'}}{p_{0,t'}} \cdot 100\% = \left(\frac{p_{0,t}}{p_{0,t'}} - 1 \right) \cdot 100\% . \quad (\text{I4: 19})$$

Von **Deflation** sprechen wir bei einem *kontinuierlichen Rückgang des Preisniveaus* bzw. einer anhaltenden Steigerung der Kaufkraft. Die Inflationsrate wäre in einem solchen Fall also negativ. Des Weiteren wird häufig der Begriff der **Disinflation** gebraucht. Dabei handelt es sich um eine Phase noch *positiver*, jedoch *rückläufiger Inflationsraten*.

Gemessen wird die Inflation in *Deutschland* anhand des *Verbraucherpreisindex* (VPI, vgl. Abschnitt I 4.2.5.1), der vom Statistischen Bundesamt berechnet wird. Mit diesem Index lässt sich zwar nur ein Teil der gesamtwirtschaftlichen Preisentwicklung abbilden, doch da die private Bedürfnisbefriedigung als finaler Zweck des Wirtschaftens gilt, ist eine Verwendung dieses Index dennoch zweckmäßig. Im *Eurosystem* bezieht sich das Interesse vornehmlich auf den sog. *Harmonisierten Verbraucherpreisindex* (HVPI, vgl. Abschnitt I 4.2.5.2), welcher vom Statistischen Amt der EU berechnet wird. Beide Indizes sind vom *Laspeyres-Typ*. Wie wir bereits wissen, wird bei einer derartigen Indexberechnung unterstellt, dass sich die Verbrauchsgewohnheiten der Konsumenten über einen mehr oder weniger langen Zeitraum nicht verändern. Ein Warenkorb ist bei seiner Einführung normalerweise bereits vier Jahre, bei seiner Ablösung neun Jahre alt. Dieses lange Festhalten an

einem Warenkorb impliziert eine Reihe von **Fehlerquellen bei der Inflationsmessung**:³

- Product Substitution Bias

Die Tatsache, dass relativ teurer gewordene Güter von den Haushalten weniger nachgefragt werden, kann mit einem fixen Warenkorb nicht erfasst werden.

- Quality Change Bias

Qualitätsverbesserungen werden nur unzureichend aus den Preiserhöhungen herausgerechnet. Dies liegt vor allem daran, dass Qualitätsveränderungen bei gewissen Gütern und Dienstleistungen nur äußerst schwer auszumachen sind.

- New Product Bias

Neue Produkte, deren Preise in der Regel nach Markteinführung fallen, werden oft erst mit einer Verzögerung von mehreren Jahren in die Preisbeobachtung einbezogen. Ein Warenkorb, der die neuen Produkte nicht einbezieht, überzeichnet also die Teuerung.

- Outlet Substitution Bias

Basiert der Preisindex auf einer festen Auswahl von Berichtsstellen, die die Preise der im Warenkorb enthaltenen Güter an das statistische Bundesamt melden, können Strukturänderungen im Handel, die zu günstigeren Einkaufsmöglichkeiten für die Haushalte führen, nur unzureichend berücksichtigt werden.

Insgesamt kommt es durch diese Probleme i.d.R. zu einer systematischen Übertreibung der tatsächlichen Inflationsrate. Dies hat zur Konsequenz, dass das reale Wachstum der Volkswirtschaft insgesamt, aber auch etwa die reale Zunahme der verfügbaren Haushaltseinkommen unterschätzt wird (vgl. Abschnitt I 4.2.1.6). Auch Leistungen, die an den Verbraucherpreisindex gekoppelt sind, fallen zu hoch aus.

4.2.1.6 Preisbereinigung und reale Größen

Grundsätzlich wird in der Wirtschaftstheorie zwischen nominalen und realen Größen unterschieden. Die Ermittlung eines *nominalen Wertaggregates* zur Berichtszeit t stellt in der Regel kein Problem dar. Ein solches ist definiert als

$$\text{nominales Wertaggregat} = \sum_{i=1}^n p_t^{(i)} \cdot q_t^{(i)} = W_{t,t} . \quad (\text{I4: 20})$$

Da *reale Größen* streng genommen in physischen Mengeneinheiten gemessen werden sollten, dies aber in der Praxis oftmals nicht möglich oder sinnvoll ist, behelfen wir uns in der Statistik durch die Approximation

$$\text{reales Wertaggregat} \approx \sum_{i=1}^n p_0^{(i)} \cdot q_t^{(i)} = W_{0,t} . \quad (\text{I4: 21})$$

Der Quotient aus nominaler und approximativ ermittelte realer Größe ergibt, wie unschwer zu erkennen ist, einen Preisindex nach Paasche:

³ Vgl. Görgens, E., Ruckriegel, K., Seitz, F. (2009).

$$\frac{\text{nominales Wert aggregat}}{\text{reales Wert aggregat}} = \frac{W_{t,t}}{W_{0,t}} = p_{0,t}^p \quad (\text{I4: 22})$$

Durch Umformung und Verallgemeinerung dieses Zusammenhangs (I4: 22) lässt sich zeigen, dass eine preisbereinigte oder deflationierte Größe durch Division einer nominalen Größe durch einen entsprechenden Preisindex ermittelt werden kann. Diese bereinigte Größe ist dann eine gute *Approximation für die reale Größe*.

$$W_{0,t} = \frac{W_{t,t}}{p_{0,t}} \quad (\text{I4: 23})$$

Der dabei verwendete Preisindex sollte genau auf das Aggregat zugeschnitten sein. Finden wir keinen Index, der diese Bedingung erfüllt, so kann hilfsweise ein anderer Index verwendet werden, der diesem am nächsten kommt.

Beispiel I 4-5

BIP-Deflator

Angenommen eine Volkswirtschaft produziert nur die 2 Güter Computer und Autos. Unter Verwendung der Daten nachfolgender Tabelle sollen für das Jahr 2005 zunächst das nominale und das reale BIP (Bruttoinlandsprodukt) ermittelt werden.

Jahr	Computerpreis (in Euro)	Verkaufte Computer (in Stück)	Autopreis (in Euro)	Verkaufte Autos (in Stück)
2000	10.000	200.000	15.000	1.000.000
2005	2.000	1.500.000	25.000	1.500.000

Damit ergeben sich folgende nominale und reale Größen:

$$BIP_{2005}^{\text{nom}} = 2.000 \cdot 1.500.000 + 25.000 \cdot 1.500.000 = 40.500.000.000$$

$$BIP_{2005}^{\text{real}} = 10.000 \cdot 1.500.000 + 15.000 \cdot 1.500.000 = 37.500.000.000$$

Aus diesen kann über (I4: 22) ein Inflationsmaß abgeleitet werden, das in der Praxis auch als BIP-Deflator bezeichnet wird. Dieser ergibt sich als

$$\text{BIP-Deflator} = \frac{BIP_{2005}^{\text{nom}}}{BIP_{2005}^{\text{real}}} = \frac{40.500.000.000}{37.500.000.000} = 1,08$$

und gibt in diesem Beispiel Auskunft darüber, dass die Inflationsrate für den gesamten Be trachtungszeitraum bei 8 % lag.

Beispiel I 4-6

Preisbereinigung

Der Zusammenhang (I4: 23) wird in der Praxis häufig dazu genutzt aus einem nominalen Index durch Preisbereinigung einen realen Index zu gewinnen. Typisches Beispiel hierfür ist der reale Lohn- und Gehaltsindex, der sich als

$$\text{Realer Lohn- und Gehaltsindex} = \frac{\text{Nominaler Lohn- und Gehaltsindex}}{\text{Verbraucherpreisindex (VPI)}} \cdot 100$$

ergibt. Wichtig ist dabei, dass sich die eingesetzten Werte auf dieselbe Basis beziehen. Welchen Sinn nun eine solche Preisbereinigung macht, wollen wir durch ein einfaches Zahlenbeispiel verdeutlichen:

Jahr	Lohn- und Gehaltsindex nominal	VPI	Lohn- und Gehaltsindex real
2004 (Basis)	100,0	100,0	100,0
2005	101,5	102,0	99,5

Wird nur die nominale Lohn- und Gehaltsentwicklung betrachtet, könnte die Aussage gemacht werden, dass es von 2004 bis 2005 zu einer Lohn- und Gehaltserhöhung von 1,5 % gekommen ist. Wird aber die gleichzeitige Preiserhöhung von 2,0 % in die Betrachtung einbezogen, so stellen wir *real* eine Lohn- und Gehaltsabnahme von 0,5 % fest. Wir sagen auch, dass die Kaufkraft der Einkommen gesunken ist.

4.2.1.7 Interregionale Kaufkraftvergleiche

Die bisher behandelten Preisindizes dienten dem intertemporalen Kaufkraftvergleich. Ein Anstieg des Preisindex bedeutete hier bei gleichzeitiger Konstanz oder Abnahme des nominalen Einkommens einen Verlust an Kaufkraft. Neben diesem zeitlichen Vergleich kann nun aber auch ein Vergleich zwischen Ländern mit verschiedenen Währungen von Interesse sein. Im Folgenden widmen wir uns dazu einem stark vereinfachten Modell des bilateralen Kaufkraftvergleichs (d.h. zwischen nur zwei Ländern), um die Grundidee dieses Konzepts zu verdeutlichen. Wie bei allen Kaufkraftvergleichen sollen auch in diesem vereinfachten Modell nur die Preise der für die Lebenshaltung relevanten Güter betrachtet werden.

Bevor auf ein konkretes Beispiel eingegangen wird, müssen wir zunächst einige grundlegende Begriffe aus dem Bereich Außenhandel klären: Bei nominalen **Wechselkursen**, welche den relativen Preis zweier Währungen beschreiben, ist zwischen sog. Preis- und Mengennotierung zu unterscheiden. Von *Preisnotierung* sprechen wir, wenn wir zum Ausdruck bringen, wie viele Euro wir z.B. für einen US-Dollar aufwenden müssen ($1 \$ = e €$). Im Falle der *Mengennotierung* interessieren wir uns hingegen dafür, wie viele US-Dollar wir für einen Euro erhalten ($1 € = e' \$$). Je nachdem welche Notierung betrachtet wird, kommt es bei Änderungen von e bzw. e' zu Auf- oder Abwertung der betrachteten Währungen.

Notierung	Beispiel	Veränderung	Wert des €	Wert des \$
Preis	$1 \$ = e €$	$e \uparrow$	↓	↑
		$e \downarrow$	↑	↓
Menge	$1 € = e' \$$	$e' \uparrow$	↑	↓
		$e' \downarrow$	↓	↑

Beide Notierungen stehen in der Beziehung

$$e = \frac{1}{e'} \quad \text{bzw.} \quad e \cdot e' = 1.$$

Wichtig wird im folgenden Beispiel vor allem die Mengennotierung sein.

Beispiel I 4-7**Kaufkraftindizes**

Es soll ein Kaufkraftvergleich zwischen Deutschland (D) und den Vereinigten Staaten (USA) durchgeführt werden. Dies erfolgt in zwei Schritten. Zunächst wird die sog. **Kaufkraftparität** ("Purchasing Power Parity", PPP, vom Statistischen Bundesamt auch als Verbrauchergeldparität bezeichnet) ermittelt. Es handelt sich dabei u.a. um einen Korrekturfaktor, der volkswirtschaftliche Größen (z.B. das Bruttoinlandsprodukt) international vergleichbar machen soll. Eine bloße Umrechnung mit aktuellen Wechselkursen wäre nicht genau, da die Kaufkraft in unterschiedlichen Währungsräumen oft unterschiedlich stark ist und dieser Sachverhalt (über die PPP) bei der Umrechnung berücksichtigt werden muss. Ist die PPP ermittelt, wird sie zusammen mit dem Devisenkurs benutzt um einen **Index für die Kaufkraft** aufzustellen.

1. Schritt: Ermittlung der Kaufkraftparität

Im Folgenden wollen wir drei Möglichkeiten der Bestimmung der PPP betrachten.

a) Deutsche Sicht:

$$\text{PPP}_{D,USA} = \frac{\text{Wert des deutschen Standardwarenkorb in USA}}{\text{Wert des deutschen Standardwarenkorb in D}} = \frac{\sum_{i=1}^n p_{USA}^{(i)} \cdot q_D^{(i)}}{\sum_{i=1}^n p_D^{(i)} \cdot q_D^{(i)}}$$

Der deutsche Standardwarenkorb ist dabei eine repräsentative Güterauswahl und Verbrauchsstruktur für die privaten Haushalte in Deutschland. Angenommen der deutsche Standardwarenkorb koste in Deutschland 1.500 € und in den Vereinigten Staaten 1.700 \$, so ergibt sich

$$\text{PPP}_{D,USA} = \frac{1.700 \$}{1.500 €} = 1,1333 \$/€,$$

was eine Kaufkraftparität von $1 € = 1,1333 \$$ bzw. $1 \$ = 0,8824 €$ (Kehrwert) liefert.

b) Amerikanische Sicht:

$$\text{PPP}_{USA,D} = \frac{\text{Wert des amerik. Standardwarenkorb in D}}{\text{Wert des amerik. Standardwarenkorb in USA}} = \frac{\sum_{i=1}^n p_D^{(i)} \cdot q_{USA}^{(i)}}{\sum_{i=1}^n p_{USA}^{(i)} \cdot q_{USA}^{(i)}}$$

Der amerikanische Standardwarenkorb ist dabei eine repräsentative Güterauswahl und Verbrauchsstruktur für die privaten Haushalte in den Vereinigten Staaten. Angenommen der amerikanische Standardwarenkorb koste in den Vereinigten Staaten 1.300 \$ und in Deutschland 1.400 €, so ergibt sich

$$\text{PPP}_{USA,D} = \frac{1.400 €}{1.300 \$} = 1,0769 €/$,$$

was zu einer Kaufkraftparität von $1 \$ = 1,0769 €$ bzw. $1 € = 0,9286 \$$ (Kehrwert) führt.

c) Durchschnittliche PPP:

Zur Ermittlung einer durchschnittlichen PPP empfiehlt sich die Berechnung des geometrischen Mittels (Fisherscher Idealindex) der unter a) und b) ermittelten Werte, d.h.

$$1 \$ = \sqrt{0,8824 \cdot 1,0769} € = 0,9748 €$$

oder gleichbedeutend

$$1 € = \sqrt{1,1333 \cdot 0,9286} \$ = 1,0259 \$.$$

Es ist wichtig, dass diese ermittelten Kaufkraftparitäten in der Regel nicht mit dem Wechselkurs übereinstimmen, da für diesen nicht nur die Verbraucherpreise sondern auch andere Faktoren wie z.B. Import- und Exportgüterpreise oder Zinssätze maßgeblich sind.

2. Schritt: Ermittlung des Index für die Kaufkraft

Kaufkraftparitäten sind grundsätzlich nur dann aussagekräftig, wenn der (aktuelle) Wechselkurs bekannt ist. Im Folgenden soll die Kaufkraft des Euro in den USA anhand eines geeigneten Index untersucht werden. Wir wollen dafür den folgenden Index verwenden, der sich der *Mengennotierung* bedient:

$$\text{Index für die Kaufkraft des Euro in den USA} = \frac{\text{Devisenkurs}}{\text{Kaufkraftparität}}$$

Hier stellt sich nun die Frage, welche Kaufkraftparität (a oder b) in diese Formel eingesetzt werden soll. Betrachten wir als Zielgruppe Deutsche, die in die USA gezogen sind und ihre deutschen Verbrauchsgewohnheiten beibehalten, so gelten für diese die unter a) ermittelten Werte. Unter der Annahme, der € koste 1,1629 \$ (Devisenkurs in Mengennotierung), so ergibt sich der Indexwert

$$\text{Index für die Kaufkraft des Euro in den USA} = \frac{1,1629 \text{ \$/€}}{1,1333 \text{ \$/€}} = 1,0261.$$

Da die Kaufkraft des Euro in Deutschland gleich 1 ist (Basiswert des Index), bedeutet dieses Ergebnis, dass der "Kaufkraftvorteil" des Euro in den USA 2,61 % beträgt. Ein steigender Wechselkurs (Mengennotierung) bedeutet also in einer solchen Kaufkraftparitäts situation einen Kaufkraftgewinn (mehr \$ für einen €), ein sinkender Wechselkurs einen Kaufkraftverlust (weniger \$ für einen €) des Euro in den Vereinigten Staaten.

4.2.1.8 Umbasierung und Verknüpfung

Stehen wir vor der Aufgabe einen bestimmten Index von der alten Indexbasis auf eine neue Indexbasis umzustellen, sprechen wir von **Umbasierung**. Dieses Verfahren wollen wir direkt anhand eines einfachen Beispiels erläutern:

Beispiel I 4-8

Umbasierung von Preisindizes

Für die Jahre 2003 bis 2006 seien für einen Konsumentenpreisindex folgende (bereits mit 100 multiplizierte) Indexwerte zur Basis 2003 bekannt:

Jahr t	2003	2004	2005	2006
Index $p_{2003,t}$	100	103	102	105

Als neue Basis soll nun das Jahr 2005 dienen und eine entsprechende Umbasierung vorgenommen werden. Die Ermittlung der neuen Indexwerte $p_{2005,t}$ erfolgt dabei über

$$p_{2005,t} = \frac{p_{2003,t}}{p_{2003,2005}} \cdot 100,$$

d.h. durch Division aller Indexwerte zur alten Basis durch den Indexwert des neuen Basisjahres zur alten Basis. Wir können so für t = 2004 beispielweise

$$p_{2005,2004} = \frac{p_{2003,2004}}{p_{2003,2005}} \cdot 100 = \frac{103,00}{102,00} \cdot 100 = 100,98$$

bestimmen. Die gesamte neue Indexreihe hat damit folgendes Aussehen:

Jahr t	2003	2004	2005	2006
Index p _{2005,t}	98,04	100,98	100,00	102,94

Wir erkennen deutlich, dass die Vorgehensweise bei der Umbasierung dieser Indexwerte identisch ist mit der bei der Umbasierung von Messzahlen (vgl. Abschnitt I 4.1).

Bei Indexreihen, denen ein konstantes Gewichtungsschema zugrunde liegt, wird dieses im Allgemeinen im Laufe der Zeit an Aktualität einbüßen (Veränderung der Verbrauchsgewohnheiten, Qualitätsverbesserungen, neue Güter, usw.) und muss durch ein neues ersetzt werden. Beim Preisindex nach Laspeyres wird dazu einfach der Warenkorb der Basisperiode durch einen aktuelleren Warenkorb ersetzt. Wir sprechen dabei auch von einer *Neubasierung* des Index. Sind wir allerdings an der Preisentwicklung über einen längeren Zeitraum interessiert, müssen wir den neuen Index (aktuellerer Warenkorb) mit dem alten Index (veralterter Warenkorb) **verknüpfen**. Wir werden sehen, dass dazu lediglich Umbasierungen notwendig sind. Betrachten wir folgende Situation:

Es sind zwei Preisindexreihen

$$P_{0,0}, P_{0,1}, \dots, P_{0,s} \quad \text{mit} \quad P_{0,0} = 1$$

$$P_{s,s}, P_{s,s+1}, \dots, P_{s,T} \quad \text{mit} \quad P_{s,s} = 1$$

mit den Basen $t = 0$ und $t = s$ gegeben. Der Preisindex wurde also zunächst zur Basiszeit $t = 0$ berechnet, im Zeitpunkt $t = s$ erfolgte dann eine Neubasierung.

Wir erhalten eine verknüpfte Indexreihe zur Basiszeit $t = 0$ über

$$\dot{P}_{0,t} = \begin{cases} P_{0,t} & \text{für } t = 0, 1, \dots, s-1 \\ P_{0,s} \cdot P_{s,t} & \text{für } t = s, s+1, \dots, T \end{cases}, \quad (\text{I4: 24})$$

wobei die Werte für $t = s, s+1, \dots, T$ umbasierte Indexzahlen sind. Alternativ erhalten wir gemäß

$$\ddot{P}_{s,t} = \begin{cases} \frac{P_{0,t}}{P_{0,s}} & \text{für } t = 0, 1, \dots, s-1 \\ P_{s,t} & \text{für } t = s, s+1, \dots, T \end{cases} \quad (\text{I4: 25})$$

die verknüpfte Indexreihe zur Basis $t = s$, wobei nun die Werte für $t = 0, 1, \dots, s-1$ umbasierte Indexzahlen sind.

Wichtig ist in diesem Zusammenhang, dass umbasierte Indexzahlen nicht vom selben Indextyp wie der Ausgangsindex sein müssen. Dies gilt insbesondere für Indizes vom Laspeyres-, Paasche- und Fisher-Typ. Beim Preisindex nach Laspeyres erkennt man das leicht daran, dass der Warenkorb nicht über dem gesamten Betrachtungszeitraum konstant bleibt. Die Umbasierung liefert nur einen Näherungswert für den Ausgangsindex.

Beispiel I 4-9

Verknüpfung von Preisindizes

Ein Index (vgl. folgende Tabelle), der bisher zur Basis 1999 geführt wurde (Spalte 2), wird im Jahr 2003 mit der neuen Basis 2003 fortgeführt (Spalte 3). Wollen wir nun die Preisentwicklung von 1999 bis 2006 in einer einzigen Indexreihe zusammenfassen, so bieten sich zur Verknüpfung des alten mit dem neuen Index zwei Möglichkeiten: a) Fortführung des alten Index und b) Rückrechnung des neuen Index. Hierbei wird unterstellt, dass sich der neue Index in der Vergangenheit proportional zum alten Index entwickelt hat bzw. dass sich der alte Index in der Zukunft proportional zum neuen Index entwickelt.

Jahr t	originale Indexreihen	verknüpfte Indexreihen
	$P_{1999,t}$	$P_{2003,t}$
1999	100,00	-
2000	101,00	-
2001	102,00	-
2002	104,00	-
2003	106,00	100,00
2004	-	102,00
2005	-	103,00
2006	-	107,00

a) b)

Zur Fortführung und Rückrechnung können wir nun auf die Erkenntnisse aus (I4: 24) und (I4: 25) zurückgreifen. Wir können daraus nämlich zwei Faktoren ableiten, die sich aus den Indexwerten der überlappenden Periode (2003) ergeben. Diese berechnen wir, wie folgt:

$$\text{Fortführungsfaktor } F = \frac{P_{1999,2003}}{P_{2003,2003}} = \frac{106,00}{100,00} = 1,06$$

$$\text{Rückrechnungsfaktor } R = \frac{P_{2003,2003}}{P_{1999,2003}} = \frac{100,00}{106,00} = 0,9434$$

Die Ermittlung der fortgeführten Werte erfolgt nun durch Multiplikation der Werte des neuen Index mit dem Faktor F. Die zurückgerechneten Werte ergeben sich durch Multiplikation der Werte des alten Index mit dem Faktor R. Für das Jahr 2002 ergibt sich so z.B. der rückgerechnete Wert $104,00 \cdot 0,9434 = 98,11$ und für das Jahr 2006 der fortgeführte Wert $107,00 \cdot 1,06 = 113,42$. Alle anderen Werte lassen sich analog ermitteln und sind in obiger Tabelle kursiv dargestellt.

4.2.2 Mengenindizes

Die Bausteine von Mengenindizes (auch als Volumenindizes bezeichnet) sind die Mengenmesszahlen

$$\frac{q_l^{(t)}}{q_0^{(t)}}$$

der n im Index enthaltenen Güter. Gewichten wir diese mit den von Laspeyres und Paasche vorgeschlagenen Wägungsschemata, dann erhalten wir durch Aggregation der gewichteten Messzahlen die Mengenindizes nach Laspeyres und Paasche. Sie erlauben uns Aussagen über *durchschnittliche Mengenentwicklung* zwischen Basis- und Berichtsperiode.



Den **Mengenindex nach Laspeyres** erhalten wir in seiner Aggregatform zu

$$q_{0,t}^L = \frac{\sum_{i=1}^n p_0^{(i)} \cdot q_t^{(i)}}{\sum_{i=1}^n p_0^{(i)} \cdot q_0^{(i)}} = \frac{W_{0,t}}{W_{0,0}} . \quad (\text{I4: 26})$$

Im liegt die Fiktion zugrunde, dass die in der Basisperiode vorherrschenden Preise unverändert auch in der Berichtsperiode gelten. Der Index spiegelt somit eine reine Mengenentwicklung wider.

Die Aggregatform des **Mengenindex nach Paasche** ergibt sich zu

$$q_{0,t}^P = \frac{\sum_{i=1}^n p_t^{(i)} \cdot q_t^{(i)}}{\sum_{i=1}^n p_t^{(i)} \cdot q_0^{(i)}} = \frac{W_{t,t}}{W_{t,0}} . \quad (\text{I4: 27})$$

Er unterstellt, dass die in der Berichtsperiode vorherrschenden Preise auch bereits in der Basisperiode Gültigkeit hatten. Von Periode zu Periode der Indexberechnung werden neue Preise verwendet, was im Vergleich zum Index nach Laspeyres eine größere Aktualität gewährleistet.

Die im Abschnitt I 4.2.1.4 behandelten weiteren Formen von Preisindizes können wir analog auch auf Mengenindizes übertragen. So ergibt sich etwa der Mengenindex nach Fisher als geometrisches Mittel der Mengenindizes nach Laspeyres und Paasche. Unsere Ausführungen zur Umbasierung und Verknüpfung unter I 4.2.1.8 können wir außerdem übertragen.

Wichtige von der amtlichen Statistik veröffentlichte Mengenindizes sind etwa der *Index der industriellen Nettoproduktion* und der *Index des Auftragseingangs im Bauhauptgewerbe*. Ersterer liefert Informationen über die Entwicklung und die Situation im produzierenden Gewerbe. Zweiterer gibt Auskunft über das Volumen der zu errichteten Bauwerke bzw. die Veränderung dieses Volumens im Zeitablauf.

Beispiel I 4-10

Mengenindizes

Gegeben sei folgende Arbeitstabelle, die sich auch bei der Berechnung von Preis- und Wertindizes als äußerst hilfreich erweisen kann. Aus dieser sollen die Mengenindizes nach Laspeyres und Paasche berechnet werden.

Gut i	Basiszeit		Berichtszeit		$W_{0,0}$ $p_0 \cdot q_0$	$W_{1,1}$ $p_1 \cdot q_1$	$W_{1,0}$ $p_1 \cdot q_0$	$W_{0,1}$ $p_0 \cdot q_1$
	p_0	q_0	p_1	q_1				
1	4	10	7	4	40	28	70	16
2	5	12	9	7	60	63	108	35
3	6	14	9	3	84	27	126	18
Σ	-	-	-	-	184	118	304	69

Die Mengenindizes belaufen sich damit auf

$$q_{0,1}^L = \frac{69}{184} \cdot 100 = 37,50$$

und

$$q_{0,1}^P = \frac{118}{304} \cdot 100 = 38,82 \text{ .}$$

Da die Basis beider Mengenindizes bei 100 liegt, kam es zwischen Basis- und Berichtsperiode zu durchschnittlichen Mengenänderungen von $(37,50 - 100) \% = -62,50 \% \text{ bzw. } (38,82 - 100) \% = -61,18 \% \text{, d.h. zu einem starken Mengenrückgang.}$



4.2.3 Wertindex

Soll die Veränderung des Warenkorbwerts zwischen Basis- und Berichtszeit erfasst werden, wird der sog. **Wertindex** (auch Umsatzindex genannt) herangezogen. Dieser ist in (I4: 28) als Verhältnis aus Warenkorbwert der Berichtszeit und Warenkorbwert der Basiszeit definiert. Er stellt jedoch genau genommen keinen echten Index dar, da er nur die Entwicklung einer Variablen widerspiegelt. Er ist daher eine *reine Messzahl*.

$$u_{0,t} = \frac{\sum_{i=1}^n p_t^{(i)} \cdot q_t^{(i)}}{\sum_{i=1}^n p_0^{(i)} \cdot q_0^{(i)}} = \frac{W_{t,t}}{W_{0,0}} \quad (\text{I4: 28})$$

Wir können leicht zeigen, dass sich der Umsatzindex auch wie folgt aus Indizes vom Laspeyres-, Paasche- und Fisher-Typ bestimmen lässt:

$$u_{0,t} = p_{0,t}^L \cdot q_{0,t}^L = p_{0,t}^P \cdot q_{0,t}^L \quad (\text{I4: 29})$$

$$u_{0,t} = p_{0,t}^R \cdot q_{0,t}^R \quad (\text{I4: 30})$$

Es gilt aber

$$u_{0,t} \neq p_{0,t}^L \cdot q_{0,t}^L \quad \text{und} \quad u_{0,t} \neq p_{0,t}^P \cdot q_{0,t}^P.$$

Während also die Fisher-Indizes die sog. Faktorumkehrprobe (Wert = Menge · Preis) bestehen, scheitern die Indizes vom Laspeyres- und Paasche-Typ.

Beispiel I 4-11

Umsatzindex

Folgende Tabelle spiegelt die Umsatzentwicklung eines Unternehmens von 2002 bis 2006 wider und zeigt den dazugehörigen Umsatzindex zur Basis 2004.

Jahr	2002	2003	2004	2005	2006
Umsatz (p · q) in Mio. Euro	55.000	40.000	50.000	60.000	65.000
Umsatzindex	110,00	80,00	100,00	120,00	130,00

Der Wertindex des Jahres 2006 zur Basis 2004 ergab sich darin z.B. als

$$u_{2004,2006} = \frac{65.000}{50.000} \cdot 100 = 130,00,$$

was eine Umsatzsteigerung von $(130 - 100) \% = 30 \% \text{ im Zeitraum 2004 bis 2006 aufzeigt.}$

4.2.4 Wichtige Indizes aus der Wirtschaftspraxis

4.2.4.1 Verbraucherpreisindex (VPI)



Vom Statistischen Bundesamt wird regelmäßig und aktuell eine Vielzahl verschiedener Preisindizes berechnet, die detailliert Preisentwicklungen auf branchen- oder gesamtwirtschaftlicher Ebene messen.⁴ Der vor dem Jahr 2000 als "Preisindex für die Lebenshaltung aller privaten Haushalte" bezeichnete *Verbraucherpreisindex* (VPI) ist ein wesentlicher Bestandteil des preisstatistischen Berichtssystems in Deutschland, zu dem u.a. auch der Index der Einzelhandelspreise, Indizes der Ein- und Ausfuhrpreise, der Index für Erzeugerpreise gewerblicher Produkte, der Preisindex für Bauwerke, der Erzeugerpreisindex für land- und forstwirtschaftliche Produkte und der Index für Großhandelsverkaufspreise gehören. Da der private Konsum der dominante Zweck des Wirtschaftens ist, ist der VPI in Deutschland

- ein *Indikator* für die Beurteilung der *Geldwertstabilität* und wird als *Inflationsmaßstab* verwendet. Aus diesem Grund wird seine Veränderungsrate oft als Inflationsrate bezeichnet (vgl. Abschnitt I 4.2.1.5).
- Grundlage für die *Deflationierung* von *Wertgrößen der volkswirtschaftlichen Gesamtrechnung*, z.B. zur Berechnung der realen Lohn- und Gehaltssummen (vgl. Abschnitt I 4.2.1.6).
- Anker zur Wertsicherung wiederkehrender Zahlungen in *Wertsicherungsklauseln* (z.B. Realzinsgarantie bei gewissen Zertifikaten) in längerfristigen Vertragsbeziehungen.

Der VPI misst die *durchschnittliche Preisveränderung aller Waren und Dienstleistungen, die von privaten Haushalten zu Konsumzwecken gekauft werden*. Berücksichtigt werden dabei Güter des täglichen Bedarfs (z.B. Lebensmittel, Bekleidung), Mieten und langlebige Gebrauchsgüter (z.B. Kraftfahrzeuge, Kühlschränke) und Dienstleistungen (z.B. Friseur, Reinigung, Versicherungen). Damit ist sichergestellt, dass der Index Preisveränderungen umfassend abbildet. Die Indexwerte werden monatlich ermittelt und veröffentlicht. Das aktuelle Basisjahr ist 2005.

Wie fast alle bedeutsamen Indizes des preisstatistischen Berichtssystems ist der VPI ein *Preisindex nach Laspeyres*. Er beschreibt die reinen Preisänderungen eines hinsichtlich Güterauswahl und zugehöriger Verbrauchsmengen konstanten Warenkorbs im Zeitverlauf, der rund 700 Waren und Dienstleistungen als sog. *Preisrepräsentanten* enthält. Diese werden funktional in 12 *Abteilungen* gegliedert, für die jeweils Teilindizes ermittelt werden. Diese Abteilungsindizes gehen aktuell, d.h. mit Basisjahr 2005, mit den folgenden Gewichten (in Promille) in den VPI ein:

⁴ Vgl. dazu Homepage des Statistischen Bundesamtes www.destatis.de.

1	Nahrungsmittel und alkoholfreie Getränke	103,55
2	Alkoholische Getränke und Tabakwaren	38,99
3	Bekleidung und Schuhe	48,88
4	Wohnung, Wasser, Strom, Gas und andere Brennstoffe	308,00
5	Einrichtungsgegenstände u.ä. für den Haushalt und deren Instandhaltung	55,87
6	Gesundheitspflege	40,27
7	Verkehr	131,90
8	Nachrichtenübermittlung	31,00
9	Freizeit, Unterhaltung und Kultur	115,68
10	Bildungswesen	7,40
11	Beherbergungs und Gaststättendienstleistungen	43,99
12	Andere Waren und Dienstleistungen	74,47

Bezeichnen wir die einzelnen Abteilungsindizes mit $p_{0,t}^{L(j)}$ und ihre Gewichte mit $g^{(j)}$ mit $j = 1, 2, \dots, 12$, ermittelt sich also der Gesamtindex formal nach

$$VPI_{0,t}^L = \sum_{j=1}^{12} p_{0,t}^{L(j)} \cdot g^{(j)} \quad \text{mit} \quad \sum_{j=1}^{12} g^{(j)} = 1.$$

Die Gewichte $g^{(j)}$ geben für die einzelnen Abteilungen die Ausgabenanteile an den Gesamtausgaben der Basisperiode an. Sind im Gesamtindex also insgesamt n Güter und in den Abteilungen n_j Güter berücksichtigt, so ergeben sich die Gewichte zu

$$g^{(j)} = \frac{\sum_{k=1}^{n_j} p_0^{(k)} \cdot q_0^{(k)}}{\sum_{i=1}^n p_0^{(i)} \cdot q_0^{(i)}} \quad \text{mit} \quad \sum_{j=1}^{12} n_j = n.$$

Beispiel I 4-12

Unbekannte Abteilungsgewichtung

Sind in einer Veröffentlichung Indexwerte für das Gesamttaggregat sowie für Abteilungen gegeben, nicht aber die Gewichte, so lassen sich die Gewichte dennoch eindeutig bestimmen. Zur Veranschaulichung der Vorgehensweise betrachten wir einen Laspeyres-Preisindex, der aus nur drei Abteilungen bestehe und für den zu den Berichtszeiten $t = 1, 2$ bei einer Basiszeit $t = 0$ folgende Indexwerte vorliegen:

t	$p_{0,t}^{L(1)}$	$p_{0,t}^{L(2)}$	$p_{0,t}^{L(3)}$	$p_{0,t}^L$
1	110	120	140	126
2	120	130	150	136

Die (zunächst unbekannten) Gewichte $g^{(1)} = 0,2$, $g^{(2)} = 0,4$ und $g^{(3)} = 0,4$ erhalten wir aus dem gegebenen Datenmaterial durch Lösung des Gleichungssystems

$$110 \cdot g^{(1)} + 120 \cdot g^{(2)} + 140 \cdot g^{(3)} = 126$$

$$120 \cdot g^{(1)} + 130 \cdot g^{(2)} + 150 \cdot g^{(3)} = 136$$

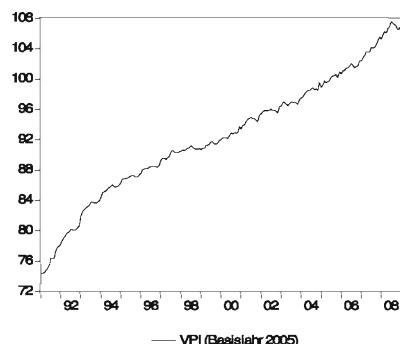
$$g^{(1)} + g^{(2)} + g^{(3)} = 1.$$

Zur Berechnung des VPI werden für die repräsentativen Güter repräsentative Preise benötigt. Preise variieren nun jedoch von Ort zu Ort und von Anbieter zu Anbieter, nicht zuletzt aufgrund unterschiedlicher Qualität. Um einen repräsentativen Preis zu erhalten müssen daher viele Preise je Preisrepräsentant erhoben werden, um daraus einen "typischen" Preis für ein Gut "mittlerer" Qualität bei einem "typischen" Anbieter bestimmen zu können. Rund 600 Preisermittler erfassen dazu in 188 Städten und Gemeinden Monat für Monat Preise der gleichen Produkte in den gleichen Geschäften. Zusätzlich werden für viele Güterarten Preise zentral erhoben (z.B. über das Internet). Insgesamt werden so über 300.000 Einzelpreise erfasst. Die Einzeldaten werden dann zu gewogenen Durchschnittspreisen für die Preisrepräsentanten verdichtet, die schließlich in die Indexberechnung eingehen.

Die Bestimmung des repräsentativen Warenkorbs (Güter und deren Gewicht) erfolgt auf Basis der Ausgaben der privaten Haushalte für Waren und Dienstleistungen. Diese Ausgaben werden in Haushaltsbefragungen auf Stichprobenbasis erhoben. In Deutschland zeichnen rund 60.000 Teilnehmer freiwillig einige Monate lang ihre Einnahmen und Ausgaben im Rahmen der sog. Einkommens- und Verbrauchsstichprobe (EVS) auf. Zusammen mit Ergebnissen der Volkswirtschaftlichen Gesamtrechnung (VGR), der Steuerstatistik, anderer amtlicher Statistiken, der Zahlungsbilanzstatistik der Deutschen Bundesbank und anderer Quellen, können damit wesentliche Güterarten identifiziert und anhand ihrer Ausgabenanteile auch deren Gewichte (Promilleangaben in obiger Auflistung) bestimmt werden.

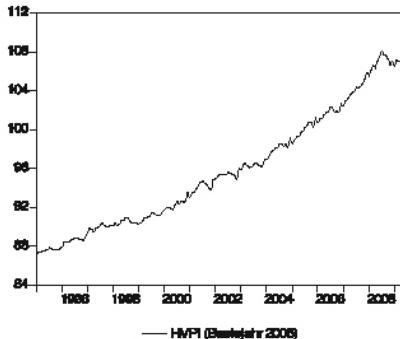
Die Verbrauchsgewohnheiten der privaten Haushalte ändern sich aber im Zeitablauf. Neue Güter erweitern die Konsummöglichkeiten, alte Güter scheiden aus dem Marktgeschehen aus. Folge hiervon ist, dass ein bestimmter Warenkorb nur für begrenzte Zeit den Verbrauch der Haushalte repräsentieren kann. Diesem Wandel versucht das Statistische Bundesamt durch *Neubasierung* des VPI Rechnung zu tragen, welche neben der Festlegung eines neuen Basisjahres, der Erhebung neuer Preisrepräsentanten und Bestimmung eines neuen Wägungsschemas auch die Umbasierung alter Indexwerte einschließt. In der Vergangenheit wurde der VPI ca. alle 5 Jahre neubasiert. Die Basisjahre waren 1950, 1958, 1962, 1970, 1976, 1980, 1985, 1991, 1995 und 2000.

Für die Jahre 1991 bis 2009 zeigte der VPI (Basisjahr 2005) den im Folgenden abgebildeten Verlauf. Wir erkennen daran für Deutschland einen deutlichen Preissteigerungstrend.



4.2.4.2 Harmonisierter Verbraucherpreisindex (HVPI)

Primär für europäische Zwecke berechnet das Statistische Bundesamt den Harmonisierten Verbraucherpreisindex (HVPI). Für diesen Index steht für Deutschland bisher nur eine Zeitreihe ab Januar 1995 (zur Basis 2005) zur Verfügung. Sie zeigt folgende Gestalt:



Wichtigste Zielsetzung des HVPI ist der *Vergleich der Preisveränderungen zwischen den Mitgliedstaaten der Europäischen Union* (EU). Da sich die nationalen Verbraucherpreisindizes der einzelnen EU-Länder in vielfältiger Weise unterscheiden, ist die Berechnung des HVPI unerlässlich. Diese Unterschiede sind zum Teil historisch bedingt, resultieren aber auch aus unterschiedlichen gesellschaftlichen Rahmenbedingungen oder abweichenden Strukturen des statistischen Systems. Auch sind die Ziele der VPI-Berechnung oft weiter gefasst als die des HVPI. Während der HVPI primär das Ziel der Inflationsmessung verfolgt, dienen die nationalen Verbraucherpreisindizes zusätzlich als Deflator (z.B. zur Berechnung des realen Wachstums) und als Kompensationsmaßstab (Wertsicherungsklauseln). Da die nationalen Indizes darauf ausgelegt sind, ihr Umfeld möglichst gut zu beschreiben, führt die Verwendung unterschiedlicher Berechnungsformeln und Konzepte zu nicht vergleichbaren Ergebnissen. Daher wurden gemeinsame Methoden, Konzepte und Verfahren für den HVPI verbindlich festgelegt. Die Erfassungsbereiche (Auswahl der Waren und Dienstleistungen für die Berechnung des HVPI) des deutschen VPI und des deutschen HVPI unterscheiden sich im Wesentlichen durch die zusätzliche Einbeziehung des vom Eigentümer selbst genutzten Wohneigentums (durch Mietäquivalente) im deutschen VPI.

Das Statistische Amt der Europäischen Union (Eurostat) überwacht die Einhaltung dieser Regeln und berechnet anhand der nationalen HVPIs Verbraucherpreisindizes für die Europäische Union, für die Europäische Währungsunion und für den Europäischen Wirtschaftsraum. Diese Verbraucherpreisindizes werden als *gewogene Durchschnitte der nationalen HVPI* berechnet. Als Gewicht dient der private Verbrauch aus den Volkswirtschaftlichen Gesamtrechnungen. Zur *Umrechnung der unterschiedlichen Währungseinheiten innerhalb der Europäischen Union und dem Europäischen Wirtschaftsraum* werden *Kaufkraftparitäten* herangezogen.

Der HVPI dient unter anderem zur **Messung des Konvergenzkriteriums Preisstabilität** (Referenzwert: ungewogenes arithmetisches Mittel der jährlichen Inflationsraten der drei preisstabilsten EU-Länder plus 1,5 Prozentpunkte). Somit war er

gemeinsam mit den anderen Konvergenzkriterien eine wichtige Entscheidungshilfe für die Zulassung der einzelnen Mitgliedstaaten der EU zur Währungsunion und wird dies bei der Aufnahme weiterer EU-Staaten in die Europäische Währungsunion wieder sein. Für die Europäische Zentralbank ist die Verwendung des Verbraucherpreisindex für die Europäische Währungsunion (VPI-EWU), der aus den nationalen HVPI der Mitglieder der Euro-Zone aggregiert wird, der Maßstab für die Geldwertstabilität des Euro.

4.2.4.3 Deutscher Aktienindex (DAX)

Als wichtige Informationsinstrumente auf dem Kapitalmarkt werden von Land zu Land sog. Aktienindizes berechnet und der Öffentlichkeit zur Verfügung gestellt. Sie sollen kurzfristig, repräsentativ und im Zeitablauf vergleichbar über Marktentwicklungen bei Aktienwerten informieren. Allgemein zu unterscheiden sind dabei Kurs- und Performance-Indizes. *Kursindizes* messen die eigentliche Kursentwicklung und werden lediglich um Erträge aus Bezugsrechten und Sonderzahlungen bereinigt. Bei den *Performance-Indizes* werden darüber hinaus sämtliche Erträge aus Dividenden- und Bonuszahlungen in das Indexportfolio reinvestiert.

Zu den weltweit wichtigsten Aktienindizes zählen folgende:

- *Deutschland*: CDAX, DAX, MDAX, SDAX, TecDAX, VDAX, HDAX
- *Europa*: FT-SE 100 (London Stock Exchange), CAC 40 (Bourse de Paris), AEX (Amsterdam), ATX (Wien), Dow Jones Euro Stoxx/Stoxx 50, Dow Jones Stoxx/Stoxx 50, Mib 30 (Mailand), Ibex 35 (Spanien)
- *USA*: Dow Jones 30 Industrial, Nasdaq, Nasdaq 100, S&P 500
- *Asien*: Nikkei 225 (Tokio), Hang Seng (Hongkong), Singapore Times, Taiwan Weighted, Thai Stock Exchange
- *Lateinamerika*: MCSI Mexico, MCSI Brazil, Chilean Stock Market Select

Der meist beachtete Aktienindex für Deutschland ist der **Deutsche Aktienindex** (DAX). Es handelt sich dabei um den Bluechip-Index, der die 30 größten und umsatzstärksten deutschen Aktienwerte enthält, die im Prime Standard (spezielles Marktsegment) der Frankfurter Wertpapierbörsen (FWB) gelistet sind und fortlaufend im Xetra (Computersystem zum elektronischen Handel aller an der FWB notierten Wertpapiere) gehandelt werden sowie einen Mindest-Freefloat (Aktien eines Unternehmens, die am Markt frei gehandelt werden und nicht im Besitz von Anlegern sind, welche sich dauerhaft an der Unternehmung beteiligen wollen) von 5 % aufweisen. Er ist ein *Performanceindex* und wird neben einer Reihe anderer Indizes von der Trägergesellschaft der FWB, der Deutschen Börse AG, ermittelt. Der DAX beruht auf dem Konzept der Indexberechnung nach Laspeyres.⁵ Die Indexwerte werden mit 1.000 multipliziert und als Promillezahlen veröffentlicht. Die Basiszeit mit dem Basiswert 1.000 ist der 30.12.1987. Der DAX wird während der Handelszeit der FWB sekündlich berechnet und misst die Kursentwicklung eines kapitalgewichteten Portfolios aus den 30 enthaltenen Aktien.

⁵ Genau genommen ist er ein sog. modifizierter Laspeyres-Kettenindex. Für Details zur Indexberechnung vgl. man den "Leitfaden zu den Aktienindizes der Deutschen Börse", der unter www.deutsche-boerse.com zu finden ist.

Die Entwicklung des DAX seit seiner Einführung im Jahr 1988 (bzw. im Dezember 1987) ist in der folgenden Grafik, die aktuelle Zusammensetzung (Stand: 31. Mai 2009) in der nachfolgenden Tabelle zu sehen:



Name	Anteil	Anzahl	Börsenwert
Adidas AG	01,25 %	193.515.512	4.648
Allianz SE	07,81 %	453.050.000	28.991
BASF SE	05,93 %	918.478.694	21.998
Bayer AG	07,77 %	764.344.530	28.846
Beiersdorf AG	00,68 %	252.000.000	2.539
BMW AG	01,90 %	601.995.196	7.066
Commerzbank AG	00,49 %	886.014.693	1.825
Daimler AG	04,91 %	964.557.432	18.229
Deutsche Bank AG	04,29 %	620.859.015	15.905
Deutsche Börse AG	02,10 %	195.000.000	7.798
Deutsche Lufthansa AG	00,99 %	457.937.572	3.691
Deutsche Post AG	01,71 %	1.209.015.874	6.331
Deutsche Telekom AG	07,34 %	4.361.319.993	27.226
E.ON AG	10,00 %	1.916.028.663	37.112
Fresenius SE (Vz)	00,68 %	80.571.867	2.530
Fresenius Medical Care AG & Co. KGaA	01,32 %	293.932.036	4.884
Hannover Rückversicherung AG	00,40 %	120.597.134	1.480
Henkel KGaA	00,92 %	178.162.875	3.415
K+S AG	01,16 %	165.000.000	4.316
Linde AG	02,40 %	168.489.496	8.908
MAN AG	00,85 %	140.974.350	3.142
Merck KGaA	01,10 %	64.621.126	4.098
Metro AG	00,72 %	324.109.563	2.671
Münchener Rück AG	05,02 %	206.403.804	18.628
RWE AG	05,78 %	523.405.000	21.437
Salzgitter AG	00,45 %	60.097.000	1.655
SAP AG	06,20 %	1.225.844.696	23.003
Siemens AG	09,73 %	914.203.421	36.121
ThyssenKrupp AG	01,22 %	514.489.044	4.514
Volkswagen AG	04,88 %	294.925.087	18.112
	100,00 %	19.069.943.673	371.119

Die Zusammensetzung des DAX wird vierteljährlich überprüft und ggf. durch Austausch von Aktiengesellschaften geändert. Wir sprechen dabei auch von *ordentlicher Anpassung* des Index. Zusätzlich wird auch das Wägungsschema vierteljährlich aktualisiert, auch dann, wenn die Zusammensetzung des Index unverändert bleibt. Kriterien bei der Aufnahme neuer Aktien bzw. dem Ausschluss bereits enthaltener Aktien sind dabei:

- Orderbuchumsatz = Summe des an der FWB (inkl. Xetra) in den letzten zwölf Monaten festgestellten Umsatzes der Aktie einer Gesellschaft
- Freefloat-Marktkapitalisierung = Produkt aus Anzahl der frei verfügbaren Aktien und dem umsatzgewichteten Durchschnittspreises der vergangenen 20 Handelstage

In gewissen Ausnahmefällen (z.B. kurzfristig angekündigte Übernahmen oder signifikante Veränderung des Freefloats, Eröffnung des Insolvenzverfahrens, ...) kann der Vorstand der Deutschen Börse AG in Abstimmung mit dem Arbeitskreis Aktienindizes auch eine *außerordentliche Anpassung* des DAX vornehmen.

4.

Aufgaben

Grundlagen

Aufgabe G1

Geben Sie zu folgenden statistischen Massen an, ob es sich um Bestands- oder Bewegungsmassen handelt:

- a) Todesfälle in einer Großstadt
- b) Maschinenausfälle in einem Unternehmen
- c) Wartende Kunden vor einem Schalter
- d) Arbeitnehmer in einem Betrieb
- e) Anmeldungen in einem Einwohnermeldeamt

Aufgabe G2

Geben Sie zu folgenden Bestandsmassen mögliche korrespondierende Bewegungsmassen an:

- a) Einwohner der Stadt München
- b) Kontostand
- c) Rentenempfänger
- d) Auftragsbestand

Aufgabe G3

Auf welchen Skalen können folgende Merkmale gemessen werden?

- a) Jahresumsatz
- b) Temperatur
- c) Handelsklasse von Obst
- d) Nationalität
- e) Tarifklassen bei der Kfz-Haftpflichtversicherung
- f) Zensuren

Eindimensionale Häufigkeitsverteilungen

Aufgabe EH1

Ein Anleger kaufte am 02.01.2003 Aktien bei einem Kurs von 100 €. Am 02.01.2004 lag der Kurs bei 160 €, am 02.01.2005 bei 80 €.

Wie hoch war die durchschnittliche Wertveränderung pro Jahr?

Aufgabe EH2

Der durchschnittliche Bruttomonatsverdienst von Arbeitnehmern, die der Sozialversicherungspflicht unterliegen, betrug im Jahr 2001 in Westdeutschland 2.816 € in Ostdeutschland 2.013 €. Im Westen waren 23 Mio. und im Osten 4,8 Mio. Arbeitnehmer beschäftigt.

Berechnen Sie den durchschnittlichen Monatsverdienst aller Deutschen!

Aufgabe EH3

Die Arbeitslosenquoten betrugen im Jahr 2000 in Ost- und Westdeutschland 17,40 % bzw. 7,80 %. Im Osten gab es 7,806 Mio. und im Westen 32,505 Mio. Erwerbspersonen. Berechnen Sie die gesamtdeutsche Arbeitslosenquote und beachten Sie dabei, dass die folgenden Definitionen gelten:

$$\text{Arbeitslosenquote} = \frac{\text{Arbeitslose}}{\text{Arbeitslose} + \text{Erwerbstätige}}$$

$$\text{Erwerbspersonen} = \text{Arbeitslose} + \text{Erwerbstätige}$$

Aufgabe EH4

Für die Preise von PC-Monitoren eines bestimmten Typs ergab sich ein Durchschnittspreis von $\bar{x}_M = 300$ € und eine Standardabweichung von $s_M = 50$ €. Für Flachbildschirme ergab sich hingegen $\bar{x}_F = 550$ € und $s_F = 100$ €.

Kann nun die Aussage gemacht werden, dass Flachbildschirmspreise doppelt so stark streuen wie die Preise für PC-Monitore?

Aufgabe EH5

Betrachten wir drei theoretische Häufigkeitsverteilungen A, B und C:

Verteilung A		Verteilung B						
a_j	14	a_j	12	13	14	15	16	
$h(a_j)$	10	$h(a_j)$	1	2	4	2	1	
Verteilung C								
a_j	9	12	13	15	16	17	18	19
$h(a_j)$	2	2	1	1	1	1	1	1

Berechnen Sie aus diesem Zahlenmaterial arithmetisches Mittel, Standardabweichung und Spannweite! Machen Sie außerdem ohne Kennzahlenberechnung Aussagen zur Schiefe der Verteilungen!

Aufgabe EH6

Es wurden 30 Energiesparlampen einer Lebensdauerprüfung unterzogen. Die Lebensdauern in Stunden x_i für diese Lampen sind in folgender Tabelle angegeben:

i	x_i	i	x_i	i	x_i
1	251,7	11	755,8	21	897,6
2	295,1	12	767,5	22	906,2
3	301,1	13	772,3	23	934,0
4	378,7	14	789,1	24	951,3
5	489,0	15	799,4	25	957,8
6	546,9	16	800,5	26	965,9
7	599,2	17	834,0	27	999,3
8	632,6	18	851,3	28	1.004,4
9	711,9	19	877,1	29	1.055,1
10	734,7	20	884,9	30	1.156,6

Berechnen Sie den Median, den Quartilsabstand und den Quartilskoeffizienten dieser Verteilung!

Aufgabe EH7

Die Polizei greift an einem Abend 14 Studenten auf dem Heimweg von einer Semesteranfangsparty auf und stellt dabei folgende Blutalkoholkonzentrationen fest:

Männer	0,082	0,091	0,079	0,070	0,084	0,092	0,089
Frauen	0,052	0,049	0,032	0,041	0,052	0,061	0,059

Vergleichen Sie die Streuung der Alkoholkonzentration bei Männern und Frauen anhand geeigneter Maßzahlen!

Aufgabe EH8

Aus den Statistiken einer Bank ergaben sich für 64 Kreditkunden folgende Tilgungszahlungseingänge (in % der Gesamtforderung). Der Rest der jeweiligen Forderung wird von der Bank als endgültig verloren betrachtet und abgeschrieben.

0	10	30	5	35	12	12	8
12	5	3	30	5	35	5	5
6	40	7	6	0	7	5	5
12	5	3	6	12	5	7,5	9
6	7,5	5	5	12	12	6	5,5
12	21	15	3,5	15	3	17	12
21	17	21	12	11,5	15	11	3,5
17	12	5	40	12	12	15	15

Zu bearbeiten sind nun folgende vier Teilaufgaben. Versuchen Sie a), c) und d) mit Hilfe von *MS-Excel* zu lösen, da Ihnen hier das Programm einen erheblichen Rechenaufwand abnimmt.

- a) Geben Sie die kummulierten relativen Häufigkeiten der Zahlungsausfälle an, die in die empirische Verteilungsfunktion eingehen!
- b) Wie groß ist der Anteil der Schuldner, bei denen
 - b1) höchstens 70 % der Forderung ausfällt?
 - b2) mehr als 85 % und weniger als 90 % der Forderung ausfällt?
- c) Berechnen Sie arithmetisches Mittel, Modus und Median der Zahlungsausfälle!
- d) Berechnen Sie Spannweite, Varianz, Standardabweichung und Variationskoeffizient der Verteilung!

Aufgabe EH9

Von 100 Unternehmen einer Branche wurden folgende Reingewinne erzielt:

Gewinn (in Mio. Euro)	Anzahl der Unternehmen
5-15	20
15-25	10
25-35	30
35-45	5
45-55	5
55-65	10
65-75	20

Ermitteln Sie mit Hilfe von *MS-Excel* den durchschnittlichen Umsatz, der von den betrachteten Unternehmen erzielt wurde, sowie Varianz, Standardabweichung und den Variationskoeffizienten dieser Häufigkeitsverteilung!

Aufgabe EH10

Die Aufzeichnungen einer Wetterstation ergaben über einen Zeitraum von 2339 Tagen folgende Windgeschwindigkeiten:

j	a_j^u bis unter a_j^o	h_j	f_j
1	0,00	2,50	231
2	2,50	5,00	403
3	5,00	7,50	570
4	7,50	10,00	321
5	10,00	12,50	200
6	12,50	15,00	201
7	15,00	17,50	198
8	17,50	20,00	110
9	20,00	22,50	105
Σ		2339	1,0000

Was lässt sich aus diesem Zahlenmaterial über die Schiefe der Häufigkeitsverteilung der Windgeschwindigkeiten aussagen? Nutzen Sie für Ihre Berechnungen *MS-Excel*! Interpretieren Sie Ihr Ergebnis inhaltlich!

Aufgabe EH11

Ein Unternehmen hat für das abgelaufene Geschäftsjahr sämtliche Rechnungsbezüge (in Euro) für Beschaffungsgüter in Größenklassen eingeteilt. Hinsichtlich Anzahl und Beträge der Rechnungen ergab sich zum Jahresende folgendes Bild:

Großenklassen der Rechnungsbeträge	Anzahl der Rechnungen je Klasse	Beschaffungskosten je Klasse
[0, 500[255	74.000
[500, 1.000[490	343.000
[1.000, 2.500[165	295.000
[2.500, 5.000[60	216.000
[5.000, 10.000[24	187.000
[10.000, 20.000[6	85.000
	1.000	1.200.000

Beschaffungsgüter können im Rahmen einer ABC-Analyse in A-, B- und C-Güter eingeteilt werden. Diejenigen Güter, die mit den 20 % der Rechnungen mit den größten Beträgen verbunden sind, werden als A-Güter klassifiziert. Die Güter aus den 50 % kleinsten Rechnungen sind C-Güter. Die Güter dazwischen werden B-Güter genannt.

Skizzieren Sie für dieses Datenmaterial die Lorenz-Kurve und kennzeichnen Sie darin die A-, B- und C-Güter. Geben Sie außerdem an, welchen Anteil an den Gesamtkosten die A-Güter ausmachen!

Aufgabe EH12

Zwei Märkte weisen folgende Strukturen auf:

Markt A	Markt B
4 Firmen mit je 5 % Marktanteil	5 Firmen mit je 3 % Marktanteil
2 Firmen mit je 10 % Marktanteil	5 Firmen mit je 17 % Marktanteil
4 Firmen mit je 15 % Marktanteil	

Bestimmen Sie anhand des normierten Gini-Koeffizienten, welcher Markt die höhere Konzentration aufweist!

Mehrdimensionale Häufigkeitsverteilungen

Aufgabe MH1

Eine Befragung von 500 Studenten nach dem Studiengang (Merkmal X) und der sportlichen Betätigung (Merkmal Y) liefert folgende Kontingenztabelle:

X	Y	nie	gelegentlich	regelmäßig	Σ
BWL	30	40	200	270	
VWL	20	50	20	90	
Mathematik	80	40	20	140	
Σ	130	130	240	500	

Bestimmen Sie die bedingten Verteilungen des Merkmals Y und interpretieren Sie ihre Werte exemplarisch!

Aufgabe MH2

Ein Automobilhersteller liefert Klein-, Mittelklasse- und Oberklassewagen in die drei Regionen A, B und C. An einem bestimmten Tag wurden insgesamt 1.000 Fahrzeuge ausgeliefert. Davon gingen je 300 in die Regionen A und C. Die Lieferung enthielt zu 50 % Kleinwagen, zu 30 % Mittelklassewagen und zu 20 % Oberklassewagen. In die Region A wurden 150 Kleinwagen und 100 Mittelklassewagen, in die Region B 100 Mittelklassewagen und in die Region C 50 Oberklassewagen geliefert.

- Erstellen Sie hierzu eine zweidimensionale Häufigkeitstabelle der absoluten und relativen Häufigkeiten mit den entsprechenden Randverteilungen!
- Bestimmen Sie die Häufigkeitstabelle der bedingten Verteilung des Merkmals Pkw-Typ, die durch Beschränkung auf die Region A entsteht!
- Bestimmen Sie die Häufigkeitstabelle der bedingten Verteilung des Merkmals Region, die durch Beschränkung auf Kleinwagen entsteht!
- Geben Sie den Modus bzgl. beider Merkmale an!

Aufgabe MH3

Für 500 Mietwohnungen ergab eine Untersuchung folgende Daten über Mietpreis Y (in Euro) in Abhängigkeit von der Wohnfläche X (in m²):

X	Y	0 - 300	300 - 600	600 - 900	900 - 1200
0 - 40	84	43	0	0	
40 - 80	76	160	23	12	
80 - 120	3	14	25	60	

Berechnen Sie ein geeignetes Maß zur Charakterisierung des Zusammenhangs zwischen Mietpreis und Wohnfläche!

Aufgabe MH4

Nehmen Sie an, eine UNICEF-Studie liefert die folgenden Kindersterblichkeitsraten X in 9 Staaten des Nahen Ostens und zudem die prozentualen Anteile der Lese- und Schreibkundigen der Gesamtbevölkerung der jeweiligen Länder. Diese Alphabetisierungsrate Y kann als Indikator für den Entwicklungsstand des betreffenden Landes interpretiert werden.

Land	x_i (in %)	y_i (in %)
Kuwait	1,9	67,0
Vereinigte Arabische Emirate	3,0	38,0
Jordanien	5,2	70,0
Libanon	5,6	73,0
Iran	5,9	43,0
Syrien	5,9	51,0
Irak	8,6	49,0
Saudi-Arabien	9,1	48,0
Jemen	18,7	21,0

Untersuchen Sie mittels eines geeigneten Korrelationsmaßes den Zusammenhang zwischen Kindersterblichkeits- und Alphabetisierungsrate!

Aufgabe MH5

Die folgende Vierfeldertafel weise eine vom Statistischen Bundesamt ermittelte Gliederung der in der Bundesrepublik lebenden Personen nach dem Geschlecht (Merkmals X) und der Beteiligung an der Erwerbstätigkeit (Merkmals Y) aus (Angaben in Tausend Personen):

	Erwerbspersonen	Nichterwerbspersonen	Σ
Männlich	22.232	18.098	40.330
Weiblich	17.814	24.346	42.160
Σ	40.046	42.444	82.490

- Um welche spezielle Art von Merkmalen handelt es sich bei X und Y?
- Beurteilen Sie den statistischen Zusammenhang der Merkmale X und Y anhand einer geeigneten Maßzahl!
- Betrachten Sie den Fall, dass alle Männer erwerbstätig und alle Frauen nicht erwerbstätig sind, und machen Sie Aussagen zur Abhängigkeit zwischen X und Y in diesem Fall!

Aufgabe MH6

Beweisen Sie die Gültigkeit der Beziehung (I3: 23) bzw. des Verschiebungssatzes für die Kovarianz!

Preisindizes

Aufgabe P1

In welchem Fall liefern die Preisindizes nach Laspeyres und Paasche identische Ergebnisse? Begründen Sie Ihre Aussage!

Aufgabe P2

Ein Laspeyres-Index zeigt folgendes Verhalten: Steigt der Preis eines Gutes X um 8 % und bleiben alle restlichen Preise unverändert, so resultiert ausgehend von einem Startwert 1 ein Indexwert von 1,004. Mit welchem Gewicht ist das Gut X im Warenkorb vertreten?

Aufgabe P3

Folgende Tabelle zeigt die Durchschnittswerte für Preis (in Euro) und Verbrauch (in Liter) von Kraftstoff für zwei Jahre A und B:

Kraftstoff	Preis Jahr A	Verbrauch Jahr A	Preis Jahr B	Verbrauch Jahr B
Super	0,96	400	1,15	700
Normal	0,90	700	1,10	600
Diesel	0,87	350	1,01	250

- a) Berechnen Sie den Preisindex nach Laspeyres und nach Paasche! Basisjahr ist das Jahr A.
- b) Ermitteln Sie den Wertindex!

Aufgabe P4

Gegeben ist die folgende Indexreihe (Preisindex nach Laspeyres für die Lebenshaltungskosten):

Jahr	2001	2002	2003	2004	2005	2006
Index	98,1	100,0	122,5	130,7	146,6	149,4

Entscheiden Sie für folgende Aussagen jeweils, ob sie richtig oder falsch sind! Begründen Sie dabei stets Ihre Aussage!

- a) Der Preisindex nach Laspeyres ist ein gewogenes arithmetisches Mittel aus Preisen.
- b) Der Preisindex nach Laspeyres ist ein gewogenes arithmetisches Mittel aus Preisverhältnissen mit Gewichten aus der Basisperiode.
- c) Im Jahr 2002 waren die Kosten für die Lebenshaltung im Durchschnitt 100 % höher als zu Beginn der statistischen Berechnung.
- d) Die Steigerung der durchschnittlichen Lebenshaltungskosten im Jahr 2006 bezogen auf 2005 belief sich auf 2,8 %.
- e) Die Steigerung der durchschnittlichen Lebenshaltungskosten im Jahr 2004 bezogen auf 2001 beträgt 33,23 %.

Aufgabe P5

Gegeben sind zwei Indexreihen A und B. Bei A ist 2003, bei B 2000 als Basisjahr zu betrachten. Beide Indexreihen sollen nun miteinander verglichen werden. Wie lässt sich dies bewerkstelligen? Führen Sie die erforderlichen Operationen durch!

	2003	2004	2005	2006
Index I ^A	100	104	115	120
Index I ^B	140	150	155	157

Aufgabe P6

Für eine bestimmte Menge von Warengruppen und denselben Basis- und Berichtszeitraum (Basisjahr 2002, Berichtsjahr 2006) sind der Preisindex nach Laspeyres mit $P_{2002,2006}^L = 125,7$ und der Mengenindex nach Paasche mit $q_{2002,2006}^P = 110,8$ gegeben.

- Welchen Wert nimmt der Umsatzindex $u_{2002,2006}$ an? Beweisen Sie in diesem Zusammenhang die Gültigkeit der von Ihnen zur Berechnung verwendeten formalen Beziehung!
- Wie hoch ist das durchschnittliche Umsatzwachstum während dieser 4 Jahre?

II WAHRSCHEINLICHKEITS- RECHNUNG

Thema dieses Kapitels ist die Wahrscheinlichkeitstheorie, die sich mit der mathematischen Beschreibung von Zufallserscheinungen und der Analyse von Gesetzmäßigkeiten befasst, die diesen innewohnen. Wir legen dazu im Abschnitt II 1 zunächst die begrifflichen Grundlagen und beschäftigen uns mit der Messung der "Neigung" oder "Chance" des Eintretens bestimmter Ereignisse durch Zahlen, die man Wahrscheinlichkeiten nennt. Im Abschnitt II 2 führen wir das zentrale Konzept der Zufallsvariablen ein und analysieren deren Wahrscheinlichkeitsverteilungen sowie Maßzahlen zur Charakterisierung dieser Verteilungen. Wir werden hier starke Analogien zur ein- und mehrdimensionalen Häufigkeitsverteilung feststellen, weshalb wir häufig auf erneute Beweisführungen verzichten können. Abschnitt II 3 beschließt das Kapitel mit einem Überblick über wichtige spezielle Wahrscheinlichkeitsverteilungen, die uns insbesondere in der induktiven Statistik des Kapitels III nützliche Dienste erweisen werden.

1.

Grundlagen der Wahrscheinlichkeitstheorie

Viele Ergebnisse wirtschaftlicher Entscheidungen sind nicht streng vorherbestimmbar, sondern besitzen Zufallscharakter. So lässt sich beispielsweise nicht genau im Voraus bestimmen, welche Rendite die Investition in eine Aktie oder eine andere risikobehaftete Anlageform letztendlich liefert. Mit derartigen Zufallssituationen werden wir uns in diesem Kapitel auseinandersetzen. Da das Verständnis der Wahrscheinlichkeitstheorie anhand realer Entscheidungssituationen aus dem Wirtschaftsleben jedoch oft schwer fällt, erläutern wir die Sätze und Regeln der Wahrscheinlichkeitsrechnung zumeist anhand einfacher Glücksspiele (z.B. Münzwurf, Würfeln, etc.). Sie können nämlich als die Wurzeln der Wahrscheinlichkeitstheorie betrachtet werden.

1.1 Grundbegriffe

Ausgangspunkt der Wahrscheinlichkeitsrechnung ist der sog. **Zufallsvorgang**. Er ist definiert als Geschehen, bei dem aus einer gegebenen Ausgangssituation heraus mehrere sich gegenseitig ausschließende Folgesituationen möglich sind, wobei ungewiss ist, welche dieser Folgesituationen eintreten wird. Typische Beispiele für Zufallsvorgänge sind etwa das Werfen einer Münze oder eines Würfels, das Ziehen einer Karte aus einem Kartenspiel, die Ziehung der Lottozahlen, das Drehen eines Glücksrades oder die Entnahme einer Zufallsstichprobe (z.B. zur Bestimmung des Ausschussanteils einer Produktionsserie).

Lassen wir einen Zufallsvorgang mehrfach ablaufen, sprechen wir von einem **Zufallsexperiment**. Ein solches ist dadurch charakterisiert, dass der Zufallsvorgang

- nach einer exakt bestimmten Vorschrift durchgeführt wird und
- unter identischen Bedingungen beliebig oft wiederholbar ist.

Bei einem Zufallsexperiment setzt man also voraus, dass sich die einzelnen Wiederholungen des Zufallsvorgangs nicht gegenseitig beeinflussen, also *unabhängig voneinander* sind.

Die einzelnen, nicht mehr weiter zerlegbaren, möglichen Ausgänge oder Ergebnisse eines Zufallsexperiments werden als **Elementarereignisse** ω bezeichnet. Bei jeder Durchführung des Experiments kann bzw. muss immer nur ein *einziges* dieser Elementarereignisse eintreten. Man sagt daher, dass sich die Elementarereignisse *gegenseitig ausschließen*. Das nach Ablauf des Geschehens tatsächlich eintretende Elementarereignis ist das sog. **Ergebnis** des Zufallsexperiments. Die Menge Ω aller Elementarereignisse eines Zufallsexperiments nennen wir **Ergebnisraum** (in der Literatur gelegentlich auch Ergebnismenge oder Ereignisraum) des Zufallsexperiments. Existieren n mögliche Elementarereignisse, dann gilt

$$\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}. \quad (\text{II1: 1})$$

Eine solche Darstellung ist allerdings nur möglich, wenn der Ergebnisraum *endlich* viele oder höchstens *abzählbar unendlich* viele Elementarereignisse hat. In der Literatur wird die Anzahl der Elementarereignisse des Ergebnisraumes häufig als *Mächtigkeit* $|\Omega|$ bezeichnet. Es gibt jedoch auch Zufallsexperimente, bei denen der Ergebnisraum mehr als endlich viele oder abzählbar unendlich viele Elemente hat. Wir sprechen hier auch von einem *stetigen Kontinuum* von Elementarereignissen.

Neben dieser Kategorisierung hinsichtlich der Anzahl und Abzählbarkeit der Elemente können wir Ergebnisseräume auch hinsichtlich der Struktur der Elemente unterscheiden. So kann ein Elementarereignis ein einzelnes Objekt oder ein n-Tupel von Objekten sein (vgl. dazu Beispiel II 1-1).

Beispiel II 1-1

Endliche Ergebnisseräume

- Einmaliger Wurf einer Münze:

Bezeichnen wir das Elementarereignis, dass die Münze nach dem Werfen "Kopf" zeigt, mit K, und dass sie "Zahl" zeigt, mit Z, ergibt sich der Ergebnisraum

$$\Omega = \{K, Z\}.$$

- Zweimaliger Wurf einer Münze:

Hier sind die Elementarereignisse geordnete Paare (i,j), wobei i das Resultat des ersten und j das Resultat des zweiten Münzwurfs darstellt. Der Ergebnisraum ist damit

$$\Omega = \{KK, KZ, ZK, ZZ\}.$$

- Einmaliger Wurf eines Würfels:

Da die Elementarereignisse durch die Zahl der Augen auf der oben liegenden Fläche des Würfels gekennzeichnet sind, ergibt sich folgendes Bild des Ergebnisraumes:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- Zweimaliger Wurf eines Würfels:

Auch hier sind die Elementarereignisse geordnete Paare (i,j), wobei i das Ergebnis des ersten und j das Ergebnis des zweiten Würfelwurfs bezeichnet. Bei $i = 1, 2, \dots, 6$ und $j = 1, 2, \dots, 6$ ergeben sich insgesamt $|\Omega| = 6 \cdot 6 = 36$ Elementarereignisse und der folgende Ergebnisraum:

$$\Omega = \left\{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \right\}$$

Alternativ lässt sich dieser Ergebnisraum auch folgendermaßen darstellen:

$$\Omega = \{(x_1, x_2) : x_i \in \{1, 2, 3, 4, 5, 6\} \text{ für } i = 1, 2\}$$

- **Ziehung der Lottozahlen:**

Beim Mittwochs- und Samstaglotto werden aus 49 von 1 bis 49 durchnummerierten Kugeln genau 6 Kugeln zufällig gezogen und die jeweiligen Nummern registriert. Die dabei auftretenden Nummern sind die Gewinnzahlen und werden nach der Ziehung der Größe nach geordnet, da es keine Rolle spielt, in welcher Reihenfolge sie gezogen werden. Ein Lottospieler tippt also auf ein Elementarereignis der allgemeinen Form $\omega = \{x_1, x_2, x_3, x_4, x_5, x_6\}$. Da die Nummern ohne Zurücklegen einer vorher gezogenen Kugel ermittelt werden, ist es wichtig zu beachten, dass keine Kugel mehrmals gezogen werden kann. Der Ergebnisraum zeigt sich damit wie folgt:

$$\Omega = \left\{ (x_1, x_2, x_3, x_4, x_5, x_6) : \begin{array}{l} x_i \in \{1, 2, 3, \dots, 49\} \\ \text{und } x_1 < x_2 < x_3 < x_4 < x_5 < x_6 \end{array} \right\}$$

Beispiel II 1-2

Abzählbar unendlicher Ergebnisraum

Eine Münze wird so lange geworfen bis Zahl erscheint. Der Ergebnisraum hat hier unendlich viele Elementarereignisse, ist jedoch abzählbar.

$$\Omega = \{\text{Z}, \text{KZ}, \text{KKZ}, \text{KKKZ}, \text{KKKKZ}, \dots\}$$

Beispiel II 1-3

Unendlicher Ergebnisraum

Mittels einer Zufallsstichprobe soll die in Kilometer gemessene Laufleistung x von Pkw-Motoren gemessen werden, die diese bis zu ihrem "Ausfall" erbringen. Der Ergebnisraum eines solchen Zufallsexperiments kann wie folgt dargestellt werden:

$$\Omega = \{x : x \geq 0; x \text{ reell}\}$$

1.2 Ereignisse und ihre Darstellung

Unter einem sog. **Ereignis A** verstehen wir jede beliebige Teilmenge des Ergebnisraumes Ω , d.h. $A \subset \Omega$. Dies bedeutet, dass sich jedes Ereignis aus *einem oder mehreren* Elementarereignissen zusammensetzt. Die einzige Ausnahme stellt hierbei das sog. unmögliche Ereignis dar, welches aus Null Elementarereignissen besteht. Ein Ereignis tritt allgemein dann ein, wenn ein in ihm enthaltenes Elementarereignis $\omega \in A$ eintritt.

Beispiel II 1-4

Endliche Ereignisse

- Beim Zufallsexperiment "Einmaliger Würfelwurf" sind beispielsweise folgende Ereignisse denkbar:

$$A_1: \text{"Es ergeben sich nur gerade Augenzahlen."} \quad \rightarrow \quad A_1 = \{2, 4, 6\}$$

$$A_2: \text{"Die 6 wird gewürfelt."} \quad \rightarrow \quad A_2 = \{6\}$$

- Einige Ereignisse beim Zufallsexperiment "Zweimaliger Würfelwurf":

$$A_1: \text{"Die Summe der Augenzahlen ist mindestens 10."}$$

$$\rightarrow A_1 = \{(x_1, x_2) : x_1 + x_2 \geq 10\} = \{(4, 6), (6, 4), (5, 5), (5, 6), (6, 5), (6, 6)\}$$

A_2 : "Die Summe der Augenzahlen ist höchstens 4."

$$\rightarrow A_2 = \{(x_1, x_2) : x_1 + x_2 \leq 4\} = \{(1,1), (1,2), (2,1), (2,2), (1,3), (3,1)\}$$

A_3 : "Die Summe der Augenzahlen ist größer als 10."

$$\rightarrow A_3 = \{(x_1, x_2) : x_1 + x_2 > 10\} = \{(5,6), (6,5), (6,6)\}$$

A_4 : "Es werden zwei gleiche Zahlen gewürfelt."

$$\rightarrow A_4 = \{(x_1, x_2) : x_1 = x_2\} = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$$

A_5 : "Die zuerst gewürfelte Zahl ist eine 6."

$$\rightarrow A_5 = \{(x_1, x_2) : x_1 = 6; x_2 \in \{1,2,3,4,5,6\}\} = \{(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$$

A_6 : "Es wird zweimal die 6 gewürfelt."

$$\rightarrow A_6 = \{(6,6)\}$$

Beispiel II 1-5

Unendliches Ereignis

Für das im vorangegangenen Abschnitt aufgeführte Beispiel II 1-3 lässt sich das folgende Ereignis betrachten:

A : "Die Motorlaufleistung ist größer als 150.000 km."

$$\rightarrow A = \{x : x > 150.000; x \text{ reell}\}$$

Für Ereignisse, die i.d.R. mit Großbuchstaben A , A_1 , A_2 , B , C usw. bezeichnet werden, Beziehungen zwischen Ereignissen sowie Ereignisoperationen, mit denen sich aus Ereignissen neue Ereignisse bilden lassen, existieren eine Reihe wichtiger **Schreibweisen**, die in der folgenden Schematik im Überblick dargestellt und im Anschluss daran im Detail behandelt werden. Da Ereignisse in ihrer einfachsten Form als Mengen definiert sind, werden wir im Folgenden sehen, dass wir die in der Mengenlehre verwendeten Notationen und Operationen auch auf Ereignisse übertragen können.

Bezeichnung	Darstellung
1 A ist ein sicheres Ereignis .	$A = \Omega$
2 A ist ein unmögliches Ereignis .	$A = \emptyset = \{\}$
3 A ist ein Teilereignis von B .	$A \subset B$
4 A und B sind äquivalente Ereignisse .	$A = B$
5 A ist die Vereinigung der A_j .	$A = A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{j=1}^n A_j$
6 A ist der Durchschnitt der A_j .	$A = A_1 \cap A_2 \cap \dots \cap A_n = \bigcap_{j=1}^n A_j$
7 A und B sind disjunkte Ereignisse .	$A \cap B = \emptyset$
8 A und B sind komplementäre Ereignisse .	$B = \bar{A}$
9 C ist die Differenz aus A und B (A ohne B).	$C = A \setminus B$

1. Sicheres Ereignis

Ein Ereignis A wird als sicheres Ereignis bezeichnet, wenn das Ereignis der Ergebnisraum Ω selbst ist und damit alle Elementarereignisse ω enthält. Ein solches Ereignis A tritt sicher bei jedem Ausführen des Zufallsexperimentes ein. Formal schreiben wir daher

$$A = \Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}. \quad (\text{II1: 2})$$

2. Unmögliches Ereignis

Ein Ereignis A tritt sicher nicht bzw. niemals ein und gilt als unmöglich, wenn es gleich der sog. leeren Menge \emptyset ist. Diese enthält per Definition kein Element und somit auch kein Elementarereignis aus Ω . Es gilt

$$A = \emptyset = \{\}. \quad (\text{III1: 3})$$

3. Teilereignisse

Sind alle Elementarereignisse eines Ereignisses A auch in einem Ereignis B enthalten, bezeichnet man das Ereignis A als Teilereignis von Ereignis B. Tritt das Ereignis A ein, so tritt automatisch auch das Ereignis B ein. Wir sagen auch, dass das Ereignis A das Ereignis B impliziert, und schreiben

$$A \subset B. \quad (\text{II1: 4})$$

Ereignisse und Ergebnisraum lassen sich im sog. **Venn-Diagramm** anschaulich darstellen. Die Ereignisse werden dabei als Flächen dargestellt, die sich möglicherweise überlappen. Abbildung II 1 (links) zeigt das Venn-Diagramm für den Fall eines Teilereignisses.

4. Äquivalente Ereignisse

Zwei Ereignisse A und B werden als äquivalent bezeichnet, wenn die Elemente beider Ereignisse identisch sind. Genau dann, wenn A eintritt, tritt auch B ein. Wir schreiben

$$A = B. \quad (\text{III1: 5})$$

Das dazugehörige Venn-Diagramm ist in Abbildung II 1 (rechts) zu sehen. Wir erkennen daran, dass hier lediglich ein spezieller Fall eines Teilereignisses vorliegt. Alle Elemente von A sind auch in B enthalten und umgekehrt.

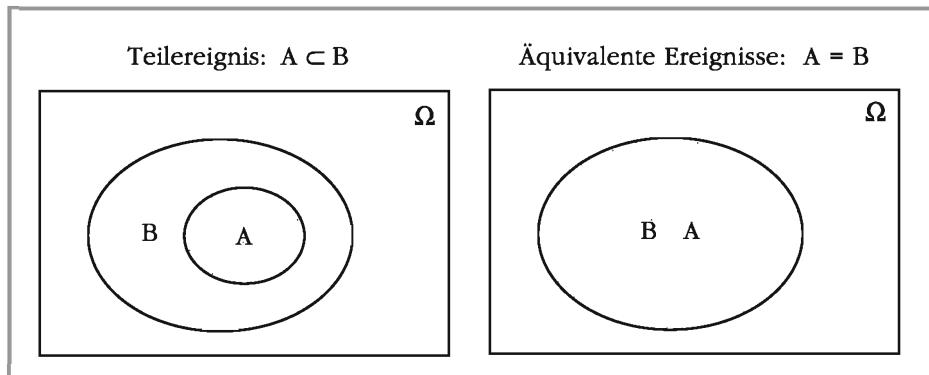


Abbildung II 1: Teilereignis und äquivalente Ereignisse

5. Vereinigung

Die Vereinigung V zweier Ereignisse A und B ist definiert als die Menge aller Elementarereignisse, die entweder zu A allein oder zu B allein oder sowohl zu A als auch zu B gemeinsam gehören. Wir schreiben

$$V = A \cup B. \quad (\text{III1: 6})$$

Das neue Ereignis V tritt also genau dann ein, wenn entweder A oder B allein oder beide Ereignisse gemeinsam eintreten.¹ Wir sprechen kurz: "A oder B". Die Vereinigung stellt sich im Venn-Diagramm wie in Abbildung II 2 (links) dar.

6. Durchschnitt

Der Durchschnitt D zweier Ereignisse A und B ist definiert als die Menge aller Elementarereignisse, die sowohl zu A als auch zu B gehören.

$$D = A \cap B \quad (\text{III1: 7})$$

Das neue Ereignis D tritt somit genau dann ein, wenn Ereignis A und Ereignis B gemeinsam eintreten.² Wir sprechen kurz: "A und B". Es ergibt sich damit das Venn-Diagramm in Abbildung II 2 (rechts). Auf die darin enthaltenen Differenzen $A \setminus B$ und $B \setminus A$ werden wir in Kürze zu sprechen kommen.

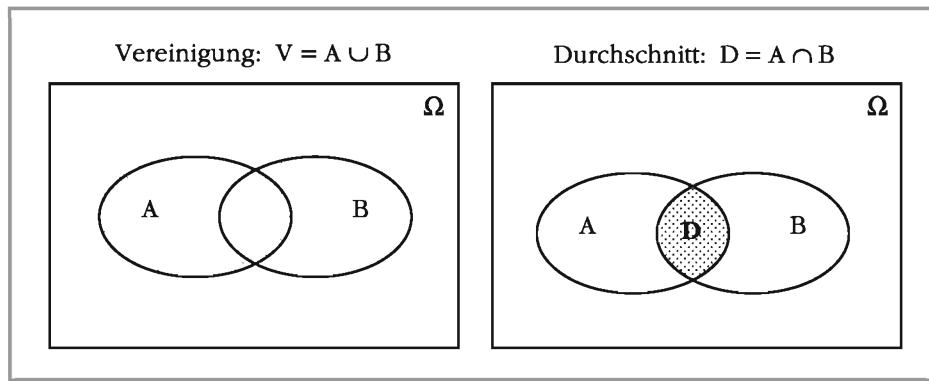


Abbildung II 2: Vereinigung und Durchschnitt

7. Disjunkte Ereignisse

Zwei Ereignisse A und B heißen disjunkt oder unvereinbar, wenn

$$A \cap B = \emptyset \quad (\text{III1: 8})$$

gilt, sie also kein Elementarereignis gemeinsam haben (sich nicht "schneiden"). Damit gilt immer, dass zwei beliebige Elementarereignisse selbst paarweise disjunkt sind.

Disjunkte Ereignisse schließen sich gegenseitig aus. Dies bedeutet, dass sie nicht zugleich eintreten können. Wenn also A eintritt, tritt B nicht ein. Wenn A

¹ In der vorhergehenden Schematik wurde die Vereinigung mehrerer Ereignisse A_i beschrieben. Für diesen Fall gilt: Genau dann, wenn mindestens ein A_i eintritt, tritt A ein.

² Für den in der Schematik beschriebenen Fall gilt: Genau dann, wenn alle A_i eintreten, tritt A ein.

nicht eintritt, bedeutet dies aber nicht automatisch, dass B eintritt. Dies ist auch im Venn-Diagramm in Abbildung II 3 (links) zu sehen.

8. Komplementäre Ereignisse

Unter dem Komplementärereignis oder Komplement \bar{A} eines Ereignisses A verstehen wir die Menge sämtlicher Elementarereignisse des Ergebnisraumes Ω , die nicht in Ereignis A enthalten sind. Das Ereignis \bar{A} tritt also immer dann ein, wenn das Ereignis A nicht eintritt. Genau dann, wenn A eintritt, tritt \bar{A} nicht ein.

Komplementäre Ereignisse sind nach (II1: 9) stets disjunkt und ergänzen sich nach (II1: 10) zum sicheren Ereignis.

$$A \cap \bar{A} = \emptyset \quad (\text{II1: 9})$$

$$A \cup \bar{A} = \Omega \quad (\text{II1: 10})$$

Es zeigt sich das Venn-Diagramm in Abbildung II 3 (rechts). Zudem ist festzuhalten, dass das komplementäre Ereignis zu \bar{A} wieder das ursprüngliche Ereignis A ergibt, also

$$\bar{\bar{A}} = A \quad (\text{II1: 11})$$

gilt, und das Komplementärereignis zum sicheren Ereignis das unmögliche Ereignis ist (II1: 12) und umgekehrt (II1: 13):

$$\bar{\Omega} = \emptyset \quad (\text{II1: 12})$$

$$\bar{\emptyset} = \Omega \quad (\text{II1: 13})$$

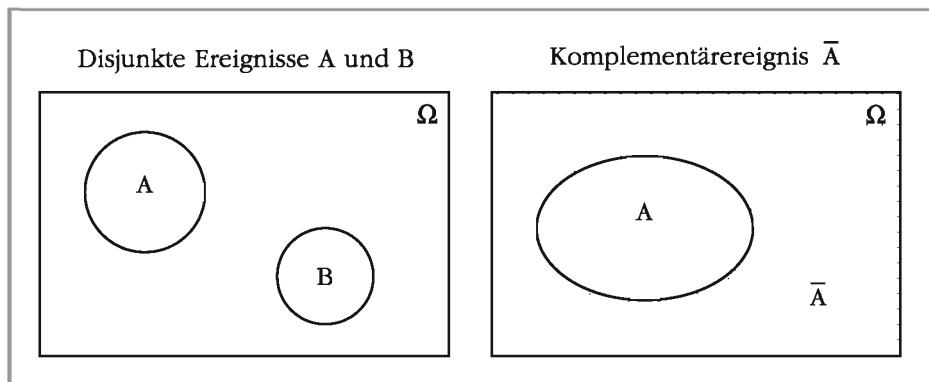


Abbildung II 3: Disjunkte und komplementäre Ereignisse

9. Differenz

Ein Ereignis C, welches sich dadurch ergibt, dass die Elemente, die die Ereignisse A und B gemeinsam haben (d.h. $A \cap B$), von den Elementen des Ereignisses A abgezogen werden, bezeichnen wir als Differenz. Das Ereignis C beschreiben wir formal als

$$C = A \setminus B \quad (\text{II1: 14})$$

und sprechen "C ist A ohne B". Das Ereignis C tritt genau dann ein, wenn zwar A, aber nicht B eintritt. Es resultiert das Venn-Diagramm in Abbildung II 4.

Das Komplement \bar{A} eines Ereignisses A können wir als spezielle Differenz betrachten. Es gilt nämlich

$$\bar{A} = \Omega \setminus A.$$

(II1: 15)

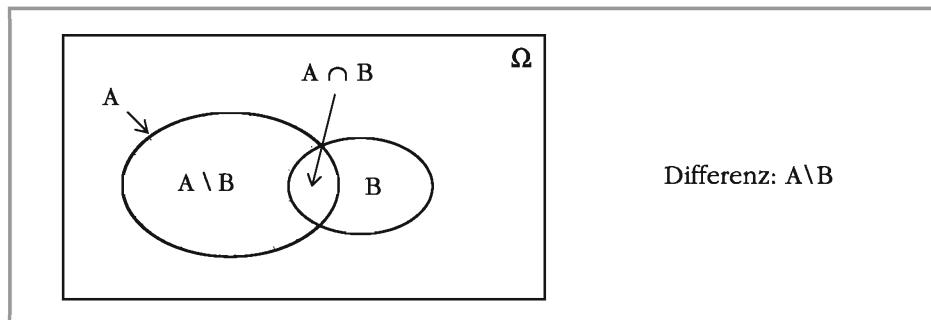
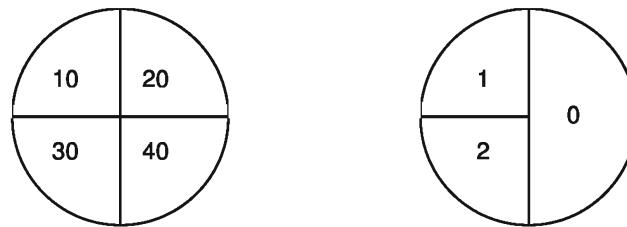


Abbildung II 4: Differenz

Beispiel II 1-6

Ereignisbeziehungen

Ein Glücksspiel bestehe aus dem Drehen der beiden folgenden Glücksräder. Der Spielteilnehmer soll dabei das Produkt der zwei sich ergebenden Zahlen als Gewinn erhalten.



Der Ergebnisraum ($4 \cdot 3 = 12$ Elemente) dieses Zufallsexperiments kann wie folgt dargestellt werden:

$$\Omega = \left\{ (10,1), (20,1), (30,1), (40,1), (10,2), (20,2), (30,2), (40,2), (10,0), (20,0), (30,0), (40,0) \right\}$$

Gegeben seien außerdem die folgenden ausgewählten Ereignisse:

E_1 : "Der Spieler macht keinen Gewinn."

$$\rightarrow E_1 = \{(10,0), (20,0), (30,0), (40,0)\}$$

E_2 : "Der Spieler macht einen Gewinn."

$$\rightarrow E_2 = \{(10,1), (20,1), (30,1), (40,1), (10,2), (20,2), (30,2), (40,2)\}$$

E_3 : "Der Spieler gewinnt mindestens 60 Euro."

$$\rightarrow E_3 = \{(30,2), (40,2)\}$$

E_4 : "Der Spieler gewinnt genau 10 Euro."

$$\rightarrow E_4 = \{(10,1)\}$$

Zur Veranschaulichung der vorhergehend behandelten Notationen, Beziehungen und Operationen wollen wir nun anhand dieser Angaben einige Sachverhalte darstellen:

- **Sicheres Ereignis:**
 U_1 : "Der Spieler gewinnt höchstens 80 Euro." $\rightarrow U_1 = \Omega$
- **Unmögliches Ereignis:**
 U_2 : "Der Spieler gewinnt 200 Euro." $\rightarrow U_2 = \emptyset$
- **Teilereignis:**
 Es gilt $E_4 \subset E_2$, da alle Elemente von E_4 auch in E_2 enthalten sind.
- **Disjunkte Ereignisse:**
 Es gilt $E_1 \cap E_2 = \emptyset$, $E_1 \cap E_3 = \emptyset$ und $E_1 \cap E_4 = \emptyset$, da die betreffenden Ereignisse jeweils keine Elemente gemeinsam haben.
- **Komplementäre Ereignisse:**
 Das Ereignis E_1 ist das komplementäre Ereignis zu E_2 ($E_1 = \bar{E}_2$), da $E_1 \cup E_2 = \Omega$.
- **Differenz:**
 $E_2 \setminus E_3 = \{(10,1), (20,1), (30,1), (40,1), (10,2), (20,2)\}$
- **Vereinigung:**
 $E_3 \cup E_4 = \{(30,2), (40,2), (10,1)\}$
- **Durchschnitt:**
 $E_2 \cap E_3 = \{(30,2), (40,2)\}$

Diese Schnittmenge lässt sich auch in Tabellenform unter Kennzeichnung der entsprechenden Mengen wie folgt darstellen:

		Glücksrad 1			
		10	20	30	40
Glücksrad 2	0	•	•	•	•
	1	•	•	•	•
	2	•	•	•	•

1.3 Wahrscheinlichkeitsregeln und -definitionen

Das Ergebnis eines Zufallsexperimentes ist nicht vorhersehbar. Es ist höchstens möglich den Ereignissen gewisse "Wahrscheinlichkeiten" zuzuordnen. Während es sich bei den Ereignissen um Mengen handelt, sind Wahrscheinlichkeiten reelle Zahlen. Eine **Wahrscheinlichkeit** ist also nichts anderes als ein Maß zur Quantifizierung des Grades der Sicherheit oder Unsicherheit des Eintretens eines bestimmten Ereignisses im Rahmen eines Zufallsexperimentes. Welche Voraussetzungen ein solches Maß erfüllen muss, wollen wir im Folgenden näher betrachten.

1.3.1 Axiome der Wahrscheinlichkeitsrechnung

Jede Funktion P , die einem Ereignis A eine Wahrscheinlichkeit $P(A)$ zuordnet, wird als **Wahrscheinlichkeitsfunktion** und $P(A)$ als **Wahrscheinlichkeit** von A bezeichnet, wenn sie die drei folgenden, von A. N. Kolmogorov (1933) formulierten **Axiome** (keiner Beweise bedürfenden Grundsätze) erfüllt:

Axiom 1

Die Wahrscheinlichkeit $P(A)$ des Eintretens eines betrachteten Ereignisses A eines Zufallsexperimentes ist eine eindeutig bestimmbarer, nichtnegative reelle Zahl, die zudem kleiner oder gleich Eins sein muss.³ Dies muss für alle $A \subset \Omega$ erfüllt sein.

$$0 \leq P(A) \leq 1 \quad (\text{II1: 16})$$

Axiom 2

Die Wahrscheinlichkeit $P(A)$ eines Ereignisses, welches alle Elementarereignisse enthält (sicheres Ereignis $A = \Omega$), ist gleich Eins.

$$P(\Omega) = 1 \quad (\text{II1: 17})$$

Axiom 3

Schließen sich zwei Ereignisse A und B gegenseitig aus (disjunkte Ereignisse), so gilt die Additionsregel

$$P(A \cup B) = P(A) + P(B) \quad \text{falls} \quad A \cap B = \emptyset. \quad (\text{II1: 18})$$

Verallgemeinern wir diese Regel für eine beliebige Anzahl paarweise disjunkter Ereignisse, so ergibt sich Folgendes:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) \quad \text{falls}$$

$$A_i \cap A_j = \emptyset \text{ für alle } i \neq j \text{ mit } i, j = 1, 2, \dots, n$$

Aus diesen Axiomen lassen sich ohne große Schwierigkeiten weitere wichtige **Wahrscheinlichkeitssätze** ableiten, die in der Wahrscheinlichkeitsrechnung von großer Bedeutung sind:

Satz 1 – Wahrscheinlichkeit von komplementären Ereignissen

Die Wahrscheinlichkeit eines zu einem Ereignis A komplementären Ereignisses \bar{A} ergibt sich stets als

$$P(\bar{A}) = 1 - P(A). \quad (\text{II1: 20})$$

Dieser Zusammenhang lässt sich relativ einfach beweisen: Da A und \bar{A} komplementäre Ereignisse sind, ergibt die Vereinigung dieser Ereignisse den Ergebnisraum Ω , d.h. es gilt $A \cup \bar{A} = \Omega$ (vgl. Abbildung II 3, rechts). Die Wahrscheinlichkeit des Eintretens des sicheren Ereignisses Ω ist nach dem Axiom 2 gleich Eins:

³ Genau genommen lautet Axiom 1 nur $P(A) > 0$. In der Literatur wird jedoch meist bereits die Normierungsfunktion von Axiom 2 in Axiom 1 berücksichtigt, was zu (II1: 16) führt.

$$P(\Omega) = 1 \rightarrow P(A \cup \bar{A}) = 1$$

Unter Umstellung von Axiom 3 kann daraus aufgrund der Disjunktheit der beiden Ereignisse der Zusammenhang

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1 \rightarrow P(\bar{A}) = 1 - P(A)$$

hergeleitet werden.

Satz 2 – Wahrscheinlichkeit des unmöglichen Ereignisses

Das unmögliche Ereignis hat die Wahrscheinlichkeit Null.

$$P(\emptyset) = 0 \quad (\text{II1: 21})$$

Auch der Beweis hierfür ist nicht schwer zu erbringen: Ω und \emptyset sind komplementäre Ereignisse ($\bar{\emptyset} = \Omega$). Die Wahrscheinlichkeit des Eintretens von Ω ist nach Axiom 2 gleich Eins. Unter Anwendung von Satz 1 und Axiom 1 ergibt sich damit

$$P(\emptyset) = 1 - P(\bar{\emptyset}) = 1 - P(\Omega) = 1 - 1 = 0.$$

Satz 3 – Wahrscheinlichkeit einer Differenz $A \setminus B$

Für die Wahrscheinlichkeit einer Differenz $A \setminus B$ gilt stets

$$P(A \setminus B) = P(A) - P(A \cap B). \quad (\text{II1: 22})$$

Erklären lässt sich dieser Zusammenhang dadurch, dass sich das Ereignis A aus den beiden disjunkten Ereignissen $A \setminus B$ und $A \cap B$ (vgl. Abbildung II 4) zusammensetzt und somit nach Axiom 3

$$P(A) = P(A \setminus B) + P(A \cap B)$$

gilt. Durch Umformung lässt sich daraus Satz 3 herleiten.

Satz 4 – Additionssatz für beliebige Ereignisse

Für *zwei* beliebige Ereignisse A und B aus Ω gilt stets

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (\text{II1: 23})$$

Zur Erklärung dieser Formel werfen wir zunächst einen Blick auf Abbildung II 2. Aus dieser lässt sich ablesen, dass sich ein Ereignis $A \cup B$ aus 3 disjunkten Ereignissen zusammensetzt, also

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$$

gilt. Nach Satz 3 gilt dann

$$P(A \setminus B) = P(A) - (A \cap B)$$

$$P(B \setminus A) = P(B) - (A \cap B).$$

Unter Anwendung der Verallgemeinerung von Axiom 3 bzw. (II1: 19) und Zusammenführung dieser Erkenntnisse ergibt sich schließlich

$$\begin{aligned} P(A \cup B) &= P(A \setminus B) + P(A \cap B) + P(B \setminus A) \\ &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

An dieser Stelle sei noch erwähnt, dass Axiom 3 (disjunkte Ereignisse) nichts anderes als ein Spezialfall dieses Satz 4 für $A \cap B = \emptyset$ bzw. $P(A \cap B) = 0$ ist.

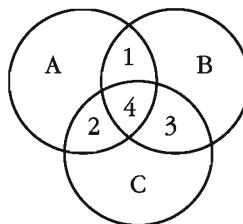
Für den Fall *dreier* beliebiger Ereignisse A, B und C lautet die Beziehung

$$P(A \cup B \cup C) = P(A) + P(\bar{A} \cap B) + P(\bar{A} \cap \bar{B} \cap C)$$

oder alternativ

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned} \quad (\text{II1: 24})$$

Wir können uns die Bedeutung von (II1: 24) relativ schnell klar machen: Würden wir einfach nur $P(A)$, $P(B)$ und $P(C)$ addieren, käme es dazu, dass wir die Wahrscheinlichkeiten gemeinsamer Elementarereignisse mehrfach berücksichtigen. So würden wir im Venn-Diagramm



die Elemente in den Bereichen 1, 2 und 3 doppelt und jene im Bereich 4 dreifach erfassen. Deswegen subtrahieren wir die Wahrscheinlichkeit jener Ereignisse, die die Elementarereignisse jeweils paarweise gemeinsam haben. Der Bereich 4, der die Elemente enthält, die A, B und C gemeinsam haben, bleibt dadurch jedoch unberücksichtigt, sodass die Wahrscheinlichkeit $P(A \cap B \cap C)$ wieder addiert werden muss.

Satz 5 – Monotonieeigenschaft des Wahrscheinlichkeitsmaßes

Ist ein Ereignis A eine Teilmenge eines Ereignisses B, d.h. impliziert das Ereignis A das Ereignis B, dann ist die Wahrscheinlichkeit von A niemals größer als die von B.

$$A \subset B \rightarrow P(A) \leq P(B) \quad (\text{II1: 25})$$

Ein formaler Beweis dieser Beziehung erübrigt sich, da ihre Gültigkeit unmittelbar durch einen Blick auf Abbildung II 1 (links) klar wird.⁴

Satz 6 – Durchschnitte bei einer Zerlegung

Betrachten wir Abbildung II 5, so erkennen wir, dass hier ein Ergebnisraum Ω vorliegt, der sich aus drei paarweise disjunkten Ereignissen A_1 , A_2 und A_3 zusammensetzt. Resultiert aus der Vereinigung von m Ereignissen A_1, A_2, \dots, A_m der Ergebnisraum Ω , d.h. $A_1 \cup A_2 \cup \dots \cup A_m = \Omega$, und sind diese m Ereignisse paarweise disjunkt, d.h. $A_i \cap A_j = \emptyset$ für $i, j = 1, 2, \dots, m$ mit $i \neq j$, sagen wir, dass die A_i eine **Zerlegung** von Ω bilden.

⁴ Es ist im Zusammenhang mit (II1: 25) anzumerken, dass zwar $A = B \rightarrow P(A) = P(B)$ gilt, jedoch *nicht* $P(A) = P(B) \rightarrow A = B$.

Bilden die A_j eine Zerlegung von Ω , so gilt für jedes Ereignis $B \subset \Omega$

$$P(B) = \sum_{j=1}^m P(B \cap A_j) \quad \text{falls} \quad \bigcup_{j=1}^m A_j = \Omega, \\ A_i \cap A_j = \emptyset \quad \text{für alle } i, j = 1, 2, \dots, m \text{ mit } i \neq j \quad (\text{III1: 26})$$

Betrachten wir zur Verdeutlichung von (III1: 26) nochmals Abbildung II 5, so erkennen wir, dass sich das Ereignis B aus den disjunkten Ereignissen X , Y und Z zusammensetzt. Für diese gilt

$$X = B \cap A_1, \quad Y = B \cap A_2, \quad Z = B \cap A_3.$$

Die Wahrscheinlichkeit des Ereignisses B lässt sich daher wie folgt ermitteln:

$$P(B) = P(X) + P(Y) + P(Z) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) = \sum_{j=1}^3 P(B \cap A_j)$$

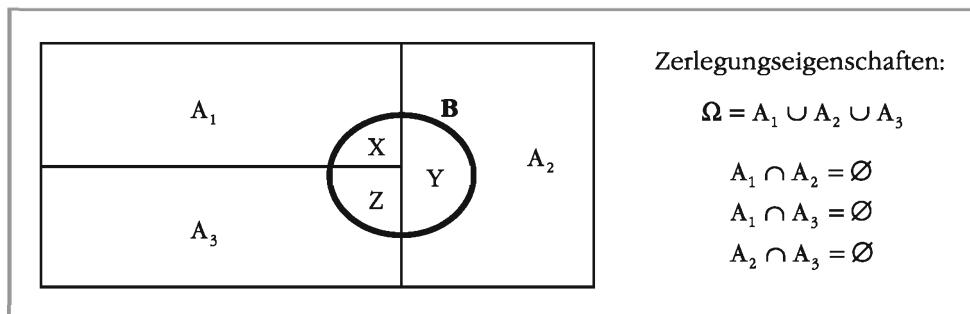


Abbildung II 5: Durchschnitte bei der Zerlegung

Die bisher vorgestellten **Axiome** und die daraus abgeleiteten **Sätze** liefern **keinen Beitrag zur Ermittlung der Zahl $P(A)$** . Sie legen lediglich formale Eigenschaften für Wahrscheinlichkeiten fest, dienen dazu, weitere Rechenregeln abzuleiten und die Wahrscheinlichkeiten komplexer Ereignisse zu berechnen. Sie bilden die Basis der Wahrscheinlichkeitstheorie. Wie sich die eigentlichen Wahrscheinlichkeiten messen oder angeben lassen, werden wir im Rahmen der folgenden Wahrscheinlichkeitsbegriffe näher betrachten.

1.3.2 Klassische Wahrscheinlichkeitsdefinition

Nach der sog. klassischen oder **Laplaceschen Definition der Wahrscheinlichkeit** ergibt sich die Wahrscheinlichkeit $P(A)$, dass bei einem bestimmten Zufallsexperiment ein Ereignis A eintritt, als der Quotient aus der Anzahl der für A günstigen Elementarereignisse und der Anzahl aller möglichen Elementarereignisse.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ günstigen Elementarereignisse}}{\text{Anzahl aller möglichen Elementarereignisse}} \quad (\text{III1: 27})$$

Diese Formel gilt nur für Zufallsexperimente mit *endlich vielen Elementarereignissen*, bei denen aufgrund der Ausgangssituation und der sonstigen Bedingungen, unter denen das Geschehen abläuft, als sicher gelten kann, dass die einzelnen Ele-

mentarereignisse dieselbe Chance des Eintretens besitzen, also *gleichwahrscheinlich* sind. Zufallsexperimente dieser Art heißen **Laplace-Experimente**.

Eigentlich sollte diese klassische Wahrscheinlichkeitsdefinition in der **Praxis** nur auf *Glücksspiele* (wie z.B. Münzwurf, Würfeln oder Roulette) und auf recht *einfach konstituierte Zufallsexperimente* angewendet werden, da wir hier wohl am ehesten von gleichwahrscheinlichen Elementarereignissen ausgehen können.⁵ Trotzdem wird sie aber häufig auch auf komplexere Sachverhalte angewendet. Das Resultat sind zwar quantifizierte Wahrscheinlichkeitsaussagen, doch sind diese vielleicht nur annähernd richtig oder zu ungenau, um als Entscheidungsgrundlage zu dienen.

Beispiel II 1-7

Laplace-Wahrscheinlichkeit

In einer Schachtel befinden sich insgesamt 10 Zündkerzen, wovon genau 2 defekt sind. Gesucht ist nun die Wahrscheinlichkeit $P(A)$ eines Ereignisses A, dass eine zufällig ausgewählte Zündkerze funktionstüchtig ist.

Da der Ergebnisraum hier insgesamt 10 Elementarereignisse und das Ereignis A genau 8 (günstige) Elementarereignisse umfasst, lässt sich $P(A)$ nach (II1: 27) als

$$P(A) = \frac{8}{10} = 0,80$$

ermitteln.

Gehen wir von einem Zufallsexperiment mit einem Ergebnisraum mit n gleichwahrscheinlichen Elementarereignissen aus, ergibt sich die Wahrscheinlichkeit dieser einzelnen Elementarereignisse als

$$P(\omega_1) = P(\omega_2) = \dots = P(\omega_i) = \dots = P(\omega_n) = \frac{1}{n}, \quad (\text{II1: 28})$$

da aufgrund von Axiom 3

$$\sum_{i=1}^n P(\omega_i) = 1 \quad (\text{II1: 29})$$

gilt.

Beispiel II 1-8

Laplace-Experimente

- Einmaliges Werfen eines Würfels

Da der Ergebnisraum $\Omega = \{1, 2, 3, 4, 5, 6\}$ dieses Zufallsexperimentes genau sechs Elemente aufweist, gilt für die Wahrscheinlichkeiten der einzelnen gleichwahrscheinlichen Elementarereignisse

$$P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = \frac{1}{|\Omega|} = \frac{1}{6}.$$

⁵ Sicherlich könnten wir behaupten, dass z.B. eine Münze nicht "perfekt" oder ein Würfel nicht "fair" ist und somit der Gleichwahrscheinlichkeit der Elementarereignisse widersprechen, doch stellen solche Manipulationen die Ausnahme dar.

Die Tatsache, dass die Summe der Wahrscheinlichkeiten der Elementarereignisse gleich 1 ist, ist hier relativ leicht zu zeigen:

$$\sum_{i=1}^n P(\omega_i) = P(\{1\}) + P(\{2\}) + \dots + P(\{6\}) = 6 \cdot \frac{1}{6} = 1$$

Berechnen wir nun für dieses Zufallsexperiment die Wahrscheinlichkeiten einiger Ereignisse (auch unter Zuhilfenahme der bisher behandelten Axiome und Sätze):

$$A_1: \text{"gerade Augenzahl"} \quad \rightarrow P(A_1) = \frac{|A_1|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$$

$$A_2: \text{"Augenzahl 5"} \quad \rightarrow P(A_2) = \frac{|A_2|}{|\Omega|} = \frac{1}{6}$$

$$A_3: \text{"Augenzahl 5 oder gerade"} \quad \rightarrow P(A_3) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \\ = \frac{1}{2} + \frac{1}{6} - 0 = \frac{2}{3}$$

- Zufälliges Ziehen einer Karte aus einem Skatspiel (32 Karten)

Ereignis: "Man erhält ein Ass (A) oder ein Karo (B)."

$$\rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ = \frac{|A|}{|\Omega|} + \frac{|B|}{|\Omega|} - \frac{|A \cap B|}{|\Omega|} = \frac{4}{32} + \frac{8}{32} - \frac{1}{32} = \frac{11}{32}$$

Hierbei ist zu beachten, dass das Karo-Ass sowohl in A als auch in B enthalten und daher 1/32 abzuziehen ist, um eine Doppelerfassung zu verhindern.

- Zweimaliges Werfen eines Würfels

Betrachten wir die beiden Ereignisse $A_1 = \{(x_1, x_2) : x_1 + x_2 > 10\} = \{(5, 6), (6, 5), (6, 6)\}$ und $A_2 = \{(x_1, x_2) : x_1 + x_2 \leq 10\}$. Für die Wahrscheinlichkeit von Ereignis A_1 gilt

$$P(A_1) = \frac{|A_1|}{|\Omega|} = \frac{3}{6 \cdot 6} = \frac{1}{12}.$$

Da zwischen den Ereignissen A_1 und A_2 die komplementäre Beziehung $A_2 = \bar{A}_1$ besteht, können wir die Wahrscheinlichkeit von A_2 bestimmen als

$$P(A_2) = 1 - P(A_1) = 1 - \frac{1}{12} = \frac{11}{12}.$$

Ist bekannt, dass bestimmte Elementarereignisse wahrscheinlicher sind als andere, liefert der klassische Wahrscheinlichkeitsbegriff meist keine korrekten Aussagen zur Wahrscheinlichkeit von Ereignissen. Auch bei Fragestellungen wie

- Wie groß ist die Wahrscheinlichkeit, dass eine in einer Brauerei abgefüllte Flasche ihre Füllmenge um mehr als 10 ml überschreitet?
- Wie ermittelt eine Versicherungsgesellschaft die Wahrscheinlichkeit dafür, dass ein Versicherungsnehmer (Kfz-Haftpflicht) im nächsten Jahr einen Unfall verursacht und damit eine Versicherungsleistung fällig wird?

stößt die klassische Wahrscheinlichkeitsdefinition an ihre Grenzen, sodass wir uns im Folgenden noch weiteren Definitionen widmen müssen.

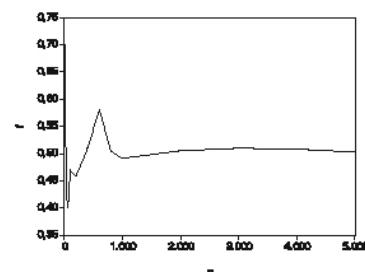
1.3.3 Statistische Wahrscheinlichkeitsdefinition

Häufigkeitstheoretiker wie J. Venn oder R. von Mises sind der Ansicht, dass die einzige zulässige Methode zur Bestimmung von Eintrittswahrscheinlichkeiten ein Experiment sei. Gehen wir zur Begründung dieser Behauptung von einem beliebig oft wiederholbaren Zufallsvorgang aus, d.h. seine Ausgangssituation lasse sich immer wieder von neuem herstellen, sodass jedes Ereignis A bei jeder Durchführung dieselbe Chance des Eintretens hat. Das Werfen einer Münze ist etwa ein solcher Zufallsvorgang, und auch beim Ziehen mit Zurücklegen stellt jede einzelne Entnahme eine derartige Wiederholung dar. Bei n-maliger Durchführung eines solchen wiederholbaren Zufallsvorganges, bei dem die n Durchführungen nach den selben Wahrscheinlichkeitsgesetzen und ohne gegenseitige Beeinflussung ablaufen, ergibt sich eine Anzahl $h_n(A)$ von Durchführungen, bei denen A eingetreten ist. Die relative Häufigkeit $f_n(A)$ des Eintretens von A ergibt sich damit zu

$$f_n(A) = \frac{1}{n} \cdot h_n(A). \quad (\text{III1: 30})$$

Für ein fortlaufendes n, d.h. für eine gegen unendlich strebende Anzahl von Wiederholungen, ist nun zu "erwarten", dass sich die Folge der $f_n(A)$ "nahe P(A) stabilisiert". Anders ausgedrückt, sollten sich $f_n(A)$ und $P(A)$ bei einer hinreichend großen Zahl an Versuchen immer mehr annähern. Ein solches Verhalten wird auch durch die "Erfahrung" bestätigt. So können wir beispielsweise bei sehr häufigem Werfen einer Münze erwarten, dass die Kopfseite ungefähr mit einer relativen Häufigkeit von 0,5 auftritt und finden diese Erwartung nachträglich fast immer bestätigt. In seinem Buch "Introduction to Statistics" beschrieb Noether, G. E. (1971) etwa folgende Versuchsreihe für die Häufigkeit des Ereignisses "Kopf" beim Zufallsexperiment "Werfen einer Münze":

n Versuche	h_n -mal "Kopf"	f_n "Kopf"
10	7	0,700
20	11	0,550
40	17	0,425
60	24	0,400
80	34	0,425
100	47	0,470
200	92	0,460
400	204	0,510
600	348	0,580
800	404	0,505
1.000	492	0,492
2.000	1.010	0,505
3.000	1.530	0,510
4.000	2.043	0,508
5.000	2.515	0,503



Wie unschwer zu erkennen ist, gibt es bei den relativen Häufigkeiten eine Konvergenz gegen den Wert 0,5. Dies ist genau der Wert, den man für das Ereignis "Kopf" auch nach der klassischen Wahrscheinlichkeitsdefinition erwarten würde. Das Wis-

sen um eine "starke" Abweichung davon (trotz zunehmenden n) wäre wohl für jeden Betrachter ein Anlass, die Münze als manipuliert zu betrachten, obgleich selbst das Ereignis, dass bei sehr häufiger Durchführung stets nur "Zahl" auftritt, auch bei Gleichwahrscheinlichkeit von "Kopf" und "Zahl" nicht völlig unmöglich ist.

Der Zusammenhang zwischen Wahrscheinlichkeiten und relativen Häufigkeiten wurde von J. Bernoulli durch das "**Gesetz der großen Zahlen**" präzisiert.⁶ Dieses besagt, dass die Wahrscheinlichkeit dafür, dass $f_n(A)$ von $P(A)$ mindestens um einen beliebigen positiven Betrag ϵ abweicht, mit wachsendem n gegen Null konvergiert:

$$\lim_{n \rightarrow \infty} P(|f_n(A) - P(A)| \geq \epsilon) = 0 \quad (\text{II1: 31})$$

Das "Gesetz der großen Zahlen" kann, wie wir bereits beim besprochenen Münzwurfexperiment gesehen haben, zur "Bestätigung" von Laplace-Wahrscheinlichkeiten herangezogen werden. Es ist insbesondere dazu geeignet, Aufschluss über unbekannte Wahrscheinlichkeiten zu erhalten; darauf beruhen viele Methoden der induktiven Statistik. Dabei nimmt die Aussagekraft, die $f_n(A)$ über $P(A)$ hat, natürlich mit wachsendem n zu. Wir sprechen in diesem Zusammenhang bei dem Wert, bei dem sich die relative Häufigkeit $f_n(A)$ bei wachsender Zahl von Versuchswiederholungen "stabilisiert" von der **statistischen Wahrscheinlichkeit** für das Eintreten eines Ereignisses A. Es ist jedoch klar, dass der statistische Wahrscheinlichkeitsbegriff keine exakten Wahrscheinlichkeiten, sondern nur Näherungswerte dafür liefern kann. In der **Praxis** wird i.d.R. für eine große, endliche Anzahl n von Versuchen die beobachtete relative Häufigkeit als **Näherungs- oder Schätzwert** $\hat{P}(A)$ für die gesuchte Wahrscheinlichkeit $P(A)$ herangezogen:

$$\hat{P}(A) = f_n(A) \quad (\text{II1: 32})$$

Dabei wird gemäß (II1: 31) das Abweichen der relativen Häufigkeit $f_n(A)$ von der Wahrscheinlichkeit $P(A)$ um mehr als den Wert ϵ mit zunehmendem n immer unwahrscheinlicher.

Abschließend sei noch erwähnt, dass sich aufgrund der hier beschriebenen Zusammenhänge die bisher behandelten Regeln für Wahrscheinlichkeiten auch auf relative Häufigkeiten übertragen lassen. So gilt etwa für zwei disjunkte Ereignisse A und B $f_n(A \cup B) = f_n(A) + f_n(B)$, für komplementäre Ereignisse $f_n(\bar{A}) = 1 - f_n(A)$, für Teilereignisse $f_n(A) \leq f_n(B)$, wenn $A \subset B$, usw.

1.3.4 Subjektive Wahrscheinlichkeitsdefinition

In vielen Entscheidungssituationen im Wirtschaftsleben lassen sich Wahrscheinlichkeiten weder durch die klassische noch durch die statistische Wahrscheinlichkeitsdefinition objektiv bestimmen. Die objektiven Wahrscheinlichkeitsbegriffe versagen häufig besonders dann, wenn es darum geht, die Wahrscheinlichkeit **einmaliger Ereignisse** zu bestimmen. Fragen wie "Wie groß war am Tag davor die Wahrscheinlichkeit für einen Reaktorunfall in Tschernobyl?" können wir höchstens da-

⁶ Zur Unterscheidung von einer anderen Form des "Gesetzes der großen Zahlen", das wir im Rahmen der induktiven Statistik noch kennenlernen werden, wird (II1: 31) auch häufig als "Bernoullis Gesetz der großen Zahlen" bezeichnet.

durch beantworten, indem wir die Wahrscheinlichkeit des Zusammentreffens der Ereignisse angeben, die einen Reaktorunfall auslösen können. Aber wie sollen wir die Wahrscheinlichkeit dafür angeben, dass ein bestimmtes Pferd ein Rennen gewinnt oder eine bestimmte Fußballmannschaft deutscher Meister wird? Wir können hier nur auf sog. **subjektive Wahrscheinlichkeiten** zurückgreifen.

Der Begriff der subjektiven Wahrscheinlichkeit wurde vor allem von B. De Finetti und L. J. Savage geprägt. Sie und andere Subjektivisten waren der Meinung, dass alle Bemühungen zur Objektivierung von Wahrscheinlichkeiten zum Scheitern verurteilt sind und letztendlich Wertangaben für Wahrscheinlichkeiten doch nur vernünftige **Glaubensaussagen** sein könnten. Anders ausgedrückt, ist die Wahrscheinlichkeit nach dieser Theorie ein Maß für das persönliche Vertrauen, das ein Individuum in das Eintreten eines bestimmten Ereignisses setzt.

Beispiel II 1-9

Subjektive Wahrscheinlichkeit

Zu Beginn eines Würfelspiels hält ein Spieler den verwendeten Würfel für "fair". Er ordnet dem Ereignis "Eine Sechs wird gewürfelt." deshalb die subjektive Wahrscheinlichkeit 1/6 zu. Nach mehreren Spieldurchläufen könnte er nun aber seine Meinung ändern und aufgrund seiner Beobachtungen zu der Überzeugung kommen, der Würfel sei ungleichmäßig. In diesem Fall würde er demselben Ereignis nun eine entsprechend höhere oder geringere subjektive Wahrscheinlichkeit zuordnen. Ist sich der Spieler absolut sicher, dass beim nächsten Wurf die Sechs fällt, würde er die Wahrscheinlichkeit sogar als Eins ansehen.

Als praktisches Verfahren, die subjektive Wahrscheinlichkeit eines Individuums für das Eintreten eines bestimmten Ereignisses herauszufinden, wollen wir im Folgenden zwei Möglichkeiten vorstellen: Die Auswertung alternativer Risikosituationen und die Analyse von Wettquotienten. Beide Verfahren setzen allerdings *Risikoneutralität* des betrachteten Individuums voraus. Zum besseren Verständnis dieser Voraussetzung betrachten wir zunächst folgendes einfache Gewinnspiel:

Mit einer Wahrscheinlichkeit von jeweils 50 % gewinnt oder verliert man 100 Euro, d.h. im Durchschnitt würde man weder gewinnen noch verlieren. Eine Person, die sich nun im Hinblick auf die Teilnahme an diesem Gewinnspiel als indifferent erweist, ist *riskoneutral*. Wer bereit ist, für die Teilnahme am Spiel einen Einsatz zu zahlen, ist *riskofreudig*. Wer nicht bereit ist, gegen Zahlung eines Einsatzes teilzunehmen, ist *riskoavers*.

Ein erstes Verfahren besteht nun darin, einem Individuum **alternative Risikosituationen** anzubieten. So wären etwa die folgenden denkbar:

Situation A: Man erhält mit einer Wahrscheinlichkeit p den Betrag 100 Euro oder mit einer Wahrscheinlichkeit $1 - p$ den Betrag 0 Euro.

Situation B: Man erhält 100 Euro, wenn der DAX innerhalb des nächsten Monats um 100 Punkte fällt. Andernfalls erhält man 0 Euro.

Die subjektive Wahrscheinlichkeit dafür, dass der DAX im nächsten Monat um 100 Punkte fällt, erhalten wir nun dadurch, dass wir p so lange variieren, bis das Individuum zwischen beiden Risikosituationen indifferent ist.

Eine weitere interessante Möglichkeit der Ableitung subjektiver Wahrscheinlichkeiten, besteht in der Analyse von **Wettquotienten** (Quoten). Betrachten wir dazu zunächst, wie in der Praxis für eine Fußballwette Wettquotienten festgelegt werden:

In einem ersten Schritt legt der Buchmacher zunächst für jedes der möglichen Ereignisse "Sieg", "Unentschieden" und "Niederlage" eine (subjektive) Wahrscheinlichkeit fest, wobei die Summe der drei Wahrscheinlichkeiten natürlich bei Eins liegen muss. Für ein Spiel FC Bayern gegen Schalke 04 könnte sich z.B.

Sieg Bayern:	0,60
Unentschieden:	0,25
Sieg Schalke:	0,15

ergeben. Aus diesen Wahrscheinlichkeiten werden dann sog. *faire Quoten* nach der Formel $Q^F = 1 / P$ (mit Q^F = faire Quote und P = Wahrscheinlichkeit) festgelegt:

Sieg Bayern:	$1 / 0,60 = 1,67$
Unentschieden:	$1 / 0,25 = 4,00$
Sieg Schalke:	$1 / 0,15 = 6,67$

Dies sind jedoch nicht die *realen Quoten*, da der Buchmacher bei den fairen Quoten das gesamte Geld der Verlierer an die Gewinner auszahlen müsste. Die realen Quoten sind daher niedriger. Sie werden durch Multiplikation mit $(1 - b)$ bestimmt, wobei b der Anteil der Einsätze ist, der vom Buchmacher als Gewinn einbehalten wird. Es gilt also $Q^R = Q^F \cdot (1 - b)$. Im Fall $b = 0,10$ erhielten wir für das obige Beispiel folgende reale Quoten:

Sieg Bayern:	$1,67 \cdot 0,90 = 1,5$
Unentschieden:	$4,00 \cdot 0,90 = 3,6$
Sieg Schalke:	$6,67 \cdot 0,90 = 6,0$

Diese Quoten können nun in verschiedenen Formaten angegeben werden:

	EU-Format	UK-Format	US-Format
Sieg Bayern:	1,5	1 / 2	- 200
Unentschieden:	3,6	13 / 5	+ 260
Sieg Schalke:	6,0	5 / 1	+ 500

Beim *EU-Format* gibt die Quote das Verhältnis von möglicher Auszahlung zum Einsatz an. Der auszuzahlende Betrag ergibt sich im Gewinnfall aus der Multiplikation des Wetteinsatzes mit der Quote. Setzen wir 100 Euro erfolgreich auf Bayern, würden wir also hier $100 \text{ Euro} \cdot 1,5 = 150 \text{ Euro}$, d.h. einen Gewinn von 50 Euro, erhalten, den wir auch über $100 \text{ Euro} \cdot (1,5 - 1) = 50 \text{ Euro}$ bestimmen können.

Beim *UK-Format* geben die Quoten das Verhältnis von möglichem Gewinn zum Einsatz an. Den möglichen Gewinn erhalten wir daher durch Multiplikation des Einsatzes mit der Quote. Beim Sieg von Bayern läge der Gewinn in unserem Beispiel also bei $100 \text{ Euro} \cdot 1 / 2 = 50 \text{ Euro}$.

Beim *US-Format* werden die Quoten im Plus- und Minusformat angegeben, wobei die Vorzeichen jedoch keine mathematische Bedeutung haben. Minusquoten zeigen den benötigten Einsatz für einen Gewinn von 100 Einheiten an. Plusquoten den Gewinn bei einem Einsatz von 100 Einheiten. Bei einem Sieg von Bayern er-

halten wir also bei einem Einsatz von 100 Euro einen Gewinn von 100 Euro · $100/200 = 50$ Euro. Bei einem erfolgreichen Tip auf den Sieg von Schalke erhalten wir hingegen einen Gewinn von 500 Euro.

Da wir nun wissen, wie Wettquotienten festgelegt und interpretiert werden, sind wir im Umkehrschluss auch in der Lage von gegebenen Wettquotienten auf die Wahrscheinlichkeiten zu schließen, die das Individuum bei der Aufstellung der Wettquotienten unterstellt hat. Bietet also der Buchmacher in unserem Beispiel eine Quote von 1,5 für den Sieg von Bayern an und wissen wir, dass er stets 10 % der Einsätze einbehält, so können wir die von ihm angenommene Wahrscheinlichkeit als $1 / (1,5 / 0,90) = 0,6$ (d.h. $1/(Q^r/(1 - b)) = 1/Q^f = P$) ermitteln.

Außerdem können wir eine Person hinsichtlich der Quote befragen, zu der sie gerade noch bereit ist, auf das Eintreten eines Ereignisses A zu wetten, und daraus subjektive Wahrscheinlichkeiten ableiten. Wir wollen dies am nachfolgenden Beispiel veranschaulichen.

Beispiel II 1-10	Wettquotienten
------------------	----------------

Jemand ist gerade noch bereit mit der Quote 5 (EU-Format) darauf zu wetten, dass Deutschland im Jahr 2014 Fußballweltmeister wird. Dies bedeutet, dass er pro eingesetzten Euro eine Auszahlung von 5 Euro bzw. einen Gewinn von 4 Euro fordert. Eine zahlenmäßige Vorstellung von der subjektiven Wahrscheinlichkeit der Ereignisse A: "Deutschland wird Weltmeister" und \bar{A} : "Deutschland wird nicht Weltmeister" ergibt sich daraus als Kehrwert der EU-Quote bzw. Komplementärwahrscheinlichkeit zu

$$P(A) = \frac{1}{5} = 0,2 \quad P(\bar{A}) = 1 - P(A) = 0,8 .$$

Da sich für solche Ereignisse wohl keine objektiven Wahrscheinlichkeiten bestimmen lassen, sind diese Werte die einzige sinnvollen Wahrscheinlichkeitsmaße.

Bei Betrachtung von Beispiel II 1-10 wird der kritische Leser feststellen, dass es wohl kein effizientes Verfahren ist, die Wahrscheinlichkeit eines Ereignisses aus der Wettbereitschaft nur einer *einzigsten Person* abzuleiten. Diese hängt nämlich davon ab, welche Informationen die Person nutzt und wie sie diese verarbeitet, und ist daher allein oft wenig aussagekräftig. In der Praxis wird daher häufig versucht subjektive Wahrscheinlichkeiten durch die *mehrmalige Befragung* einer Gruppe von Experten zu objektivieren.

1.4 Zufallsauswahl und Kombinatorik

Im Rahmen des klassischen Wahrscheinlichkeitsbegriffes wurde die Wahrscheinlichkeit eines Ereignisses A als Quotient aus der Anzahl der für A günstigen Elementarereignisse und der Anzahl aller möglichen Elementarereignisse definiert. Bei einfachen Zufallsexperimenten wie etwa dem Würfeln lassen sich diese Fälle auch recht einfach ermitteln. Bei komplexeren Experimenten ist dies jedoch etwas komplizierter. In der Vergangenheit wurden daher eine Reihe von Methoden entwickelt, die bei der Ermittlung dieser Fälle hilfreiche Dienste leisten. Genau mit diesen Verfahren wollen wir uns nun näher befassen.

1.4.1 Zufallsauswahl und Urnenmodell

Unter einer **Zufallsauswahl** verstehen wir allgemein das zufällige Ziehen von n Objekten (Stichprobe) aus einer endlichen Menge von N Objekten (Grundgesamtheit). Typische Beispiele hierfür sind etwa die Befragung von 1.000 Personen aus der deutschen Bevölkerung im Rahmen einer Meinungsumfrage oder die Entnahme von 50 Produkten aus der laufenden Fertigung. Zur Veranschaulichung dieses Auswahlprozesses und seiner verschiedenen Varianten wird in der Wahrscheinlichkeitstheorie das sog. **Urnenmodell** herangezogen. Dabei befinden sich in einer Urne N gleichartige Kugeln, die durchnummieriert sind und von denen zufällig (blind) n Kugeln mit oder ohne anschließendes Zurücklegen herausgegriffen und mit oder ohne Berücksichtigung der Reihenfolge (Anordnung) notiert werden (vgl. Abbildung II 6). Gerade von der Reihenfolge und dem Zurücklegen hängt letztendlich die Anzahl der möglichen Elementarereignisse (Auswählen, Stichproben) des Zufallsexperimentes ab.

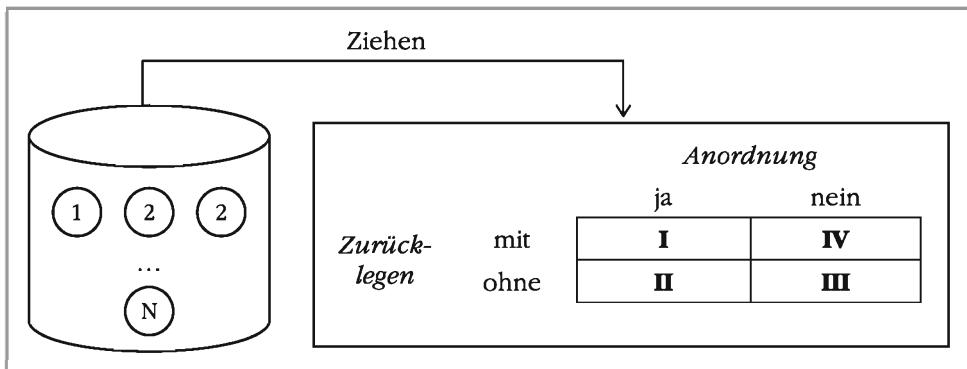


Abbildung II 6: Urnenmodell

Die in Abbildung II 6 auftauchenden römischen Ziffern sind lediglich als Platzhalter zu betrachten, die im Abschnitt II 1.4.2.2 näher behandelt werden.

1.4.2 Kombinatorik

Die Kombinatorik befasst sich allgemein mit der Bildung von Zusammenstellungen von Elementen aus einer vorgegebenen endlichen Menge. Verschiedene kombinatorische Modelle stellen dabei die Anzahl der möglichen Zusammenstellungen fest. Um diese unter II 1.4.2.2 näher erläutern und übersichtlich darstellen zu können, führen wir im Abschnitt II 1.4.2.1 zunächst einige wichtige Begrifflichkeiten ein.

1.4.2.1 n-Fakultät und Binomialkoeffizient

Unter der **n-Fakultät** versteht man das Produkt der natürlichen Zahlen von 1 bis n :

$$n! := 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1) \cdot n \quad (\text{III1: 33})$$

Für sie gelten die folgenden Zusammenhänge:

$$0! := 1 \quad (\text{III1: 34})$$

$$(n+1)! := n! \cdot (n+1) \quad (\text{III1: 35})$$

Beispiel II 1-11

Fakultäten

Bestimmen wir einige einfache Fakultäten:

$$1! = 1$$

$$2! = 1 \cdot 2 = 2$$

$$3! = 1 \cdot 2 \cdot 3 = 6$$

Unter Anwendung von (II1: 35) lässt sich 4-Fakultät wie folgt ermitteln:

$$4! = 3! \cdot 4 = 24$$

Da bereits $10!$ den Wert 3.628.800 liefert, können wir festhalten, dass Fakultäten enorm schnell ansteigen.

Der sog. **Binomialkoeffizient** ("N über n") ist allgemein definiert als

$$\binom{N}{n} := \frac{N!}{n! \cdot (N-n)!} \quad \text{mit } N > 0, n \geq 0, N \geq n, \quad (\text{II1: 36})$$

wobei N und n nicht nur positiv, sondern auch ganzzahlig sein müssen. Für die Sonderfälle $n = 0$, $n = N$, $N = 1$ und $N = N - 1$ gilt für diesen

$$\binom{N}{0} = 1, \quad \binom{N}{N} = 1, \quad \binom{N}{1} = N \quad \text{und} \quad \binom{N}{N-1} = N.$$

Außerdem gilt

$$\binom{N}{n} = \binom{N}{N-n}.^7 \quad (\text{II1: 37})$$

Beispiel II 1-12

Binomialkoeffizient

Berechnen wir den Binomialkoeffizienten "7 über 3":

$$\binom{7}{3} = \frac{7!}{3! \cdot (7-3)!} = \frac{7!}{3! \cdot 4!} = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35$$

Wie wir hier deutlich erkennen können, ist ein Teil der Faktoren des Nenners, nämlich $(N - n)!$, auch im Zähler enthalten, sodass sich dieser Teil weglassen lässt und sich die Berechnung des Binomialkoeffizienten erheblich vereinfacht. Besonders für große N wird daher auch häufig auf die Formel

$$\binom{N}{n} = \frac{N \cdot (N-1) \cdot (N-2) \cdot \dots \cdot (N-n+1)}{n!}$$

zurückgegriffen.

⁷ Weitere wesentliche Eigenschaften von Binomialkoeffizienten werden beschrieben in Auer, B.R., Seitz, F. (2009), Kapitel 1.

Bei Berechnungen mit Binomialkoeffizienten kann sich insbesondere

$$\binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \dots + \binom{N}{N} = 2^N \quad (\text{II1: 38})$$

als nützlich erweisen. Das Ergebnis dieser Formel kann nämlich für den Fall, dass ein Ergebnisraum Ω genau N Elementarereignisse umfasst, als die *Anzahl aller möglichen Teilmengen von Ω* interpretiert werden.

Beispiel II 1-13

Teilmengen

Berechnen wir (II1: 38) für $N = 4$, so erhalten wir

$$\binom{4}{0} + \binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 1 + 4 + 6 + 4 + 1 = 16$$

bzw.

$$2^4 = 16.$$

Enthält also ein Ergebnisraum genau 4 Elementarereignisse, so können wir daraus genau 16 Teilmengen bilden. Nehmen wir etwa an, die Elementarereignisse sind A, B, C und D, so erhalten wir die folgenden möglichen Teilmengen:

1	4	6	4	1
{ }	{A}	{AB}	{ABC}	{ABCD}
	{B}	{AC}	{ABD}	
	{C}	{AD}	{ACD}	
	{D}	{BC}	{BCD}	
		{BD}		
		{CD}		

1.4.2.2 Prinzipien der Kombinatorik

Werden aus einer Grundgesamtheit (z.B. einer Urne) mit N Elementen (z.B. Kugeln), welche von 1 bis N "durchnummieret" sind, n Elemente zufällig entnommen, so hängen die Anzahlen $|\Omega|$ der möglichen Stichprobenergebnisse von der Art der Entnahme ab. Diese Entnahmearten haben wir bereits in Abbildung II 6 mittels der Ziffern I bis IV klassifiziert und wollen sie nun im Detail betrachten:

Situation I

Ziehen *mit* Zurücklegen und *mit* Berücksichtigung der Anordnung

Da im Urnenfall aufgrund des Zurücklegens bei jedem Zug erneut auf N sich in der Urne befindliche Kugeln zugegriffen und die Reihenfolge des Ziehens berücksichtigt wird, ist die Anzahl der Möglichkeiten $|\Omega|$ definiert durch

$$|\Omega^I| = N \cdot N \cdot \dots \cdot N = N^n. \quad (\text{II1: 39})$$

Beispiel II 1-14

Anwendungsbeispiele Situation I

1. Fußballtoto

Beim sog. Fußballtoto (11-er Wette) tippt ein Spieler bei 11 Begegnungen jeweils auf "Sieg der Heimmannschaft", "Sieg des Gastes" oder "Unentschieden". Somit gilt in diesem Fall $N = 3$ und $n = 11$. Damit hat er insgesamt $|\Omega^I| = 3^{11} = 177.147$ Möglichkeiten seinen Tipp-schein auszufüllen.

2. Bits und Bytes

Ein Byte besteht aus 8 Bit, wobei jedes Bit mit 0 oder 1 besetzt werden kann. In diesem Fall ist also $N = 2$ und $n = 8$. Es ergeben sich daraus genau $|\Omega^I| = 2^8 = 256$ verschiedene Byte-Muster.

3. Zahlenschlösser

Wie groß ist die Wahrscheinlichkeit P , dass ein Dieb durch Raten die Kombination eines Zahlenschlosses mit 3 Ringen (jeweils Ziffern von 0 bis 9) herausfindet?

$$|\Omega^I| = 10^3 = 1.000 \rightarrow P = \frac{1}{1.000} = 0,001$$

Liegen k Urnen mit unterschiedlichen Kugelanzahlen N_i ($i = 1, 2, \dots, k$) vor, wobei aus jeder eine Kugel gezogen wird, lässt sich die Anzahl der Möglichkeiten $|\Omega^I|$ als

$$|\Omega^I| = N_1 \cdot N_2 \cdot \dots \cdot N_k \quad (\text{II1: 40})$$

bestimmen.

Beispiel II 1-15

Weiteres Anwendungsbeispiel Situation I

Stellenbesetzung

Ein Unternehmen beabsichtigt jeweils einen neuen Mitarbeiter in den Bereichen Controlling, Finanzen und Personal einzusetzen. Beworben haben sich 3 Controller, 5 Personaler und 2 ehemalige Banker, sodass das Unternehmen $|\Omega^I| = 3 \cdot 5 \cdot 2 = 30$ Möglichkeiten hat die Stellen zu besetzen.

Situation II

Ziehen *ohne* Zurücklegen und *mit* Berücksichtigung der Anordnung

Durch die Tatsache, dass die Kugeln nach dem Ziehen nicht mehr zurückgelegt werden, reduziert sich die Anzahl der Kugeln, die für einen zweiten, dritten, usw. Zug zur Verfügung stehen. Dementsprechend ergibt sich die Anzahl der Möglichkeiten unter Berücksichtigung der Reihenfolge folgendermaßen:

$$|\Omega^{II}| = N \cdot (N - 1) \cdot (N - 2) \cdot \dots \cdot (N - n + 1) = \frac{N!}{(N - n)!} \quad (\text{II1: 41})$$

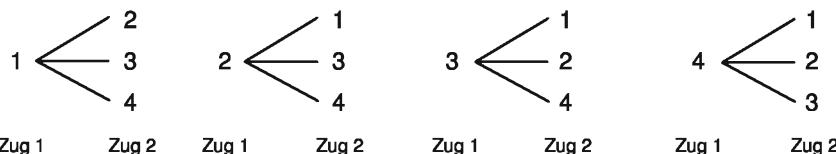
Es ergeben sich also N Möglichkeiten den ersten Platz der Zusammensetzung zu besetzen. Danach bleiben für die Besetzung des zweiten Platzes noch $N-1$ Elemente, für den dritten Platz noch $N-2$ Elemente usw.

Beispiel II 1-16

Anwendungsbeispiele Situation II

1. Urnenspiel

In einer Urne befinden sich 4 Kugeln, die von 1 bis 4 durchnummeriert und daher unterscheidbar sind. Es werden nacheinander bzw. unter Berücksichtigung der Reihenfolge zwei Kugeln entnommen, wobei eine vorher gezogene Kugel nicht mehr in die Urne zurückgelegt wird. Die möglichen Ausgänge dieses Auswahlvorganges können wie folgt dargestellt werden:



Dabei ist die erste Zahl jeweils das Ergebnis des ersten und die zweite das Ergebnis des zweiten Zuges. Es ergeben sich insgesamt $|\Omega^{\text{II}}| = 4 \cdot 3 = 12$ mögliche Ziehungsergebnisse.

2. Pferderennen

Wie viele Tippscheine gibt es, wenn bei einem Pferderennen insgesamt 20 Pferde antreten und die 3 schnellsten Pferde der Reihe nach zu tippen sind?

$$|\Omega^{\text{II}}| = \frac{20!}{(20-3)!} = 20 \cdot 19 \cdot 18 = 6.840$$

Werden alle N Elemente der Grundgesamtheit gezogen, ergeben sich die möglichen Anordnungen der Elemente gemäß

$$|\Omega^{\text{II}}| = N \cdot (N-1) \cdot (N-2) \cdots 3 \cdot 2 \cdot 1 = N! . \quad (\text{III1: 42})$$

Man spricht hier in der Literatur auch häufig von sog. **Permutationen** (Anzahl der unterschiedlichen Anordnungen von N *unterschiedlichen* Objekten).

Beispiel II 1-17

Permutation

Wieviele Möglichkeiten (verschiedene Reihenfolgen) hat ein Bibliothekar 12 verschiedene Bücher nebeneinander ins Regal zu stellen?

$$|\Omega^{\text{II}}| = 12! = 479.001.600$$

In der Permutationsdefinition (III1: 42) wird davon ausgegangen, dass die betrachteten N Elemente alle voneinander verschieden sind. Sind nun aber nicht alle N Elemente unterschiedlich, werden aus ihnen k Gruppen gleicher Elemente gebildet und jeweils die Anzahl darin enthaltener Elemente N_i ($i = 1, 2, \dots, k$) festgestellt. Diese Gruppen reduzieren die Anzahl der unterscheidbaren Permutationen, da sich gleiche Elemente nicht unterscheiden lassen und deswegen bei ihnen auch keine Reihenfolge beim Ziehen festgestellt werden kann. Die Anzahl der unterscheidbaren Permutationen ergibt sich in einem solchen Fall als

$$|\Omega^{\text{II}}| = \frac{N!}{N_1! \cdot N_2! \cdot \dots \cdot N_k!} \quad \text{mit} \quad N_1 + N_2 + \dots + N_k = N . \quad (\text{III1: 43})$$

$N_i \geq 1 \quad \text{für } i = 1, 2, \dots, k$

Beispiel II 1-18**Permutation - Sonderfall**

Wie groß ist die Anzahl der unterschiedlichen Permutationen der $N = 11$ Buchstaben des Wortes "Abrakadabra"?

$$|\Omega^{\text{II}}| = \frac{11!}{5! \cdot 2! \cdot 2! \cdot 1! \cdot 1!} = 83.160$$

Wie aus diesem letzten Beispiel gut abgeleitet werden kann, tauchen für den Fall, dass alle N Elemente verschieden sind, im Nenner von (II1: 43) nur Einsen auf. Es ergäbe sich also dann die Formel (II1: 42).

Situation III

Ziehen *ohne* Zurücklegen und *ohne* Berücksichtigung der Anordnung

Im Gegensatz zu den bereits behandelten Kombinationen in Situation I und II werden bei Kombinationen der Situationen III und IV die Anordnungen außer Acht gelassen. Dies bedeutet, dass beispielsweise "abc" und "bca" gleichwertig sind. Es muss also zwangsläufig weniger Kombinationen ohne Berücksichtigung der Anordnung als Kombinationen mit Berücksichtigung der Anordnung geben.

Werden gezogene Elemente nicht mehr in die Ursprungsmenge N eingelegt und ist bei den Ergebnissen die Reihenfolge nicht von belang, können die möglichen Kombinationen anhand des in Abschnitt II 1.4.2.1 behandelten Binomialkoeffizienten berechnet werden:

$$|\Omega^{\text{III}}| = \binom{N}{n} = \frac{N!}{n! \cdot (N-n)!} \quad (\text{II1: 44})$$

Vergleichen wir dies mit der Situation II bzw. (II1: 41), so erkennen wir den Zusammenhang

$$|\Omega^{\text{III}}| = \frac{|\Omega^{\text{II}}|}{n!} \quad (\text{II1: 45})$$

Es gibt nämlich jeweils $n!$ Möglichkeiten die n gezogenen verschiedenenartigen Elemente anzutragen.

Beispiel II 1-19**Anwendungsbeispiele Situation III****1. Urnenspiel**

Betrachten wir eine Urne mit $N = 3$ Kugeln A, B und C, aus der $n = 2$ Kugeln gezogen werden. Wieviele Kombinationen sind bei Situation III möglich (a) und wie lauten diese (b)?

a) $|\Omega^{\text{III}}| = \binom{3}{2} = \frac{3 \cdot 2}{2 \cdot 1} = 3$ b) AB, AC, BC

2. Sechs Richtiges beim Lotto

Wie groß ist die Chance sechs Richtiges im Lotto zu haben ($n = 6$ aus $N = 49$)? Dabei ist zu beachten, dass beim Lotto die Reihenfolge der gezogenen Zahlen nicht wichtig ist und einmal gezogene Zahlen nicht erneut gezogen werden können. Es gibt insgesamt



$$|\Omega^{\text{III}}| = \binom{49}{6} = 13.983.816$$

mögliche Ziehungsergebnisse. Da die Gewinnkombination aber nur eine dieser vielen Möglichkeiten ist, gilt für die gesuchte Gewinnwahrscheinlichkeit

$$P = \frac{1}{13.983.816} = 0,0000000715.$$

3. Vier Richtige beim Lotto

Wie groß ist die Wahrscheinlichkeit im Lotto mindestens vier Richtige zu haben? Für einen "Vierer" im Lotto benötigen wir 4 von den 6 Richtigen und 2 beliebige von den 43 Falschen. Es gibt damit folgende Anzahl von Viererkombinationen:

$$\binom{6}{4} \cdot \binom{43}{2} = 15 \cdot 903 = 13.545$$

Zudem sind in diesem Fall auch die Fünfer- und Sechserkombinationen zu berücksichtigen (da "mindestens vier Richtige").

$$\text{Fünferkombinationen: } \binom{6}{5} \cdot \binom{43}{1} = 6 \cdot 43 = 258$$

$$\text{Sechserkombinationen: } \binom{6}{6} \cdot \binom{43}{0} = 1 \cdot 43 = 43$$

Es gilt daher für die gesuchte Wahrscheinlichkeit

$$P = \frac{\binom{6}{4} \cdot \binom{43}{2} + \binom{6}{5} \cdot \binom{43}{1} + \binom{6}{6} \cdot \binom{43}{0}}{\binom{49}{6}} = \frac{13.545 + 258 + 43}{13.983.816} = 0,00099.$$

Eine Addition der Wahrscheinlichkeiten ist hier zulässig, da sich die einzelnen Ereignisse gegenseitig ausschließen bzw. keine Elemente gemeinsam haben.

Situation IV

Ziehen *mit* Zurücklegen und *ohne* Berücksichtigung der Anordnung

Beim Ziehen von n aus N Elementen mit Zurücklegen und ohne Berücksichtigung der Reihenfolge gilt für die Anzahl der möglichen Kombinationen

$$|\Omega^{\text{IV}}| = \binom{N+n-1}{n} = \frac{(N+n-1)!}{n! \cdot (N-1)!}. \quad (\text{II1: 46})$$

Beispiel II 1-20

Anwendungsbeispiel Situation IV

In einem Topf befinden sich $N = 3$ Kugeln A, B und C. Es werden $n = 2$ Kugeln ohne Beachtung der Reihenfolge und mit anschließendem Zurücklegen gezogen. Wieviele Kombinationen sind möglich (a) und wie lauten diese (b)?

$$\text{a)} \quad |\Omega^{\text{IV}}| = \binom{3+2-1}{2} = \frac{(3+2-1)!}{2! \cdot (3-1)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2! \cdot 2!} = 6$$

b) AA, BB, CC, AB, AC, BC

Im Vergleich zu Beispiel II 1-19 (Nr. 1) erkennen wir deutlich, dass sich durch das Zurücklegen drei zusätzliche Kombinationen ergeben.

Die im Beispiel II 21 aufgeführten Anwendungsbeispiele stellen mögliche Mischungen der vier beschriebenen Situationen dar. Sie sollen die Vielfalt Kombinatorischer Überlegungen veranschaulichen.

Beispiel II 1-21

Zusammenfassende Anwendungsbeispiele

1. Kfz-Kennzeichen

Wieviele Kfz-Kennzeichen kann die Stadt Weiden unter Berücksichtigung folgender Bedingungen ausgeben?

- Zwei unterschiedliche Buchstaben, wobei die Kombinationen AH, SA, HJ und KZ verboten sind
- Drei unterschiedliche und von Null verschiedene Ziffern

Für den ersten Buchstaben bestehen 26 Möglichkeiten, für den zweiten eine weniger, da die Buchstaben voneinander verschieden sein sollen. Von den damit möglichen Buchstabenkombinationen $26 \cdot 25$ sind schließlich noch die 4 verbotenen zu subtrahieren. Für die erste Ziffer des Kennzeichens bestehen 9 Besetzungsmöglichkeiten (1, ..., 9). Diese sind für die beiden nachfolgenden Ziffern jeweils um eins geringer, da eine bereits verwendete Ziffer nicht noch einmal benutzt werden darf. Für die möglichen Kennzeichen gilt daher

$$|\Omega^1| = (26 \cdot 25 - 4) \cdot 9 \cdot 8 \cdot 7 = 325.584.$$

2. Doppelgeburtstage

Wie groß ist die Wahrscheinlichkeit, dass in einer zufällig zusammengestellten Gruppe von 10 Personen mindestens 2 Personen am selben Tag Geburtstag haben?

Beschreiben wir die Tatsache, dass 2 Personen am selben Tag Geburtstag haben, mit dem Begriff "Doppelgeburtstag", so gilt

$$P(\text{mindestens ein Doppelgeburtstag}) = 1 - P(\text{kein Doppelgeburtstag})$$

mit

$$P(\text{kein Doppelgeburtstag}) = \frac{|\text{kein Doppelgeburtstag}|}{|\Omega|}$$

bzw. konkret

$$P(\text{kein Doppelgeburtstag}) = \frac{365 \cdot 364 \cdot 363 \cdot 362 \cdot 361 \cdot 360 \cdot 359 \cdot 358 \cdot 357 \cdot 356}{365^{10}} = 0,8831.$$

Daraus folgt

$$P(\text{mindestens ein Doppelgeburtstag}) = 1 - 0,8831 = 0,1169.$$

1.4.2.3 Zusammenfassung und Vergleich

Zum besseren Verständnis stellen wir nun einen direkten Vergleich der von uns behandelten Situationen I bis IV an. Wir gehen dabei von einer Urne mit drei Kugeln 1, 2 und 3 aus, aus der wir zwei Kugeln zufällig entnehmen. Betrachten wir zunächst das *Ziehen ohne Zurücklegen*:

Mit Berücksichtigung der Anordnung erhalten wir aus dem Urnenspiel die folgenden 6 möglichen Ergebnisse (Wahrscheinlichkeiten in Klammern):

1-2	(1/6)	2-1	(1/6)
1-3	(1/6)	3-1	(1/6)
2-3	(1/6)	3-2	(1/6)

Die Wahrscheinlichkeit für jedes Elementarereignis beträgt also 1/6, sodass eine *Gleichwahrscheinlichkeit* selbiger gegeben ist.

Ohne Berücksichtigung der Anordnung reduziert sich die Anzahl der Elementarereignisse auf 3, da die oben in einer Zeile dargestellten Ergebnisse jeweils als gleichwertig betrachtet werden:

1-2	(1/3)
1-3	(1/3)
2-3	(1/3)

Wir erkennen, dass sich beim Ziehen ohne Zurücklegen durch Vernachlässigung der Reihenfolge lediglich die Anzahl der Elementarereignisse und damit auch ihre Wahrscheinlichkeit halbiert haben. Ihre *Gleichwahrscheinlichkeit* ist jedoch weiterhin gegeben.

Ziehen wir mit Zurücklegen unter Berücksichtigung der Reihenfolge, erhalten wir folgende 9 mögliche Zweierkombinationen:

1-1	(1/9)	2-1	(1/9)	3-1	(1/9)
1-2	(1/9)	2-2	(1/9)	3-2	(1/9)
1-3	(1/9)	2-3	(1/9)	3-3	(1/9)

Auch hier ist eine *Gleichwahrscheinlichkeit* der Elementarereignisse gegeben, die stets bei 1/9 liegt.

Vernachlässigen wir nun beim *Ziehen mit Zurücklegen* hingegen die *Reihenfolge*, so reduziert sich die Zahl der Elementarereignisse zunächst auf 6, d.h. es liegen konkret die Ausprägungen

1-1	(1/9)
1-2	(2/9)
1-3	(2/9)

2-2	(1/9)
2-3	(2/9)
3-3	(1/9)

vor. Eine *Gleichwahrscheinlichkeit* der Elementarereignisse ist nun aber **nicht mehr gegeben**. Während Kombinationen gleicher Zahlen eine Wahrscheinlichkeit von 1/9 besitzen, weisen verschiedenartige Kombinationen eine Wahrscheinlichkeit von 2/9 auf.

In der Statistik sind daher vor allem das Ziehen *ohne* Zurücklegen und *ohne* Berücksichtigung der Anordnung (Situation III) und das Ziehen *mit* Zurücklegen und *mit* Berücksichtigung der Reihenfolge (Situation I) und manchmal auch Situation II von Bedeutung. Der Fall IV ist wenig relevant, da hier eine Gleichwahrscheinlichkeit der Elementarereignisse nicht gegeben ist und deswegen die Laplace-Formel nicht angewendet werden kann.

Die folgende Abbildung soll zum Abschluss die wichtigsten Zusammenhänge der Kombinatorik noch einmal übersichtlich darstellen.

		Berücksichtigung der Reihenfolge	
		ja	nein
		I	IV
Zurücklegen	ja	N^n	$\binom{N+n-1}{n} = \frac{(N+n-1)!}{n! \cdot (N-1)!}$
	nein	$\binom{N}{n} \cdot n! = \frac{N!}{(N-n)!}$	$\binom{N}{n} = \frac{N!}{n! \cdot (N-n)!}$
		II	III

Abbildung II 7: Überblick Kombinatorik

1.5 Bedingte Wahrscheinlichkeiten

1.5.1 Definition und Interpretation

Oftmals hängt bei einem Zufallsexperiment die Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses A vom Eintreten eines anderen Ereignisses B ab. Die Wahrscheinlichkeit für A unter der Voraussetzung (Annahme, Bedingung oder sogar mit dem Wissen), dass ein anderes Ereignis B bereits eingetreten ist oder gleichzeitig mit A eintritt, wird als **bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B** bezeichnet. Sie ist formal definiert als

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad \text{mit } P(B) > 0. \quad (\text{II1: 47})$$

Analog ergibt sich die *bedingte Wahrscheinlichkeit des Ereignisses B unter der Bedingung A* als

$$P(B | A) = \frac{P(A \cap B)}{P(A)} \quad \text{mit } P(A) > 0. \quad (\text{II1: 48})$$

Bedingte Wahrscheinlichkeiten werden häufig von unbedingten verschieden sein, da sie zusätzliche Informationen über den Zufallsvorgang nutzen.

Beispiel II 1-22

Bedingte Wahrscheinlichkeiten

Für das Zufallsexperiment "Zweimaliges Werfen eines Würfels" wollen wir die folgenden beiden Ereignisse

A: "Augensumme ≤ 4 "

B: "erster Wurf = 1"

näher betrachten. Grafisch lassen sich diese folgendermaßen veranschaulichen:

6						
5						
4						
3	A					
2	A	A				
1	AB	AB	AB	B	B	B
	1	2	3	4	5	6
	Zweiter Wurf					
Erster Wurf						

Es ergibt sich daraus

$$P(A) = \frac{6}{36}, \quad P(B) = \frac{6}{36} \quad \text{und} \quad P(A \cap B) = \frac{3}{36}.$$

Die Wahrscheinlichkeit dafür, dass die Augensumme höchstens 4 ist, wobei der erste Wurf eine 1 ergab, ergibt sich nun nach (II1: 47) als

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{3}{36}}{\frac{6}{36}} = \frac{3}{6} = \frac{1}{2}$$

Es gilt also hier, dass die unbedingte Wahrscheinlichkeit kleiner als die bedingte ist.

Wie dieses Beispiel sehr schön zeigt, führen **Bedingungen** stets zu einer **Reduzierung des relevanten Ergebnisraumes**. Eine Bedingung stellt nämlich eine Beschränkung auf Ereignisse dar, die auftreten, während ein anderes Ereignis vorher, gleichzeitig oder hypothetisch aufgetreten ist. Aufgrund der Bedingung übernimmt jetzt eine Teilmenge der Elementarereignisse die Rolle des sicheren Ereignisses.

Bedingte Wahrscheinlichkeiten gehorchen ebenso den Axiomen von Kolmogorov. Es gilt für alle $B \subset \Omega$ also $P(A | B) \geq 0$ für jedes mögliche $A \subset \Omega$, $P(\Omega | B) = 1$ und falls $A_1 \cap A_2 = \emptyset$ auch $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$. Zudem lassen sich die unter II 1.3.1 definierten Sätze auf bedingte Wahrscheinlichkeiten übertragen.

1.5.2 Multiplikationssatz

Der **Multiplikationssatz** ergibt sich durch einfache Umformung der Definition der bedingten Wahrscheinlichkeit. Er besagt, dass die **Wahrscheinlichkeit dafür, dass zwei Ereignisse A und B gleichzeitig eintreten**, bei $P(A) > 0$ als

$$P(A \cap B) = P(A) \cdot P(B | A) \tag{II1: 49}$$

ermittelt werden kann. Es gilt natürlich bei $P(B) > 0$ auch

$$P(A \cap B) = P(B) \cdot P(A | B). \tag{II1: 50}$$

Beispiel II 1-23

Multiplikationssatz

In einer Urne befinden sich 2 rote und 3 schwarze Kugeln. Es werden nacheinander ohne Zurücklegen zwei Kugeln gezogen. Betrachten wir in diesem Zusammenhang die folgenden beiden Ereignisse:

A: "Die erste gezogene Kugel ist rot."

B: "Die zweite gezogene Kugel ist rot."



Die Ergebnisse bzw. der Ablauf dieses zweistufigen Zufallsexperimentes lassen sich in einem sog. Baumdiagramm (vgl. nachfolgende Grafik) anschaulich darstellen. Die Wahrscheinlichkeiten der auf jeder Stufe möglichen Ereignisse werden darin an den entsprechenden Ästen abgetragen. Auf der ersten Verzweigung finden sich so die Wahrscheinlichkeit von A und die seines Komplementärereignisses:

$$P(A) = \frac{2}{5} \rightarrow P(\bar{A}) = 1 - P(A) = 1 - \frac{2}{5} = \frac{3}{5}$$

Die Wahrscheinlichkeit des Ereignisses B hängt nun vom Ausgang des ersten Zuges ab und nimmt daher, wie wir im Baumdiagramm erkennen können, bedingte Werte an. Die bedingte Wahrscheinlichkeit $P(B | A)$ beschreibt z.B. die Wahrscheinlichkeit, beim zweiten Zug eine rote Kugel zu erwischen, wenn bereits beim ersten Zug eine rote Kugel entnommen wurde. Da sowohl die Zahl der roten Kugeln und damit auch die Gesamtkugelzahl in der Urne durch das Ziehen einer roten Kugel beim ersten Zug um Eins abnimmt (Ziehen ohne Zurücklegen), gilt

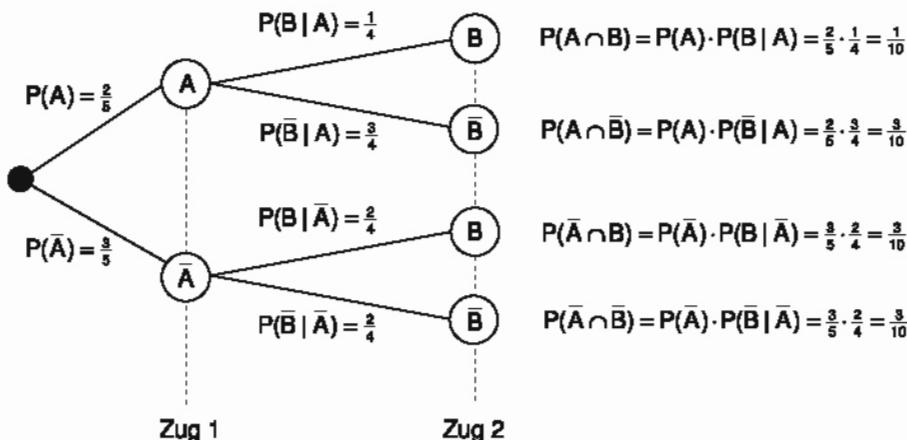
$$P(B | A) = \frac{2-1}{5-1} = \frac{1}{4}.$$

Analoges gilt für die weiteren bedingten Wahrscheinlichkeiten im Diagramm.

Mittels dieser Daten lässt sich nun z.B. die Wahrscheinlichkeit, zweimal hintereinander eine rote Kugel zu ziehen, folgendermaßen angeben:

$$P(A \cap B) = P(A) \cdot P(B|A) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$$

Diese nach dem Multiplikationssatz entstehenden Wahrscheinlichkeiten geben die Wahrscheinlichkeiten der vier Elementarereignisse dieses Zufallsexperimentes an und werden an den Endpunkten des Baumdiagrammes abgetragen:



Ob die Berechnung der Werte des Baumdiagrammes korrekt durchgeführt wurde, kann ganz einfach geprüft werden. Es muss nämlich

$$P(A \cap B) + P(A \cap \bar{B}) + P(\bar{A} \cap B) + P(\bar{A} \cap \bar{B}) = 1$$

gelten.

Für den Fall von *drei Ereignissen A, B und C* kann der Multiplikationssatz folgendermaßen dargestellt werden:

$$P(C \cap B \cap A) = P(C | B \cap A) \cdot P(B | A) \cdot P(A) \quad (\text{III1: 51})$$

Diesen Zusammenhang können wir recht einfach beweisen: Bezeichnen wir zur einfacheren Darstellung zunächst den Durchschnitt $B \cap A$ mit D, d.h.

$$D = B \cap A .$$

Es ergibt sich damit

$$P(C \cap B \cap A) = P(C \cap D) = P(C | D) \cdot P(D) .$$

Ersetzen wir nun D wieder durch den Durchschnitt $B \cap A$, erhalten wir

$$P(C \cap B \cap A) = P(C | B \cap A) \cdot P(B \cap A) .$$

Für $P(B \cap A)$ kann dann schließlich das Ergebnis des Multiplikationssatzes für zwei Ereignisse A und B eingesetzt werden:

$$P(C \cap B \cap A) = P(C | B \cap A) \cdot P(B | A) \cdot P(A)$$

Beispiel II 1-24

Multiplikationssatz für drei Ereignisse

Ein Fertigungslos umfasst genau 100 Bauteile. Davon sind in der Regel 10 % fehlerhaft. Wie groß ist nun die Wahrscheinlichkeit, beim Ziehen (ohne Zurücklegen) einer Stichprobe von 3 Stück genau 3 einwandfreie Bauteile zu erhalten?

Zur Beantwortung dieser Frage definieren wir die folgenden drei Ereignisse:

A: "Erstes Bauteil fehlerfrei."

B: "Zweites Bauteil fehlerfrei."

C: "Drittes Bauteil fehlerfrei."

Es gilt für die Wahrscheinlichkeit beim ersten Zug ein einwandfreies Bauteil zu erhalten

$$P(A) = \frac{90}{100} .$$

Die Wahrscheinlichkeit beim zweiten Zug ein einwandfreies Bauteil zu erhalten, unter der Bedingung, dass bereits beim ersten Zug ein fehlerfreies Stück entnommen wurde, ergibt sich als

$$P(B | A) = \frac{89}{99} ,$$

da durch die Entnahme sowohl die Gesamtzahl der Bauteile als auch die Anzahl der fehlerfreien Stücke um Eins abnimmt. Entsprechend ergibt sich die bedingte Wahrscheinlichkeit des Ereignisses C unter der Bedingung, dass vorher die Ereignisse A und B eingetreten sind als

$$P(C | A \cap B) = \frac{88}{98} .$$

Die Wahrscheinlichkeit, dass alle drei Bauteile in Ordnung sind, ergibt sich damit als

$$P(C \cap B \cap A) = P(C | B \cap A) \cdot P(B | A) \cdot P(A) = \frac{88}{98} \cdot \frac{89}{99} \cdot \frac{90}{100} = \frac{178}{245} = 0,7265 .$$

Alternativ könnte hier auch mit der Laplace-Formel gerechnet werden. Sie liefert

$$P(C \cap B \cap A) = \frac{\text{Anzahl der günstigen Elementarereignisse}}{\text{Anzahl der möglichen Elementarereignisse}} = \frac{\binom{90}{3}}{\binom{100}{3}} = 0,7265.$$

Zusatz:

Wie groß wäre die Wahrscheinlichkeit im obigen Beispiel beim Ziehen mit Zurücklegen?

$$P(C \cap B \cap A) = \frac{90}{100} \cdot \frac{90}{100} \cdot \frac{90}{100} = \frac{90^3}{100^3} = 0,9^3 = 0,729$$

Für eine beliebige Anzahl m von Ereignissen $A_1, A_2, \dots, A_m \subset \Omega$ eines Zufallsvorganges gilt der allgemeine Multiplikationssatz

$$P(A_1 \cap \dots \cap A_m) = P(A_1) \cdot P(A_2 | A_1) \cdot \dots \cdot P(A_m | A_1 \cap \dots \cap A_{m-1}), \quad (\text{II1: 52})$$

wobei $P(A_1 \cap \dots \cap A_{m-1}) > 0$ gelten muss.

1.5.3 Unabhängigkeit von Ereignissen

Zwei Ereignisse A und B eines Zufallsvorganges werden als unabhängig bezeichnet, wenn das Eintreten des einen keine Informationen über die Wahrscheinlichkeiten des Eintretens des anderen liefert bzw. die Wahrscheinlichkeit des Eintretens des einen vollkommen unbeeinflusst davon ist, ob das Ereignis B eingetreten ist oder nicht. Man spricht hier auch häufig von **stochastischer Unabhängigkeit**.

Ist das Ereignis A von Ereignis B unabhängig, so gilt (sofern $P(A) > 0$)

$$P(A | B) = P(A | \bar{B}) = P(A). \quad (\text{II1: 53})$$

Damit ist natürlich auch das Ereignis B von A unabhängig und es gilt (bei $P(B) > 0$)

$$P(B | A) = P(B | \bar{A}) = P(B). \quad (\text{II1: 54})$$

Aus diesen Zusammenhängen und dem bereits unter 1.5.2 behandelten Multiplikationssatz für zwei Ereignisse A und B lässt sich durch einfaches Einsetzen der sog. **Multiplikationssatz für unabhängige Ereignisse** ableiten. Dieser ergibt sich als

$$P(A \cap B) = P(A) \cdot P(B). \quad (\text{II1: 55})$$

Dieser Satz ist in zweierlei Hinsicht hilfreich. Er kann zum einen dazu verwendet werden Wahrscheinlichkeiten zu bestimmen, wenn vom Sachverhalt her klar ist, dass zwei Ereignisse unabhängig sind. Zum anderen kann er dazu angewendet werden, zu prüfen, ob zwei Ereignisse A und B bei gegebenen Werten von $P(A)$, $P(B)$ und $P(A \cap B)$ unabhängig sind.

In diesem Zusammenhang wird in der Literatur häufig die sog. **Vierfeldertafel** aufgeführt, mit der sich für zweistufige Zufallsexperimente Wahrscheinlichkeiten anschaulich beschreiben bzw. darstellen lassen. Das typische Aussehen einer solchen Tafel ist in Abbildung II 8 dargestellt. Ihre Anwendung wird in Beispiel II 1-25 veranschaulicht.

	Ereignis B	Ereignis \bar{B}	Σ
Ereignis A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A)$
Ereignis \bar{A}	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A})$
Σ	$P(B)$	$P(\bar{B})$	$P(\Omega) = 1$

Abbildung II 8: Vierfeldertafel

Beispiel II 1-25**Unabhängigkeitsprüfung**

Von den Kunden eines Supermarktes ist bekannt, ob sie mit der Parkplatzsituation (P) und dem Service (S) zufrieden sind. 22 % der Kunden sind mit der Parkplatzsituation und 50 % mit dem Service unzufrieden. Insgesamt 48 % der Kunden waren sowohl mit dem Service als auch mit der Parkplatzsituation zufrieden. Es ergibt sich daraus die Vierfeldertafel:

	S	\bar{S}	Σ
P	0,48	0,30	0,78
\bar{P}	0,02	0,20	0,22
Σ	0,50	0,50	1,00

Anhand dieses Zahlenmaterials wollen wir nun prüfen, ob die Ereignisse S und P voneinander unabhängig sind. Im Unabhängigkeitsfall müsste $P(S \cap P) = P(S) \cdot P(P)$ gelten. Für die vorliegenden Befragungsergebnisse ergibt sich nun aber $P(S) \cdot P(P) = 0,50 \cdot 0,78 = 0,39$. Da dieser Wert nicht mit 0,48 übereinstimmt, sind die beiden Ereignisse S und P *nicht stochastisch unabhängig* bzw. *stochastisch abhängig*.

Beispiel II 1-26**Wahrscheinlichkeitsermittlung**

Wie groß ist beim zweimaligen Werfen einer idealen Münze die Wahrscheinlichkeit, zweimal "Kopf" zu erhalten?

Bei jeder einzelnen Durchführung dieses Zufallsexperimentes ist die Wahrscheinlichkeit "Kopf" zu werfen gleich 0,50. Die beiden Würfe sind eindeutig voneinander unabhängig. Für die beiden Ereignisse

A: "Der erste Wurf zeigt Kopf"

B: "Der zweite Wurf zeigt Kopf"

liefert der Multiplikationssatz für unabhängige Ereignisse daher

$$P(A \cap B) = P(A) \cdot P(B) = 0,50 \cdot 0,50 = 0,25.$$

Für den Fall von *drei unabhängigen Ereignissen A, B und C* stellt sich der Multiplikationssatz für unabhängige Ereignisse wie folgt dar:

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) \quad (\text{II1: 56})$$

Die Ereignisse A, B und C sind also nur dann voneinander unabhängig, wenn

$$\begin{aligned} P(A) &= P(A | B) = P(A | C) = P(A | B \cap C) \\ P(B) &= P(B | A) = P(B | C) = P(B | A \cap C) . \\ P(C) &= P(C | A) = P(C | B) = P(C | A \cap B) \end{aligned} \quad (\text{II1: 57})$$

Beispiel II 1-27

Drei unabhängige Ereignisse

Wie groß ist die Wahrscheinlichkeit beim dreimaligen Würfeln dreimal hintereinander die "6" zu erhalten?

Für die im Falle eines fairen Würfels eindeutig unabhängigen Ereignisse

A: "Erster Wurf = 6"

B: "Zweiter Wurf = 6"

C: "Dritter Wurf = 6"

gilt

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216} .$$

Für eine beliebige Anzahl m unabhängiger Ereignisse A_i gilt analog

$$P(A_1 \cap \dots \cap A_m) = P(A_1) \cdot \dots \cdot P(A_m) . \quad (\text{II1: 58})$$

Beispiel II 1-28

m unabhängige Ereignisse

Wie groß ist die Wahrscheinlichkeit, dass beim europäischen Roulette (Glücksrad mit 18 roten, 18 schwarzen und einem grünen Feld) 10-mal in Folge "rot" erscheint?

Bezeichnen wir mit A_i das Ereignis, dass das Rouletterad beim i-ten Drehen "rot" zeigt, dann gilt für die gesuchte Wahrscheinlichkeit

$$P(A_1 \cap \dots \cap A_{10}) = P(A_1) \cdot \dots \cdot P(A_{10}) = \left(\frac{18}{37}\right)^{10} = 0,00074.$$

In Abschnitt 1.5.2 dieses Kapitels wurde im Zusatz zu Beispiel II 1-24 eine Stichprobenentnahme mit Zurücklegen behandelt. Nur beim **Ziehen mit Zurücklegen** erfolgen die Entnahmen **unabhängig** voneinander. Deswegen konnten wir im Beispiel einfach $0,9 \cdot 0,9 \cdot 0,9$ rechnen und können bei derartigen Stichprobenentnahmen stets den Multiplikationssatz für unabhängige Ereignisse heranziehen.

Abschließend ist noch zu erwähnen, dass unabhängige Ereignisse nicht mit disjunkten Ereignissen verwechselt werden dürfen. Für zwei disjunkte Ereignisse A und B gilt nämlich immer $P(A \cap B) = 0$. Zwei disjunkte Ereignisse sind also nur dann un-

abhängig, wenn mindestens eines der Ereignisse die Wahrscheinlichkeit Null besitzt. *Disjunkte Ereignisse mit positiver Wahrscheinlichkeit sind in höchstem Maße abhängige Ereignisse*, denn wenn eines von ihnen eintritt, dann kann das andere gar nicht mehr eintreten.

1.5.4 Satz der totalen Wahrscheinlichkeit

Hinter dem Begriff der totalen Wahrscheinlichkeit verbirgt sich nichts anderes als eine einfache Wahrscheinlichkeit. Der Zusatz "total" dient nur als Abgrenzung zur bedingten Wahrscheinlichkeit, die sich nur auf eine Teilmenge von Ω bezieht.

Die einfache bzw. **totale Wahrscheinlichkeit** eines Ereignisses B können wir berechnen, wenn wir seine bedingten Wahrscheinlichkeiten und die Wahrscheinlichkeiten der Bedingungen kennen. Sie ergibt sich dann als

$$P(B) = P(B | A) \cdot P(A) + P(B | \bar{A}) \cdot P(\bar{A}). \quad (\text{III1: 59})$$

Den Beweis dieses Zusammenhangs wollen wir im Folgenden unter Nutzung von Abbildung II 9 erbringen. Der Ergebnisraum Ω teilt sich hier auf in die disjunkten Ereignisse A und \bar{A} . Das Ereignis B setzt sich zusammen aus den disjunkten Ereignissen $B \cap A$ und $B \cap \bar{A}$. Nach dem Multiplikationssatz ist die Wahrscheinlichkeit dieser Ereignisse durch

$$P(B \cap A) = P(B | A) \cdot P(A)$$

und

$$P(B \cap \bar{A}) = P(B | \bar{A}) \cdot P(\bar{A})$$

definiert. Nach Axiom 3 folgt dann schließlich

$$P(B) = P(B | A) \cdot P(A) + P(B | \bar{A}) \cdot P(\bar{A}).$$

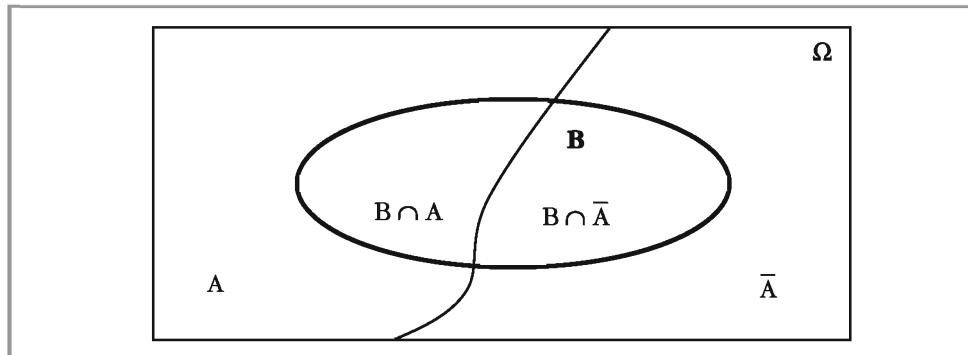


Abbildung II 9: Herleitung der totalen Wahrscheinlichkeit I

Beispiel II 1-29

Satz der totalen Wahrscheinlichkeit

An einer Fachhochschule lesen zwei Professoren die Fächer Mathematik und Statistik. Aus Erfahrungswerten ergaben sich folgende Wahrscheinlichkeiten (M: "Mathematikklausur bestanden", S: "Statistikklausur bestanden"):

$$P(M) = 0,65, \quad P(S | M) = 0,75, \quad P(S | \bar{M}) = 0,30$$

Wie hoch ist nun die Wahrscheinlichkeit die Statistikklausur zu bestehen?

$$\begin{aligned} P(S) &= P(S | M) \cdot P(M) + P(S | \bar{M}) \cdot P(\bar{M}) \\ &= 0,75 \cdot 0,65 + 0,30 \cdot (1 - 0,65) \\ &= 0,5925 \end{aligned}$$

Analog zur obigen Herleitung kann die totale Wahrscheinlichkeit auch ermittelt werden, wenn eine *Aufteilung des Ergebnisraumes in mehr als zwei Teile* vorliegt. Besteht eine vollständige Zerlegung des Ergebnisraumes in m Ereignisse A_1, A_2, \dots, A_m ,

- die sich gegenseitig ausschließen ($A_i \cap A_j = \emptyset$ für alle $j, i = 1, 2, \dots, m$ mit $i \neq j$),
- aber zusammen den Ergebnisraum ganz auffüllen ($A_1 \cup A_2 \dots \cup A_m = \Omega$),

und ist zusätzlich $P(A_j) > 0$ für alle $j = 1, 2, \dots, m$, so gilt für ein jedes Ereignis $B \subset \Omega$ folgender allgemeiner Satz der totalen Wahrscheinlichkeit:

$$P(B) = \sum_{j=1}^m P(B | A_j) \cdot P(A_j) \quad (\text{II1: 60})$$

Zur Veranschaulichung dieses Zusammenhangs betrachten wir auch hier zunächst eine einfache grafische Darstellung, und zwar Abbildung II 10. Wie wir erkennen, ergibt sich das Ereignis B als Vereinigung der disjunkten Ereignisse $B \cap A_1, B \cap A_2, \dots, B \cap A_m$. Nach dem Multiplikationssatz (II1: 52) gilt für die Wahrscheinlichkeiten dieser Ereignisse aber gerade $P(B \cap A_j) = P(B | A_j) \cdot P(A_j)$, sodass ihre Summe den allgemeinen Satz der totalen Wahrscheinlichkeit in (II1: 60) ergibt.

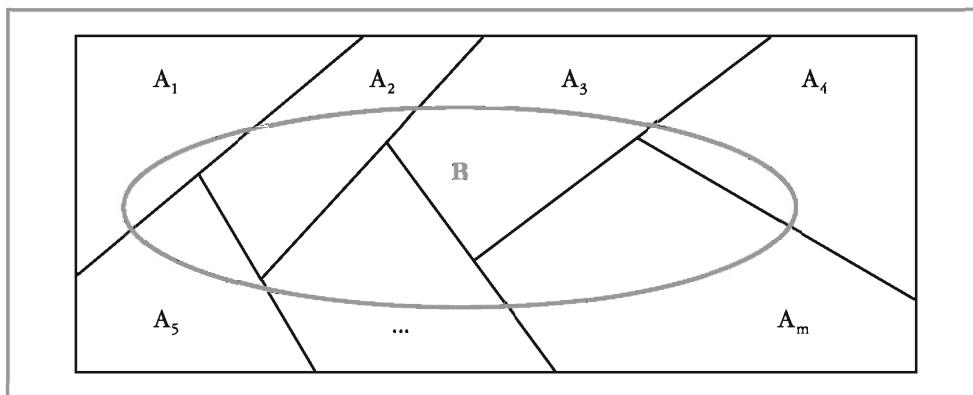


Abbildung II 10: Herleitung der totalen Wahrscheinlichkeit II

Beispiel II 1-30

Verallgemeinerter Satz der totalen Wahrscheinlichkeit

Ein Unternehmen stellt täglich 1.000 Konsumartikel auf drei Maschinen unterschiedlicher Kapazitäten und Ausschusswahrscheinlichkeiten her. Die Daten je Maschine sind in folgender Tabelle enthalten:

Maschine	M_1	M_2	M_3
Tagesproduktion	200 St.	100 St.	700 St.
Ausschusswahrscheinlichkeit	2 %	4 %	1 %

Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Stück aus der Tagesproduktion fehlerhaft ist?

Bezeichnen wir mit A_i ($i = 1, 2, 3$) das Ereignis, dass ein ausgewähltes Stück auf der Maschine M_i hergestellt wurde und mit B , dass es fehlerhaft ist, so lassen sich folgende Wahrscheinlichkeiten angeben:

$$P(A_1) = \frac{200}{1.000} = 0,20 \quad P(B | A_1) = 0,02$$

$$P(A_2) = \frac{100}{1.000} = 0,10 \quad P(B | A_2) = 0,04$$

$$P(A_3) = \frac{700}{1.000} = 0,70 \quad P(B | A_3) = 0,01$$

Nach dem Satz der totalen Wahrscheinlichkeit kann nun die gesuchte Wahrscheinlichkeit $P(B)$ damit wie folgt ermittelt werden:

$$\begin{aligned} P(B) &= P(B | A_1) \cdot P(A_1) + P(B | A_2) \cdot P(A_2) + P(B | A_3) \cdot P(A_3) \\ &= 0,02 \cdot 0,20 + 0,04 \cdot 0,10 + 0,01 \cdot 0,70 \\ &= 0,015 \end{aligned}$$

1.5.5 Formel von Bayes

Die sog. **Formel von Bayes** bildet die Brücke zwischen zwei bedingten Wahrscheinlichkeiten, bei denen die Rollen von Bedingung und Ereignis vertauscht sind, also beispielsweise zwischen $P(A | B)$ und $P(B | A)$. Der bereits behandelte Multiplikationssatz für zwei Ereignisse A und B stellt einen ersten Schritt zur Herleitung dieser Formel dar. Er besagte folgendes:

$$P(A \cap B) = P(A | B) \cdot P(B)$$

$$P(A \cap B) = P(B | A) \cdot P(A)$$

Setzen wir die beiden rechten Seiten dieser Gleichungen gleich und nehmen anschließend eine einfache Umformung vor, so ergibt sich

$$P(A | B) \cdot P(B) = P(B | A) \cdot P(A)$$

bzw.

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}. \quad (\text{II1: 61})$$

Nach dem Satz der totalen Wahrscheinlichkeit können wir für den Fall, dass sich der Ergebnisraum Ω aus nur zwei Ereignissen A und \bar{A} zusammensetzt, für $P(B)$

$$P(B) = P(B | A) \cdot P(A) + P(B | \bar{A}) \cdot P(\bar{A})$$

schreiben und in (II1: 61) einsetzen. Das Resultat davon ist die **Formel von Bayes**, bei der aber zu beachten ist, dass neben $P(A) > 0$ und $P(\bar{A}) > 0$ auch $P(B) > 0$ gelten muss:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B | A) \cdot P(A) + P(B | \bar{A}) \cdot P(\bar{A})} \quad (\text{II1: 62})$$

Analog gilt natürlich auch

$$P(\bar{A} | B) = \frac{P(B | \bar{A}) \cdot P(\bar{A})}{P(B | A) \cdot P(A) + P(B | \bar{A}) \cdot P(\bar{A})}. \quad (\text{II1: 63})$$

Bei Berechnungen mit der Formel von Bayes sind oft die komplementären Beziehungen $P(A | B) + P(\bar{A} | B) = 1$ und $P(A | \bar{B}) + P(\bar{A} | \bar{B}) = 1$ wertvoll. Es gilt jedoch *nicht* $P(A | B) + P(A | \bar{B}) = 1$ oder $P(\bar{A} | B) + P(\bar{A} | \bar{B}) = 1$.

Die Formel von Bayes ist besonders bei einer Situation interessant, bei der zwei sich gegenseitig ausschließende Ereignisse A und \bar{A} als möglich erachtet werden und über die Chancen der einzelnen Zustände (nämlich "wahr" zu sein) Vermutungen bestehen, welche sich in Gestalt von Wahrscheinlichkeiten beschreiben lassen. Die Werte $P(A)$ und $P(\bar{A})$ bezeichnen wir in diesem Zusammenhang auch als **a-priori-Wahrscheinlichkeiten** der Ereignisse A und \bar{A} . Wird nun ein Zufallsvorgang realisiert, wobei für mögliche Ereignisse B dieses Zufallsvorganges die bedingten Wahrscheinlichkeiten $P(B | A)$ bzw. $P(B | \bar{A})$ bestimmbar sind, so können für das tatsächlich eintretende Ereignis B gemäß der Formel von Bayes die bedingten Wahrscheinlichkeiten $P(A | B)$ und $P(\bar{A} | B)$ berechnet werden. Diese werden auch als **a-posteriori-Wahrscheinlichkeit** der Ereignisse A und \bar{A} bezeichnet und als "Verbesserung" gegenüber der a-priori-Wahrscheinlichkeit interpretiert, da sie auf mehr Informationen über das Eintreten von A und \bar{A} fußen (gegeben durch das Eintreten oder Nichteintreten von B) als $P(A)$ und $P(\bar{A})$. Wir können also sagen, dass die Formel von Bayes der *Informationsverarbeitung* dient.

Beispiel II 1-31

Formel von Bayes

Bei der Herstellung eines Produktes ergibt sich aus Erfahrungswerten ein Anteil einwandfreier Teile von $P(E) = 0,97$. Nachdem die Erzeugnisse die Fertigungsstraße durchlaufen haben, werden sie alle auf Funktionsfähigkeit getestet, wobei das dabei verwendete Testverfahren die Eigenschaften $P(F | E) = 0,02$ und $P(F | \bar{E}) = 0,999$ aufweist. Dabei gilt F : "Prüfverfahren zeigt Defekt an".

Wie groß ist nun die Wahrscheinlichkeit dafür, dass ein Erzeugnis defekt ist, obwohl bei der Qualitätsprüfung kein Mangel festgestellt wurde?

$$P(\bar{E} | \bar{F}) = \frac{P(\bar{F} | \bar{E}) \cdot P(\bar{E})}{P(\bar{F} | \bar{E}) \cdot P(\bar{E}) + P(\bar{F} | E) \cdot P(E)} = \frac{(1 - 0,999) \cdot (1 - 0,97)}{(1 - 0,999) \cdot (1 - 0,97) + 0,98 \cdot 0,97} = 0,00003$$

Wie groß ist hingegen die Wahrscheinlichkeit dafür, dass ein Erzeugnis einwandfrei ist, wenn bei der Kontrolle kein Mangel festgestellt wurde?

$$P(E | \bar{F}) = 1 - P(\bar{E} | \bar{F}) = 1 - 0,00003 = 0,99997$$

Während die a-priori-Wahrscheinlichkeit für ein einwandfreies Produkt bei 0,97 lag, ist die a-posteriori Wahrscheinlichkeit, die zusätzliche Informationen aus dem Prüfverfahren nutzt, mit 0,99997 hier höher und verbessert die anfänglichen Erfahrungswerte. Die Wahrscheinlichkeit dafür, dass ein defektes Produkt an einen Kunden verkauft wird, sinkt somit drastisch.

In der Praxis wird die Formel von Bayes häufig zur Berechnung der Wahrscheinlichkeiten von **Hypothesen** (Annahmen) benutzt. Hier kommt der Informationsverarbeitungsgedanke der Formel von Bayes besonders zum Tragen. Veranschaulichen wollen wir dies direkt anhand eines häufig in der Literatur zu findenden Beispiels zum Zufallsvorgang Münzwurf.

Beispiel II 1-32

Hypothesen und Formel von Bayes

Beim Glücksspiel Münzwurf vermutet ein Mitspieler, dass die Münze zwei Zahlseiten besitzt. Dieser Annahme bzw. *Hypothese*, nämlich dem Ereignis A: "Münze besitzt zwei Zahlseiten", räumt er die subjektive Wahrscheinlichkeit $P(A) = 0,70$ ein. Der Gegenannahme bzw. *Alternativhypothese*, es handle sich um eine faire Münze, wird daher die Wahrscheinlichkeit $P(\bar{A}) = 0,30$ zugeordnet.

Fällt nun beim ersten Wurf tatsächlich "Zahl" (Ereignis B), sieht sich der Spieler in seiner Befürchtung bestätigt und bestimmt unter Verwendung dieser neuen Information erneut die Wahrscheinlichkeit dafür, dass es sich tatsächlich um eine Münze mit zwei Zahlseiten handelt. Da für eine Münze mit zwei Zahlseiten $P(B|A) = 1$ und eine faire Münze $P(B|\bar{A}) = 0,5$ gilt, wird er daher die Wahrscheinlichkeit des Ereignisses A unter der Bedingung, dass vorher das Ereignis B eingetreten ist, wie folgt berechnen:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})} = \frac{1 \cdot 0,70}{1 \cdot 0,70 + 0,5 \cdot 0,30} = 0,8235$$

Wir können also sagen, dass sich durch den ersten Wurf die Wahrscheinlichkeit der Hypothese, es liege eine falsche Münze mit zwei Zahlseiten vor, von 0,70 auf 0,8235 erhöht.

Die oben definierte **Formel von Bayes** gilt nicht nur für den Fall, dass sich der Ergebnisraum Ω in zwei disjunkte Mengen A und \bar{A} zerlegt, sondern entsprechend auch für *allgemeinere Zerlegungen* von Ω . Bilden die m Ereignisse A_1, A_2, \dots, A_m eine vollständige Zerlegung des Ergebnisraumes Ω , liegt $P(A_i) > 0$ vor und ist B ein Ereignis mit $P(B) > 0$, dann gilt für jedes A_i

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^m P(B|A_j) \cdot P(A_j)}. \quad (\text{II1: 64})$$

Zur Herleitung dieser allgemeinen Formel müssen wir lediglich den verallgemeinerten Satz der totalen Wahrscheinlichkeit (II1: 60) in (II1: 61) einsetzen und das Ereignis A entsprechend indizieren.

Beispiel II 1-33

Verallgemeinerte Formel von Bayes

Auf drei verschiedenen Fertigungsanlagen F_1, F_2 und F_3 wird der gleiche DVD-Player-Typ produziert. Für diese Fertigungsanlagen liegt folgendes Datenmaterial vor:

Fertigungsanlage	F_1	F_2	F_3
Anteil an der Produktion	0,30	0,30	0,40
Ausschussanteil	0,05	0,03	0,07

Aus der Gesamtproduktion wird nun zufällig ein DVD-Player entnommen. Wie groß ist die Wahrscheinlichkeit, dass dieser auf F_1 , F_2 bzw. F_3 gefertigt wurde, wenn festgestellt wird, dass er von einwandfreier Qualität (a) bzw. defekt (b) ist?

Aus dieser Aufgabenstellung lässt sich zunächst

$$P(F_1) = 0,30, \quad P(D | F_1) = 0,05 \rightarrow P(\bar{D} | F_1) = 1 - 0,05 = 0,95$$

$$P(F_2) = 0,30, \quad P(D | F_2) = 0,03 \rightarrow P(\bar{D} | F_2) = 1 - 0,03 = 0,97$$

$$P(F_3) = 0,40, \quad P(D | F_3) = 0,07 \rightarrow P(\bar{D} | F_3) = 1 - 0,07 = 0,93$$

ableiten, womit wir beide Fragestellungen beantworten können:

- a) Für die Berechnung der gesuchten Wahrscheinlichkeiten empfiehlt es sich zunächst die totale Wahrscheinlichkeit $P(\bar{D})$ zu ermitteln:

$$\begin{aligned} P(\bar{D}) &= P(\bar{D} | F_1) \cdot P(F_1) + P(\bar{D} | F_2) \cdot P(F_2) + P(\bar{D} | F_3) \cdot P(F_3) \\ &= 0,95 \cdot 0,30 + 0,97 \cdot 0,30 + 0,93 \cdot 0,40 = 0,948 \end{aligned}$$

Die bedingten Wahrscheinlichkeiten ergeben sich damit als

$$P(F_1 | \bar{D}) = \frac{P(\bar{D} | F_1) \cdot P(F_1)}{P(\bar{D})} = \frac{0,95 \cdot 0,30}{0,948} = 0,3006$$

$$P(F_2 | \bar{D}) = \frac{P(\bar{D} | F_2) \cdot P(F_2)}{P(\bar{D})} = \frac{0,97 \cdot 0,30}{0,948} = 0,3070$$

$$P(F_3 | \bar{D}) = \frac{P(\bar{D} | F_3) \cdot P(F_3)}{P(\bar{D})} = \frac{0,93 \cdot 0,40}{0,948} = 0,3924$$

und ergeben in der Summe natürlich Eins.

- b) Die hier zu berechnenden bedingten Wahrscheinlichkeiten erfordern zunächst die Bestimmung von $P(D)$. Der Wert dafür ergibt sich als

$$P(D) = P(D | F_1) \cdot P(F_1) + P(D | F_2) \cdot P(F_2) + P(D | F_3) \cdot P(F_3)$$

oder

$$P(D) = 1 - P(\bar{D}) = 1 - 0,948 = 0,052.$$

Die gesuchten bedingten Wahrscheinlichkeiten, die auch hier in der Summe wieder Eins ergeben müssen (ggf. Rundungsungenauigkeiten), ergeben sich damit zu:

$$P(F_1 | D) = \frac{P(D | F_1) \cdot P(F_1)}{P(D)} = \frac{0,05 \cdot 0,30}{0,052} = 0,2885$$

$$P(F_2 | D) = \frac{P(D | F_2) \cdot P(F_2)}{P(D)} = \frac{0,03 \cdot 0,30}{0,052} = 0,1731$$

$$P(F_3 | D) = \frac{P(D | F_3) \cdot P(F_3)}{P(D)} = \frac{0,07 \cdot 0,40}{0,052} = 0,5385$$

Diese Teilaufgabe können wir dazu verwenden eine wichtige Interpretationsmöglichkeit der Formel von Bayes zu beschreiben: "Betrachten wir die Ereignisse F_i als Ursachen, die eine gewisse Wirkung, nämlich das Ereignis D , auslösen, liefert die Formel von Bayes die Möglichkeit, die Wahrscheinlichkeit dafür auszurechnen, dass der beobachteten Wirkung eine der Ursachen F_i zugrunde liegt." So ist beispielsweise die Wahrscheinlichkeit dafür, dass das gefundene defekte Stück (Wirkung) von F_3 stammt (Ursache), mit 0,5385 am größten.

Beispiel II 1-34

Dreitorspiel

Die früher sehr beliebte Gameshow "Geh auf's Ganze" mit Moderator Jörg Dräger endete gewöhnlich mit dem sog. "Superdeal". Dabei konnte ein Spieler seinen in der Show gewonnenen Preis setzen und hatte dadurch die Möglichkeit in einem "Drei-Tore-Spiel" ein Auto zu gewinnen. Dem Spieler standen dabei drei Tore zur Auswahl. In einem befand sich ein Auto und in den beiden anderen ein "Zonk" (Stoffpuppe). Nachdem sich der Spieler für ein Tor entschieden hatte, öffnete der Showmaster eines, das der Spieler nicht gewählt hatte. Anschließend hatte der Spieler erneut die Gelegenheit sich für ein anderes Tor zu entscheiden oder bei dem bereits gewählten zu bleiben.

Nehmen wir an, dass sich der Spieler für das Tor 1 entschieden hat und betrachten dazu folgende drei Fälle. Das x kennzeichnet dabei jeweils die Position des Autos.

	Tor 1	Tor 2	Tor 3
Fall 1	x		
Fall 2		x	
Fall 3			x

Welche Strategie sollte dieser Spieler, der sich anfänglich für Tor 1 entschieden hat, nun nach Öffnung eines Tores durch den Showmaster verfolgen? Bei einer *Bleibestrategie* gewinnt der Spieler nur im Fall 1. Im Fall 2 und 3 würde er das Auto nicht gewinnen, falls er weiterhin Tor 1 wählt. Deswegen führt die *Wechselstrategie* in den Fällen 2 und 3 zum Gewinn des Autos. Dies gilt, da der Showmaster grundsätzlich ein "leeres Tor" (Tor ohne Auto bzw. mit "Zonk") öffnet bzw. aus dem Spiel nimmt. Die nachfolgenden Berechnungen sollen diese Strategien nun noch näher beleuchten. Wir verwenden dabei die Ereignisse S_i : "Der Showmaster öffnet Tor i." und T_i : "Das Auto befindet sich hinter Tor i.":

Die Wahrscheinlichkeit, das Tor mit dem Auto bei der ersten Auswahl eines Tores zu erwischen, liegt bei

$$P(T_1) = P(T_2) = P(T_3) = \frac{1}{3}.$$

Hat sich der Spieler für Tor 1 entschieden, bleiben dem Showmaster nur noch zwei Tore, die er öffnen kann. Berechnen wir nun die unbedingte Wahrscheinlichkeit dafür, dass der Showmaster Tor 3 öffnet. Wir erhalten sie über den Satz der totalen Wahrscheinlichkeit zu

$$\begin{aligned} P(S_3) &= P(S_3 | T_1)P(T_1) + P(S_3 | T_2)P(T_2) + P(S_3 | T_3)P(T_3) \\ &= \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2}. \end{aligned}$$

Die darin verwendeten bedingten Wahrscheinlichkeiten ergeben sich wie folgt: Die Wahrscheinlichkeit dafür, dass der Showmaster Tor 3 öffnet, wenn das Auto in Tor 1 ist, liegt bei $P(S_3 | T_1) = 0,5$, da es in diesem Fall egal ist, welches der Tore 2 oder 3 er aus dem Spiel nimmt. Ist das Auto in Tor 2, liegt die Wahrscheinlichkeit dafür, dass er Tor 3 öffnet, bei $P(S_3 | T_2) = 1$, da er das Auto nicht frühzeitig aus dem Spiel nimmt und natürlich auch nicht das Tor des Spielers öffnen kann. Aus den gleichen Gründen wird er nicht Tor 3 öffnen, wenn das Auto in Tor 3 ist, sodass $P(S_3 | T_3) = 0$ gilt.

Nach den gleichen Überlegungen können wir auch die unbedingte Wahrscheinlichkeit $P(S_2) = 0,5$ bestimmen.

Es gilt nun nach der Formel von Bayes

- für die Wahrscheinlichkeit, dass sich das Auto in Tor 2 befindet, nachdem der Showmaster Tor 3 geöffnet hat

$$P(T_2 | S_3) = \frac{P(S_3 | T_2) \cdot P(T_2)}{P(S_3)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

- für die Wahrscheinlichkeit, dass sich das Auto in Tor 3 befindet, nachdem der Showmaster Tor 2 geöffnet hat

$$P(T_3 | S_2) = \frac{P(S_2 | T_3) \cdot P(T_3)}{P(S_2)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

- für die Wahrscheinlichkeit, dass sich das Auto in Tor 1 befindet, nachdem der Showmaster Tor 3 geöffnet hat

$$P(T_1 | S_3) = \frac{P(S_3 | T_1) \cdot P(T_1)}{P(S_3)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

- für die Wahrscheinlichkeit, dass sich das Auto in Tor 1 befindet, nachdem der Showmaster Tor 2 geöffnet hat

$$P(T_1 | S_2) = \frac{P(S_2 | T_1) \cdot P(T_1)}{P(S_2)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

Die Wahrscheinlichkeitsrechnung empfiehlt dem Spieler also einen Wechsel des Tores, nachdem der Showmaster eines der beiden anderen Tore geöffnet hat, da die Gewinnwahrscheinlichkeit der Wechselstrategie bei 2/3 und nicht wie bei der Bleibestrategie bei 1/3 liegt. Auch hieran ist die Bedeutung der Formel von Bayes wieder sehr schön erkennbar. Es wird nämlich die zusätzliche Information berücksichtigt, die sich durch die Öffnung des Tores durch den Showmaster ergibt.

Diese Berechnungen können analog für die Fälle, in denen sich der Spieler anfänglich für Tor 2 oder 3 entscheidet, durchgeführt werden. Auch hier zeigt sich, dass der Spieler mit einer *Wechselstrategie die bessere Chance zum Gewinnen des Autos hat*.

2.

Zufallsvariablen

Ein wichtiges Thema der Wahrscheinlichkeitsrechnung stellen Zufallsvariablen dar, da sie die praktische Anwendung von stochastischen Modellen erheblich erleichtern. Wir widmen uns zunächst zwei Klassen von eindimensionalen Zufallsvariablen (stetigen und diskreten). Für beide Klassen beschäftigen wir uns näher mit der dazugehörigen Wahrscheinlichkeits- und Verteilungsfunktion sowie der Berechnung von Erwartungswert und Varianz. Des Weiteren gehen wir ausführlich auf mehrdimensionale Zufallsvariablen ein und legen hier den Schwerpunkt insbesondere auf die Korrelation und Linearkombination von Zufallsvariablen.

2.1 Begriff der Zufallsvariable

Im vorhergehenden Kapitel wurden den Elementarereignissen und anderen Ereignissen eines Zufallsexperimentes Wahrscheinlichkeiten zugeordnet. Bei der Beobachtung eines Zufallsexperimentes treten nun aber häufig Situationen auf, die es erforderlich machen, den Elementarereignissen zunächst reelle Zahlen und erst diesen reellen Zahlen Wahrscheinlichkeiten zuzuordnen. Typische Fälle dafür sind:

- Als Elementarereignisse werden qualitative Größen beobachtet (z.B. "Mann", "Frau" oder "Europäer", "Amerikaner", "Afrikaner") für die man eine Codierung in ein Zahlensystem wünscht.
- Es interessiert nicht die gesamte Vielfalt der möglichen Versuchsausgänge, sondern nur bestimmte Merkmale (z.B. die Anzahl defekter Geräte aus einer Lieferung oder die Augensumme beim dreimaligen Würfeln).

Nicht bei allen Zufallsexperimenten ist eine Codierung (Zuordnung reeller Zahlen zu den Elementarereignissen) erforderlich. Bei manchen Zufallsexperimenten sind die Ergebnisse bereits reelle Zahlen.

Beispiel II 2-1	Codierung
<ul style="list-style-type: none">▪ Ergebnisse sind bereits reelle Zahlen: Einmaliger Würfelwurf (Augenzahl), Stichprobe aus der laufenden Produktion (Anzahl fehlerhafter Stücke), Rendite eines Wertpapiers in einer Periode	
<ul style="list-style-type: none">▪ Beispiele für Codierung: Europäisches Roulette: "grün" → 0, "schwarz" → 1, "rot" → 2 Befragung nach Geschlecht: "Mann" → 0, "Frau" → 1	

Eine Funktion X , die jedem Elementarereignis $\omega \in \Omega$ genau eine reelle Zahl $X(\omega) \in \mathbb{R}$ zuordnet, wird als **Zufallsvariable** oder stochastische Variable bezeichnet. Anders ausgedrückt, bildet also die Zufallsvariable X die Elemente von Ω auf die reelle Zahlengerade ab (vgl. Abbildung II 11). Zufallsvariablen werden im Allgemeinen mit großen lateinischen Buchstaben bezeichnet. Man verwendet dabei bevorzugt jene am Ende des Alphabets, d.h. ..., X, Y, Z. Die einzelnen Werte, die eine Zufallsvariable X bei Durchführung des Zufallsexperimentes annimmt, heißen **Realisationen** (Ausprägungen) der Zufallsvariablen und werden mit den entsprechenden kleinen lateinischen Buchstaben x_1, x_2, \dots bezeichnet. Wir können daher

$$X(\omega_i) = x_i \quad (\text{II2: 1})$$

schreiben. Den Ergebnisraum Ω nennen wir *Definitionsbereich* der Funktion X . Die Menge der reellen Zahlen ist ihr *Wertebereich*. Der Wertebereich muss jedoch nicht zwangsläufig alle reellen Zahlen umfassen.

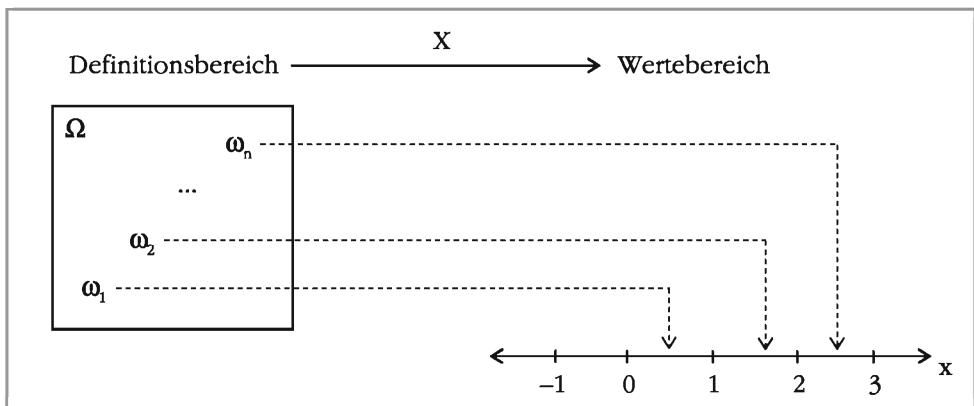
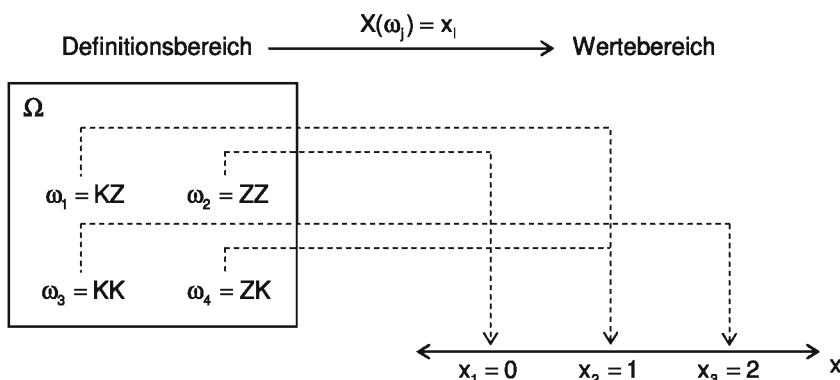


Abbildung II 11: Zufallsvariablen als Abbildung von Ω auf die reelle Achse

Beispiel II 2-2

Definitions- und Wertebereich

Beim Zufallsexperiment "Zweimaliges Werfen einer Münze" wollen wir die Frage untersuchen, wie oft "Kopf" erscheint. Wir definieren dazu die Zufallsvariable X : "Anzahl Kopf", die drei mögliche Realisationen $x_1 = 0, x_2 = 1$ und $x_3 = 2$ besitzt. Der Wertebereich der Zufallsvariable X ergibt sich dabei durch folgende Zuordnung:



Beispiel II 2-3**Weitere Zufallsvariablen**

Wird ein fairer Würfel zweimal geworfen, ergibt sich ein Ergebnisraum mit 36 Elementarereignissen der Form $\Omega = \{(i,j) : i = 1, 2, \dots, 6 \text{ und } j = 1, 2, \dots, 6\}$. Typische Zufallsvariablen für ein solches Zufallsexperiment sind z.B.

- X: "Augensumme" mit den 11 Ausprägungen $x = 2, 3, \dots, 12$ und
- Y: "Absolutbetrag der Augendifferenz" mit den 6 Ausprägungen $y = 0, 1, 2, \dots, 5$.

Unsere bisherigen Ausführungen lassen erkennen, dass bei der Definition einer Zufallsvariable, Wahrscheinlichkeiten keine Rolle spielen. Weder ist die Zuordnung der $X(\omega)$ zu den $\omega \in \Omega$ zufällig, noch ordnen Zufallsvariablen Wahrscheinlichkeiten zu. Den möglichen Realisationen einer Zufallsvariable werden hingegen Wahrscheinlichkeiten zugeordnet. Wie dies geschieht, werden wir im Folgeabschnitt II 2.2 im Detail behandeln.

Die meisten in der Praxis vorkommenden Zufallsvariablen können in zwei Typen aufgeteilt werden. Es ist zu unterscheiden zwischen stetigen und diskreten Zufallsvariablen. Kriterium für die Aufteilung sind die Ergebnisräume, über die die jeweiligen Zufallsvariablen definiert sind.

- **Diskrete Zufallsvariablen**

Wird eine Zufallsvariable über einem endlichen Ergebnisraum definiert, umfasst sie auch nur endlich viele Realisationen. Solche Zufallsvariablen, deren *Wertevorrat nur endlich oder abzählbar unendlich viele Ausprägungen* besitzen, werden als diskrete Zufallsvariablen bezeichnet.¹ Ihre Ausprägungen werden über einen Zählvorgang ermittelt. Es gilt also

$$W = \{x_1, x_2, x_3, \dots\} \text{ mit } W \subset \mathbb{R}.$$

Typische Beispiele für diskrete Zufallsvariablen sind etwa X: "Anzahl Kopf beim dreimaligen Münzwurf" oder Y: "Anzahl defekter Glühbirnen in einer Stichprobe von 10 Stück".

- **Stetige (kontinuierliche) Zufallsvariablen**

Diese Zufallsvariablen können zumindest in einem *bestimmten Bereich* der reellen Zahlen *jeden beliebigen Wert* annehmen. Ihre Ausprägungen werden über einen Messvorgang bestimmt. Es gilt

$$W = \{x \mid a \leq x \leq b\} \text{ mit } W \subset \mathbb{R}.$$

Theoretisch kann der Wertevorrat auch aus der ganzen reellen Achse bestehen. Es gilt dann $W = \mathbb{R}$. Als typische Beispiele für stetige Zufallsvariablen sind z.B. X: "Zetaufwand zur Produktion eines Werkstücks" oder Y: "Gewicht einer zufällig aus einer Abfüllanlage entnommenen 1L-Flasche" zu nennen.

Wie diese Beispiele erkennen lassen, können wir jedes Zufallsexperiment mit reellwertigen Elementarereignissen (z.B. Aktienrenditen) automatisch als Zufallsvariable auffassen, indem wir $X(\omega) = \omega$ für jedes $\omega \in \Omega$ setzen. Es soll auch nicht unerwähnt bleiben, dass sich *stetige Zufallsvariablen* aufgrund nicht beliebig erhöbarer Messgenauigkeit in der **Praxis** streng genommen häufig nur *diskret erfassen* lassen.

¹ Eine Menge wird dann abzählbar unendlich genannt, wenn sie zwar unendlich viele Elemente besitzt, sich aber jedem einzelnen Element eine natürliche Zahl zuordnen lässt.

Die Nützlichkeit von Zufallsvariablen begründet sich zum einen dadurch, dass durch sie das umständliche Rechnen mit Elementarereignissen und Ereignissen weitgehend entfällt. Zum anderen können mit ihrer Hilfe bestimmte Ereignisse statt als Teilmengen von Ω kürzer dargestellt werden. Typische Beispiele dafür sind:

- "X nimmt (genau, mindestens oder höchstens) den Wert x an."
- Neue Darstellung: $X = x$, $X \geq x$, $X \leq x$
- "X nimmt Werte zwischen a und b an."
- Neue Darstellung: $a \leq X \leq b$

Besonders bei der Berechnung von Wahrscheinlichkeiten wird sich diese neue Schreibweise im Folgenden noch als sehr vorteilhaft erweisen.

2.2 Diskrete Zufallsvariablen

2.2.1 Wahrscheinlichkeitsfunktion

Ist X eine diskrete Zufallsvariable, die endlich viele Werte x_1, x_2, \dots, x_n annehmen kann, so ergibt sich die Wahrscheinlichkeit, dass X eine spezielle Ausprägung x_i annimmt, d.h. $P(X = x_i)$, durch Addition der Wahrscheinlichkeiten derjenigen Elementarereignisse ω_j , denen die Realisation x_i zugeordnet ist. Es gilt also formal

$$P(X = x_i) = \sum_{X(\omega_j)=x_i} P(\omega_j). \quad (\text{II2: 2})$$

Beispiel II 2-4

Wahrscheinlichkeiten und Zufallsvariablen

Ein idealer Würfel werde zweimal geworfen. Die betrachtete Zufallsvariable X sei die Augensumme. Der Ergebnisraum und die einzelnen Realisationen der Zufallsvariablen (hochgestellte Zahlen) sind durch folgende Tabelle beschrieben:

2	(1,1)	3	(1,2)	4	(1,3)	5	(1,4)	6	(1,5)	7	(1,6)
3	(2,1)	4	(2,2)	5	(2,3)	6	(2,4)	7	(2,5)	8	(2,6)
4	(3,1)	5	(3,2)	6	(3,3)	7	(3,4)	8	(3,5)	9	(3,6)
5	(4,1)	6	(4,2)	7	(4,3)	8	(4,4)	9	(4,5)	10	(4,6)
6	(5,1)	7	(5,2)	8	(5,3)	9	(5,4)	10	(5,5)	11	(5,6)
7	(6,1)	8	(6,2)	9	(6,3)	10	(6,4)	11	(6,5)	12	(6,6)

Jedem dieser 36 gleichwahrscheinlichen Elementarereignisse wird nach der klassischen Wahrscheinlichkeitsdefinition die Wahrscheinlichkeit $1/36$ zugeordnet. Die Wahrscheinlichkeiten der unterschiedlichen Ausprägungen der Zufallsvariablen X ergeben sich durch Addition der Wahrscheinlichkeiten der dazugehörigen Elementarereignisse wie folgt:

x_i	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Mit diesem Beispiel können wir sehr schön veranschaulichen, warum wir in Formel (II2: 2) zwei Indizes i und j benötigen. Für das Zufallsexperiment, das wir hier betrachten, existieren nämlich genau 36 Elementarereignisse ω_j mit $j = 1, \dots, 36$. Für die Augensumme gibt es jedoch nur 11 mögliche Werte x_i mit $i = 1, \dots, 11$. Jedem i wird also mehr als ein j zugeordnet, sodass wir nicht allein mit einem Index auskommen.

Eine Funktion $f(x)$, die jeder reellen Zahl x die Wahrscheinlichkeit zuordnet, mit der sie bei dieser diskreten Zufallsvariablen X auftritt, heißt **Wahrscheinlichkeitsfunktion** der Zufallsvariablen. Für sie gilt

$$f(x) = \begin{cases} P(X = x_i) & \text{falls } x = x_i \\ 0 & \text{sonst} \end{cases}. \quad (\text{II2: 3})$$

Die Wahrscheinlichkeitsfunktion ist auf der gesamten reellen Achse $-\infty < x < \infty$ definiert und hat nur an den Stellen $x = x_i$, die zum Wertevorrat der Zufallsvariablen gehören, nichtnegative Werte $P(X = x_i)$. Dazwischen ist sie Null.

Jede Wahrscheinlichkeitsfunktion muss die Eigenschaften

$$0 \leq f(x_i) \leq 1 \text{ und } \sum_i f(x_i) = 1 \text{ für jeweils } i = 1, 2, \dots \quad (\text{II2: 4})$$

erfüllen. Wie aus diesen Eigenschaften zu erkennen ist, haben die Wahrscheinlichkeiten $f(x_i)$ die gleichen Eigenschaften wie die im Rahmen der deskriptiven Statistik behandelten relativen Häufigkeiten. Die **Wahrscheinlichkeitsverteilung** ist damit ein Analogon zur Verteilung der relativen Häufigkeiten. Auch die grafische Darstellung der Verteilung in Form eines *Stabdiagrammes* ist ähnlich.

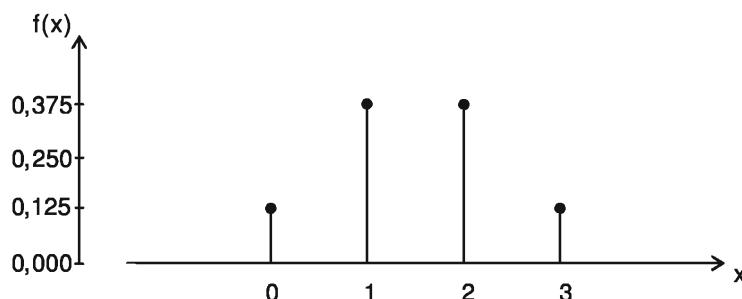
Beispiel II 2-5

Wahrscheinlichkeitsfunktion

Für eine Zufallsvariable X : "Anzahl defekter Zündkerzen in einer 3er-Box" ergaben sich folgende Realisationen und zugehörige Wahrscheinlichkeiten:

x_i	0	1	2	3	Σ
$f(x_i)$	0,125	0,375	0,375	0,125	1,000

Grafisch lässt sich diese Wahrscheinlichkeitsfunktion wie folgt skizzieren:



Mit Hilfe der Wahrscheinlichkeitsfunktion kann natürlich auch angegeben werden, wie groß die Wahrscheinlichkeit dafür ist, dass die Zufallsvariable einen Wert in einem Intervall $[a, b]$ annimmt. Es gilt nämlich

$$P(a \leq X \leq b) = \sum_{a \leq x_i \leq b} P(X = x_i) = \sum_{a \leq x_i \leq b} f(x_i). \quad (\text{II2: 5})$$

Beispiel II 2-6

Anwendung von (II2: 5)

Stellen wir im Zusammenhang mit Beispiel II 2-5 die Frage, wie groß die Wahrscheinlichkeit dafür ist, dass 1 oder 2 defekte Stücke auftreten, so kann diese folgendermaßen beantwortet werden:

$$P(1 \leq X \leq 2) = \sum_{1 \leq x_i \leq 2} f(x_i) = 0,375 + 0,375 = 0,75$$

Die einfache Addition lässt sich damit begründen, dass die beiden Ereignisse $X = 1$ und $X = 2$ disjunkt sind.

2.2.2 Verteilungsfunktion

Eine Funktion, die die Wahrscheinlichkeit dafür angibt, dass eine Zufallsvariable X höchstens den Wert x annimmt, bezeichnet man als **Verteilungsfunktion** $F(x)$ der Zufallsvariablen X . Die Werte der Verteilungsfunktion lassen sich durch Aufsummieren derjenigen Funktionswerte der Wahrscheinlichkeitsfunktion ermitteln, die zu den Ausprägungen $x_i \leq x$ gehören.

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i) \quad (\text{II2: 6})$$

Wie wir hieraus gut erkennen können, ist die Verteilungsfunktion ein Analogon zur Summenhäufigkeitsfunktion aus der deskriptiven Statistik. Das Bild der Verteilungsfunktion einer diskreten Zufallsvariable ist daher auch das einer *Treppenfunktion*. Für eine Verteilungsfunktion ergeben sich aus ihrer Definition die *Eigenschaften*:

- Der niedrigste Wert von $F(x)$ ist Null, der höchste Eins. $F(x)$ nimmt also nur Werte zwischen $F(-\infty) = 0$ und $F(+\infty) = 1$ an, d.h. es gilt

$$0 \leq F(x) \leq 1 \quad \text{für alle } x \in \mathbb{R}.$$

Die Funktion $F(x)$ besitzt also die beiden Grenzwerte $F(x) \rightarrow 0$ für $x \rightarrow -\infty$ und $F(x) \rightarrow 1$ für $x \rightarrow +\infty$.

- $F(x)$ steigt für wachsendes x monoton an oder bleibt zumindest auf gleicher Höhe. Dies bedeutet, dass für alle $-\infty \leq a \leq b \leq +\infty$ Folgendes gilt:

$$F(a) \leq F(b)$$

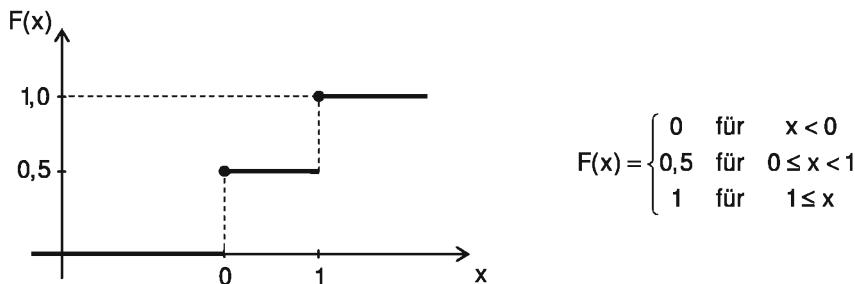
- Die Verteilungsfunktionen diskreter Zufallsvariablen weisen eine typische *Treppenform* auf. Ihr Funktionswert verharret auf dem gleichen Niveau und springt an den Stellen x_i um den Betrag der Wahrscheinlichkeit $P(X = x_i)$ nach oben. An den Sprungstellen selbst gilt der "obere" Wert.

Beispiel II 2-7**Verteilungsfunktion**

Betrachten wir die Zufallsvariable X : "Anzahl der Köpfe beim einmaligen Münzwurf". Da die Wahrscheinlichkeit Kopf zu werfen 0,5 beträgt, beschreibt die folgende Tabelle die Wahrscheinlichkeits- und Verteilungsfunktion dieser Zufallsvariablen.

x_i	0	1
$f(x_i)$	0,5	0,5
$F(x_i)$	0,5	1,0

Die Werte $F(x)$ entstehen, einfach ausgedrückt, immer durch Addition von $f(x_i)$ und aller "vorhergehenden" Werte der Wahrscheinlichkeitsfunktion (Beispiel: $F(1) = f(1) + f(0) = 0,5 + 0,5 = 1,0$). Die Verteilungsfunktion ist hier konkret durch die folgende Grafik und Funktion gegeben:



Unter Zuhilfenahme der Verteilungsfunktion $F(x)$ lässt sich die Wahrscheinlichkeit, dass eine diskrete Zufallsvariable X einen Wert im Intervall $[a,b]$ annimmt, einfach berechnen. Es gilt nämlich für alle a,b mit $-\infty \leq a \leq b \leq +\infty$ stets die Beziehung

$$P(a < X \leq b) = F(b) - F(a).^2 \quad (\text{II2: 7})$$

Allgemein erleichtert die Verwendung von $F(x)$ die Berechnung von Wahrscheinlichkeiten spezieller Ereignisse erheblich, was folgendes Beispiel verdeutlichen soll.

Beispiel II 2-8**Wahrscheinlichkeitsberechnung und Verteilungsfunktion**

Die Anzahl der in einer Reparaturwerkstatt pro Stunde abgefertigten Pkw besitzt folgende Verteilungsfunktion. Wie sich damit zunächst zeigen lässt, können wir aus der jeweiligen Sprunghöhe von $F(x)$ die Werte $f(x_i)$ ermitteln:

x_i	$F(x_i)$	$f(x_i)$
0	0,4	0,4
1	0,7	0,3
2	0,9	0,2
3	1,0	0,1

² Es gilt auch $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$, $P(X < a) = P(X \leq a) - P(X = a) = F(a) - P(X = a)$, $P(X \geq a) = 1 - P(X < a) = 1 - F(a) + P(X = a)$. Daraus folgt $P(a \leq X \leq b) = F(b) - F(a) + P(X = a)$, $P(a < X < b) = F(b) - P(X = b) - F(a)$ und $P(a \leq X < b) = F(b) - P(X = b) - F(a) + P(X = a)$.

Gesucht sind nun die Wahrscheinlichkeiten dafür, dass

- genau 1 Pkw abgefertigt wird,
- weniger als 2 Pkw die Werkstatt verlassen,
- mindestens 1, aber höchstens 3 Pkw in einer Stunde abgefertigt werden, und
- die Wahrscheinlichkeit dafür, dass mindestens 2 Pkw repariert werden, sowie
- die grafische Darstellung der Verteilungsfunktion.

Lösungen:

a) $P(X = 1) = P(X \leq 1) - P(X \leq 0) = F(1) - F(0) = 0,7 - 0,4 = 0,3$

Alternativ: $f(1) = 0,3$

b) $P(X < 2) = P(X \leq 1) = F(1) = 0,7$

Alternativ: $f(0) + f(1) = 0,4 + 0,3 = 0,7$

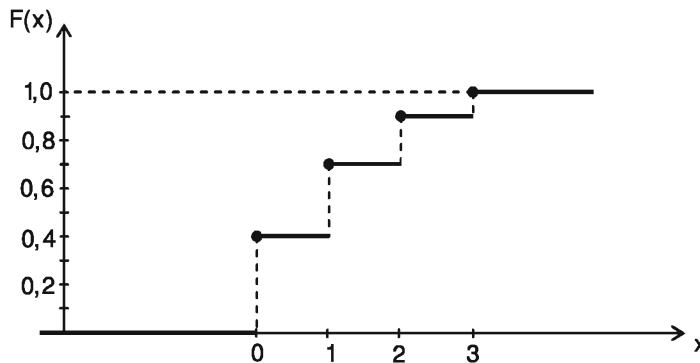
c) $P(1 \leq X \leq 3) = P(0 < X \leq 3) = F(3) - F(0) = 1,0 - 0,4 = 0,6$

Alternativ: $f(1) + f(2) + f(3) = 0,3 + 0,2 + 0,1 = 0,6$

d) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - F(1) = 1 - 0,7 = 0,3$

Alternativ: $f(2) + f(3) = 0,2 + 0,1 = 0,3$

e) Grafische Darstellung:



2.2.3 Zusammenfassende Gegenüberstellung

Diesen ersten Abschnitt zu diskreten Zufallsvariablen wollen wir mit einer Zusammenfassung der Zusammenhänge zwischen Wahrscheinlichkeits- und Verteilungsfunktion abschließen (vgl. insbesondere die Abbildung auf der nächsten Seite). Das Verständnis dieser Beziehungen ist für die folgenden Abschnitte besonders wichtig.

Die Verteilungsfunktion $F(x)$ lässt sich leicht aus der Wahrscheinlichkeitsfunktion $f(x)$ durch Aufsummieren derjenigen Werte von $f(x)$ gewinnen, die zu den Ausprägungen $x_i \leq x$ gehören. Es gilt nämlich nach (II2: 6)

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i).$$

Dieses Vorgehen ist der Grund für die Sprünge von Verteilungsfunktionen diskreter Zufallsvariablen an den Stellen $x = x_i$.

Aus der Verteilungsfunktion $F(x)$ lässt sich umgekehrt auch wieder die Wahrscheinlichkeitsfunktion $f(x)$ ableiten, indem wir an jeder Stelle x von ihrem Funktionswert den linksseitigen Grenzwert abziehen:

$$f(x) = F(x) - \lim_{\Delta x \rightarrow 0} F(x - \Delta x) \quad (\text{II2: 8})$$

Da diese Differenz bzw. $f(x)$ nur an den Sprungstellen x_i einen von Null verschiedenen Wert annimmt, können wir alternativ auch

$$f(x_i) = F(x_i) - F(x_{i-1}) \quad (\text{II2: 9})$$

schreiben. Um die Wahrscheinlichkeit der Realisation x_i zu bestimmen, müssen wir also lediglich vom Funktionswert der Verteilungsfunktion an der Stelle x_i ihren Funktionswert für die nächstkleinere Realisationsmöglichkeit abziehen. Diese Vorgehensweise ist uns bereits im Beispiel II 2-8 begegnet.

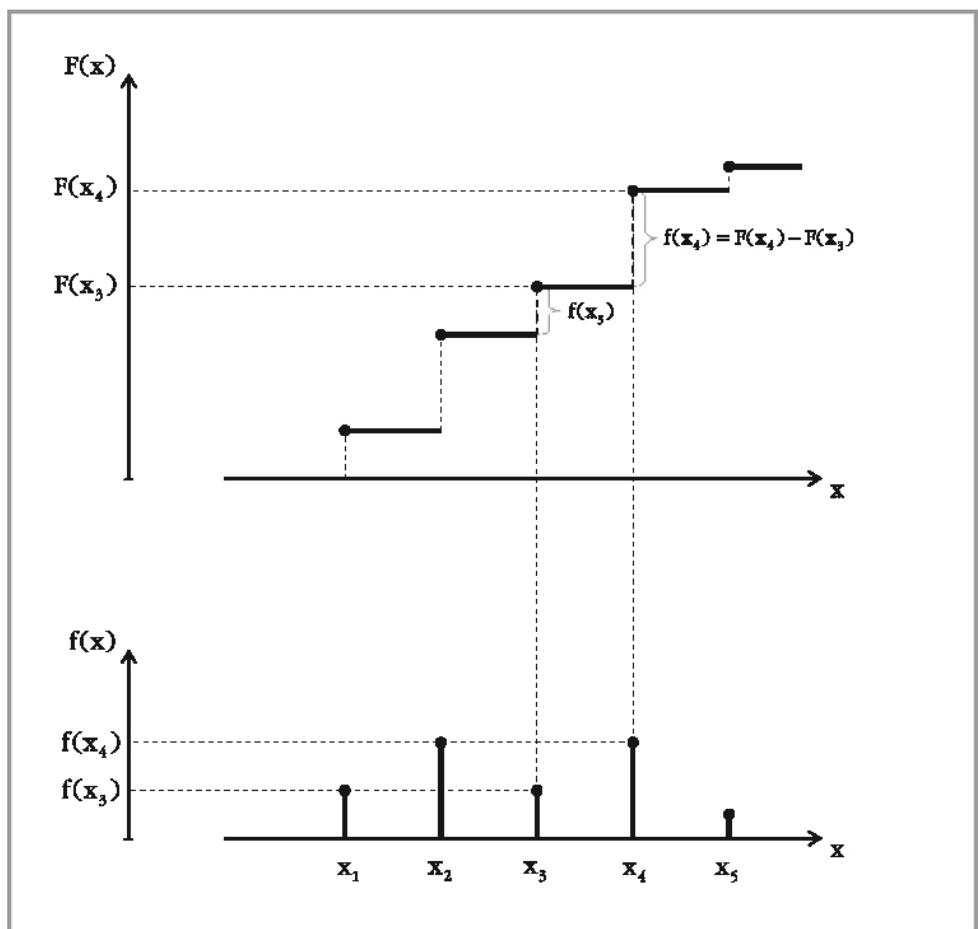


Abbildung II 12: Zusammenhang Wahrscheinlichkeits- und Verteilungsfunktion

Beispiel II 2-9**Abschließendes Beispiel zu diskreten Zufallsvariablen**

Für eine diskrete Zufallsvariable X ergab sich folgende lückenhafte Wahrscheinlichkeits- und Verteilungsfunktion. Es gilt nun die fehlenden Werte a , b , c und d zu ermitteln, wobei bekannt ist, dass $P(3 < X \leq 5) = 0,5$ gilt.

x_i	1	2	3	4	5	Σ
$f(x_i)$	0,1	0,2	b	c	0,2	1,0
$F(x_i)$	0,1	a	d	0,8	1,0	-

Die gesuchten Werte ergeben sich wie folgt:

$$a = F(2) = f(2) + f(1) = 0,2 + 0,1 = 0,3$$

$$d = F(3) \rightarrow P(3 < X \leq 5) = F(5) - F(3) = 0,5 \rightarrow F(3) = 1,0 - 0,5 = 0,5$$

$$b = f(3) = F(3) - F(2) = 0,5 - 0,3 = 0,2$$

$$0,1 + 0,2 + 0,2 + c + 0,2 = 1 \rightarrow c = 0,3$$

Der Wert von c ließe sich auch schneller über den folgenden Zusammenhang ermitteln:

$$P(3 < X \leq 5) = f(4) + f(5) = 0,5 \rightarrow c = f(4) = 0,3$$

2.3 Stetige Zufallsvariablen

2.3.1 Verteilungsfunktion

Ist eine Zufallsvariable X stetig, so kann sie in einem bestimmten Intervall jeden beliebigen Wert annehmen. Dies hat zur Folge, dass ihre **Verteilungsfunktion** $F(x) = P(X \leq x)$ keine Treppenfunktion mehr ist, sondern eine *stetige Funktion* (vgl. Abbildung II 13) darstellt.

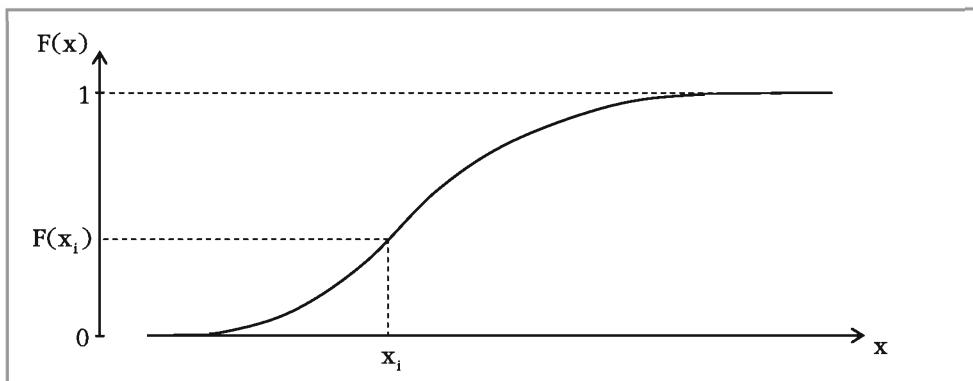


Abbildung II 13: Verteilungsfunktion einer stetigen Zufallsvariable

Die Eigenschaften der Verteilungsfunktionen stetiger Zufallsvariablen sind denen diskreter Zufallsvariablen sehr ähnlich und weichen nur in einem Aspekt voneinander ab. Die Verteilungsfunktion $F(x)$ stetiger Zufallsvariablen hat im Allgemeinen folgende Eigenschaften:

1. Die Funktionswerte liegen im Intervall $0 \leq F(x) \leq 1$.
2. $F(x)$ besitzt die Grenzwerte $F(x) \rightarrow 0$ für $x \rightarrow -\infty$ und $F(x) \rightarrow 1$ für $x \rightarrow +\infty$.
3. $F(x)$ ist monoton wachsend, d.h. für $a < b$ gilt $F(a) \leq F(b)$. Wir können alternativ auch $F(x) \leq F(x+c)$ für $c > 0$ schreiben.

Im Gegensatz zu diskreten Zufallsvariablen kommt folgende Eigenschaft hinzu:

4. $F(x)$ ist keine Treppenfunktion, sondern überall stetig.

Ein typisches Beispiel für eine stetige Zufallsvariable, deren Verteilungsfunktion einen Verlauf wie in Abbildung II 13 zeigt, ist die Lebensdauer eines Gerätes. Die Verteilungsfunktion gibt hier die Wahrscheinlichkeit dafür an, dass ein Gerät eine Lebensdauer von höchstens x Zeiteinheiten besitzt.

2.3.2 Dichtefunktion

Die erste Ableitung der Verteilungsfunktion $F(x)$ nach x wird als **Wahrscheinlichkeitsdichte** oder **Dichtefunktion** $f(x)$ bezeichnet. Sie ist bis auf höchstens endlich viele Stellen ebenfalls eine stetige Funktion und entspricht formal (nicht jedoch in ihrer Interpretation) der Wahrscheinlichkeitsfunktion im diskreten Fall.

$$f(x) = \frac{d F(x)}{d x} = F'(x) \quad (\text{II2: 10})$$

Umgekehrt lässt sich aus der Dichtefunktion durch Integration ($t =$ Integrationsvariable) wieder die Verteilungsfunktion bestimmen.

$$F(x) = \int_{-\infty}^x f(t) dt \quad (\text{II2: 11})$$

Die Dichtefunktion $f(x)$ besitzt folgende wichtige *Eigenschaften*:

1. Die Dichtefunktion verläuft stets oberhalb oder auf der reellen Zahlengeraden (x -Achse). Dies folgt unmittelbar aus der Tatsache, dass die Verteilungsfunktion monoton steigend ist, also nirgends eine negative Ableitung haben kann.

$$f(x) \geq 0 \quad (\text{II2: 12})$$

2. Die gesamte Fläche, die zwischen x -Achse und Dichtefunktion eingeschlossen wird, besitzt den Flächeninhalt Eins.

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad (\text{II2: 13})$$

Wie bereits erwähnt wurde, sind Dichte- und Wahrscheinlichkeitsfunktion miteinander verwandt. Es ist aber zu beachten, dass die Dichtefunktion im Gegensatz zur Wahrscheinlichkeitsfunktion auch ohne weiteres Funktionswerte annehmen kann, die größer als 1 sind. Sie gibt nämlich keine Wahrscheinlichkeiten an. Dichtefunktionen geben *Wahrscheinlichkeiten* nur mittelbar an. Wahrscheinlichkeiten sind lediglich als Flächen *unter der Dichtefunktion* und für bestimmte *Intervalle* zu finden. Wie z.B. Abbildung II 14 zeigt, kann die Wahrscheinlichkeit $P(X \leq x_i)$ bzw. $F(x_i)$ als Fläche unter der Kurve $f(x)$ dargestellt werden.

$$P(X \leq x_i) = F(x_i) = \int_{-\infty}^{x_i} f(t) dt \quad (\text{II2: 14})$$

Außerdem lässt sich die Wahrscheinlichkeit dafür, dass die stetige Zufallsvariable X einen Wert im Intervall $[a,b]$ mit $-\infty \leq a \leq b \leq +\infty$ annimmt, als Fläche unter der Dichtefunktion in den Grenzen a und b interpretieren. Es gilt

$$P(a \leq X \leq b) = \int_a^b f(x) dx . \quad (\text{II2: 15})$$

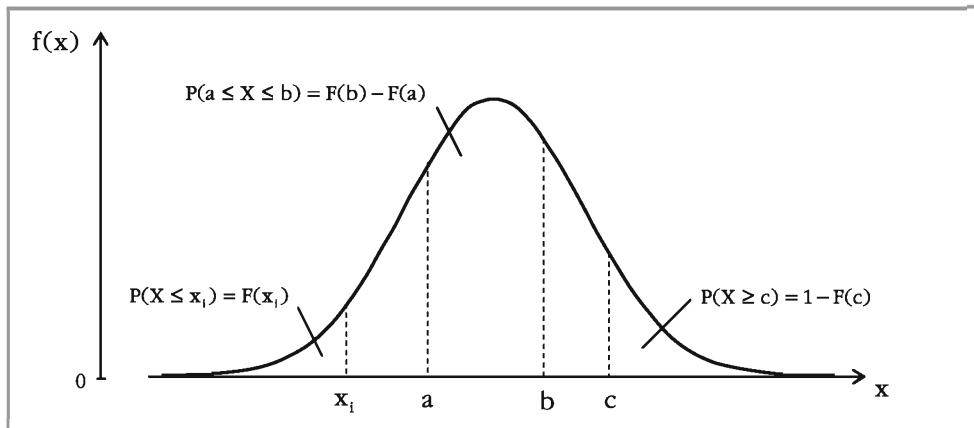


Abbildung II 14: Dichtefunktion einer stetigen Zufallsvariable

Besonders wichtig ist bei stetigen Zufallsvariablen, dass die Wahrscheinlichkeit dafür, dass ein spezieller Wert x angenommen wird, immer Null ist.³

$$P(X = x) = 0 \quad \text{für alle } x \in \mathbb{R} \quad (\text{II2: 16})$$

Dies können wir wie folgt beweisen:

$$P(X = x) = \int_x^x f(t) dt = F(x) - F(x) = 0$$

Aus diesem Grund ist es bei der Bestimmung von $P(a \leq X \leq b)$ unerheblich, ob die beiden Grenzen a und b in das Intervall eingeschlossen werden oder nicht. Es gilt also bei stetigen im Gegensatz zu diskreten Zufallsvariablen immer

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) . \quad (\text{II2: 17})$$

Dies schließt natürlich auch die Gültigkeit von

$$P(X \leq x) = P(X < x) \quad (\text{II2: 18})$$

mit ein.

³ Trotzdem ist das Ereignis $X = x$ nicht für alle x unmöglich. Diese auf den ersten Blick verwirrend scheinende Aussage wird verständlich, wenn wir uns klar machen, dass irgend eine der Realisationen x zwangsläufig angenommen werden muss und in der Praxis oft nicht zwischen der Realisation x (z.B. $x = 0,5$) selbst und einem je nach Messgenauigkeit mehr oder weniger kleinem Intervall um x (z.B. $[0,45; 0,55]$) unterschieden wird. Sofern ein Intervall im Wertebereich liegt, fällt ein stetiges X bereits mit positiver Wahrscheinlichkeit in dieses Intervall, auch wenn das Intervall noch so klein ist.

Neben der Integration der Dichtefunktion $f(x)$ bietet die Verteilungsfunktion $F(x)$ eine weitere Möglichkeit zur Berechnung der Wahrscheinlichkeit $P(a \leq X \leq b)$. Da $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$, $P(X \leq b) = F(b)$ und $P(X \leq a) = F(a)$ gilt, kann

$$P(a \leq X \leq b) = F(b) - F(a) \quad (\text{II2: 19})$$

geschrieben werden.

Beispiel II 2-10

Charakterisierung stetiger Zufallsvariablen

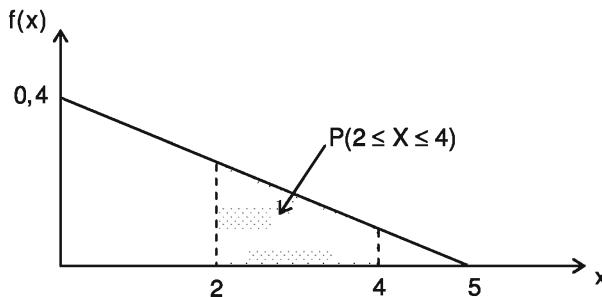
Für eine stetige Zufallsvariable X , die die Verspätung einer Straßenbahn an einer bestimmten Haltestelle (in Minuten) beschreibt, sei folgende vermeintliche Dichtefunktion gegeben:

$$f(x) = \begin{cases} 0,4 - 0,08x & \text{für } 0 \leq x \leq 5 \\ 0 & \text{sonst} \end{cases}$$

Wir wollen zunächst prüfen, ob es sich dabei tatsächlich um eine Dichtefunktion handeln kann. Dazu müssen die beiden Voraussetzungen (II2: 12) und (II2: 13) erfüllt sein.

1. $f(x) \geq 0$ für alle x

Diese Voraussetzung ist klar erfüllt, was ein Blick auf die grafische Darstellung der Funktion (grobe Skizze) bestätigt:



2. Die Fläche unter der Funktion muss gleich 1 sein.

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} (0,4 - 0,08x) dx = \int_0^5 (0,4 - 0,08x) dx = [0,4x - 0,04x^2]_0^5 = (2 - 1) - (0 - 0) = 1$$

Es handelt sich also hier klar um eine echte Dichtefunktion. Die dazugehörige Verteilungsfunktion ergibt sich zu

$$F(x) = \int_{-\infty}^x f(t) dt = \int_0^x (0,4 - 0,08t) dt = [0,4t - 0,04t^2]_0^x = 0,4x - 0,04x^2$$

bzw.

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0,4x - 0,04x^2 & \text{für } 0 \leq x \leq 5 \\ 1 & \text{für } x > 5 \end{cases}$$

Wir berechnen nun die Wahrscheinlichkeit dafür, dass die Zufallsvariable X einen Wert zwischen 2 und 4 annimmt, d.h. die Verspätung zwischen 2 und 4 Minuten beträgt. Uns stehen dazu die beiden folgenden Möglichkeiten zur Verfügung:

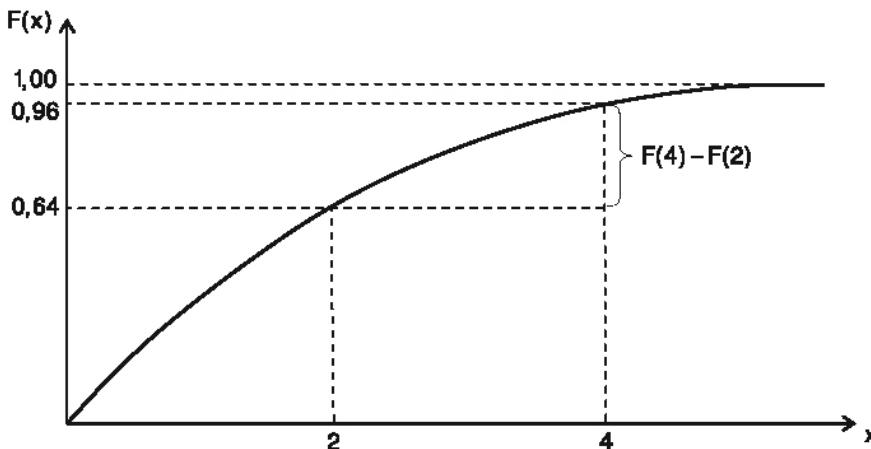
1. Möglichkeit (Integral):

$$P(2 \leq X \leq 4) = \int_2^4 f(x) dx = \int_2^4 (0,4 - 0,04x) dx = [0,4x - 0,04x^2]_2^4 = 0,96 - 0,64 = 0,32$$

2. Möglichkeit (Verteilungsfunktion):

$$P(2 \leq X \leq 4) = F(4) - F(2) = (0,4 \cdot 4 - 0,04 \cdot 4^2) - (0,4 \cdot 2 - 0,04 \cdot 2^2) = 0,96 - 0,64 = 0,32$$

Skizziert zeigt sich die Lösung anhand der Verteilungsfunktion folgendermaßen:



Beispiel II 2-11

Bestimmung der Dichtefunktion

Gegeben sei folgende Verteilungsfunktion einer stetigen Zufallsvariable X:

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{20}x & \text{für } 0 \leq x \leq 20 \\ 1 & \text{für } x > 20 \end{cases}$$

Die Dichtefunktion kann daraus durch Ableitung aller drei Teilstücke ermittelt werden:

$$f(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{20} & \text{für } 0 \leq x \leq 20 \\ 0 & \text{für } x > 20 \end{cases} \text{ bzw. } f(x) = \begin{cases} \frac{1}{20} & \text{für } 0 \leq x \leq 20 \\ 0 & \text{für sonst} \end{cases}$$

Wie diese Beispiele zeigen, sind beim Umgang mit stetigen Zufallsvariablen Kenntnisse der Differential- und der Integralrechnung hilfreich. Wir haben daher die Formelsammlung in unserem Online-Service auch um die wichtigsten Regeln aus diesem Gebiet ergänzt.



2.3.3 Zusammenfassende Gegenüberstellung

Da der Zusammenhang zwischen Verteilungs- und Dichtefunktion bei stetigen Zufallsvariablen ebenfalls sehr wichtig für die folgenden Kapitel ist, wollen wir ihn zum Abschluss dieses Abschnitts wie bereits bei diskreten Zufallsvariablen noch einmal übersichtlich grafisch darstellen.

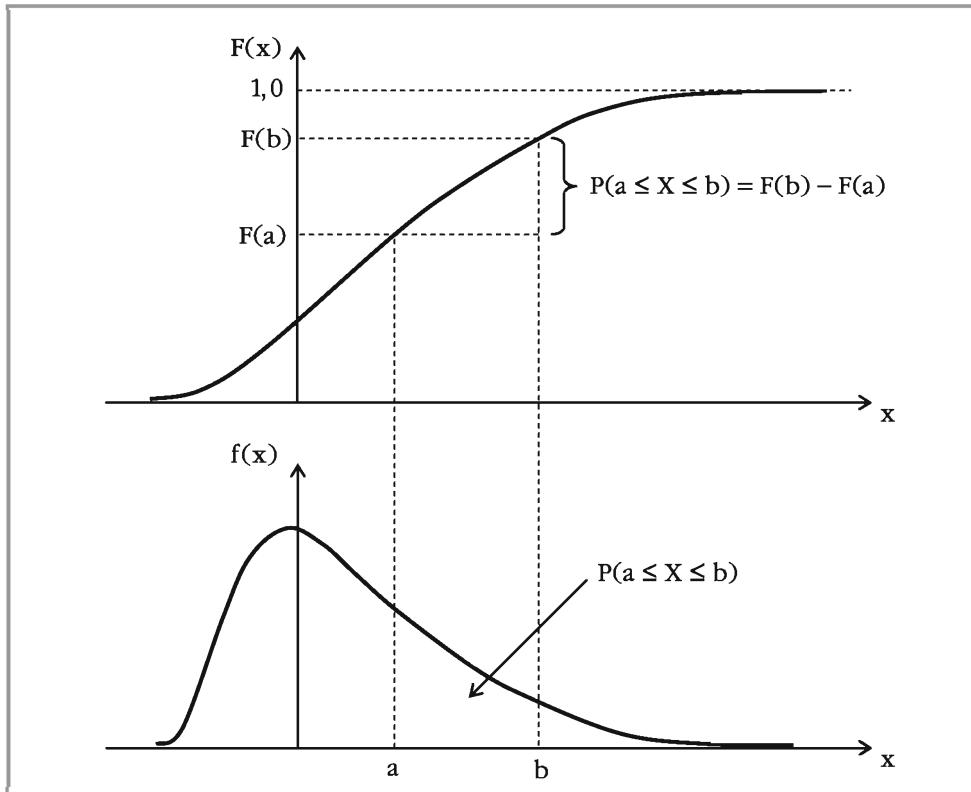


Abbildung II 15: Zusammenhang Verteilungs- und Dichtefunktion

2.4 Kennzahlen von Wahrscheinlichkeitsverteilungen

Wie auch bei den Häufigkeitsverteilungen der deskriptiven Statistik lassen sich für die Charakterisierung von *Wahrscheinlichkeitsverteilungen* von Zufallsvariablen Maßzahlen (Verteilungsparameter) berechnen. Es können Lage- und Streuungsparameter sowie weitere auf Momenten basierende Parameter bei eindimensionalen und Kovarianzen und Korrelationskoeffizienten bei mehrdimensionalen Zufallsvariablen (vgl. Abschnitt II 2.7) ermittelt werden.

2.4.1 Erwartungswert

2.4.1.1 Definition

Der wohl wichtigste Lageparameter der Wahrscheinlichkeitsverteilung einer Zufallsvariablen X ist der analog zum arithmetischen Mittel einer Häufigkeitsverteilung gebildete **Erwartungswert** $E(X)$. Dieser dient insbesondere dazu, den zu erwartenden Wert zu bestimmen, den eine Zufallsvariable bei einem Zufallsexperiment annimmt. Präziser formuliert: Der Parameter beschreibt den Durchschnitt/Mittelwert, den eine Zufallsvariable bei unendlich vielen Durchführungen eines Zufallsexperiments annimmt. Der Erwartungswert charakterisiert damit den Schwerpunkt

der Wahrscheinlichkeitsverteilung (z.B. erwartete Lebensdauer eines Produktes, erwarteter Gewinn bei einem Glücksspiel).⁴

Der Erwartungswert ist für eine **diskrete Zufallsvariable** X mit der Wahrscheinlichkeitsfunktion $f(x)$ als

$$E(X) = \sum_i x_i \cdot P(X = x_i) = \sum_i x_i \cdot f(x_i) \quad (\text{II2: 20})$$

definiert. Diese Formel zeigt deutlich die Verwandtschaft zum *arithmetischen Mittel* einer Häufigkeitsverteilung. Der einzige Unterschied besteht darin, dass die Gewichtung der einzelnen Werte x_i mit der Wahrscheinlichkeit erfolgt, mit der die Zufallsvariable diese Werte annimmt. Die Wahrscheinlichkeiten $f(x_i)$ treten also an die Stelle der relativen Häufigkeiten.

Sind für alle Werte x_i einer diskreten Zufallsvariablen X die dazugehörigen Wahrscheinlichkeiten $P(X = x_i)$ gleich, so kann $E(X)$ vereinfacht als

$$E(X) = x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + \dots + x_n \cdot \frac{1}{n} = \frac{1}{n} \cdot (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (\text{II2: 21})$$

berechnet werden. n ist hierbei die Anzahl der Realisationsmöglichkeiten der Zufallsvariablen X. $E(X)$ ist also interpretierbar als der Mittelwert einer endlichen Grundgesamtheit.

Für eine **stetige Zufallsvariable** X mit der Dichtefunktion $f(x)$ kann der Erwartungswert folgendermaßen ermittelt werden:

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad (\text{II2: 22})$$

Sollte die Dichtefunktion nur in einem Intervall $x_u \leq x \leq x_o$ positive Werte annehmen, müssen wir nur über diesem Intervall die Integration ausführen.

$$E(X) = \int_{x_u}^{x_o} x \cdot f(x) dx \quad (\text{II2: 23})$$

Der Erwartungswert $E(X)$ oder dessen Zahlenwert wird in der Literatur auch häufig mit dem griechischen Buchstaben μ (gesprochen: mü) bezeichnet. Um konkret auszudrücken, zu welcher Zufallsvariable der Erwartungswert gehört, ist dann natürlich eine Form wie μ_x erforderlich.

Beispiel II 2-12

Erwartungswert diskreter Zufallsvariablen

Die folgende Tabelle zeigt die Wahrscheinlichkeit dafür, dass ein Notebook einer bestimmten Marke nach X Jahren funktionsuntüchtig wird.

Jahr x_i	1	2	3	4	5	6	7	8
$P(X = x_i) = f(x_i)$	0	0	0,02	0,18	0,30	0,28	0,10	0,02
$x_i \cdot f(x_i)$	0	0	0,06	0,72	1,50	1,68	0,70	0,16

⁴ Betrachtet man Körper an den Orten x_i mit den Massen $m_i = f(x_i)$ oder eine Massenverteilung mit der Massendichte $f(x)$, dann ist der Erwartungswert tatsächlich der physikalische Schwerpunkt der Massenverteilung.

Wie hoch ist die erwartete Lebensdauer eines Notebooks dieser Marke?

$$E(X) = \sum_i x_i \cdot f(x_i) = 0 + 0 + 0,06 + 0,72 + 1,50 + 1,68 + 0,70 + 0,16 = 4,82 \text{ Jahre}$$

Beispiel II 2-13

Erwartungswert diskreter Zufallsvariablen

Eine Frau schließt eine Risiko-Lebensversicherung über eine Summe von 100.000 Euro und einer Laufzeit von einem Jahr ab. Die Prämie hierfür beträgt 800 Euro. Wie hoch ist der "durchschnittliche Reingewinn" für die Versicherung, wenn die Sterblichkeitswahrscheinlichkeit der Frau mit 5 Promille kalkuliert wird?

Gesucht ist hier der Erwartungswert der Zufallsvariablen X = "Reingewinn der Versicherung". Die Wahrscheinlichkeitsfunktion von X lässt sich wie folgt tabellarisch darstellen:

	Frau stirbt	Frau stirbt nicht
Reingewinn (x_i)	-99.200 (= -100.000 + 800)	800
$P(X = x_i) = f(x_i)$	0,005	0,995

Der erwartete Reingewinn ergibt sich damit als

$$E(X) = \sum_i x_i \cdot f(x_i) = -99.200 \cdot 0,005 + 800 \cdot 0,995 = 300 \text{ Euro.}$$

Beispiel II 2-14

Erwartungswert stetiger Zufallsvariablen

In Beispiel II 2-10 hatten wir eine stetige Zufallsvariable X : "Verspätung einer Straßenbahn in Minuten" mit der Dichtefunktion

$$f(x) = \begin{cases} 0,4 - 0,08x & \text{für } 0 \leq x \leq 5 \\ 0 & \text{sonst} \end{cases}$$

betrachtet. Zur Berechnung ihres Erwartungswertes gehen wir wie folgt vor:

$$\begin{aligned} E(X) &= \int_{x_0}^{x_1} x \cdot f(x) \, dx = \int_0^5 x \cdot (0,4 - 0,08x) \, dx = \int_0^5 (0,4x - 0,08x^2) \, dx \\ &= \left[\frac{0,4}{2} x^2 - \frac{0,08}{3} x^3 \right]_0^5 = 5 - \frac{10}{3} = \frac{5}{3} \text{ Minuten} \end{aligned}$$

Inhaltlich bedeutet dieses Ergebnis, dass bei der Straßenbahn im Durchschnitt von einer Verspätung von 100 Sekunden auszugehen ist.

2.4.1.2 Eigenschaften

Für die Berechnung von Erwartungswerten bzw. den Vergleich von Erwartungswerten verschiedener Zufallsvariablen sind folgende *Regeln und Eigenschaften* von $E(X)$ sehr nützlich:

- Der Erwartungswert der Abweichungen jeder Zufallsvariablen X von ihrem Erwartungswert μ_x ist stets Null.

$$E(X - \mu_x) = 0 \quad (\text{II2: 24})$$

Dies entspricht der im Abschnitt I 2.2.1.3 besprochenen Zentraleigenschaft des arithmetischen Mittels.

2. Die Erwartungswertbildung lässt sich auch auf Funktionen von X ausdehnen. Ist die Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ eine reellwertige Funktion, sodass mit X auch Y = g(X) eine Zufallsvariable ist, so kann der Erwartungswert von Y ohne (direkte) Kenntnis der Verteilung von Y aus der Verteilung von X ermittelt werden.

Im *diskreten Fall* gilt

$$E(Y) = E[g(X)] = \sum_i g(x_i) \cdot f(x_i). \quad (\text{II2: 25})$$

Im *stetigen Fall* gilt

$$E(Y) = E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx. \quad (\text{II2: 26})$$

Beispiel II 2-15

Veranschaulichung von (II2: 25)

Bei einer Maschine treten pro Tag X Störfälle auf. Die Wahrscheinlichkeitsfunktion der Zufallsvariable X: "Anzahl der Störfälle pro Tag" kann durch folgende Tabelle beschrieben werden:

x _i	0	1	2	3
f(x _i)	0,35	0,50	0,10	0,05

Die Kosten (in Tausend Euro), die für die Behebung dieser Störfälle entstehen, werden durch die folgende Kostenfunktion beschrieben:

$$g(x) = 6 - \frac{5}{x+2}$$

Die Ermittlung der erwarteten Kosten erfolgt unter Zuhilfenahme der folgenden Tabelle:

x _i	0	1	2	3
g(x _i)	3,50	4,33	4,75	5,00
f(x _i)	0,35	0,50	0,10	0,05

Da die Kosten eines jeden einzelnen Störfalles mit ihrer Wahrscheinlichkeit bewertet werden müssen, wäre es falsch einfach $E(X) = 0,85$ in die Kostenfunktion einzusetzen ($g[E(X)] = 4,25$). Die korrekte Berechnung der erwarteten Kosten der Störfallbeseitigung pro Tag ergibt stattdessen

$$E[g(x)] = 3,50 \cdot 0,35 + 4,33 \cdot 0,50 + 4,75 \cdot 0,10 + 5,00 \cdot 0,05 = 4,12.$$

Der Zusammenhang zwischen $E[g(X)]$ und $g[E(X)]$ wird im Rahmen der *Jensenschen Ungleichung* (II2: 27) bzw. (II2: 28) noch näher betrachtet.

Für einige wichtige **lineare Transformationen von Zufallsvariablen** sind die Erwartungswerte in folgender Tabelle zusammengestellt. Die dargestellten Zusammenhänge gelten für alle $a, b \in \mathbb{R}$.

	Y bzw. g(X)	E(Y) bzw. E(g(X))
(A)	a	a
(B)	b · X	b · E(X)
(C)	a + X	a + E(X)
(D)	a + b · X	a + b · E(X)

Wir erkennen daran sehr schön die gleiche Transformationseigenschaft wie beim arithmetischen Mittel (vgl. Abschnitt I 2.2.1.3) und können daher auf die Erbringung eines formalen Beweises verzichten.

Beispiel II 2-16

Erwartungswert und Lineartransformation

Betrachten wir die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen X:

x_i	-2	-1	0	1	2	3
$f(x_i)$	0,05	0,10	0,30	0,40	0,05	0,10

Für den Erwartungswert dieser Zufallsvariablen gilt

$$E(X) = (-2) \cdot 0,05 + (-1) \cdot 0,10 + 0 \cdot 0,30 + 1 \cdot 0,40 + 2 \cdot 0,05 + 3 \cdot 0,10 = 0,6.$$

Multiplizieren wir nun die Zufallsvariable X bzw. deren Realisationen mit der Konstanten $b = 2$ und addieren die Konstante $a = 1$, so ergibt sich der neue Erwartungswert

$$\begin{aligned} E(1+2 \cdot X) &= [1+2 \cdot (-2)] \cdot 0,05 + [1+2 \cdot (-1)] \cdot 0,10 + [1+2 \cdot 0] \cdot 0,30 + [1+2 \cdot 1] \cdot 0,40 \\ &\quad + [1+2 \cdot 2] \cdot 0,05 + [1+2 \cdot 3] \cdot 0,10 = 2,2. \end{aligned}$$

Unter Verwendung des Zusammenhangs $E(a + b \cdot X) = a + b \cdot E(X)$ lässt sich dieses Ergebnis natürlich viel schneller bestimmen:

$$E(1+2 \cdot X) = 1 + 2 \cdot E(X) = 1 + 2 \cdot 0,6 = 2,2$$

Beispiel II 2-17

Beweis der Zentraleigenschaft von $E(X)$

Die Beziehung $E(a + X) = a + E(X)$ erlaubt für $a = -\mu_x$ die Herleitung der Zentraleigenschaft des Erwartungswertes (II2: 25). Es gilt nämlich

$$E(X - \mu_x) = E(X) - \mu_x = 0.$$

- Ist die Wahrscheinlichkeits- bzw. Dichtefunktion $f(x)$ von X *symmetrisch* bezüglich eines Punktes $x = a$, d.h. ist $f(a - x) = f(a + x)$ für alle x , so gilt $E(X) = a$ (falls der Erwartungswert existiert). Abbildung II 16 veranschaulicht dies.

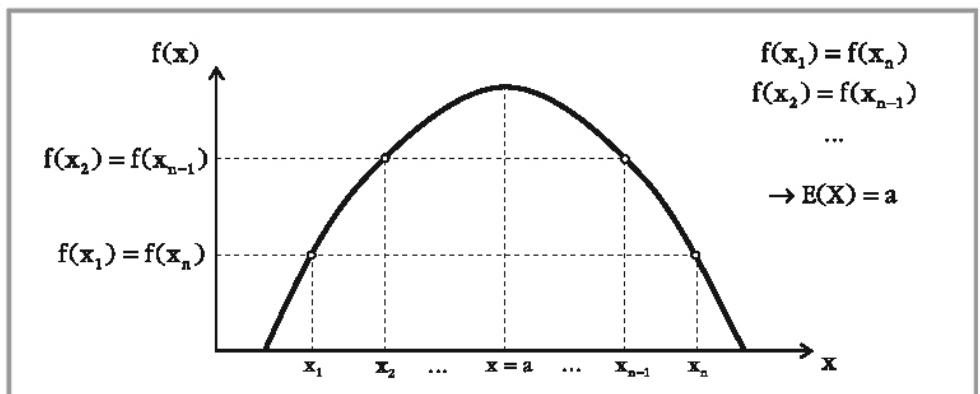


Abbildung II 16: Symmetrie und Erwartungswert (stetige ZV)

4. Gilt für jedes Elementarereignis ω der zugrunde liegenden Zufallsexperimente zweier Zufallsvariablen X und Y die Ungleichung $X(\omega) \leq Y(\omega)$, so folgt auch für die Erwartungswerte $E(X) \leq E(Y)$.
5. Ist g(X) eine konvexe oder konkave Funktion der Zufallsvariablen X, so gilt die sog. **Jensensche Ungleichung**. Diese ist definiert als

$$E[g(X)] \geq g[E(X)], \text{ falls } g(X) \text{ konvex} \quad (\text{II2: 27})$$

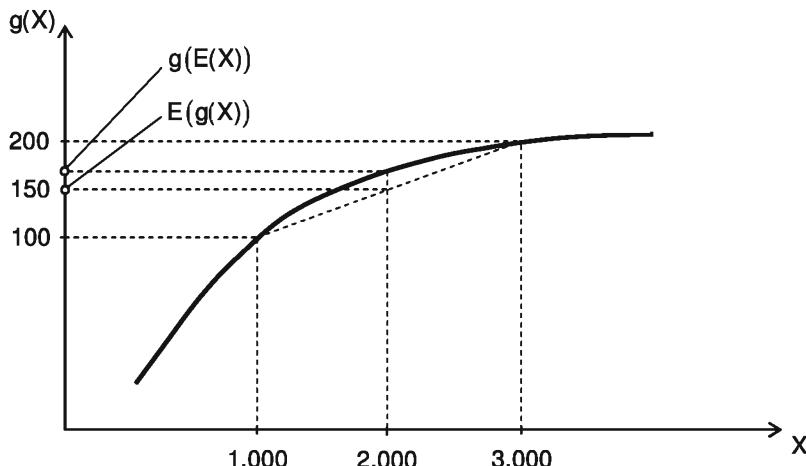
bzw.

$$E[g(X)] \leq g[E(X)], \text{ falls } g(X) \text{ konkav}. \quad (\text{II2: 28})$$

Beispiel II 2-18

Jensensche Ungleichung

Eine Person nimmt an einem Glücksspiel teil, bei dem sie jeweils mit einer Wahrscheinlichkeit von 50 % entweder 1.000 Euro oder 3.000 Euro gewinnen kann. Die Zufallsvariable X, welche ihren Gewinn beschreiben soll, kann daher nur diese zwei Werte annehmen. Der Erwartungswert des Gewinns liegt bei $E(X) = 1.000 \cdot 0,5 + 3.000 \cdot 0,5 = 2.000$ Euro. Die konkave Funktion g(X) beschreibe das Nutzenniveau, welches diese (risikoaverse) Person aus einem Gewinn zieht, und stellt sich wie folgt dar:



Untersuchen wir nun für dieses Beispiel die Gültigkeit der Jensenschen Ungleichung (II2: 28). Halten wir zunächst fest, dass der Erwartungswert des Nutzens

$$E(g(X)) = 100 \cdot 0,5 + 200 \cdot 0,5 = 150$$

und der Nutzen des erwarteten Einkommens offensichtlich

$$g(E(X)) = g(2.000)$$

ist. Betrachten wir diese Werte und die obige Grafik, können wir klar feststellen, dass

$$E(g(X)) = 150 < g(E(X)) = g(2.000).$$

2.4.2 Varianz und Standardabweichung

2.4.2.1 Definition

Wie stark die Verteilung einer Zufallsvariablen X um ihren Mittelwert (Erwartungswert) streut, kann durch ihre Streuungsparameter angegeben werden. Die Streuungsparameter, mit denen wir uns im Folgenden beschäftigen, sind die Varianz $\text{Var}(X)$ und die daraus zu gewinnende Standardabweichung.

Ist X eine Zufallsvariable und μ_x ihr Erwartungswert, so wird der Erwartungswert der quadrierten Abweichungen der Zufallsvariablen von ihrem Erwartungswert als **Varianz** der Zufallsvariablen X bezeichnet. Es gilt also

$$\text{Var}(X) = E[(X - \mu_x)^2]. \quad (\text{II2: 29})$$

An Stelle der Notation $\text{Var}(X)$ oder um ihren Zahlenwert zu bezeichnen, verwendet man häufig das Symbol σ^2 (gesprochen: sigma quadrat). Um auszudrücken, auf welche Zufallsvariable sich die Varianz bezieht, schreibt man oft auch σ_x^2 .

Konkret berechnet wird die Varianz im *diskreten Fall* mit der Formel

$$\text{Var}(X) = \sum_i [x_i - E(X)]^2 \cdot f(x_i). \quad (\text{II2: 30})$$

Aus dieser Formel ist deutlich zu erkennen, dass die Varianz einer Zufallsvariablen X in ihrer Konstruktion der *mittleren quadratischen Abweichung* einer Häufigkeitsverteilung entspricht (vgl. Abschnitt I 2.2.2.3).

Ist X eine diskrete Zufallsvariable, die alle Werte x_1, x_2, \dots, x_n mit gleicher Wahrscheinlichkeit annimmt, so gilt für die Varianz vereinfacht

$$\text{Var}(X) = \frac{1}{n} \cdot \sum_i [x_i - E(X)]^2. \quad (\text{II2: 31})$$

Im *stetigen Fall* errechnet sich die Varianz konkret als

$$\text{Var}(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 \cdot f(x) dx \quad (\text{II2: 32})$$

bzw.

$$\text{Var}(X) = \int_{x_u}^{x_o} [x - E(X)]^2 \cdot f(x) dx. \quad (\text{II2: 33})$$

Beispiel II 2-19

Varianz diskreter Zufallsvariablen

Die Zufallsvariable X : "Anzahl der Köpfe beim zweimaligen Münzwurf" besitzt folgende Wahrscheinlichkeitsfunktion:

x_i	0	1	2
$f(x_i)$	0,25	0,50	0,25

Wegen der Symmetrie der Verteilung besitzt die Zufallsvariable X den Erwartungswert $E(X) = 1$. Für ihre Varianz gilt

$$\text{Var}(X) = (0 - 1)^2 \cdot 0,25 + (1 - 1)^2 \cdot 0,50 + (2 - 1)^2 \cdot 0,25 = 0,50.$$

Beispiel II 2-20**Varianz stetiger Zufallsvariablen**

In Beispiel II 2-14 hatten wir den Erwartungswert einer stetigen Zufallsvariable X: "Verspätung einer Straßenbahn in Minuten" mit der Dichtefunktion

$$f(x) = \begin{cases} 0,4 - 0,08x & \text{für } 0 \leq x \leq 5 \\ 0 & \text{sonst} \end{cases}$$

als $E(X) = \frac{5}{3}$ Minuten bestimmt. Die zugehörige Varianz erhalten wir wie folgt:

$$\begin{aligned} \text{Var}(X) &= \int_{x_0}^{x_1} (x - E(X))^2 \cdot f(x) dx = \int_0^5 (x - \frac{5}{3})^2 \cdot (0,4 - 0,08x) dx \\ &= \int_0^5 (x^2 - \frac{10}{3}x + \frac{25}{9}) \cdot (0,4 - 0,08x) dx \\ &= \int_0^5 (0,4x^2 - 0,08x^3 - \frac{4}{3}x + \frac{0,8}{3}x^2 + \frac{10}{9} - \frac{2}{9}x) dx \\ &= \int_0^5 (-0,08x^3 + \frac{2}{3}x^2 - \frac{14}{9}x + \frac{10}{9}) dx \\ &= \left[-0,02x^4 + \frac{2}{9}x^3 - \frac{7}{9}x^2 + \frac{10}{9}x \right]_0^5 = -12,5 + \frac{250}{9} - \frac{175}{9} + \frac{50}{9} \approx 1,39 \text{ Minuten}^2 \end{aligned}$$

2.4.2.2 Eigenschaften

Für die Berechnung und Interpretation der Varianz sind folgende *Regeln und Eigenschaften* von besonderem Nutzen:

1. Die Varianz ist abgesehen von dem trivialen Sonderfall einer Zufallsvariablen, die stets denselben Wert annimmt, *positiv*. Beim erwähnten Sonderfall nimmt sie einen Wert von 0 an.
2. Die positive Wurzel aus der Varianz einer Zufallsvariablen X bezeichnet man als *Standardabweichung* σ_x der Zufallsvariablen X.

$$\sigma_x = \sqrt{\text{Var}(X)} \quad (\text{II2: 34})$$

An Stelle dieser Notation können wir auch kurz σ verwenden, wenn ein Bezug zur Variable X nicht relevant ist.

3. Der für die mittlere quadratische Abweichung bekannte *Verschiebungssatz* (vgl. Abschnitt I 2.2.2.3) gilt auch für die Varianz von Zufallsvariablen. Er besagt hier

$$\text{Var}(X) = E(X^2) - \mu_x^2 \quad (\text{II2: 35})$$

Konkret bedeutet dies für diskrete Zufallsvariablen

$$\text{Var}(X) = \sum_i x_i^2 \cdot f(x_i) - [E(X)]^2 \quad (\text{II2: 36})$$

bzw. stetige Zufallsvariablen

$$\text{Var}(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x) dx - [E(X)]^2. \quad (\text{II2: 37})$$

Beispiel II 2-21**Vereinfachte Varianzberechnung diskreter ZV**

Kurz vor Saisonende hat eine Fußballmannschaft noch 3 Spiele zu absolvieren. Dem Trainer und begeisterten Statistiker gelang es, die Wahrscheinlichkeitsfunktion für die Zufallsvariable X: "Anzahl der Siege bis Saisonende" aufzustellen:

x_i	0	1	2	3
$f(x_i)$	0,1	0,2	0,3	0,4

Wie stark streut die Anzahl der Siege um die erwartete Anzahl $E(X) = 2$?

$$\text{Var}(X) = \sum_i x_i^2 \cdot f(x_i) - [E(X)]^2 = 0^2 \cdot 0,1 + 1^2 \cdot 0,2 + 2^2 \cdot 0,3 + 3^2 \cdot 0,4 - 2^2 = 5 - 4 = 1$$

Beispiel II 2-22**Vereinfachte Varianzberechnung stetiger ZV**

Berechnen wir die Varianz der stetigen Zufallsvariable X: "Verspätung einer Straßenbahn in Minuten" aus Beispiel II 2-20 nun mittels der aus dem Verschiebungssatz resultierenden Formel:

$$\begin{aligned} \text{Var}(X) &= \int_{x_0}^{x_5} x^2 \cdot f(x) dx - [E(X)]^2 = \int_0^5 x^2 \cdot (0,4 - 0,08x) dx - (\frac{5}{3})^2 \\ &= \int_0^5 (0,4x^2 - 0,08x^3) dx - \frac{25}{9} = \left[\frac{0,4}{3}x^3 - 0,02x^4 \right]_0^5 - \frac{25}{9} \\ &= \frac{50}{3} - 12,5 - \frac{25}{9} \approx 1,39 \text{ Minuten}^2 \end{aligned}$$

Wie zu erwarten war, stimmen die Ergebnisse beider Berechnungsvarianten überein.

4. Die Bildung von Varianzen lässt sich auch auf Funktionen von X ausdehnen. Ist die Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ eine reellwertige Funktion, sodass mit X auch Y = g(X) eine Zufallsvariable ist, so kann die Varianz von Y ohne Kenntnis der Verteilung von Y aus der Verteilung von X ermittelt werden. Für einige wichtige lineare Funktionen sind die Varianzen (inklusive Standardabweichungen) in der folgenden Tabelle zusammengestellt:

Y bzw. g(X)	Var(Y) bzw. Var(g(X))	σ_y bzw. $\sigma_{g(x)}$
(A) a	0	0
(B) b · X	$b^2 \cdot \text{Var}(X)$	$ b \cdot \sigma_x$
(C) a + X	$\text{Var}(X)$	σ_x
(D) a + b · X	$b^2 \cdot \text{Var}(X)$	$ b \cdot \sigma_x$

Wir erkennen daran die gleiche Transformationseigenschaft wie auch bei der Varianz bzw. der Standardabweichung aus der deskriptiven Statistik (vgl. Abschnitt I 2.2.2.3) und können daher auf formale Beweise verzichten.

Wir können aus obiger Tabelle folgende allgemeine Schlüsse ziehen: (1) Eine Konstante kann unmöglich streuen. Ihre Varianz muss daher Null sein. (2) Addieren wir zu jedem Wert x eine Konstante a, dann führt dies zu einer Verschiebung der Verteilung nach links ($a < 0$) oder rechts ($a > 0$). Die Verschie-

bung der Verteilung um den konstanten Betrag a verändert aber ihre Streuung nicht, da auch der Erwartungswert um a verschoben wird. (3) Wird jeder Wert x einer Zufallsvariable mit einem konstanten Faktor b multipliziert, so führt dies zu einer Streckung oder Stauchung der Verteilung. Ist $|b| > 1$ spricht man von *Streckung*, bei $|b| < 1$ von *Stauchung* der Verteilung.

Beispiel II 2-23**Varianz und lineare Transformation**

Betrachten wir eine diskrete Zufallsvariable X mit folgender Verteilung:

x_i	490	500	510
$f(x_i)$	0,2	0,5	0,3

Um die Varianz von X einfacher berechnen zu können, könnte nun X mittels

$$Y = 0,1 \cdot X - 50$$

transformiert werden. Die neue Zufallsvariable Y hätte damit die folgende Verteilung:

y_i	-1	0	+1
$f(y_i)$	0,2	0,5	0,3

Erwartungswert und Varianz können wir nun für diese neue Verteilung bestimmen als

$$E(Y) = (-1) \cdot 0,2 + 0 \cdot 0,5 + 1 \cdot 0,3 = 0,1$$

$$\text{Var}(Y) = (-1)^2 \cdot 0,2 + 0^2 \cdot 0,5 + 1^2 \cdot 0,3 - 0,1^2 = 0,49.$$

Daraus wird schließlich $\text{Var}(X)$ durch Rücktransformation ermittelt:

$$\text{Var}(X) = \text{Var}(10 \cdot Y + 500) = 10^2 \cdot \text{Var}(Y) = 100 \cdot 0,49 = 49$$

Ein wichtiger Anwendungsfall dieser linearen Transformation ist die sog. *Standardisierung von Zufallsvariablen*, auf welche im Abschnitt II 2.4.2.3 noch separat eingegangen wird.

5. Auch für die Varianz gilt die sog. *Minimaleigenschaft*, die wir bereits von der mittleren quadratischen Abweichung kennen. Sie besagt, dass immer

$$E[(X - \mu_x)^2] < E[(X - h)^2] \quad \text{für} \quad h \neq \mu_x \quad (\text{II2: 38})$$

gilt, egal ob h größer oder kleiner als μ_x ist. Für $h = \mu_x$ nimmt $E[(X - h)^2]$ also sein Minimum an.

2.4.2.3 Standardisierung von Zufallsvariablen

Wird eine stetige oder diskrete Zufallsvariable X mit dem Erwartungswert μ_x und der Varianz σ_x^2 bzw. der Standardabweichung σ_x in der Form

$$Z = a + b \cdot X \quad \text{mit} \quad a = -\frac{\mu_x}{\sigma_x} \quad \text{und} \quad b = \frac{1}{\sigma_x}$$

bzw.

$$Z = \frac{X - \mu_x}{\sigma_x} \quad \text{mit} \quad \sigma_x > 0 \quad (\text{II2: 39})$$

linear transformiert, dann heißt die dabei entstehende transformierte Zufallsvariable Z **standardisiert**. Die gleiche Art von Standardisierung ist uns bereits in der deskriptiven Statistik begegnet (vgl. Abschnitt I 2.2.2.3), wobei wir hier nun \bar{x} durch μ_x und s_x durch σ_x ersetzen. Die besondere Eigenschaft einer standardisierten Zufallsvariable Z ist ein *Erwartungswert von Null* und eine *Varianz bzw. eine Standardabweichung von Eins*. Die folgende Tabelle stellt die besonders im Zusammenhang mit der Normalverteilung (vgl. Abschnitt II 3.2.3) wichtige Standardisierung noch einmal kompakt dar. Der erste Schritt $X \rightarrow Y$ (Verschiebung) dient dabei der Generierung eines Erwartungswertes von Null, während der zweite Schritt $Y \rightarrow Z$ (Streckung) die Varianz bzw. die Standardabweichung von Eins herbeiführt.

Zufallsvariable	X	\rightarrow	$Y = X - \mu_x$	\rightarrow	$Z = \frac{X - \mu_x}{\sigma_x}$
Erwartungswert	μ_x		0		0
Varianz	σ_x^2		σ_x^2		1
Standardabweichung	σ_x		σ_x		1

Beispiel II 2-24

Standardisierung von Zufallsvariablen

Betrachten wir die Wahrscheinlichkeitsfunktion der folgenden diskreten Zufallsvariablen X :

x_i	10	11	12
$f(x_i)$	0,1	0,5	0,4

Ihr Erwartungswert μ_x liegt bei 11,30 und ihre Standardabweichung σ_x bei 0,6403. Wird nun eine Standardisierung von X mittels

$$Z = \frac{X - 11,30}{0,6403}$$

durchgeführt, so ergibt sich folgende neue Verteilung mit $\mu_z = 0$ und $\sigma_z^2 = 1$ bzw. $\sigma_z = 1$:

z_i	-2,0303	-0,4685	1,0932
$f(z_i)$	0,1	0,5	0,4

Wie das Beispiel II 2-24 gezeigt hat, ändert sich im Falle einer *diskreten Zufallsvariablen* X durch die Standardisierung das Wahrscheinlichkeitsmuster nicht, d.h. es gilt für die Wahrscheinlichkeitsfunktion $f(z)$ der standardisierten Zufallsvariablen Z

$$f(z_i) = f(x_i) \quad \text{für alle } i. \quad (\text{II2: 40})$$

Handelt es sich bei der zu standardisierenden Zufallsvariablen X nun aber um eine *stetige Zufallsvariable*, so gilt für deren Dichtefunktion nach der Kettenregel

$$f(x) = \frac{dF(x)}{dx} = \frac{dF(z)}{dz} \cdot \frac{dz}{dx} = f(z) \cdot \frac{1}{\sigma_x}.$$

Die *Dichtefunktion* $f(z)$ der neuen Zufallsvariablen Z ergibt sich daher als

$$f(z) = f(x) \cdot \frac{1}{\sigma_x}. \quad (\text{II2: 41})$$

Beispiel II 2-25**Standardisierung und Dichtefunktion**

Die Standardabweichung einer stetigen Zufallsvariablen X sei $\sigma_x = 4$. Wird diese Zufallsvariable nun standardisiert, führt dies dazu, dass die Dichtefunktion in x -Richtung auf ein Viertel zusammengedrückt wird. Damit die Fläche unter der Dichtefunktion gleich Eins bleiben kann, wird $f(z)$ gleichzeitig viermal so hoch sein.

Der Vorteil einer Standardisierung liegt darin, dass die entstehenden Werte z_i eine Reihe nützlicher *Interpretationen* zulassen, die wir bereits aus der deskriptiven Statistik kennen. Ist z_i nämlich positiv, so zeigt dies, dass x_i über dem Erwartungswert μ_x liegt, d.h. $z_i > 0 \rightarrow x_i > \mu_x$. Analog gilt auch $z_i < 0 \rightarrow x_i < \mu_x$. Zudem gibt z_i an, um welches Vielfache der Standardabweichungen σ_x die Realisation x_i von μ_x abweicht.

2.4.3 Höhere Momente

Ist X eine Zufallsvariable mit einer gegebenen Verteilung und k eine natürliche Zahl, so heißt der Erwartungswert der k -ten Potenz

$$E(X^k) \quad (\text{II2: 42})$$

das k -te *Moment* der Verteilung. Der Erwartungswert der k -ten Potenz der Abweichung vom Erwartungswert

$$E[(X - \mu)^k] \quad (\text{II2: 43})$$

heißt k -tes *zentrales Moment*.⁵ Bei der Charakterisierung von Verteilungen sind vor allem das erste Moment $E(X)$ und das zweite zentrale Moment $E[(X - \mu)^2] = \text{Var}(X)$ von Bedeutung. Der Erwartungswert beschreibt die Lage, die Varianz die Streuung der Verteilung. Höhere Momente messen andere Eigenschaften.

Wie in der deskriptiven Statistik (vgl. Abschnitt I 2.2.3) wollen wir nun auch für Zufallsvariablen die auf zentralen Momenten basierenden Kennzahlen der Schiefe γ (engl. skewness) und der Wölbung κ (engl. kurtosis) einführen:

$$\gamma = \frac{E[(X - \mu)^3]}{\sigma^3} \quad (\text{II2: 44})$$

$$\kappa = \frac{E[(X - \mu)^4]}{\sigma^4} \quad (\text{II2: 45})$$

Das *dritte zentrale Moment* $E[(X - \mu)^3]$ lässt erkennen, ob die Verteilung *symmetrisch oder schief* ist. Wie alle ungeraden zentralen Momente nimmt auch dieses bei einer symmetrischen Verteilung den Wert Null an. Dies liegt daran, dass durch eine ungerade Potenz das Vorzeichen einer Abweichung nicht verloren geht und sich die Abweichungen so gegenseitig aufheben können, wenn die Verteilung symmetrisch zu ihrem Mittelwert ist. Das *vierte zentrale Moment* $E[(X - \mu)^4]$ gibt Auskunft über den Grad der *Wölbung* einer Verteilung.

⁵ Zur konkreten Berechnung von Momenten und zentralen Momenten sind insbesondere die Formeln (II2: 25) und (II2: 26) mit $g(X) = X^k$ bzw. $g(X) = (X - \mu)^k$ relevant.

Da das dritte und vierte zentrale Moment von der Maßeinheit, in der eine Zufallsvariable gemessen wird, und ihrer Streuung beeinflusst wird, sind sie allein jedoch keine sehr geeigneten Maße für die Asymmetrie und die Wölbung einer Verteilung. Die zentralen Momente werden daher zur Ausschaltung dieser Einflüsse durch die dritte bzw. vierte Potenz der Standardabweichung dividiert.

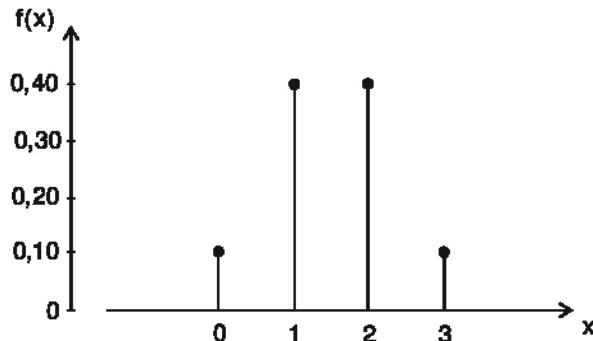
Für symmetrische Verteilungen ist $\gamma = 0$, für rechtsschiefe $\gamma > 0$ und für linksschiefe $\gamma < 0$.⁶ Für die besonders in der induktiven Statistik bedeutende Normalverteilung, die wir unter II 3.2.3 ausführlich diskutieren werden, gilt $\gamma = 0$ und $\kappa = 3$. Verteilungen mit größeren κ -Werten haben schlankere Dichtefunktionen mit dickeren (fetten) Enden (engl. sog. fat tails), solche mit kleineren κ -Werten besitzen dünneren Enden als die Normalverteilung.

Beispiel II 2-26

Höhere Momente



Betrachten wir die einfache diskrete Zufallsvariable X , die nur die Ausprägungen 0, 1, 2 und 3 annehmen kann und deren Wahrscheinlichkeitsfunktion sich wie folgt darstellt:



Diese Verteilung besitzt einen Erwartungswert von $\mu = 1,5$ und eine Varianz bzw. Standardabweichung von $\sigma^2 = 0,65$ bzw. $\sigma = 0,806$. Schiefe und Kurtosis ergeben sich zu

$$\gamma = \frac{(0-1,5)^3 \cdot 0,1 + (1-1,5)^3 \cdot 0,4 + (2-1,5)^3 \cdot 0,4 + (3-1,5)^3 \cdot 0,1}{0,806^3} = 0$$

$$\kappa = \frac{(0-1,5)^4 \cdot 0,1 + (1-1,5)^4 \cdot 0,4 + (2-1,5)^4 \cdot 0,4 + (3-1,5)^4 \cdot 0,1}{0,806^4} = 2,51.$$

Die Verteilung ist damit symmetrisch und weist eine kleinere Wölbung als die Normalverteilung auf.

2.4.4 Quantile

Neben den aus den Momenten und den zentralen Momenten resultierenden Maßzahlen dienen auch Quantile als Kenngrößen von Wahrscheinlichkeitsverteilungen.⁷ Als Spezialfall ist hierbei vor allem der Median (50 %-Quantil) interessant.

⁶ Vgl. dazu insbesondere auch die Ausführungen unter I 2.2.1.6 und I 2.2.3.

⁷ Gelegentlich ist auch der Modus x_{Mod} einer Zufallsvariable X relevant. Er ist der Wert x , für den die Wahrscheinlichkeits- bzw. Dichtefunktion $f(x)$ maximal wird.

Der **Median** x_{Med} einer stetigen oder diskreten Zufallsvariablen X ist dadurch definiert, dass X jeweils mit mindestens 50 % Wahrscheinlichkeit größer gleich bzw. kleiner gleich x_{Med} ist. Es gilt also die allgemeine Mediandefinition

$$P(X \geq x_{\text{Med}}) \geq 0,5 \quad \text{und} \quad P(X \leq x_{\text{Med}}) \geq 0,5. \quad (\text{II2: 46})$$

Die Bestimmung des Median bei *diskreten Zufallsvariablen* erfolgt wie in der deskriptiven Statistik (vgl. Abschnitt I 2.2.1.2), d.h. es wird der kleinste x -Wert mit $F(x) > 0,5$ gewählt. Es kann dabei jedoch wieder der Fall auftreten, dass alle Werte zwischen zwei Sprungstellen die Bedingung (II2: 46) erfüllen. Um einen eindeutigen Wert zu erhalten, wählt man dann in den meisten Fällen das arithmetische Mittel aus beiden Sprungstellen.

Im Falle einer *stetigen Zufallsvariable* kann der Median wie in Abbildung II 17 aus der Verteilungsfunktion $F(x)$ ermittelt werden. Existiert die Umkehrfunktion F^{-1} der Verteilungsfunktion bei $F(x) = 0,5$, so ist zur Bestimmung des Median also nur zu prüfen, an welcher Stelle x die Verteilungsfunktion $F(x)$ den Wert 0,5 annimmt:

$$P(X \leq x_{\text{Med}}) = F(x_{\text{Med}}) = 0,5 \quad (\text{II2: 47})$$

Wir sehen also, dass sich die Bedingungen (II2: 46) im Falle einer stetigen Zufallsvariablen zu (II2: 47) vereinfachen.

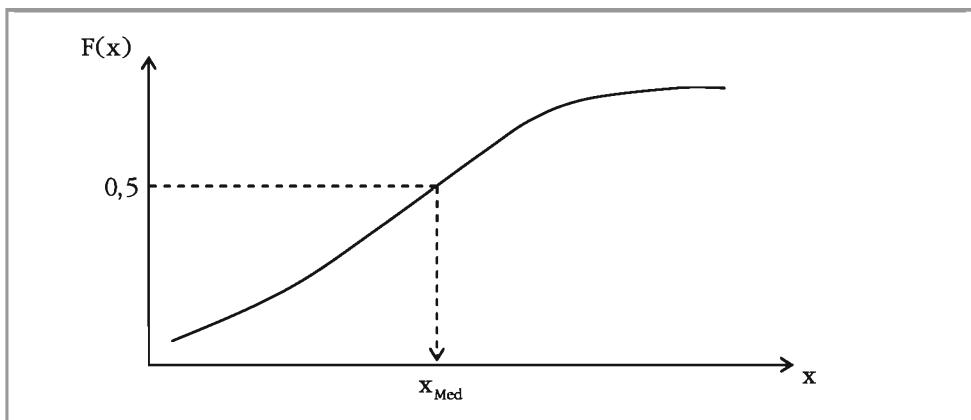


Abbildung II 17: Bestimmung des Median (stetige ZV)

Beispiel II 2-27

Median bei stetigen ZV

In Beispiel II 2-10 hatten wir für die stetige Zufallsvariable X : "Verspätung einer Straßenbahn in Minuten" die Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0,4x - 0,04x^2 & \text{für } 0 \leq x \leq 5 \\ 1 & \text{für } x > 5 \end{cases}$$

bestimmt. Ermitteln wir nun den Median für diese Verteilung, indem wir den relevanten Ast von $F(x)$ gleich 0,5 setzen und die resultierende Gleichung nach x auflösen:

$$0,4x - 0,04x^2 = 0,5 \quad \rightarrow \quad -0,04x^2 + 0,4x - 0,5 = 0$$

Über die Lösungsformel für quadratische Gleichungen erhalten wir daraus

$$x_{1,2} = \frac{-0,4 \pm \sqrt{0,4^2 - 4 \cdot (-0,04) \cdot (-0,5)}}{2 \cdot (-0,04)} \rightarrow x_1 = 1,46 \quad x_2 = 8,54.$$

Da nur x_1 innerhalb des Definitionsbereiches der "Teilfunktion" $0,4x - 0,04x^2$ liegt, ist die Lösung $x_1 = 1,46$ der Median.

Ist X eine *stetige Zufallsvariable*, so gibt es zu jeder Zahl α mit $0 < \alpha < 1$ mindestens einen Wert x_α , sodass $X \leq x_\alpha$ genau mit der Wahrscheinlichkeit α eintritt. Ein solches x_α wird als **α -Fraktil** oder **α -Quantil** der Verteilung von X bezeichnet:

$$P(X \leq x_\alpha) = F(x_\alpha) = \alpha \quad (\text{II2: 48})$$

Fraktilswerte spielen besonders in der induktiven Statistik bei Konfidenzintervallen und statistischen Tests eine große Rolle, weshalb ihrer Bestimmung aus speziellen Verteilungen (vgl. Kapitel II 3) eine besondere Bedeutung zukommt.

Beispiel II 2-28

Quantile bei stetigen ZV

Bestimmen wir für die stetige Zufallsvariable X aus Beispiel II 2-27 das 98 %-Fraktil der Verteilung von X . Wir müssen dazu (analog zum Median) lediglich den relevanten Ast der Verteilungsfunktion $F(x)$ gleich 0,98 setzen und nach x auflösen. Wir erhalten konkret

$$0,4x - 0,04x^2 = 0,98 \rightarrow x_1 = 4,29 \quad x_2 = 5,71.$$

Da nur x_1 im Definitionsbereich der "Teilfunktion" $0,4x - 0,04x^2$ liegt, gilt $x_{0,98} = 4,29$.

Da die Quantilsdefinition (II2: 48) i.d.R. nur für stetige Zufallsvariablen einen wohldefinierten Wert liefert, müssen wir die Definition zur Anwendung auf *diskrete Zufallsvariablen* verallgemeinern. In Anlehnung an die allgemeine Mediandefinition (II2: 46) gilt

$$P(X \leq x_\alpha) \geq \alpha \quad \text{und} \quad P(X \geq x_\alpha) \geq 1 - \alpha. \quad (\text{II2: 49})$$

Auch hier kann sich nun wieder der Fall einstellen, dass alle Werte zwischen zwei Sprungstellen a und b diese Bedingung erfüllen. Jeder Wert im Intervall $a \leq x_\alpha \leq b$ ist also ein α -Quantil. Um einen eindeutigen Wert zu erhalten, wählt man auch hier in den meisten Fällen das arithmetische Mittel aus beiden Sprungstellen. Wir verzichten an dieser Stelle auf ein konkretes Beispiel, da die Vorgehensweise derjenigen der Quantilsbestimmung aus der deskriptiven Statistik (vgl. Abschnitt I 2.1.5) gleicht.

2.5 Ungleichung von Tschebyscheff

Ist X eine stetige oder diskrete Zufallsvariable mit dem Erwartungswert μ_x und der Standardabweichung σ_x , so gilt für jedes $k > 0$ und *unabhängig von der Verteilungsfunktion* stets die *Ungleichung von Tschebyscheff*. Diese besagt

$$P(\mu_x - k \cdot \sigma_x < X < \mu_x + k \cdot \sigma_x) \geq 1 - \frac{1}{k^2}. \quad (\text{II2: 50})$$

In der Literatur häufig zu findende alternative Schreibweisen zum linken Term im Ausdruck (II2: 50) sind

$$P(-k \cdot \sigma_x < X - \mu_x < k \cdot \sigma_x)$$

und

$$P(|X - \mu_x| < k \cdot \sigma_x).$$

Mit der Ungleichung von Tschebyscheff lässt sich ohne Kenntnis der Verteilung von X abschätzen, wie groß die Mindestwahrscheinlichkeit dafür ist, dass die Zufallsvariable X um einen Wert von kleiner als $k \cdot \sigma_x$ von ihrem Erwartungswert abweicht.⁸ Es sei aber erwähnt, dass die Ungleichung nur *grobe Abschätzungen* liefert und so keine praktische Schätzmethode für Wahrscheinlichkeiten darstellt. Für eine stetige Zufallsvariable X lässt sich der Bereich (bzw. die Wahrscheinlichkeit), der mit der Ungleichung von Tschebyscheff abgeschätzt werden kann, wie in Abbildung II 18 grafisch darstellen.

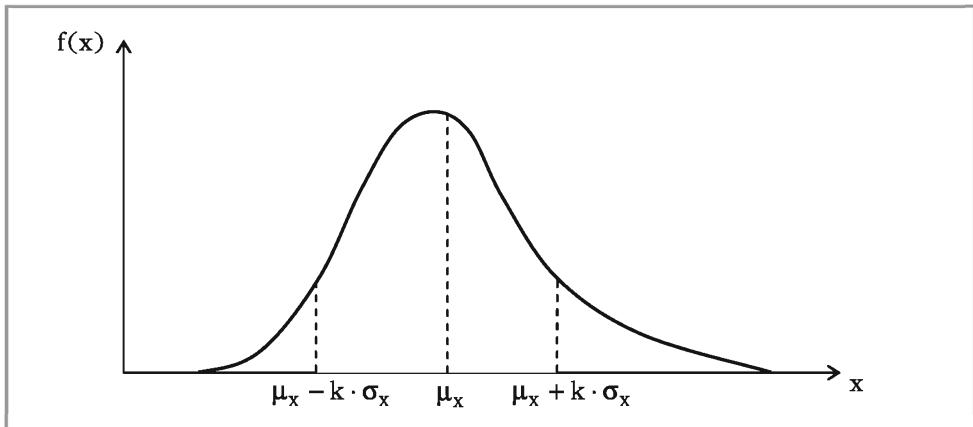


Abbildung II 18: Ungleichung von Tschebyscheff

Beispiel II 2-29

Ungleichung von Tschebyscheff I

Für ausgewählte Werte von k erhalten wir die folgenden Abschätzungen:

$k = 1,5 :$	$P(\mu_x - 1,5 \cdot \sigma_x < X < \mu_x + 1,5 \cdot \sigma_x)$	\geq	0,5555...
$k = 2 :$	$P(\mu_x - 2 \cdot \sigma_x < X < \mu_x + 2 \cdot \sigma_x)$	\geq	0,7500
$k = 2,5 :$	$P(\mu_x - 2,5 \cdot \sigma_x < X < \mu_x + 2,5 \cdot \sigma_x)$	\geq	0,8400
$k = 3 :$	$P(\mu_x - 3 \cdot \sigma_x < X < \mu_x + 3 \cdot \sigma_x)$	\geq	0,8888...

Wir erkennen daran deutlich, dass die Wahrscheinlichkeit, dass X innerhalb des offenen Intervalls $(\mu_x - k \cdot \sigma_x, \mu_x + k \cdot \sigma_x)$ liegt, mit zunehmendem k ebenfalls ansteigt. Dieser Sachverhalt wird auch bei Betrachtung von Abbildung II 18 schnell deutlich. Nimmt k zu, wird nämlich auch die in der Abbildung markierte Fläche größer.

⁸ Streng genommen liefert die Ungleichung erst für $k > 1$ echte Erkenntnisse, da sich für $k = 1$ eine Wahrscheinlichkeit von Null ergeben würde.

Interessieren wir uns für die Wahrscheinlichkeit, dass X außerhalb des offenen Intervalls $\left] \mu_x - k \cdot \sigma_x; \mu_x + k \cdot \sigma_x \right[$ liegt, können wir diese ganz einfach als Komplementärwahrscheinlichkeit zu (II2: 50) bestimmen. Es gilt daher

$$\left[P(|X - \mu_x| \geq k \cdot \sigma_x) = 1 - P(|X - \mu_x| < k \cdot \sigma_x) \right] \leq \frac{1}{k^2}. \quad (\text{II2: 51})$$

Beispiel II 2-30

Ungleichung von Tschebyscheff II

Für verschiedene Werte von k erhalten wir folgende Abschätzungen:

$$\begin{aligned} k = 1,5 : \quad P(|X - \mu_x| \geq 1,5 \cdot \sigma_x) &\leq 0,4444... \\ k = 2 : \quad P(|X - \mu_x| \geq 2 \cdot \sigma_x) &\leq 0,2500 \\ k = 2,5 : \quad P(|X - \mu_x| \geq 2,5 \cdot \sigma_x) &\leq 0,1600 \\ k = 3 : \quad P(|X - \mu_x| \geq 3 \cdot \sigma_x) &\leq 0,1111... \end{aligned}$$

Wir erkennen deutlich, dass die Wahrscheinlichkeit dafür, dass X außerhalb des offenen Intervalls $\left] \mu_x - k \cdot \sigma_x; \mu_x + k \cdot \sigma_x \right[$ liegt, mit zunehmendem k abnimmt.

2.6 Anwendungsbeispiele

2.6.1 Renditen als Zufallsvariablen

Nach der Portfoliotheorie von H. Markowitz⁹ entscheidet sich ein rationaler Investor bei der Wahl zwischen zwei alternativen Anlagemöglichkeiten für diejenige, die bei gleicher erwarteter Rendite das geringere Risiko aufweist. Unter der *Rendite* r_t ist dabei die relative Wertänderung zwischen einer Anfangsinvestition K_{t-1} zum Zeitpunkt $t-1$ und dem Endwert (inkl. Dividenden, etc.) der Anlage K_t zum Zeitpunkt t zu verstehen:

$$r_t = \frac{K_t - K_{t-1}}{K_{t-1}}$$

Nehmen wir beispielsweise an, dass es sich bei der Anlage um eine Investition in Aktien handelt, so ist klar, dass K_t zum Investitionszeitpunkt $t-1$ ungewiss ist. r_t kann daher als Ausprägung einer Zufallsvariable R_t angesehen werden, die sich durch ihren Erwartungswert und ihre Varianz charakterisieren lässt. Die Varianz bzw. die Standardabweichung von Renditen, ist nun aber, wie wir bereits unter I 2.2.2.4 gesehen haben, das in der Praxis am häufigsten eingesetzte Risikomaß. Dies bedeutet, dass uns mit diesen beiden Kennzahlen von Renditeverteilungen sämtliche Informationen vorliegen, die uns ein Treffen von Investitionsentscheidungen nach der Theorie von H. Markowitz erlauben. Das nachfolgende Beispiel II 2-31 veranschaulicht dies.

⁹ Vgl. Markowitz, H. (1952, 1959).

Beispiel II 2-31**Erwartungswert und Standardabweichung von Renditen**

Ein Investor habe die Möglichkeit zwischen zwei Aktien A und B zu wählen. Für drei mögliche gesamtwirtschaftliche Entwicklungstendenzen wurden von Analysten folgende Renditen geschätzt:

i	Entwicklung	Wahrscheinlichkeit p_i	Rendite r_i^A	Rendite r_i^B
1	äußerst positiv	20 %	30 %	50 %
2	normal	60 %	10 %	10 %
3	äußerst negativ	20 %	-10 %	-30 %

Die erwartete Rendite der Aktie A errechnet sich daraus wie folgt:

$$\begin{aligned} E(R^A) &= \sum_{i=1}^3 p_i \cdot r_i^A = 0,20 \cdot 0,30 + 0,60 \cdot 0,10 + 0,20 \cdot (-0,10) \\ &= 0,10 = 10 \% \end{aligned}$$

Analog ergibt sich für Aktie B ebenfalls eine erwartete Rendite von 10 %.

Für die Standardabweichung der Renditen von Aktie A ergibt sich:

$$\begin{aligned} \sigma(R^A) &= \sqrt{\sum_{i=1}^3 p_i \cdot (r_i^A - E(R^A))^2} \\ &= \sqrt{0,20 \cdot (0,30 - 0,10)^2 + 0,60 \cdot (0,10 - 0,10)^2 + 0,20 \cdot (-0,10 - 0,10)^2} \\ &= 0,1265 \end{aligned}$$

Da wir bei analoger Berechnung für die Volatilität von Aktie B einen höheren Wert von 0,2530 erhalten, wird ein rationaler Investor (da gleiche erwartete Renditen) in Aktie A investieren.

2.6.2 Zufallsvariablen beim Roulette

Beim *amerikanischen Roulette* wird ein Rad mit 38 Zahlenschlitten gedreht. Dieses Rad weist 18 rote und 18 schwarze Zahlen sowie die grünen 0 und 00 auf.¹⁰ Ziel ist es, die Zahl vorauszusagen, auf der die Kugel landet. Mit einer richtigen Voraussage erhält man dann auf diese Wette eine Auszahlung. Die Auszahlung hängt wiederum von der Art der Wette ab. Es gibt allgemein neun verschiedene *Wettarten*. Jede Wettart umfasst eine bestimmte Spanne von Zahlen und hat ihre eigene Auszahlungsrate. Auf der Folgeseite sind ein typischer "Wetttisch"¹¹ und eine Tabelle mit den verschiedenen möglichen Wettarten abgebildet. Die darin enthaltenen Wettarten 1-6 werden auch als "*inside bets*", die letzten 7-12 als "*outside bets*" bezeichnet.

¹⁰ Im Gegensatz dazu fehlt beim europäischen Roulette die 00.

¹¹ Es ist dabei zu beachten, dass die hellgraue Färbung die Farbe rot und die dunkelgraue die Farbe schwarz darstellen soll.

Typischer Roulettetisch:

		0	00	
1 to 18	1 st 12	1	2	3
		4	5	6
		7	8	9
		10	11	12
	2 nd 12	13	14	15
		16	17	18
		19	20	21
		22	23	24
ODD	3 rd 12	25	26	27
		28	29	30
		31	32	33
		34	35	36
		2 to 1	2 to 1	2 to 1

Übersicht und Erklärung der Wettarten:

Wettart	Erklärung	Beispiel
Straight Up	Der Einsatz wird direkt auf eine einzelne Zahl (einschließlich 0 und 00) gesetzt.	4
Split Bet	Es wird auf zwei benachbarte Zahlen gesetzt.	8,9
Street Bet	Man setzt auf eine Zahlenreihe von 3 Zahlen.	13,14,15
Corner Bet	Die Spielmarke wird auf einer Ecke platziert, auf der sich vier Zahlen treffen.	1,2,4,5
Five Bet	Diese umfasst die ersten beiden Zahlenreihen. Man setzt damit auf 0,00,1,2,3.	-
Line Bet	Die Wette umfasst alle Zahlen von zwei aneinanderliegenden Reihen, also 6 Zahlen.	22,23,24,25,26,27
Column Bet	Es wird auf die 12 Zahlen einer Spalte (ohne 0 und 00) gesetzt.	2,5,8,11,14,17, 20,23,26,29,32,35
Dozen Bet	Setzt man seine Spielmarken in eine der Boxen "1 st 12", "2 nd 12" oder "3 rd 12", umfasst die Wette diese 12 Zahlen.	25,26,27,28,29,30, 31,32,33,34,35,36
Red/Black	Es wird auf alle roten oder alle schwarzen Zahlen (jeweils 18 Zahlen, ohne 0 und 00) gesetzt.	-
Even/Odd	Man setzt auf alle geraden oder alle ungeraden Zahlen (jeweils 18 Zahlen, ohne 0 und 00).	-
1 to 18	Der Spieler setzt auf die Zahlen von 1 bis 18.	-
19 to 36	Der Spieler setzt auf die Zahlen von 19 bis 36.	-

Für die Auszahlungen und die dazugehörigen Wahrscheinlichkeiten gilt:

Wette auf	Möglicher Gewinn	Trefferchance
1 Zahl	35 zu 1	$\frac{1}{38} = 2,63\%$
2 Zahlen	17 zu 1	$\frac{2}{38} = 5,26\%$
3 Zahlen	11 zu 1	$\frac{3}{38} = 7,89\%$
4 Zahlen	8 zu 1	$\frac{4}{38} = 10,53\%$
5 Zahlen	6 zu 1	$\frac{5}{38} = 13,16\%$
6 Zahlen	5 zu 1	$\frac{6}{38} = 15,79\%$
12 Zahlen	2 zu 1	$\frac{12}{38} = 31,58\%$
18 Zahlen	1 zu 1	$\frac{18}{38} = 47,37\%$

Es ist besonders zu beachten, dass wenn die *Kugel auf 0 oder 00* landet, Wetteinsätze auf rot, schwarz, ungerade, gerade, 1 to 18 und 19 to 36 nicht komplett verlorengehen. Stattdessen verliert jeder Spieler nur die Hälfte des ursprünglichen Wetteinsatzes. Die andere Hälfte kann vom Spieltisch wieder zurückgenommen oder erneut eingesetzt werden.¹²

Nach Beschreibung dieser allgemeinen Rouletteregeln wollen wir uns nun mit dem Gewinn beschäftigen, den wir bei Setzen eines Betrages B auf bestimmte, ausgewählte Zahlenkombinationen erwarten können. Wir erhalten im Gewinnfall auch den Einsatz B zurück, sodass wir diesen im Folgenden nicht als echten Gewinn (Zufallsvariable) betrachten. Konkret untersuchen wir drei einfache Tippstrategien: Setzen eines Betrages B auf eine einzelne Zahl k , auf "1" "12" oder "rot/schwarz" bzw. "even/odd".

1. Setzen eines Betrages B auf eine einzelne Zahl k

Wahrscheinlichkeit	Gewinn	Wahrscheinlichkeit · Gewinn
$P(\{k\}) = \frac{1}{38}$	$35 \cdot B$	$\frac{35}{38} \cdot B$
$P(\{\bar{k}\}) = \frac{37}{38}$	$-B$	$-\frac{37}{38} \cdot B$
$E(G) = -\frac{1}{19} \cdot B$		

Das Risiko einer solchen Wette kann zahlenmäßig wie folgt ausgedrückt werden:

$$\text{Var}(G) = (35 \cdot B)^2 \cdot \frac{1}{38} + (-B)^2 \cdot \frac{37}{38} - (-\frac{1}{19} \cdot B)^2 = \frac{631}{19} \cdot B^2 - \frac{1}{361} \cdot B^2 = 33,2078 \cdot B^2$$

$$\sigma_G = 5,7626 \cdot B$$

¹² Beim europäischen Roulette bedeutet eine 0 hingegen einen 100 %igen Verlust, sofern nicht auf 0 selbst gesetzt wurde.

2. Setzen eines Betrages B auf "1" 12"

Wahrscheinlichkeit	Gewinn	Wahrscheinlichkeit · Gewinn
$P(\{1^{\text{st}} \text{ 12}\}) = \frac{12}{38}$	$2 \cdot B$	$\frac{24}{38} \cdot B$
$P(\overline{\{1^{\text{st}} \text{ 12}\}}) = \frac{26}{38}$	$-B$	$-\frac{26}{38} \cdot B$
$E(G) = -\frac{1}{19} \cdot B$		

Für das Risiko gilt hier:

$$\text{Var}(G) = (2 \cdot B)^2 \cdot \frac{12}{38} + (-B)^2 \cdot \frac{26}{38} - \left(-\frac{1}{19} \cdot B\right)^2 = \frac{37}{19} \cdot B^2 - \frac{1}{361} \cdot B^2 = 1,9446 \cdot B^2$$

$$\sigma_G = 1,3945 \cdot B$$

3. Setzen eines Betrages B auf "rot/schwarz" oder "even/odd"

Wahrscheinlichkeit	Gewinn	Wahrscheinlichkeit · Gewinn
$P(\{\text{rot}\}) = \frac{18}{38}$	B	$\frac{18}{38} \cdot B$
$P(\{\text{schwarz}\}) = \frac{18}{38}$	$-B$	$-\frac{18}{38} \cdot B$
$P(\{0\}) = \frac{1}{38}$	$-\frac{1}{2} \cdot B$	$-\frac{1}{76} \cdot B$
$P(\{00\}) = \frac{1}{38}$	$-\frac{1}{2} \cdot B$	$-\frac{1}{76} \cdot B$
$E(G) = -\frac{1}{38} \cdot B$		

Das Risiko ergibt sich hier als:

$$\begin{aligned} \text{Var}(G) &= B^2 \cdot \frac{18}{38} + (-B)^2 \cdot \frac{18}{38} + \left(-\frac{1}{2} \cdot B\right)^2 \cdot \frac{1}{38} + \left(-\frac{1}{2} \cdot B\right)^2 \cdot \frac{1}{38} - \left(-\frac{1}{38} \cdot B\right)^2 \\ &= 0,9605 \cdot B^2 - \frac{1}{1444} \cdot B^2 = 0,9598 \cdot B^2 \end{aligned}$$

$$\sigma_G = 0,9797 \cdot B$$

Wir sehen, dass alle Wettstrategien einen negativen Erwartungswert aufweisen und sich hinsichtlich des Risikos stark unterscheiden. Generell können wir festhalten, dass Roulette langfristig zu Gunsten des Kasinos ausgelegt ist.

2.7 Mehrdimensionale Zufallsvariablen

2.7.1 Begriff

Bisher haben wir uns nur mit Verteilungen einzelner Zufallsvariablen (*univariate Verteilungen*) beschäftigt. Bei vielen praktischen und theoretischen Fragestellungen betrachtet man als Ergebnis eines Zufallsexperimentes nicht nur eine einzige, sondern mehrere Zufallsvariablen. So kann beispielsweise die Wirtschaftslage eines Landes am Ende einer Periode als Ergebnis eines Zufallsvorganges betrachtet werden, bei dem jedes mögliche Elementarereignis (wirtschaftliche Situation) durch eine Reihe von Zufallsvariablen wie etwa das Bruttoinlandsprodukt X, die Inflations-

rate Y oder die Arbeitslosenquote Z beschrieben wird. Von besonderem Interesse werden bei der Analyse nun aber nicht die Verteilungen von X, Y und Z für sich sein, sondern Beziehungen und Abhängigkeitsverhältnisse zwischen den Zufallsvariablen. Es liegt also nahe, das Ergebnis des Zufallsvorganges als eine dreidimensionale Zufallsvariable (X, Y, Z) aufzufassen.

Die gemeinsame Wahrscheinlichkeitsverteilung zweier Zufallsvariablen bzw. die Verteilung einer zweidimensionalen Zufallsvariable wird als *bivariate Verteilung* bezeichnet. Bei drei- und mehrdimensionalen Zufallsvariablen hingegen spricht man von *multivariaten Verteilungen*. Aus Gründen der rechnerischen Einfachheit und der Tatsachen, dass sich dieser Fall grafisch noch anschaulich darstellen lässt und ein Transfer auf mehr als zwei Zufallsvariablen leicht möglich ist, beschränken wir uns im Folgenden auf diskrete zweidimensionale Zufallsvariablen.¹³

2.7.2 Wahrscheinlichkeits- und Verteilungsfunktion

Die bei eindimensionalen Zufallsvariablen eingeführten Begriffe der Verteilungsfunktion, Wahrscheinlichkeitsfunktion und -dichte können wir auch auf mehrdimensionale Zufallsvariablen übertragen. Auch können wir hier wieder eine weitgehende Analogie zur zweidimensionalen Häufigkeitsverteilung beobachten.

2.7.2.1 Gemeinsame Wahrscheinlichkeitsfunktion

Sind X und Y zwei *diskrete* Zufallsvariablen mit den Ausprägungen x_i und y_j , so gibt die *gemeinsame Wahrscheinlichkeitsfunktion* $f(x,y)$ von X und Y (zweidimensionale Wahrscheinlichkeitsverteilung) die Wahrscheinlichkeit dafür an, dass die Zufallsvariable X den Wert x und die Zufallsvariable Y *gleichzeitig* den Wert y annimmt. Sie ist definiert als

$$f(x,y) = \begin{cases} P(X = x_i, Y = y_j) = f(x_i, y_j) & \text{falls } x = x_i, y = y_j \text{ für } i, j = 1, 2, \dots \\ 0 & \text{sonst} \end{cases} \quad (\text{II2: 52})$$

Wie im eindimensionalen Fall nimmt die Funktion also nur an den Stellen konkrete Werte an, die zum Wertebereich der Zufallsvariablen X und Y gehören. Andernfalls ist sie Null. Analog zur zweidimensionalen Häufigkeitsverteilung muss immer

$$0 \leq f(x_i, y_j) \leq 1 \quad \text{für } i, j = 1, 2, \dots \quad (\text{II2: 53})$$

$$\sum_i \sum_j f(x_i, y_j) = 1 \quad \text{für } i, j = 1, 2, \dots \quad (\text{II2: 54})$$

gelten. Für den Fall, dass die Zufallsvariablen X und Y eine endliche Anzahl von Ausprägungen annehmen können, lässt sich die gemeinsame Wahrscheinlichkeitsverteilung tabellarisch mittels Abbildung II 19 darstellen. Man beachte dabei, dass die Anzahlen k und r der Ausprägungen von X und Y gleich sein können, jedoch nicht müssen.

¹³ Unter II 2.7.6 folgt ein kurzer Überblick über die wichtigsten Zusammenhänge bei zweidimensionalen stetigen Zufallsvariablen. Eine ausgezeichnete allgemeine Einführung in multivariate Verteilungen bieten Schmid, F., Trede, M. (2005) im Kontext von Renditen.

X \ Y	y_1	y_2	...	y_j	...	y_r
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$...	$f(x_1, y_j)$...	$f(x_1, y_r)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$...	$f(x_2, y_j)$...	$f(x_2, y_r)$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots
x_i	$f(x_i, y_1)$	$f(x_i, y_2)$...	$f(x_i, y_j)$...	$f(x_i, y_r)$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots
x_k	$f(x_k, y_1)$	$f(x_k, y_2)$...	$f(x_k, y_j)$...	$f(x_k, y_r)$

Abbildung II 19: Gemeinsame Wahrscheinlichkeitsfunktion

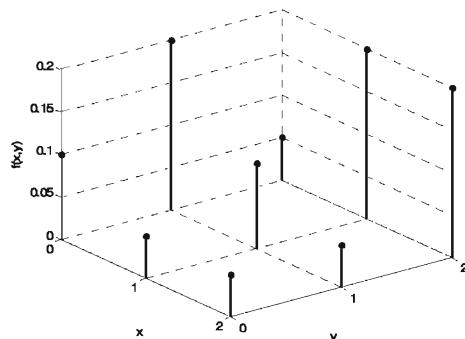
Beispiel II 2-32**Zweidimensionale Wahrscheinlichkeitsverteilung**

Bei einer bestimmten Automarke werden regelmäßig Defekte in der Lichtmaschine und im Steuergerät festgestellt. In einer Vertragswerkstatt wurde diesbezüglich eine Statistik geführt. Es ließ sich aus den Erfahrungswerten dieser Statistik eine gemeinsame Wahrscheinlichkeitsfunktion bilden, in der X die Anzahl pro Tag "vorgefahrener" Pkws mit defekter Lichtmaschine und Y die Anzahl derjeniger mit defektem Steuergerät mit Y bezeichnet.

		Steuergerätdefekte (y)		
		0	1	2
Lichtmaschinen-defekte (x)	0	0,10	0,20	0,05
	1	0,05	0,10	0,20
	2	0,05	0,05	0,20

Aus dieser Tabelle können wir nun z.B. ablesen, dass die Wahrscheinlichkeit dafür, dass an einem Tag 1 Pkw mit Lichtmaschinendefekt und gleichzeitig 2 Pkws mit Steuergerätedefekt in der Werkstatt repariert werden, bei $P(X = 1, Y = 2) = f(1; 2) = 0,20$ liegt.

Grafisch lässt sich die gemeinsame Wahrscheinlichkeitsfunktion (bei Nutzung des Befehls "stem3" in der Software Matlab 2009a) wie folgt darstellen:



Als Alternative zu diesem Stabdiagramm könnte auch ein Streudiagramm verwendet werden, in dem alle x-y-Kombinationen abgetragen und zu jedem Punkt die Zahlenwerte $f(x,y)$ angegeben werden.

2.7.2.2 Gemeinsame Verteilungsfunktion

Die *gemeinsame Verteilungsfunktion* $F(x,y)$ gibt an, mit welcher Wahrscheinlichkeit die Zufallsvariable X höchstens den Wert x und zugleich die Zufallsvariable Y höchstens den Wert y annimmt. Sie ergibt sich bei diskreten Zufallsvariablen durch Summation der entsprechenden Werte der gemeinsamen Wahrscheinlichkeitsfunktion $f(x,y)$. Es gilt also

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j). \quad (\text{II2: 55})$$

Für die gemeinsame Verteilungsfunktion gelten ähnliche Regeln, wie sie im Abschnitt II 2.2.2 für den eindimensionalen Fall angegeben wurden. Wir wollen sie daher hier nicht im Einzelnen formulieren.

Beispiel II 2-33

Zweidimensionale Verteilungsfunktion

Für das Beispiel II 2-32 zur defektfälligen Automarke zeigt die folgende Tabelle (links) die gemeinsame Verteilungsfunktion $F(x,y)$:

$F(x,y):$		$f(x,y):$		
x	y	0	1	2
y				
0	0	0,10	0,30	0,35
1	0	0,15	0,45	0,70
2	0	0,20	0,55	1,00
	1			
	2			
	x	0	1	2
	0	0,10	0,20	0,05
	1	0,05	0,10	0,20
	2	0,05	0,05	0,20

Wie wir hier sehr schön erkennen können, wird also, vereinfacht ausgedrückt, nichts anderes getan als jeweils zum "aktuellen" Zellwert all diejenigen Werte zu addieren, die links und oberhalb davon liegen. Der Fettdruck veranschaulicht dies. (Die weiteren Markierungen in den Tabellen benötigen wir im Folgenden für weitere Beispiele.)

Zwischen gemeinsamer Verteilungsfunktion und gemeinsamer Wahrscheinlichkeitsfunktion lässt sich die Beziehung

$$F(x_i, y_j) = F(x_i, y_{j-1}) + F(x_{i-1}, y_j) - F(x_{i-1}, y_{j-1}) + f(x_i, y_j) \quad (\text{II2: 56})$$

ableiten. Mit den Tabellen aus Beispiel II 2-33 (Graufärbung) können wir so z.B. den Wert $F(1,2) = F(1,1) + F(0,2) - F(0,1) + f(1,2) = 0,45 + 0,35 - 0,30 + 0,20 = 0,70$ bestimmen.¹⁴

2.7.2.3 Randverteilungen

Sind X und Y diskrete Zufallsvariablen, so kann aus der gemeinsamen Wahrscheinlichkeitsfunktion für X eine *Randverteilung* mit den Wahrscheinlichkeiten $f_x(x_i)$ und für Y eine Randverteilung mit den Wahrscheinlichkeiten $f_y(y_j)$ abgeleitet werden.

¹⁴ Mittels Beispiel II 2-33 lässt sich außerdem veranschaulichen, dass für die Wahrscheinlichkeit eines Ereignisses ($a < X \leq b, c < Y \leq d$) mit $a < b$ und $c < d$ stets der Zusammenhang $P(a < X \leq b, c < Y \leq d) = F(b,d) - F(b,c) - F(a,d) + F(a,c)$ gilt.

$f_x(x_i)$ gibt dabei an, wie groß die Wahrscheinlichkeit dafür ist, dass X einen speziellen Wert x_i annimmt, wobei es gleichgültig ist, welchen Wert die Zufallsvariable Y annimmt. Analoges gilt für $f_y(y_j)$. Die beiden Randverteilungen für X und Y lassen sich aus den Werten $f(x_i, y_j)$ der tabellarischen gemeinsamen Wahrscheinlichkeitsfunktion durch die Bildung von Zeilen- bzw. Spaltensummen errechnen. Es gilt

$$f_x(x_i) = P(X = x_i) = f(x_i, y_1) + f(x_i, y_2) + \dots = \sum_j f(x_i, y_j) \quad \text{für } i = 1, 2, \dots \quad (\text{II2: 57})$$

und

$$f_y(y_j) = P(Y = y_j) = f(x_1, y_j) + f(x_2, y_j) + \dots = \sum_i f(x_i, y_j) \quad \text{für } j = 1, 2, \dots \quad (\text{II2: 58})$$

Durch "Elimination" der jeweils anderen Zufallsvariable werden also die eindimensionalen Verteilungen von X und Y abgebildet.

Beispiel II 2-34

Randverteilungen

Für unser Automobilbeispiel II 2-32 erhalten wir nach (II2: 57) und (II2: 58) folgende Randverteilungen von X und Y:

x	y	0	1	2	$f_x(x_i)$
0	0,10	0,20	0,05	→	0,35
1	0,05	0,10	0,20	→	0,35
2	0,05	0,05	0,20	→	0,30
$f_y(y_j)$	0,20	0,35	0,45		1,00

Die Wahrscheinlichkeit, dass etwa an einem Tag 2 Pkws mit Lichtmaschinendefekt "vorgefahren" werden, liegt also hier bei $f_x(2) = 0,30$.

Unsere bisherigen Ausführungen zeigen, dass die gemeinsame Wahrscheinlichkeitsfunktion $f(x,y)$ alle Informationen über die Wahrscheinlichkeitsfunktionen $f_x(x)$ und $f_y(y)$ der eindimensionalen Zufallsvariablen X und Y enthält und daher $f_x(x)$ und $f_y(y)$ aus $f(x,y)$ abgeleitet werden können. Umgekehrt ist die Kenntnis der Randverteilungen im Allgemeinen jedoch *nicht* ausreichend, um die gemeinsame Wahrscheinlichkeitsfunktion zu ermitteln. Dies ist nur im Falle stochastisch unabhängiger Zufallsvariablen möglich (vgl. Abschnitt II 2.7.3).

2.7.2.4 Bedingte Verteilungen

Bedingte Verteilungen können ebenfalls aus der gemeinsamen Wahrscheinlichkeitsfunktion gewonnen werden. Solche Verteilungen geben die Wahrscheinlichkeit dafür an, dass eine Variable einen bestimmten Wert annimmt, unter der Bedingung, dass die jeweils andere bereits einen bestimmten Wert angenommen hat.

Die Wahrscheinlichkeit $P(X = x_i | Y = y_j)$, dass die Zufallsvariable X einen bestimmten Wert x_i annimmt, unter der Bedingung, dass Y einen Wert y_j angenommen hat, ermittelt sich im *diskreten* Fall, sofern $P(Y = y_j) > 0$ ist, als

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{f(x_i, y_j)}{f_Y(y_j)} = f(x_i | y_j). \quad (\text{II2: 59})$$

Es gilt natürlich analog auch

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{f(x_i, y_j)}{f_X(x_i)} = f(y_j | x_i). \quad (\text{II2: 60})$$

Beispiel II 2-35

Bedingte Verteilungen

Für das Beispiel II 2-32 auf Seite 227 ergeben sich für die Zufallsvariable X: "Anzahl der Pkws mit defekter Lichtmaschine" über

$$f(x_i | 0) = \frac{f(x_i, 0)}{f_Y(0)}, \quad f(x_i | 1) = \frac{f(x_i, 1)}{f_Y(1)} \quad \text{und} \quad f(x_i | 2) = \frac{f(x_i, 2)}{f_Y(2)}$$

die drei folgenden bedingten Verteilungen:

x	f(x,0)	f(x 0)	f(x,1)	f(x 1)	f(x,2)	f(x 2)
0	0,10	0,10/0,20 = 0,50	0,20	0,20/0,35 = 0,57	0,05	0,05/0,45 = 0,11
1	0,05	0,05/0,20 = 0,25	0,10	0,10/0,35 = 0,29	0,20	0,20/0,45 = 0,44
2	0,05	0,05/0,20 = 0,25	0,05	0,05/0,35 = 0,14	0,20	0,20/0,45 = 0,44
Σ	0,20	1,00	0,35	1,00	0,45	1,00

Für die Zufallsvariable Y: "Anzahl der Pkws mit defektem Steuergerät" ergeben sich über

$$f(y_j | 0) = \frac{f(0, y_j)}{f_X(0)}, \quad f(y_j | 1) = \frac{f(1, y_j)}{f_X(1)} \quad \text{und} \quad f(y_j | 2) = \frac{f(2, y_j)}{f_X(2)}$$

ebenfalls drei bedingte Verteilungen:

y	f(0,y)	f(y 0)	f(1,y)	f(y 1)	f(2,y)	f(y 2)
0	0,10	0,10/0,35 = 0,29	0,05	0,05/0,35 = 0,14	0,05	0,05/0,30 = 0,17
1	0,20	0,20/0,35 = 0,57	0,10	0,10/0,35 = 0,29	0,05	0,05/0,30 = 0,17
2	0,05	0,05/0,35 = 0,14	0,20	0,20/0,35 = 0,57	0,20	0,20/0,30 = 0,66
Σ	0,35	1,00	0,35	1,00	0,30	1,00

2.7.3 Stochastische Unabhängigkeit

Zwei Zufallsvariablen X und Y sind grob gesprochen dann voneinander *stochastisch unabhängig*, wenn sie keinerlei Einfluss aufeinander haben. Ausgehend vom Multiplikationssatz für unabhängige Ereignisse (II1: 55) können wir stochastische Unabhängigkeit daher über die folgende Beziehung konkretisieren:

$$\underbrace{P(X = x_i, Y = y_j)}_{f(x_i, y_j)} = \underbrace{P(X = x_i)}_{f_X(x_i)} \cdot \underbrace{P(Y = y_j)}_{f_Y(y_j)} \quad \text{für } i, j = 1, 2, \dots \quad (\text{II2: 61})$$

Die gemeinsamen Wahrscheinlichkeiten müssen also dem Produkt der jeweiligen Randverteilungswahrscheinlichkeiten entsprechen.¹⁵ Ist (II2: 61) nicht für alle i und j erfüllt, dann sind die beiden Zufallsvariablen stochastisch abhängig.

Beispiel II 2-36

Stochastische Unabhängigkeit

Bezeichnen wir beim zweimaligen Wurf einer fairen Münze das Ergebnis des ersten Wurfs mit X und das des zweiten mit Y, wobei wir die Codierung "Kopf" = 1 und "Zahl" = 0 vornehmen, so erhalten wir die gemeinsame Wahrscheinlichkeitsverteilung durch:

x	y	1	0	$f_X(x)$
1		0,25	0,25	0,50
0		0,25	0,25	0,50
	$f_Y(y)$	0,50	0,50	1,00

Die Zufallsvariablen X und Y sind klar stochastisch unabhängig, da hier nach (II2: 61) die gemeinsamen Wahrscheinlichkeiten gerade als Produkte der jeweiligen Randverteilungswahrscheinlichkeiten entstehen. Wie wir bereits unter II 2.7.2.3 erwähnt hatten, steckt also im Falle stochastischer Unabhängigkeit die vollständige Information zur gemeinsamen Verteilung der Zufallsvariablen X und Y in den Randverteilungen von X und Y.

Unter Verwendung von (II2: 61) ergibt sich bei den bedingten Verteilungen von unabhängigen Zufallsvariablen

$$f(x_i | y_j) = \frac{f(x_i, y_j)}{f_Y(y_j)} = \frac{f_X(x_i) \cdot f_Y(y_j)}{f_Y(y_j)} = f_X(x_i)$$

und

Unabhängigkeit

$$f(y_j | x_i) = \frac{f(x_i, y_j)}{f_X(x_i)} = \frac{f_X(x_i) \cdot f_Y(y_j)}{f_X(x_i)} = f_Y(y_j).$$

Eine äquivalente Bedingung zur Unabhängigkeit besteht also in der *Übereinstimmung der bedingten Verteilungen mit den jeweiligen Randverteilungen*:

$$f(x_i | y_j) = f_X(x_i) \quad \text{für } i, j = 1, 2, \dots \quad (\text{II2: 62})$$

¹⁵ Bei stochastischer Unabhängigkeit gilt natürlich auch für die gemeinsame Verteilungsfunktion $F(x_i, y_j) = F_X(x_i) \cdot F_Y(y_j)$ für $i, j = 1, 2, \dots$.

$$f(y_j | x_i) = f_{Y|X}(y_j) \quad \text{für } i, j = 1, 2, \dots \quad (\text{II2: 63})$$

Sind die bedingten Verteilungen für verschiedene Bedingungen unterschiedlich, so ist eine Variable davon abhängig, welchen Wert die andere annimmt. Die Variablen sind stochastisch abhängig.

2.7.4 Kennzahlen zweidimensionaler Wahrscheinlichkeitsverteilungen

2.7.4.1 Erwartungswert und Varianz

Die Erwartungswertbildung für eindimensionale Zufallsvariablen lässt sich auch auf zweidimensionale Wahrscheinlichkeitsverteilungen ausdehnen. Sind X und Y zwei *diskrete* Zufallsvariablen mit der gemeinsamen Wahrscheinlichkeitsfunktion $f(x_i, y_j)$, so ergibt sich der *Erwartungswert* von X als

$$E(X) = \sum_i \sum_j x_i \cdot f(x_i, y_j). \quad (\text{II2: 64})$$

Da aufgrund allgemeiner Summationsregeln und nach (II2: 57) die Beziehungen

$$\sum_i \sum_j x_i \cdot f(x_i, y_j) = \sum_i x_i \sum_j f(x_i, y_j) = \sum_i x_i \cdot f_X(x_i)$$

gelten, können wir den Erwartungswert der Zufallsvariable X also auch mit Hilfe der Randverteilung von X bestimmen:

$$E(X) = \sum_i x_i \cdot f_X(x_i) \quad (\text{II2: 65})$$

Für die Zufallsvariable Y erhalten wir analog

$$E(Y) = \sum_j y_j \cdot f_Y(y_j). \quad (\text{II2: 66})$$

Beispiel II 2-37

Erwartungswerte aus Randverteilungen

Für unser Beispiel der Lichtmaschinen- und Steuergerätdefekte bzw. die Tabelle in Beispiel II 2-34 ergeben sich nach (II2: 65) und (II2: 66) folgende Erwartungswerte:

$$E(X) = \sum_{i=1}^3 x_i \cdot f_X(x_i) = 0 \cdot 0,35 + 1 \cdot 0,35 + 2 \cdot 0,30 = 0,95$$

$$E(Y) = \sum_{j=1}^3 y_j \cdot f_Y(y_j) = 0 \cdot 0,20 + 1 \cdot 0,35 + 2 \cdot 0,45 = 1,25$$

Dies bedeutet nun, dass im Durchschnitt täglich 0,95 Autos mit Lichtmaschinen- und 1,25 Autos mit Steuergerätdefekt "vorgefahren" werden. Da diese Werte vielleicht etwas abstrakt scheinen (es sind ja wohl ganze Autos), empfiehlt sich in diesem Fall eine Hochrechnung der Ergebnisse auf andere Zeiträume (etwa Wochen, Monate).

Analog zur Herleitung des Erwartungswertes lässt sich auch die *Varianz* von X aufgrund von

$$\text{Var}(X) = E[X - E(X)]^2 = \sum_i \sum_j [x_i - E(X)]^2 \cdot f(x_i, y_j) = \sum_i [x_i - E(X)]^2 \sum_j f(x_i, y_j)$$

mittels der Randverteilung von X als

$$\text{Var}(X) = \sum_i [x_i - E(X)]^2 \cdot f_X(x_i) \quad (\text{II2: 67})$$

berechnen. Analog ergibt sich für die Zufallsvariable Y

$$\text{Var}(Y) = \sum_j [y_j - E(Y)]^2 \cdot f_Y(y_j) . \quad (\text{II2: 68})$$

Beispiel II 2-38

Varianzen aus Randverteilungen

Für unser Autobeispiel (Tabelle in Beispiel II 2-34) erhalten wir unter Beachtung der Ergebnisse von Beispiel II 2-37 und Anwendung des Verschiebungssatzes die Varianzen:

$$\text{Var}(X) = \sum_{i=1}^3 x_i^2 \cdot f_X(x_i) - [E(X)]^2 = 0^2 \cdot 0,35 + 1^2 \cdot 0,35 + 2^2 \cdot 0,30 - 0,95^2 = 0,65$$

$$\text{Var}(Y) = \sum_{j=1}^3 y_j^2 \cdot f_Y(y_j) - [E(Y)]^2 = 0^2 \cdot 0,20 + 1^2 \cdot 0,35 + 2^2 \cdot 0,45 - 1,25^2 = 0,59$$

Für die Standardabweichung ergibt sich hier:

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{0,65} = 0,81$$

$$\sigma_Y = \sqrt{\text{Var}(Y)} = \sqrt{0,59} = 0,77$$

Wir können also sagen, dass die Einzelwerte der Verteilung von X stärker um ihren Erwartungswert streuen als die von Y.

Neben Erwartungswerten und Varianzen aus Randverteilungen lassen sich mit Hilfe von bedingten Verteilungen bedingte Erwartungswerte und bedingte Varianzen bestimmen. *Bedingte Erwartungswerte* ergeben sich wie folgt:

$$E(X | y_j) = \sum_i x_i \cdot f(x_i | y_j) \quad (\text{II2: 69})$$

$$E(Y | x_i) = \sum_j y_j \cdot f(y_j | x_i) \quad (\text{II2: 70})$$

Bedingte Varianzen ergeben sich als

$$\text{Var}(X | y_j) = \sum_i [x_i - E(X | y_j)]^2 \cdot f(x_i | y_j) \quad (\text{II2: 71})$$

$$\text{Var}(Y | x_i) = \sum_j [y_j - E(Y | x_i)]^2 \cdot f(y_j | x_i) , \quad (\text{II2: 72})$$

wobei zur Rechenvereinfachung auch vom Verschiebungssatz Gebrauch gemacht werden kann.

Beispiel II 2-39

Bedingte Erwartungswerte und Varianzen

Unter Rückgriff auf unsere Ergebnisse aus Beispiel II 2-35 können wir z.B. bestimmen, wie viele Autos mit defekter Lichtmaschine im Durchschnitt an Tagen "vorgefahren" werden, an denen genau Y = 1 Wagen mit Steuergerätdefekt gewartet wird:

$$E(X|1) = \sum_{i=1}^3 x_i \cdot f(x_i|1) = 0 \cdot 0,57 + 1 \cdot 0,29 + 2 \cdot 0,14 = 0,57$$

Die zugehörige Varianz der Anzahl dieser Fahrzeuge ergibt sich dabei zu

$$\text{Var}(X|1) = \sum_{i=1}^3 x_i^2 \cdot f(x_i|1) - [E(X|1)]^2 = 0^2 \cdot 0,57 + 1^2 \cdot 0,29 + 2^2 \cdot 0,14 - 0,57^2 = 0,53.$$

Interessieren wir uns für die durchschnittliche tägliche Anzahl an Fahrzeugen mit Steuergerätdefekten (Y) an Tagen, an denen genau X = 2 Fahrzeuge mit Lichtmaschinendefekten "vorgefahren" werden, und die zugehörige Varianz, erhalten wir folgende Ergebnisse:

$$E(Y|2) = \sum_{j=1}^3 y_j \cdot f(y_j|2) = 0 \cdot 0,17 + 1 \cdot 0,17 + 2 \cdot 0,66 = 1,49$$

$$\text{Var}(Y|2) = \sum_{j=1}^3 y_j^2 \cdot f(y_j|2) - [E(Y|2)]^2 = 0^2 \cdot 0,17 + 1^2 \cdot 0,17 + 2^2 \cdot 0,66 - 1,49^2 = 0,59$$

Erwartungswerte und Varianzen liefern in ihrer unbedingten Form keine Information über den stochastischen Zusammenhang zwischen zwei Zufallsvariablen X und Y. Lediglich bedingte Größen können zur Feststellung eines solchen Zusammenhangs verwendet werden. Sind zwei Zufallsvariablen voneinander *unabhängig*, so sind die bedingten und unbedingten Erwartungswerte identisch, d.h.

$$E(X|y_j) = E(X) \quad \text{für } j = 1, 2, \dots, \quad (\text{II2: 73})$$

$$E(Y|x_i) = E(Y) \quad \text{für } i = 1, 2, \dots. \quad (\text{II2: 74})$$

Dies ist darauf zurückzuführen, dass im Falle der Unabhängigkeit nach (II2: 62) bzw. (II2: 63) die bedingten Wahrscheinlichkeiten den Randverteilungswahrscheinlichkeiten entsprechen.

2.7.4.2 Kovarianz und Korrelationskoeffizient

Der im Abschnitt I 3.2 für zweidimensionale Häufigkeitsverteilungen behandelte Begriff der Korrelation ist auch auf zweidimensionale Wahrscheinlichkeitsverteilungen übertragbar. Gehen großen Werte von X tendenziell mit größeren (kleineren) Werten von Y einher, sind die Zufallsvariablen X und Y positiv (negativ) korreliert. Eine Maßzahl für die Stärke dieses Zusammenhangs ist die *Kovarianz* $\text{Cov}(X,Y)$ oder kurz σ_{xy} , welche sich als der Erwartungswert des Produktes der Abweichungen der einzelnen Zufallsvariablen von ihrem jeweiligen Erwartungswert ergibt:

$$\text{Cov}(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))] \quad (\text{II2: 75})$$

Für *diskrete* Zufallsvariablen erhalten wir unter Verwendung der gemeinsamen Wahrscheinlichkeitsfunktion $f(x,y)$ konkret

$$\begin{aligned} \text{Cov}(X, Y) &= \sum_i \sum_j [x_i - E(X)] \cdot [y_j - E(Y)] \cdot f(x_i, y_j) \\ &= \sum_i \sum_j x_i \cdot y_j \cdot f(x_i, y_j) - E(X) \cdot E(Y) \\ &= E(XY) - E(X) \cdot E(Y) \end{aligned} \quad (\text{II2: 76})$$

Zufallsvariablen mit positiver Kovarianz werden als *positiv*, solche mit negativer Kovarianz als *negativ korreliert* bezeichnet. Nimmt die Kovarianz einen Wert von Null an, so sind die Variablen *unkorreliert*. Die Kovarianz einer Zufallsvariable X mit sich selbst entspricht ihrer Varianz, d.h. $\text{Cov}(X,X) = \text{Var}(X)$. Analog zur deskriptiven Statistik können wir außerdem $\text{Cov}(a + bX, c + dY) = bd\text{Cov}(X,Y)$ festhalten. Nützlich ist für drei Zufallsvariablen X, Y und Z häufig auch die Kenntnis der Beziehung $\text{Cov}(X + Y, Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$.

Da die Kovarianz $\text{Cov}(X,Y)$ von den Einheiten abhängt, in denen die Zufallsvariablen X und Y gemessen werden, wird in der Praxis durch die Division der Kovarianz durch das Produkt der Standardabweichungen von X und Y ein dimensionsloses Korrelationsmaß, der sog. *Korrelationskoeffizient* ρ_{xy} , generiert. Er gibt für den Fall, dass $\text{Var}(X) \neq 0$ und $\text{Var}(Y) \neq 0$, die Stärke des *linearen* stochastischen Zusammenhangs der beiden Zufallsvariablen an.

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_x \cdot \sigma_y} \quad (\text{II2: 77})$$

Wie der Bravais-Pearson-Korrelationskoeffizient aus der deskriptiven Statistik liegt auch ρ_{xy} im Intervall $[-1;1]$ und ist so für Vergleiche vorteilhafter. ρ_{xy} ist genau dann gleich 1, wenn $Y = a + bX$ mit $b > 0$ ist. Für $b < 0$ nimmt ρ_{xy} den Wert -1 an.

Beispiel II 2-40

Korrelation von Zufallsvariablen

Für unser Automobilbeispiel II 2-32 erhalten wir unter Verwendung der Ergebnisse aus den Beispielen II 2-37 und II 2-38 die nachfolgenden Werte:

$$\begin{aligned} E(XY) &= \sum_{i=1}^3 \sum_{j=1}^3 x_i \cdot y_j \cdot f(x_i, y_j) \\ &= 0 \cdot 0 \cdot 0,10 + 0 \cdot 1 \cdot 0,20 + 0 \cdot 2 \cdot 0,05 \\ &\quad + 1 \cdot 0 \cdot 0,05 + 1 \cdot 1 \cdot 0,10 + 1 \cdot 2 \cdot 0,20 \\ &\quad + 2 \cdot 0 \cdot 0,05 + 2 \cdot 1 \cdot 0,05 + 2 \cdot 2 \cdot 0,20 = 1,4 \text{ (Stück}^2\text{)}} \end{aligned}$$

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - E(X) \cdot E(Y) \\ &= 1,4 - 0,95 \cdot 1,25 = 0,21 \text{ (Stück}^2\text{)}} \end{aligned}$$

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_x \cdot \sigma_y} = \frac{0,21 \text{ (Stück}^2\text{)}}{0,81 \text{ (Stück)} \cdot 0,77 \text{ (Stück)}} = 0,34$$

Dieses Ergebnis deutet darauf hin, dass nur ein schwacher, positiver Zusammenhang zwischen Lichtmaschinen- und Steuergerätausfällen besteht.

Wie wir wissen, gilt bei stochastischer Unabhängigkeit der in (II2: 61) festgelegte Zusammenhang $f(x_i, y_j) = f_x(x_i) \cdot f_y(y_j)$. Setzen wir diesen in die allgemeine Formel für $E(XY)$ aus (II2: 76) ein, so ergibt sich

$$\begin{aligned} E(XY) &= \sum_i \sum_j x_i \cdot y_j \cdot f(x_i, y_j) = \sum_i \sum_j x_i \cdot y_j \cdot f_x(x_i) \cdot f_y(y_j) \\ &= \sum_i x_i \cdot f_x(x_i) \cdot \sum_j y_j \cdot f_y(y_j) = E(X) \cdot E(Y). \end{aligned}$$

Bei **Unabhängigkeit** gilt also stets

$$E(XY) = E(X) \cdot E(Y). \quad (\text{II2: 78})$$

Mit dieser neuen Erkenntnis können wir nun näher betrachten, wie sich die Kovarianz und der Korrelationskoeffizient bei Unabhängigkeit verhalten. Wir setzen dazu (II2: 78) in (II2: 76) ein:

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - E(X) \cdot E(Y) \\ &= E(X) \cdot E(Y) - E(X) \cdot E(Y) = 0 \end{aligned}$$

Unabhängige Zufallsvariablen besitzen also immer eine Kovarianz von Null und folglich auch einen Korrelationskoeffizienten von Null. *Unabhängigkeit* hat stets *Unkorreliertheit zur Folge*.

$$X, Y \text{ unabhängig} \rightarrow \text{Cov}(X, Y) = 0 \rightarrow \rho_{XY} = 0 \quad (\text{II2: 79})$$

Umgekehrt dürfen wir aber bei einer *Kovarianz von Null nicht* allgemein *auf Unabhängigkeit schließen*.¹⁶ Unkorrelierte Zufallsvariablen können stochastisch Abhängig sein, denn es kann ein nichtlinearer Zusammenhang bestehen. Ist die *Kovarianz ungleich Null*, so folgt daraus sicher die *Abhängigkeit*.

Die Formel (II2: 76) liefert durch einfache Umstellung den **Multiplikationssatz für Erwartungswerte**. Dieser besagt, dass der Erwartungswert des Produktes zweier beliebiger Zufallsvariablen X und Y als

$$E(XY) = E(X) \cdot E(Y) + \text{Cov}(X, Y) \quad (\text{II2: 80})$$

bestimmt werden kann. Sind X und Y *unabhängige* Zufallsvariablen, so vereinfacht sich der Multiplikationssatz zu (II2: 78). Liegen n unabhängige Zufallsvariablen X_1, X_2, \dots, X_n vor, so gilt stets, dass sich der Erwartungswert des Produktes der Zufallsvariablen als Produkt der einzelnen Erwartungswerte ergibt:

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i) \quad (\text{II2: 81})$$

Beispiel II 2-41

Produkte aus Zufallsvariablen

Es sei die Zufallsvariable X die Anzahl der Bestellungen einer Einkaufsabteilung pro Jahr. Die Zufallsvariable Y beschreibt die Anzahl der Artikel je Bestellung. Es ist nun unter Verwendung von $E(X) = 10.000$ Stück und $E(Y) = 1.000$ Stück die erwartete Anzahl der insgesamt pro Jahr bestellten Artikel $E(XY)$ zu bestimmen.

Unterstellen wir Unabhängigkeit, so gilt

$$E(XY) = E(X) \cdot E(Y) = 10.000 \cdot 1.000 = 10.000.000 \text{ Stück.}$$

¹⁶ Nur im Fall gemeinsam normalverteilter Zufallsvariablen ist die Unkorreliertheit mit der Unabhängigkeit äquivalent.

2.7.5 Linearkombinationen von Zufallsvariablen

Analog zu den eindimensionalen Zufallsvariablen kann auch bei zweidimensionalen die Erwartungswertbildung auf Funktionen $g(X, Y)$ der Zufallsvariablen X und Y ausgedehnt werden. Es gilt dabei allgemein

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_j) \cdot f(x_i, y_j) . \quad (\text{II2: 82})$$

Durch die Funktion $g(X, Y)$ entsteht eine neue Zufallsvariable W. Eine spezielle Möglichkeit der Generierung der Zufallsvariable W ist die sog. *Linearkombination* der Zufallsvariablen X und Y. Dies bedeutet konkret, dass

$$W = g(X, Y) = aX + bY \quad (\text{II2: 83})$$

gilt, wobei a und b von Null verschiedene Konstanten sind, die als Gewichte von X und Y in W angesehen werden können. Für den *Erwartungswert* einer solchen Linearkombination W können wir

$$E(W) = a \cdot E(X) + b \cdot E(Y) \quad (\text{II2: 84})$$

zeigen. Die *Varianz* von W ergibt sich als

$$\text{Var}(W) = a^2 \cdot \text{Var}(X) + b^2 \cdot \text{Var}(Y) + 2 \cdot a \cdot b \cdot \text{Cov}(X, Y) . \quad (\text{II2: 85})$$

Für einige spezielle Ausprägungen von a und b sind in der nachfolgenden Tabelle *wichtige Linearkombinationen* zusammengestellt:

	W	E(W)	Var(W)
Summe			
a = 1, b = 1	X + Y	E(X) + E(Y)	Var(X) + Var(Y) + 2 · Cov(X, Y)
Differenz			
a = 1, b = -1	X - Y	E(X) - E(Y)	Var(X) + Var(Y) - 2 · Cov(X, Y)
Mittelwert			
a = 0,5, b = 0,5	$\frac{1}{2} \cdot (X + Y)$	$\frac{1}{2} \cdot [E(X) + E(Y)]$	$\frac{1}{4} \cdot \text{Var}(X) + \frac{1}{4} \cdot \text{Var}(Y)$ $+ \frac{1}{2} \cdot \text{Cov}(X, Y)$

Auf die Aufführung konkreter Beweise der allgemeinen Zusammenhänge (II2: 84) und (II2: 85) wollen wir an dieser Stelle verzichten, da sie analog zu denen aus der deskriptiven Statistik in Abschnitt I 3.2.4 erfolgen. Die aufgeführten Beziehungen besitzen sowohl für diskrete als auch für stetige Zufallsvariablen Gültigkeit. Sind X und Y unabhängige Zufallsvariablen, so können die Formeln ebenfalls verwendet werden. Es ist lediglich darauf zu achten, dass die Kovarianz dann den Wert Null annimmt, also der entsprechende Term aus der Formel der Varianz herausfällt. Wir erhalten also im Fall der **Unabhängigkeit** von X und Y aus (II2: 85) allgemein

$$\text{Var}(W) = a^2 \cdot \text{Var}(X) + b^2 \cdot \text{Var}(Y) . \quad (\text{II2: 86})$$

Vor allem für die praktische Anwendung (vgl. Abschnitt II 2.7.7) hat sich eine alternative Formulierung von (II2: 85) bewährt, die den Korrelationskoeffizienten zwischen X und Y enthält. Aufgrund der Beziehung

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \Leftrightarrow \text{Cov}(X, Y) = \sigma_X \cdot \sigma_Y \cdot \rho_{XY}$$

können wir nämlich auch

$$\text{Var}(W) = a^2 \cdot \text{Var}(X) + b^2 \cdot \text{Var}(Y) + 2 \cdot a \cdot b \cdot \sigma_X \cdot \sigma_Y \cdot \rho_{XY} \quad (\text{II2: 87})$$

schreiben.

Beispiel II 2-42

Linearkombinationen beim Würfeln

Ein idealer Würfel wird zweimal geworfen. Betrachten wir die Zufallsvariablen S: "Augensumme", X_1 : "Augenzahl beim ersten Wurf" und X_2 : "Augenzahl beim zweiten Wurf", so gilt für die Erwartungswerte der beiden letzten Variablen

$$E(X_1) = E(X_2) = \frac{1}{6} \cdot (1+2+3+4+5+6) = 3,5.$$

Ohne zeitraubende Umwege kann nun daraus $E(S)$ als

$$S = X_1 + X_2 \rightarrow E(S) = E(X_1 + X_2) = E(X_1) + E(X_2) = 3,5 + 3,5 = 7$$

ermittelt werden.

Zusatzfrage:

Wie verändert sich $E(S)$, wenn die Augenzahlen auf beiden Würfeln jeweils mit 10 multipliziert werden?

1. Möglichkeit der Lösung:

$$E(10 \cdot S) = 10 \cdot E(S) = 10 \cdot 7 = 70$$

2. Möglichkeit der Lösung:

$$E(10 \cdot X_1) = 10 \cdot E(X_1) = 10 \cdot 3,5 = 35, \quad E(10 \cdot X_2) = 10 \cdot E(X_2) = 10 \cdot 3,5 = 35$$

$$\rightarrow E(10 \cdot S) = E(10 \cdot X_1) + E(10 \cdot X_2) = 35 + 35 = 70$$

Für den Fall von n Zufallsvariablen X_1, X_2, \dots, X_n und n von Null verschiedenen Konstanten a_1, a_2, \dots, a_n können die Beziehungen (II2: 84) und (II2: 85) folgendermaßen verallgemeinert werden:

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i) \quad (\text{II2: 88})$$

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j) \quad (\text{II2: 89})$$

Da die Kovarianz einer Zufallsvariablen mit sich selbst stets ihrer Varianz entspricht, können wir (II2: 89) alternativ auch als

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_i a_j \text{Cov}(X_i, X_j)$$

formulieren.

Sind die X_1, X_2, \dots, X_n zusätzlich paarweise unabhängig, so gilt

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i). \quad (\text{II2: 90})$$

Eine konkrete praktische Anwendung dieser Verallgemeinerungen werden wir im Abschnitt II 2.7.7 kennenlernen.

2.7.6 Formelzusammenstellung für stetige Zufallsvariablen

Für den Fall zweier stetiger Zufallsvariablen X und Y wollen wir im Folgenden die wichtigsten formelmäßigen Zusammenhänge kurz darstellen:

Gemeinsame Dichtefunktion:

$$P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f(x, y) dy dx \quad \text{für } a < b \text{ und } c < d$$

mit den Eigenschaften

$$f(x, y) \geq 0 \quad \text{und} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1$$

Gemeinsame Verteilungsfunktion:

$$F(x_l, y_l) = \int_{-\infty}^{x_l} \int_{-\infty}^{y_l} f(x, y) dx dy$$

Randverteilungen (-dichten):

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Bedingte Verteilungen (Dichten):

$$f(x | y) = \frac{f(x, y)}{f_y(y)} \quad f(y | x) = \frac{f(x, y)}{f_x(x)}$$

Erwartungswert:

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) dx dy = \int_{-\infty}^{\infty} x \cdot f_x(x) dx \quad E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \cdot f(x, y) dx dy = \int_{-\infty}^{\infty} y \cdot f_y(y) dy$$

Varianz:

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x - E(X)]^2 \cdot f(x, y) dx dy & \text{Var}(Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - E(Y)]^2 \cdot f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} [x - E(X)]^2 \cdot f_x(x) dx & &= \int_{-\infty}^{\infty} [y - E(Y)]^2 \cdot f_y(y) dy \end{aligned}$$

Kovarianz:

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x - E(X)] \cdot [y - E(Y)] \cdot f(x, y) dx dy$$

2.7.7 Anwendungsbeispiel: Portfoliotheorie

Wie wir unter II 2.6.1 gesehen haben, bewerten Anleger alternative Investments auf Basis erwarteter Renditen und der Renditevarianz bzw. -standardabweichung. Bei gleicher erwarteter Rendite werden Anlagen mit geringerer Varianz vorgezogen. Stellt sich nun ein Anleger ein Portfolio aus verschiedenen Anlagen zusammen, so ergeben sich erwartete Portfoliorendite und Varianz der Portfoliorendite unter Berücksichtigung der unter II 2.7.5 behandelten Vorschriften für Linearkombinationen von Zufallsvariablen. Der besondere Anreiz zur Portfoliobildung liegt für den Anleger darin, dass er dadurch Renditevarianzen (d.h. ein Risiko) erreichen kann, die unter den Varianzen einzelner Anlagen liegen.

Besteht ein Portfolio aus zwei Anlagen (1 und 2) mit den Renditen R_1 und R_2 sowie den Preisen P_1 und P_2 und den Mengen X_1 und X_2 , so kann die *erwartete Rendite des gesamten Portfolios* in drei Schritten ermittelt werden:

1. Bestimmung der Anteile w_i der Anlagen am Portfolio:

$$w_1 = \frac{P_1 \cdot X_1}{P_1 \cdot X_1 + P_2 \cdot X_2} \quad \text{und} \quad w_2 = \frac{P_2 \cdot X_2}{P_1 \cdot X_1 + P_2 \cdot X_2}$$

2. Berechnung der erwarteten Renditen der einzelnen Anlagen:

Die Werte von $E(R_1)$ und $E(R_2)$ werden wie unter II 2.6.1 ermittelt.

3. Errechnung der erwarteten Gesamtrendite des Portfolios nach (II2: 84):

$$E(R_p) = w_1 \cdot E(R_1) + w_2 \cdot E(R_2)$$

Da $w_1 + w_2 = 1$ gilt, können wir auch Folgendes schreiben:

$$E(R_p) = w_1 \cdot E(R_1) + (1 - w_1) \cdot E(R_2) \quad (\text{II2: 91})$$

Beispiel II 2-43

Erwartete Portfoliorendite

Der Investor Uwe B. entschied sich dazu, seine Altersvorsorge zu 80 % mit den Aktien eines Fußballvereins (X) und zu 20 % durch die Vermarktung seiner Biografie (Y) zu sichern. Ein unabhängiger Investmentexperte analysierte dieses mutige Portfolio und kam nach Ermittlung der erwarteten Renditen (in %) von $E(R_X) = -90$ und $E(R_Y) = -50$ zu der Gesamtrendite $E(R) = 0,80 \cdot (-90) + 0,20 \cdot (-50) = -82$. Uwe wird also voraussichtlich 82 % des von ihm investierten Kapitals verlieren. Er sollte bei seinem nächsten Investment besser den Rat von Fachleuten einholen.

Für die *Varianz der Portfoliorendite* (im Folgenden kurz *Portfoliovarianz*) gilt nach (II2: 85) bzw. (II2: 87)

$$\begin{aligned} \text{Var}(R_p) &= w_1^2 \cdot \text{Var}(R_1) + w_2^2 \cdot \text{Var}(R_2) + 2 \cdot w_1 \cdot w_2 \cdot \text{Cov}(R_1, R_2) \\ &= w_1^2 \cdot \text{Var}(R_1) + w_2^2 \cdot \text{Var}(R_2) + 2 \cdot w_1 \cdot w_2 \cdot \sigma_{R_1} \cdot \sigma_{R_2} \cdot \rho_{R_1 R_2} \end{aligned} \quad (\text{II2: 92})$$

Das Risiko des Portfolios hängt somit sowohl von den Varianzen der Einzelrenditen als auch von der Korrelation zwischen den Einzelrenditen ab.¹⁷

¹⁷ Natürlich ist auch die Gewichtung als Einflussfaktor zu betrachten.

Den größten Wert nimmt die Portfoliovarianz c. p. an, wenn die Korrelation völlig gleichgerichtet und maximal ist, d.h. $\rho_{R_1 R_2} = +1$ gilt. Dies lässt sich damit erklären, dass die Rendite einer Anlage im gleichen Ausmaß fällt (steigt), wie auch die Rendite der anderen Anlage fällt (steigt). Es gilt in einem solchen Fall für (II2: 92)

$$\begin{aligned}\text{Var}(R_p)_{\max} &= w_1^2 \cdot \text{Var}(R_1) + w_2^2 \cdot \text{Var}(R_2) + 2 \cdot w_1 \cdot w_2 \cdot \sigma_{R_1} \cdot \sigma_{R_2} \\ &= [w_1 \cdot \sigma_{R_1} + w_2 \cdot \sigma_{R_2}]^2.\end{aligned}$$

Den kleinsten Wert für die Portfoliovarianz erhalten wir analog, wenn sich die Renditen der Wertpapiere völlig gegenläufig entwickeln, also $\rho_{R_1 R_2} = -1$ gilt. Das Risiko ist minimal, da ein Renditerückgang bei einer Anlage durch eine gegengerichtete Renditesteigerung bei der anderen Anlage ausgeglichen wird.

$$\begin{aligned}\text{Var}(R_p)_{\min} &= w_1^2 \cdot \text{Var}(R_1) + w_2^2 \cdot \text{Var}(R_2) - 2 \cdot w_1 \cdot w_2 \cdot \sigma_{R_1} \cdot \sigma_{R_2} \\ &= [w_1 \cdot \sigma_{R_1} - w_2 \cdot \sigma_{R_2}]^2\end{aligned}$$

Liegt die Korrelation zwischen -1 und $+1$, so können wir sagen, dass sich die Varianz der Portfoliorendite in folgendem Intervall

$$[w_1 \cdot \sigma_{R_1} - w_2 \cdot \sigma_{R_2}]^2 \leq \text{Var}(R_p) \leq [w_1 \cdot \sigma_{R_1} + w_2 \cdot \sigma_{R_2}]^2 \quad (\text{II2: 93})$$

bewegt. Für die Portfoliostandardabweichung gilt dann natürlich

$$[w_1 \cdot \sigma_{R_1} - w_2 \cdot \sigma_{R_2}] \leq \sigma_{R_p} \leq [w_1 \cdot \sigma_{R_1} + w_2 \cdot \sigma_{R_2}]. \quad (\text{II2: 94})$$

Definieren wir das gewichtete Mittel der Standardabweichungen als

$$\bar{\sigma} = w_1 \cdot \sigma_{R_1} + w_2 \cdot \sigma_{R_2}, \quad (\text{II2: 95})$$

so können wir außerdem Folgendes feststellen:

$$\begin{array}{lll} \text{Fall A:} & \bar{\sigma} = \sqrt{\text{Var}(R_p)} & \text{falls } \rho_{R_1 R_2} = 1 \\ \text{Fall B:} & \bar{\sigma} > \sqrt{\text{Var}(R_p)} & \text{falls } \rho_{R_1 R_2} < 1 \end{array} \quad (\text{II2: 96})$$

Nur für den Fall einer perfekten Gleichläufigkeit der Anlagenrenditen kann also die Portfoliostandardabweichung einfach als gewichtetes Mittel der Einzelstandardabweichungen berechnet werden (Fall A). Ist dies nicht der Fall, fällt die Portfoliostandardabweichung immer geringer aus als das gewichtete Mittel der Einzelstandardabweichungen (Fall B).

Beispiel II 2-44

Varianz der Portfoliorendite

Zwei Wertpapiere seien durch folgende Monatsdaten gekennzeichnet:

Papier i	Erwartete Rendite $E(R_i)$ in %	Standardabweichung der Rendite σ_{R_i} in %	$\rho_{R_1 R_2} = 0,1$
1	1,5	5,5	
2	1,0	4,0	

Wir wollen nun das Risiko eines Portfolios, welches sich zu je 50 % aus beiden Aktien zusammensetzt, anhand der Portfoliostandardabweichung beurteilen. Berechnen wir dazu zunächst die erwartete Rendite des Portfolios. Sie ergibt sich zu

$$E(R_p) = 0,5 \cdot E(R_1) + 0,5 \cdot E(R_2) = 0,5 \cdot (1,5 + 1,0) = 1,25.$$

Für die Portfoliostandardabweichung ergibt sich

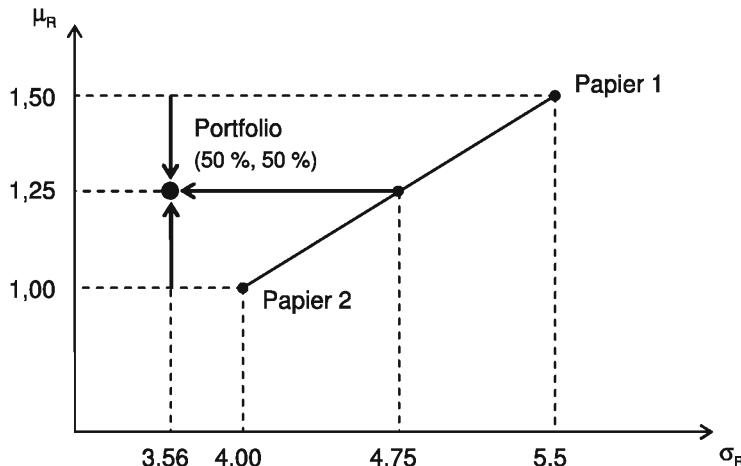
$$\begin{aligned}\text{Var}(R_p) &= w_1^2 \cdot \text{Var}(R_1) + w_2^2 \cdot \text{Var}(R_2) + 2 \cdot w_1 \cdot w_2 \cdot \sigma_{R_1} \cdot \sigma_{R_2} \cdot \rho_{R_1 R_2} \\ &= 0,5^2 \cdot 5,5^2 + 0,5^2 \cdot 4,0^2 + 2 \cdot 0,5 \cdot 5,5 \cdot 4,0 \cdot 0,1 = 12,66 \\ \sigma_{R_p} &= \sqrt{\text{Var}(R_p)} = \sqrt{12,66} = 3,56.\end{aligned}$$

Für das gewichtete Mittel der Einzelstandardabweichungen erhalten wir hingegen

$$\bar{\sigma} = w_1 \cdot \sigma_{R_1} + w_2 \cdot \sigma_{R_2} = 0,5 \cdot (5,5 + 4,0) = 4,75.$$

Investiert also ein Anleger z.B. in das Papier 2, so kann er bei einer Standardabweichung von 4,0 eine Rendite von 1,0 % erwarten. Kombiniert er dieses Papier aber mit dem Papier 1, so entsteht für ihn die Möglichkeit bei geringerem Risiko (3,56) eine höhere Rendite (1,25 %) zu erzielen. Ein solches Vorgehen, also die Aufteilung eines Investments auf verschiedene risikobehaftete Anlagen an Stelle der Investition in nur eine Anlage, wird als *Diversifikation* bezeichnet. Wünscht der Anleger eine höhere Rendite, so muss er einen höheren Anteil seines Kapitals in Papier 1 investieren. Mit der höheren Rendite hat der Anleger dann aber auch ein höheres Risiko zu tragen.

Zur grafischen Darstellung des Rendite-Risiko-Zusammenhangs wird häufig ein sog. μ - σ -Diagramm verwendet. In diesem Beispiel kann ein solches wie folgt skizziert werden:



Wie unsere Skizze zeigt, ist der Zusammenhang (II2: 96) in diesem Beispiel deutlich erfüllt. Wären die Renditen der Wertpapiere nun sogar negativ korreliert, so könnten wir das Portfoliorisiko noch weiter senken.

Häufig soll nicht nur die Varianz (oder Standardabweichung) eines bestehenden Portfolios berechnet, sondern ein Portfolio durch die Wahl der Gewichtung so zusammengestellt werden, dass die Varianz einen bestimmten Wert annimmt. Der Anteil der Anlage 1 in einem Portfolio aus 2 Anlagen, mit dem das Portfolio bei gegebenen Einzelvarianzen $\text{Var}(R_1)$ und $\text{Var}(R_2)$ sowie gegebenen Korrelationskoeffizienten $\rho_{R_1 R_2}$ die geringste Varianz aufweist, kann wie folgt ermittelt werden:

$$w_1^{\text{Var}(R_p)_{\min}} = \frac{\text{Var}(R_2) - \rho_{R_1 R_2} \cdot \sigma_{R_1} \cdot \sigma_{R_2}}{\text{Var}(R_1) + \text{Var}(R_2) - 2 \cdot \rho_{R_1 R_2} \cdot \sigma_{R_1} \cdot \sigma_{R_2}} \quad (\text{II2: 97})$$

Das damit entstehende Portfolio heißt *Minimum-Varianz-Portfolio*.

Beispiel II 2-45**Minimum-Varianz-Portfolio**

Ein Portfolio soll aus den zwei Wertpapieren des vorhergehenden Beispiels II 2-44 mit $\sigma_{R_1} = 5,5$ und $\sigma_{R_2} = 4,0$ zusammengesetzt werden. Der Korrelationskoeffizient lag bei diesen beiden Papieren bei $\rho_{R_1 R_2} = 0,1$. Mit welchem Anteil muss Wertpapier 1 in das Portfolio aufgenommen werden, um das Portfolio mit der geringsten Varianz zu erhalten?

$$w_1^{\text{Var}(R_p)_{\min}} = \frac{4,0^2 - 0,1 \cdot 5,5 \cdot 4,0}{5,5^2 + 4,0^2 - 2 \cdot 0,1 \cdot 5,5 \cdot 4,0} = \frac{13,8}{41,85} = 0,3297 = 32,97\%$$

Liegt ein *Portfolio* aus n Anlagen vor, so errechnen wir die *erwartete Rendite des Portfolios* nach (II2: 88) als

$$E(R_p) = \sum_{i=1}^n w_i \cdot E(R_i) . \quad (\text{II2: 98})$$

Für die *Portfoliovarianz* gilt nach (II2: 89)

$$\begin{aligned} \text{Var}(R_p) &= \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \text{Cov}(R_i, R_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \sigma_{R_i} \cdot \sigma_{R_j} \cdot \rho_{R_i R_j} \end{aligned} . \quad (\text{II2: 99})$$

Zum besseren Verständnis dieser Formel betrachten wir die nachfolgende Tabelle, die auch als (gewichtete) *Varianz-Kovarianz-Matrix* bezeichnet wird und die Komponenten der Varianz eines Portfolios mit n Anlagen übersichtlich darstellt. Auch der Fall mit 2 Anlagen ist darin gekennzeichnet (Grausättigung).

	1	2	3	...	n
1	$w_1^2 \cdot \text{Var}(R_1)$	$w_1 \cdot w_2 \cdot \text{Cov}(R_1, R_2)$	$w_1 \cdot w_3 \cdot \text{Cov}(R_1, R_3)$...	$w_1 \cdot w_n \cdot \text{Cov}(R_1, R_n)$
2	$w_2 \cdot w_1 \cdot \text{Cov}(R_2, R_1)$	$w_2^2 \cdot \text{Var}(R_2)$	$w_2 \cdot w_3 \cdot \text{Cov}(R_2, R_3)$...	$w_2 \cdot w_n \cdot \text{Cov}(R_2, R_n)$
3	$w_3 \cdot w_1 \cdot \text{Cov}(R_3, R_1)$	$w_3 \cdot w_2 \cdot \text{Cov}(R_3, R_2)$	$w_3^2 \cdot \text{Var}(R_3)$...	$w_3 \cdot w_n \cdot \text{Cov}(R_3, R_n)$
:	:	:	:	...	:
n	$w_n \cdot w_1 \cdot \text{Cov}(R_n, R_1)$	$w_n \cdot w_2 \cdot \text{Cov}(R_n, R_2)$	$w_n \cdot w_3 \cdot \text{Cov}(R_n, R_3)$...	$w_n^2 \cdot \text{Var}(R_n)$

Da die Kovarianz eines Wertpapieres mit sich selbst gleich seiner Varianz ist, d.h. $\text{Cov}(R_i, R_i) = \text{Var}(R_i)$ gilt (vgl. Abschnitt II 2.7.4.2), ist bei näherer Betrachtung der Tabelle festzustellen, dass die Portfoliovarianz durch die n Varianz-Terme auf der Hauptdiagonale und durch $n^2 - n$ Kovarianz-Terme bestimmt wird.

Nehmen wir an, wir besitzen derzeit ein Portfolio aus 2 Wertpapieren und beabsichtigen ein weiteres in unser Portfolio aufzunehmen. Dies bedeutet, dass bei der Berechnung der Portfoliovarianz ein weiterer Varianz-Term und vier weitere Kovarianz-Terme berücksichtigt werden müssen. Wir können also allgemein sagen, dass

die Anzahl der Kovarianz-Terme durch Hinzufügen weiterer Wertpapiere überproportional ansteigt. Die Portfoliovarianz wird daher weniger von den Einzelvarianzen, sondern vielmehr von den Kovarianzen bestimmt. Welchen Effekt das Hinzufügen eines Wertpapiers in ein Portfolio auf die Portfoliovarianz haben kann, wird im nachfolgenden Beispiel veranschaulicht.¹⁸

Beispiel II 2-46

Veränderungen der Portfoliovarianz



Gegeben seien drei Wertpapiere, deren erwartete Renditen und Varianzen (jeweils in %) und die zwischen ihnen herrschenden Korrelationen:

Wertpapier	Erw. Rendite	Varianz	Korrelation
1	1,50	5,50	1-2
2	1,00	4,00	1-3
3	2,00	6,00	2-3

Bilden wir nun ein gleichgewichtetes Portfolio aus den Papieren 1 und 2, d.h. nehmen wir jedes Papier zu einem Anteil von 50 % ins Portfolio auf, erhalten wir unter Nutzung der auf der vorhergehenden Seite abgebildeten (gewichteten) Varianz-Kovarianz-Matrix folgende erwartete Portfoliorendite und Portfoliovarianz:

	1	2	Portfoliorendite
1	1,375	0,821	1,250
2	0,821	1,000	Portfoliovarianz 4,017

Fügen wir nun zu diesem Portfolio das Wertpapier 3 hinzu. Wir wollen erneut ein gleichgewichtetes Portfolio bilden, d.h. wir müssen alle drei Wertpapiere mit 1/3 gewichten. Als Ergebnis erhalten wir dann:

	1	2	3	Portfoliorendite
1	0,611	0,365	0,128	1,500
2	0,365	0,444	0,163	Portfoliovarianz 3,034
3	0,128	0,163	0,667	

Wir erkennen, dass durch das zusätzliche Wertpapier die Varianz des Portfolios gesenkt und die erwartete Rendite sogar erhöht werden kann. Nehmen wir an, dass die Korrelation zwischen den bisherigen Wertpapieren und dem zusätzlichen Papier 3 negativ sind, d.h. konkret $\rho_{13} = -0,20$ und $\rho_{23} = -0,30$ gilt, so können wir eine noch stärkere Varianzreduktion erreichen:

	1	2	3	Portfoliorendite
1	0,611	0,365	-0,128	1,500
2	0,365	0,444	-0,163	Portfoliovarianz 1,870
3	-0,128	-0,163	0,667	

¹⁸ Eine anschauliche und kompakte Einführung in die Portfoliooptimierung bieten z.B. Auer, B.R., Schuhmacher, F. (2009).

3.

Theoretische Verteilungen

Die im Folgenden beschriebenen Wahrscheinlichkeitsverteilungen können aus zwei Perspektiven betrachtet werden: Zunächst dienen sie in der deskriptiven Statistik dazu, empirisch beobachtete Häufigkeitsverteilungen approximativ als Funktionen zu beschreiben. In der mathematischen Statistik lassen sich damit Wahrscheinlichkeiten für Ereignisse bestimmter Zufallsvariablen angeben. Wir werden uns einigen ausgewählten eindimensionalen diskreten und stetigen Verteilungen widmen. Dabei ist zu beachten, dass jedes der theoretischen Verteilungsmodelle eine Schar von unendlich vielen Einzelverteilungen repräsentiert. Erst durch die zahlenmäßige Belegung der Funktionalparameter ist eine Verteilung eindeutig festgelegt.

3.1 Diskrete Verteilungen

3.1.1 Binomialverteilung

3.1.1.1 Wahrscheinlichkeits- und Verteilungsfunktion

Die auf Jakob Bernoulli zurückgehende Binomialverteilung dient der Ermittlung von Wahrscheinlichkeiten für die Häufigkeit des Eintretens bestimmter Ereignisse bei sog. *Bernoulli-Experimenten* (z.B. Wahrscheinlichkeit für x Jungen bei n Geburten). Ein solches Experiment besteht aus einer Folge von Bernoulli-Versuchen, die folgende Eigenschaften aufweisen:

- Für jeden Versuch gibt es *nur zwei mögliche Ausgänge* (Ergebnisse) A und \bar{A} (z.B. Zahl/Wappen, Junge/Mädchen, fehlerhaft/fehlerfrei, Erfolg/Misserfolg). Für die Wahrscheinlichkeit dieser Ergebnisse gilt

$$P(A) = p \quad P(\bar{A}) = 1 - p \quad \text{mit} \quad 0 \leq p \leq 1. \quad (\text{II3: 1})$$

Man bezeichnet p häufig als Erfolgswahrscheinlichkeit und die Wahrscheinlichkeit des Gegenereignisses (komplementäres Ereignis) $1 - p$ als Misserfolgwahrscheinlichkeit.

- Das Zufallsexperiment wird *n-mal* (n ist eine natürliche Zahl) unter gleichen Bedingungen *wiederholt*. Damit sind die einzelnen Versuche voneinander *unabhängig*, d.h. das Ergebnis eines Versuches beeinflusst nicht die weiteren Ergebnisse. Außerdem sind die *Wahrscheinlichkeiten* (II3: 1) *konstant*.

Wir beschreiben dabei den i -ten Versuch durch eine Zufallsvariable X_i , die nur die Werte 0 und 1 annehmen kann (sog. *dichotome Zufallsvariable*). Es gilt für diese

$$X_i = \begin{cases} 1 & \text{wenn } A \text{ bei der } i\text{-ten Wiederholung eintritt} \\ 0 & \text{wenn } \bar{A} \text{ bei der } i\text{-ten Wiederholung eintritt.} \end{cases} \quad (\text{II3: 2})$$

Die diskrete Zufallsvariable X zählt nun, wie oft bei n Versuchen eines Bernoulli-Experimentes A eingetreten ist. X ist dann die Summe der X_i mit den Realisationen $x = 0, 1, 2, \dots, n$:

$$X = \sum_{i=1}^n X_i \quad (\text{II3: 3})$$

Beispiel II 3-1

Bernoulli-Experiment

Wird eine Münze dreimal geworfen (Kopf = 1, Zahl = 0), so kann eine Zufallsvariable X als "Anzahl der Köpfe" definiert werden. Diese Zufallsvariable kann die Realisationen $x = 0, 1, 2$ oder 3 annehmen. Welche Realisation eintritt, hängt vom Zufall ab. So kann sich z.B. beim ersten Wurf "Zahl" ($X_1 = 0$) und beim zweiten und dritten Wurf "Kopf" ($X_2 = X_3 = 1$) ergeben. Wir erhalten damit nach (II3: 3)

$$X = \sum_{i=1}^3 X_i = 0 + 1 + 1 = 2.$$

Gesucht ist nun die Wahrscheinlichkeit dafür, dass bei einem Bernoulli-Experiment mit n Versuchen das Ereignis A genau x -mal auftritt. Anders ausgedrückt, suchen wir die Wahrscheinlichkeit $P(X = x)$, dass X eine bestimmte Realisation x annimmt. Gerade diese Information liefert die Wahrscheinlichkeitsfunktion der Binomialverteilung, mit deren Herleitung wir uns nun näher befassen wollen:

Nehmen wir an, dass bei der Folge von n Versuchen das Ereignis A bei den ersten x Versuchen und bei den restlichen $n - x$ Versuchen das Gegenereignis \bar{A} eintritt.

$$\underbrace{A, A, \dots, A}_{x\text{-mal}}, \underbrace{\bar{A}, \bar{A}, \dots, \bar{A}}_{(n-x)\text{-mal}} \text{ bzw. } \underbrace{1, 1, \dots, 1}_{x\text{-mal}}, \underbrace{0, 0, \dots, 0}_{(n-x)\text{-mal}}$$

Da die einzelnen Versuche unabhängig voneinander sind, ergibt sich aus dem Zusammenhang

$$P(A) = P(X_i = 1) = p \quad \text{und} \quad P(\bar{A}) = P(X_i = 0) = 1 - p,$$

der für jedes $i = 1, 2, \dots, n$ gilt, und dem Multiplikationssatz für unabhängige Ereignisse (II1: 58) folgende Schlussfolgerung:

$$\underbrace{p \cdot p \cdot \dots \cdot p}_{x\text{-mal}} \cdot \underbrace{(1-p) \cdot (1-p) \cdot \dots \cdot (1-p)}_{(n-x)\text{-mal}} = p^x \cdot (1-p)^{n-x}$$

Bei der hier betrachteten Folge handelt es sich nun aber um eine spezielle Anordnung der x "Einsen" und $n - x$ "Nullen". Natürlich gibt es eine Reihe weiterer Folgen, bei denen ebenfalls genau x -mal Ereignis A bzw. 1, jedoch in anderer Reihenfolge, eintritt. Um die Anzahl dieser Möglichkeiten zu bestimmen, bedienen wir uns der Erkenntnisse der Kombinatorik. Zu beantworten ist die Frage, wieviele Möglichkeiten es gibt, die x "Einsen" auf die n Plätze aufzuteilen (Ziehen ohne Zurücklegen, ohne Berücksichtigung der Reihenfolge). Es existieren demnach genau " n über x " Folgen, von denen jede einzelne die Wahrscheinlichkeit $p^x \cdot (1-p)^{n-x}$ besitzt. Nach dem Additionssatz für sich gegenseitig ausschließende (disjunkte) Ereignisse ergibt sich daher

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}. \quad (\text{II3: 4})$$

Die Zufallsvariable X besitzt also folgende Wahrscheinlichkeitsfunktion $f_B(x)$:

$$f_B(x) = \begin{cases} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} & \text{für } x = 0, 1, 2, \dots, n \\ 0 & \text{sonst} \end{cases} \quad (\text{II3: 5})$$

Weist eine Zufallsvariable X die durch (II3: 5) definierte Wahrscheinlichkeitsfunktion auf, so heißt sie binomialverteilt. Wir schreiben kurz $X \sim B(n, p)$.

Zur Veranschaulichung dieser wichtigen Verteilung bedient man sich häufig des *Urnenmodells mit Zurücklegen*. Es befinden sich dabei in einer Urne mit N Kugeln genau M Kugeln, die eine bestimmte Eigenschaft A aufweisen. (Man könnte sich auch eine Warenlieferung des Umfangs N mit M defekten Produkten vorstellen.) Werden aus dieser Urne nun n Kugeln zufällig und mit Zurücklegen nacheinander gezogen, so ist gewährleistet, dass bei jedem weiteren Zug die gleiche Wahrscheinlichkeit $p = M/N$ vorliegt, eine Kugel mit der Eigenschaft A zu erhalten. Es wird schließlich durch das Zurücklegen die Urne immer wieder in den Ausgangszustand zurückversetzt. Die einzelnen Züge erfolgen also *unabhängig* voneinander. Da dies nun genau die Situation ist, die auch von der Binomialverteilung erfasst wird, kann man eine Zufallsvariable X: "Anzahl der Kugeln mit Eigenschaft A in einer Stichprobe von n Kugeln" als $B(n, M/N)$ -verteilt betrachten.

Beispiel II 3-2

Binomialverteilung und Urnenmodell

Wird aus einer Urne mit N Kugeln, von denen M eine gewisse Eigenschaft A besitzen, genau einmal ($x = 0, 1$) bzw. zweimal ($x = 0, 1, 2$) mit Zurücklegen gezogen, so können sich allgemein folgende Wahrscheinlichkeiten ergeben:

X ist $B(1, p)$ -verteilt:

$$P(X = 0) = \binom{1}{0} \cdot p^0 \cdot (1-p)^1 = 1-p$$

$$P(X = 1) = \binom{1}{1} \cdot p^1 \cdot (1-p)^0 = p$$

X ist $B(2, p)$ -verteilt:

$$P(X = 0) = \binom{2}{0} \cdot p^0 \cdot (1-p)^2 = (1-p)^2$$

$$P(X = 1) = \binom{2}{1} \cdot p^1 \cdot (1-p)^1 = 2 \cdot (p - p^2)$$

$$P(X = 2) = \binom{2}{2} \cdot p^2 \cdot (1-p)^0 = p^2$$

Für die Wahrscheinlichkeitsfunktionen ergibt sich damit:

x	0	1
$f_B(x)$	$1-p$	p

x	0	1	2
$f_B(x)$	$(1-p)^2$	$2 \cdot (p - p^2)$	p^2

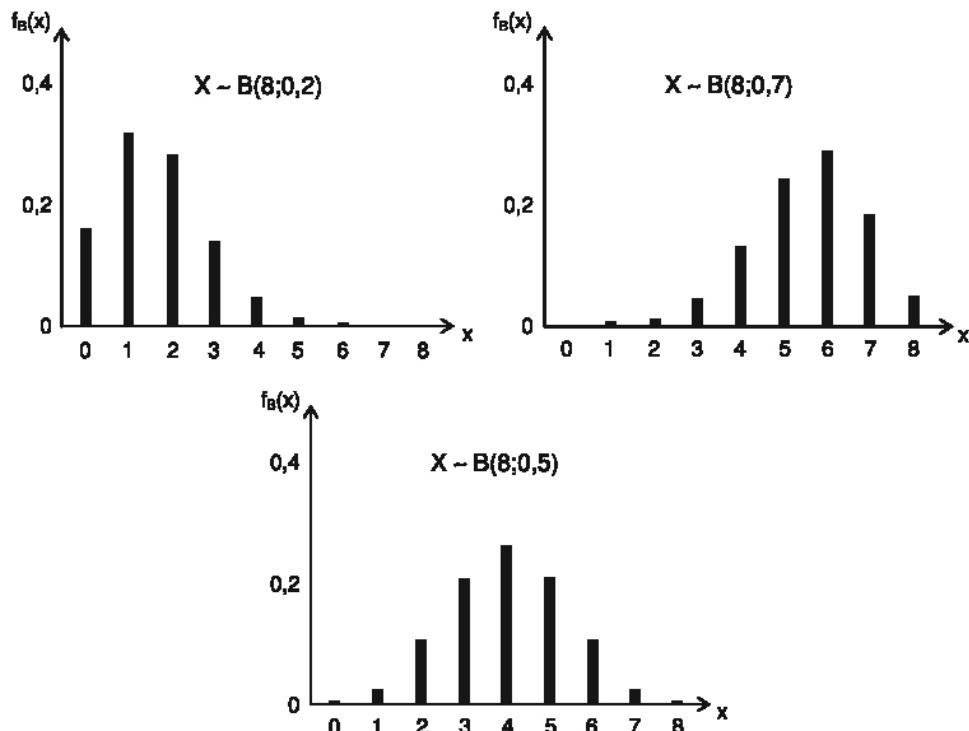
Betrachten wir nun konkret eine Urne mit 10 roten und 5 schwarzen Kugeln, aus der zufällig zwei Kugeln gezogen werden. Die Wahrscheinlichkeit, dabei genau eine schwarze Kugel zu erwischen, ergibt sich als

$$P(X = 1) = \binom{2}{1} \cdot \left(\frac{5}{15}\right)^1 \cdot \left(1 - \frac{5}{15}\right)^{2-1} = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}.$$

Die grafische Gestalt der Wahrscheinlichkeitsfunktion der Binomialverteilung hängt stets von ihren beiden **Funktionalparametern** n und p ab. Grundsätzlich gilt, dass alle Binomialverteilungen mit $p = 0,5$ symmetrisch sind. Für $p < 0,5$ sind die Verteilungen linkssteil, für $p > 0,5$ rechtssteil. Beispiel II 3-3 veranschaulicht dies.

Beispiel II 3-3

Verschiedene Binomialverteilungen



Die **Verteilungsfunktion** $F_B(x)$ der Binomialverteilung gibt die Wahrscheinlichkeit an, dass eine binomialverteilte Zufallsvariable X höchstens den Wert x annimmt, also $P(X \leq x)$. Sie besitzt, wie jede diskrete Verteilungsfunktion, die Gestalt einer **Treppenfunktion** und wird durch Summation der einzelnen binomialen Wahrscheinlichkeiten gebildet:

$$\begin{aligned} F_B(x) &= f_B(0) + f_B(1) + \dots + f_B(x) \\ &= \binom{n}{0} \cdot p^0 \cdot (1-p)^n + \binom{n}{1} \cdot p^1 \cdot (1-p)^{n-1} + \dots + \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \end{aligned}$$

Unter Verwendung eines Summationsindex v können wir auch schreiben:

$$F_B(x) = \begin{cases} 0 & \text{für } x < 0 \\ \sum_{v=0}^x \binom{n}{v} \cdot p^v \cdot (1-p)^{n-v} & \text{für } 0 \leq x < n \\ 1 & \text{für } n \leq x \end{cases} \quad (\text{II3: 6})$$

Beispiel II 3-4**Wahrscheinlichkeits- und Verteilungsfunktion**

Aus einer großen Geräteproduktion, bei der erfahrungsgemäß 30 % defekt sind, werden 6 Geräte gekauft. Wie groß ist die Wahrscheinlichkeit dafür, dass

- a) alle 6 Geräte defekt sind,
- b) alle 6 Geräte in Ordnung sind,
- c) höchstens 2 Geräte defekt sind,
- d) die Hälfte oder mehr als die Hälfte der Geräte defekt sind?

Betrachten wir die Anzahl defekter Geräte X als $B(6;0,3)$ -verteilt, so erhalten wir folgende Ergebnisse:

$$a) P(X = 6) = \binom{6}{6} \cdot 0,3^6 \cdot 0,7^0 = 0,0007$$

$$b) P(X = 0) = \binom{6}{0} \cdot 0,3^0 \cdot 0,7^6 = 0,1176$$

$$c) P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0,1176 + 0,3025 + 0,3241 = 0,7443$$

$$d) P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0,7443 = 0,2557$$

Im Anhang ist die Wahrscheinlichkeits- und die Verteilungsfunktion der Binomialverteilung für ausgewählte (gebräuchliche) Werte von n und p in tabellarischer Form abgedruckt. Daraus lassen sich in vielen Fällen Wahrscheinlichkeiten ohne formelmäßige Berechnungen recht schnell ermitteln. Dazu nutzt man, $f(x) = F(x) - F(x - 1)$, um Werte der Wahrscheinlichkeitsfunktion auch aus der Verteilungsfunktion zu errechnen.

3.1.1.2 Eigenschaften

Wie wir in (II3: 3) gesehen haben, ist X eine Summe unabhängiger, identisch $B(1,p)$ verteilter Zufallsvariablen X_i mit $E(X_i) = 0 \cdot (1-p) + 1 \cdot p = p$. Für $X \sim B(n,p)$ ergibt sich daher als *Erwartungswert* von X

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = n \cdot p. \quad (\text{II3: 7})$$

Der Erwartungswert einer binomialverteilten Zufallsvariable X ist somit nichts anderes als das Produkt der beiden Funktionalparameter n und p .

Für die Varianz eines einzelnen Zuges gilt

$$\text{Var}(X_i) = \sum_{i=1}^2 (x_i - E(X_i))^2 \cdot f(x_i) = (0 - p)^2 \cdot (1-p) + (1 - p)^2 \cdot p = \dots = p \cdot (1-p).$$

Daraus lässt sich für die *Varianz* von X der Zusammenhang

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n p \cdot (1-p) = n \cdot p \cdot (1-p) \quad (\text{II3: 8})$$

ableiten. Das zweite Gleichheitszeichen gilt dabei, da die einzelnen Züge voneinander *unabhängig* sind.

Der Sonderfall $n = 1$, der in der Literatur häufig als separate Verteilung (*Bernoulli-Verteilung*) behandelt wird, lässt $E(X)$ und $\text{Var}(X)$ auf Erwartungswert und Varianz des einzigen durchgeführten Bernoulli-Versuchs schrumpfen. Solche $B(1,p)$ -verteilte Zufallsvariablen werden häufig als Indikatorvariablen, dichotome oder binäre Zufallsvariablen oder als *Dummy* bezeichnet.

Beispiel II 3-5

Erwartungswert und Varianz

Eine faire Münze wird sechsmal geworfen. Die Zufallsvariable X soll dabei die Anzahl der erscheinenden "Kopf-Seiten" repräsentieren.

Der Erwartungswert $E(X) = 6 \cdot 0,5 = 3$ besagt nun in diesem Beispiel, dass man beim sechsmaligen Münzwurf im Durchschnitt dreimal das Ereignis "Kopf" erhält. Die Varianz nimmt hier einen Wert von $\text{Var}(X) = 6 \cdot 0,5 \cdot 0,5 = 1,5$ an.

Wird die Münze nur ein einziges Mal geworfen, ergibt sich $E(X) = 0,5$ und $\text{Var}(X) = 0,5 \cdot 0,5 = 0,25$, wodurch die 50:50 Chance des Ergebnisses zum Ausdruck kommt.

Aus den Herleitungen von Erwartungswert und Varianz binomialverteilter Zufallsvariablen folgt unmittelbar die sog. *Reproduktivität* der Binomialverteilung. Dies bedeutet, dass wenn $X \sim B(n_1, p)$ und $Y \sim B(n_2, p)$ unabhängig sind, stets $X + Y \sim B(n_1 + n_2, p)$ gilt.

3.1.1.3 Praxisanwendung: Operationscharakteristiken

Zur Entscheidung, ob eine Lieferung wieder an den Lieferanten zurückgeschickt werden soll, bedient man sich in der Praxis häufig eines statistischen Prüfplans. Nehmen wir an, ein Unternehmen erhält regelmäßig größere Lieferungen von gewissen Einzelteilen, die einen unbekannten Anteil p (Ausschussanteil) von defekten Stücken enthalten. Es werden daraus n Stücke zufällig herausgegriffen und die Anzahl X der fehlerhaften Stücke in der Stichprobe ermittelt. Es wird zwar in der Praxis ohne Zurücklegen gezogen, doch ist der Umfang n der Stichprobe in der Regel im Verhältnis zum Umfang N der Lieferungen sehr klein, sodass es für den Fehleranteil in der Grundgesamtheit beim nächsten Zug keine Rolle spielt, ob bereits ein Stück zur Prüfung entnommen wurde, und somit auch vom Ziehen mit Zurücklegen ausgegangen werden kann.¹ Ein *statistischer Prüfplan* stellt nun die Vorschrift dar, dass die Lieferung nicht zu akzeptieren ist, wenn die Anzahl X eine gewisse vorher definierte Grenze c (Annahmezahl) überschreitet. Ist $X \leq c$, wird die Lieferung angenommen.

Zur Beurteilung eines solchen Prüfplans ziehen wir seine sog. *Operationscharakteristik* $L_{n,c}(p)$ heran, die die Wahrscheinlichkeit für die Annahme einer bestimmten Lieferung in Abhängigkeit vom Ausschussanteil p dieser Lieferung angibt:

$$L_{n,c}(p) = P(X \leq c) = \sum_{v=0}^c \binom{n}{v} \cdot p^v \cdot (1-p)^{n-v} \quad (\text{II3: 9})$$

Die sog. *OC-Kurve* (vgl. Beispiel II 3-6) ist damit festgelegt durch die beiden Parameter n und c .

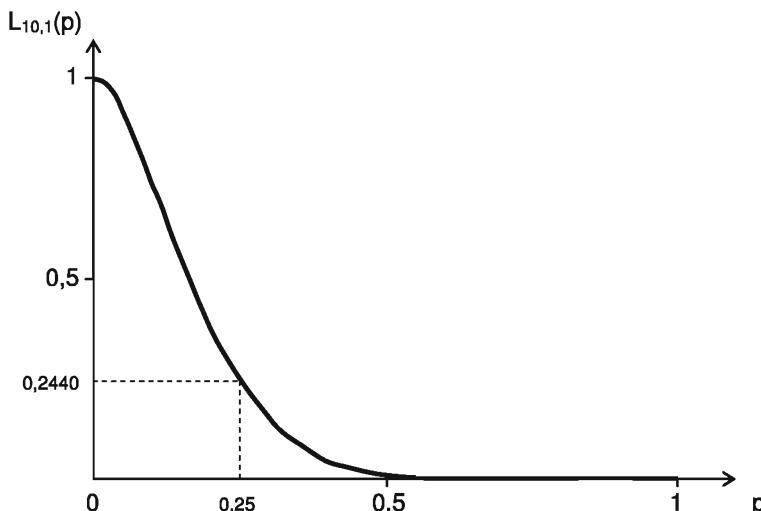
¹ Dies werden wir im Abschnitt II 3.1.2.3 zur Hypergeometrischen Verteilung bzw. ihrer Approximation durch die Binomialverteilung im Detail behandeln.

Beispiel II 3-6**Operationscharakteristik**

Für $n = 10$ und $c = 1$ ergibt sich für die Operationscharakteristik

$$L_{10,1}(p) = \sum_{v=0}^1 \binom{n}{v} \cdot p^v \cdot (1-p)^{n-v} = \binom{10}{0} \cdot p^0 \cdot (1-p)^{10} + \binom{10}{1} \cdot p^1 \cdot (1-p)^{10-1} = (1-p)^9 \cdot (1+9 \cdot p).$$

Für die minimale ($p = 0$) und maximale ($p = 1$) Ausschusswahrscheinlichkeit ergeben sich die Werte $L_{10,1}(0) = 1$ und $L_{10,1}(1) = 0$. Grafisch zeigt die OC-Kurve folgendes Bild:



Diese Funktion sagt nun aus, dass z.B. bei einem Ausschussanteil von $p = 0,25$ die Lieferung mit einer Wahrscheinlichkeit von $L_{10,1}(0,25) = 0,2440$ angenommen wird. Ein solches Ergebnis macht diese OC-Kurve für die Praxis aber untauglich, da eine Lieferungsannahme unter solchen Bedingungen nicht akzeptabel ist.

Auf einer OC-Kurve werden in der Praxis zwei Punkte vorgegeben, mit Hilfe derer entschieden werden kann, ob die jeweilige Kurve bzw. das Prüfverfahren tauglich ist. Diese Punkte sind die folgenden:

- **Ablehngrenze:** Ausschussanteil, bei dem die Lieferung mit einer kleinen Wahrscheinlichkeit von z.B. 5 % angenommen werden soll (d.h. die Annahmewahrscheinlichkeit der Ablehngrenze beträgt 5 %)
- **Annahmegrenze:** Ausschussanteil, bei dem die Lieferung mit einer großen Wahrscheinlichkeit von z.B. 95 % angenommen werden soll

Beispiel II 3-7**Ablehn- und Annahmegrenze**

Ein Unternehmen hat eine Annahmegrenze von 0,02 und eine Ablehngrenze von 0,10 definiert. Für zwei Prüfpläne A ($n = 20, c = 1$) und B ($n = 100, c = 4$) sollen nun die Annahmewahrscheinlichkeiten (Werte der OC-Kurven) für die gegebenen Annahme- und Ablehnungsgrenzen ermittelt werden.

$$A: L_{20,1}(0,02) = 0,98^{20} + \binom{20}{1} \cdot 0,02^1 \cdot 0,98^{19} = 0,940 \quad L_{20,1}(0,10) = \dots = 0,392$$

$$B: L_{100,4}(0,02) = \dots = 0,949 \quad L_{100,4}(0,10) = \dots = 0,024$$

Diese Werte zeigen, dass sich Prüfplan A nicht als praxistauglich erweist, da eine Lieferung, bei der 10 % der Einzelteile fehlerhaft sind, mit einer Wahrscheinlichkeit von 39,2 % angenommen wird. Bei Prüfplan B liegt diese Wahrscheinlichkeit bei nur 2,4 % und für die Annahmegrenze beträgt der Wert der OC auch fast 95 %, sodass dieser geeigneter ist.

3.1.2 Hypergeometrische Verteilung

3.1.2.1 Wahrscheinlichkeits- und Verteilungsfunktion

Wie wir gesehen haben, entsprach die Binomialverteilung dem Urnenmodell mit Zurücklegen, da hier die Unabhängigkeit durch das Zurücklegen der gezogenen Kugeln realisiert wurde. Bei Zufallsstichproben in der Praxis werden die einmal gezogenen Elemente jedoch i.d.R. nicht wieder zurückgelegt, bevor das nächste Stichprobenelement entnommen wird. Mit diesem *Urnenmodell ohne Zurücklegen* kann die *Hypergeometrische Verteilung* beschrieben werden:

Es werden aus einer Urne mit insgesamt N Kugeln, von denen M eine gewisse Eigenschaft A (z.B. schwarz) und N – M eine Eigenschaft \bar{A} (also nicht A, z.B. weiß) aufweisen, genau n Kugeln ($n \leq N$) rein zufällig gezogen. Eine einmal entnommene Kugel wird jedoch nicht wieder in die Urne zurückgelegt. Es vermindert sich dadurch nach jedem Zug die Anzahl der Kugeln, die sich noch in der Urne befinden. Außerdem kommt es so zu einer Veränderung der Wahrscheinlichkeiten beim nächsten Zug, eine Kugel mit einer der beiden definierten Eigenschaften zu erhalten.

Beispiel II 3-8

Ziehen ohne Zurücklegen

Wir erhalten beim ersten Zug einer Kugel aus einer mit 3 schwarzen und 3 weißen Kugeln gefüllten Urne mit einer Wahrscheinlichkeit von 0,5 eine schwarze und mit 0,5 eine weiße Kugel. Wird nun z.B. beim ersten Zug eine schwarze Kugel erwischt, die nicht mehr zurückgelegt wird, reduziert sich die Zahl der Kugeln in der Urne auf 5 (2 schwarze, 3 weiße). Die Wahrscheinlichkeit bei einem zweiten Zug erneut eine schwarze Kugel zu erwischen, ist nun nicht mehr 0,5, sondern nur noch 2/5.

Wir interessieren uns auch hier wieder für die Zufallsvariable X: "Anzahl der gezogenen Kugeln mit Eigenschaft A". Die Zufallsvariable X konnte beim Urnenmodell mit Zurücklegen die Realisationen 0, 1, ..., n annehmen. Beim Urnenmodell ohne Zurücklegen unterliegen die möglichen Werte x von X der Restriktion $0 \leq x \leq M$, da A nur M-mal in der Urne vorkommt, und der weiteren Restriktion $0 \leq n - x \leq N - M$, da insgesamt N – M Kugeln die Eigenschaft A nicht besitzen. Die genaue Bedeutung dieser Bedingungen können wir uns mittels eines einfachen Beispiels recht schnell veranschaulichen. Ist eine Urne beispielsweise mit roten und schwarzen

Kugeln gefüllt, so können sowohl nicht mehr rote als auch schwarze Kugeln in die Stichprobe gezogen werden, als in der Urne sind.

Es ist nun konkret die Wahrscheinlichkeit von Interesse, dass unter den n gezogenen Kugeln genau x mit der Eigenschaft A zu finden sind bzw. X die Realisation x annimmt: Grundsätzlich gilt, dass es bei einem Ereignis $X = x$ nicht darauf ankommt, in welcher Reihenfolge die n ohne Zurücklegen gezogenen Kugeln aus der Gesamtmenge N entnommen wurden. Zur Bestimmung der Wahrscheinlichkeitsfunktion (nach dem Laplaceschen Wahrscheinlichkeitsbegriff) kann daher für die Anzahl der möglichen Elementarereignisse die Anzahl der gleichwahrscheinlichen Möglichkeiten verwendet werden, aus N Kugeln ohne Beachtung der Reihenfolge genau n herauszunehmen:

$$\binom{N}{n} \quad (\text{II3: 10a})$$

Zur Bestimmung der Anzahl der für $X = x$ günstigen Elementarereignisse betrachten wir zunächst, dass aus M Kugeln mit der Eigenschaft A genau x ausgewählt werden. Auch hier spielt die Reihenfolge keine Rolle. Die Anzahl der Kombinationen ohne Wiederholung der einzelnen Kugeln und ohne Beachtung der Reihenfolge beträgt nun

$$\binom{M}{x}. \quad (\text{II3: 10b})$$

Analog beträgt die Anzahl der Möglichkeiten, aus den $N - M$ Kugeln, die die Eigenschaft A nicht aufweisen, die restlichen $n - x$ Kugeln der Stichprobe zu ziehen

$$\binom{N - M}{n - x}. \quad (\text{II3: 10c})$$

Da nun jede mögliche Stichprobe "x Kugeln mit Eigenschaft A aus M Kugeln" mit jeder einzelnen möglichen Stichprobe "n - x Kugeln ohne Eigenschaft A aus N - M Kugeln" kombiniert werden kann, ergibt sich die Anzahl der Möglichkeiten, dass genau x der n Kugeln die Eigenschaft A aufweisen, als

$$\binom{M}{x} \binom{N - M}{n - x}. \quad (\text{II3: 10d})$$

Fassen wir diese Erkenntnisse nach der Laplaceschen Wahrscheinlichkeitsdefinition zusammen, so ergibt sich für die gesuchte Wahrscheinlichkeit

$$P(X = x) = \frac{\binom{M}{x} \binom{N - M}{n - x}}{\binom{N}{n}}. \quad (\text{II3: 11})$$

Die *Wahrscheinlichkeitsfunktion* $f_h(x)$ der Hypergeometrischen Verteilung nimmt die in (II3: 11) angegebenen Werte für die möglichen x -Werte und sonst den Wert Null an. Besitzt eine Zufallsvariable eine solche Wahrscheinlichkeitsfunktion, so heißt sie hypergeometrisch verteilt. Wir schreiben kurz $X \sim H(N, M, n)$.

Beispiel II 3-9**Verschiedene Fallbeispiele****1. Urnenspiel**

In einer Urne seien 5 rote und 3 weiße Kugeln. Es wird eine Stichprobe von 4 Kugeln ohne Zurücklegen gezogen. Wie wahrscheinlich enthält die Stichprobe 2 rote Kugeln?

X ist H(8;5;4)-verteilt. Es gilt damit

$$P(X = 2) = \binom{5}{2} \cdot \binom{3}{2} / \binom{8}{4} = \frac{10 \cdot 3}{70} = \frac{3}{7}.$$

2. Lotto

Wie groß ist die Wahrscheinlichkeit für 5 Richtiges im deutschen Zahlenlotto?

X ist H(49;6;6)-verteilt. Wir erhalten daher

$$P(X = 5) = \binom{6}{5} \cdot \binom{43}{1} / \binom{49}{6} = \frac{6 \cdot 43}{13.983.816} = 0,00001845.$$

3. Warenprüfung

In einer WarenSendung vom Umfang 100 Stück befinden sich 10 Stücke Ausschuss. Wir entnehmen zufällig ohne Zurücklegen 5 Geräte aus dieser WarenSendung. Damit ist die Zufallsvariable X: "Anzahl fehlerhafter Stücke in der Stichprobe" H(100;10;5)-verteilt.

Beschreiben wir wie in (II3: 2) die i-te Ziehung aus der Urne durch eine Indikatorvariable X_i , die dann den Wert 1 annimmt, wenn die Eigenschaft A bei der i-ten gezogenen Kugel vorliegt, und den Wert 0 annimmt, wenn A beim i-ten Zug nicht gegeben ist, so gilt wieder (II3: 3). X ergibt sich also als Summe der X_i . Auch beim Ziehen ohne Zurücklegen besitzen die X_i , welche vom Standpunkt vor dem ersten Zug aus die i-te Entnahme beschreiben, alle dieselbe B(1,M/N)-Verteilung. Im Vergleich zum Ziehen mit Zurücklegen ist allerdings eine Unabhängigkeit der X_i nicht mehr gegeben. Wurden bis zum i-ten Zug genau k Kugeln mit der Eigenschaft A gezogen, so gilt für die Wahrscheinlichkeit, dass beim (i+1)-ten Zug A auftritt

$$P\left(X_{i+1} = 1 \mid \sum_{j=1}^i X_j = k\right) = \frac{M - k}{N - i}. \quad (\text{II3: 12})$$

Beispiel II 3-10**Bedingte Wahrscheinlichkeit**

Im Lager eines Weinhändlers befindet sich noch eine Kiste mit 30 Flaschen eines Rotweines Marke Retzbach, die allerdings schon seit langer Zeit dort lagern und deshalb die Gefahr besteht, dass sich der edle Tropfen bereits zu "Essig" gewandelt hat. Ein potenzieller Abnehmer ist nun bereit diese Kiste zu einem Vorzugspreis abzunehmen, verlangt aber, dass 5 Flaschen geöffnet werden und bei der Verprobung mindestens 3 der 5 geöffneten Flaschen einwandfrei sind. Nehmen wir nun an, dass die ersten beiden Flaschen bereits "Essig" waren. Wie groß sind nun unter dieser Bedingung die Wahrscheinlichkeiten dafür, dass

- a) die dritte Flasche, die gekostet wird, einwandfrei ist
 - b) der Kunde den Kauf letztendlich tätigt,
- wenn tatsächlich 40 % der Weinflaschen einwandfrei waren.

Definieren wir zur Lösung dieser Problemstellung die Indikatorvariable

$$X_i = \begin{cases} 1 & \text{falls } i\text{-te verprobte Weinflasche einwandfrei} \\ 0 & \text{falls } i\text{-te verprobte Weinflasche bereits "Essig".} \end{cases}$$

Für die gesuchten Wahrscheinlichkeiten gilt damit Folgendes:

a) $P(X_3 = 1 | X_1 = 0, X_2 = 0) = \frac{12 - 0}{30 - 2} = 0,4286$

b) $P\left(\sum_{i=1}^5 X_i = 3 \mid X_1 = 0, X_2 = 0\right) = \frac{\binom{12}{3} \binom{28 - 12}{0}}{\binom{28}{3}} = 0,0672$

Das Bild einer Hypergeometrischen Verteilung ist durch ihre *Funktionalparameter* N , M und n bestimmt. Die *Verteilungsfunktion* $F_n(x)$ wird analog zur Binomialverteilung durch Summation der Einzelwahrscheinlichkeiten gebildet. Auch sie gibt Auskunft über die Wahrscheinlichkeit $P(X \leq x)$, dass höchstens x Kugeln mit der Eigenschaft A in der Stichprobe enthalten sind.



3.1.2.2 Eigenschaften

Der *Erwartungswert* einer $H(N,M,n)$ -verteilten Zufallsvariable X ist identisch mit dem einer $B(n,M/N)$ -verteilten Zufallsvariable X .

$$E(X) = n \cdot \frac{M}{N} \quad (\text{II3: 13})$$

Auch die *Varianz* ist ähnlich. Sie unterscheidet sich nur durch einen zusätzlichen Faktor $(N - n)/(N - 1)$. Es gilt

$$\text{Var}(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N - n}{N - 1}. \quad (\text{II3: 14})$$

Der auch häufig als *Endlichkeitskorrekturfaktor* bezeichnete Faktor $(N - n)/(N - 1)$ ist für $n > 1$ stets kleiner als Eins. Damit ist auch die Varianz einer $B(n,M/N)$ -verteilten Zufallsvariable größer als die einer $H(N,M,n)$ -verteilten. Dies ist unmittelbar einleuchtend, da die Varianz kleiner sein muss, wenn ohne Zurücklegen gezogen wird. Erhöhen wir n bei konstantem N , so wird die Varianz immer geringer. Ist $n = N$, gibt es schließlich nur noch ein mögliches Stichprobenergebnis und die Varianz ist Null.

Ist der Auswahlsatz n/N sehr klein, ist der Faktor $(N - n)/(N - 1)$ in der praktischen Anwendung nahezu ohne Bedeutung (vgl. Abschnitt II 3.1.2.3).

Beispiel II 3-11

Erwartungswert und Varianz

Von 20 Firmenkunden einer Bank verfügen 12 über eine Eigenkapitalausstattung von mehr als 20 %. Wie groß sind Erwartungswert und Varianz einer Zufallsvariablen X : "Anzahl der Kunden in einer Stichprobe von 10 Kunden, deren Bilanz mehr als 20 % Eigenkapital aufweist", wenn mit Zurücklegen (a) oder ohne Zurücklegen (b) gezogen wird?

a) X ist $B(10;12/20)$ -verteilt.

$$E(X) = n \cdot \frac{M}{N} = 10 \cdot \frac{12}{20} = 6$$

$$\text{Var}(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) = 10 \cdot \frac{12}{20} \cdot \frac{8}{20} = 2,4$$

b) X ist $H(20;12;10)$ -verteilt.

$$E(X) = 6$$

$$\text{Var}(X) = 2,4 \cdot \frac{10}{19} = 1,3$$

3.1.2.3 Approximation durch die Binomialverteilung

Ist der Umfang N der betrachteten Grundgesamtheit im Verhältnis zum Stichprobenumfang n sehr groß, dann liegt die Vermutung nahe, dass das Nichtzurücklegen keinen wesentlichen Einfluss auf die Wahrscheinlichkeitsverteilung hat. Tatsächlich lässt sich beweisen, dass sich die Wahrscheinlichkeitsfunktion der $H(N,M,n)$ -Verteilung für $N \rightarrow \infty$ und $M/N = \text{konstant}$ für alle x einer $B(n,M/N)$ -Verteilung nähert. Beispiel II 3-12 zeigt die Güte der Verteilungsübereinstimmung auf. Zudem ist darin zu erkennen, dass für größer werdendes N bei konstantem n der Faktor $(N - n)/(N - 1)$, der die Varianz von Hypergeometrischer Verteilung und Binomialverteilung unterscheidet, gegen 1 konvergiert und so zwischen den beiden Varianzen kaum noch eine Differenz besteht.

Beispiel II 3-12

Annäherungsgüte



Zwei Lieferungen bestehen aus 100 bzw. 10 Geräten, von denen 30 bzw. 3 defekt sind. Bei einer zufälligen Entnahme von 4 Geräten mit Zurücklegen erhalten wir für beide Lieferungen dieselbe Binomialverteilung mit den Parametern $n = 4$ und $p = M/N = 0,3$, d.h. die Wahrscheinlichkeiten für das Auftreten von 0, 1, 2, 3 oder 4 defekten Geräten in der Auswahl bleiben gleich. Bei einer Entnahme ohne Zurücklegen erhalten wir zwei Hypergeometrische Verteilungen mit $N = 100$, $M = 30$ und $n = 4$ bzw. $N = 10$, $M = 3$, $n = 4$.

x	Hypergeometrische Verteilungen		Binomialverteilung $B(4;0,3)$
	$H(10;3;4)$	$H(100;30;4)$	
0	0,1667	0,2338	0,2401
1	0,5000	0,4188	0,4116
2	0,3000	0,2679	0,2646
3	0,0333	0,0725	0,0756
4	-	0,0070	0,0081
Σ	1,0000	1,0000	1,0000
$E(X)$	1,20	1,20	1,20
$(N - n)/(n - 1)$	0,67	0,97	1,00
$\text{Var}(X)$	0,56	0,81	0,84
			→

Wir erkennen hier sehr schön, dass mit zunehmendem N bzw. fallendem n/N (n konstant) der Faktor $(N - n)/(n - 1)$ gegen 1 läuft und sich die Varianzen immer mehr annähern. Außerdem ist zu erkennen, dass sich die Werte von $f(x)$ immer mehr ähneln.

Als *Faustregel* können wir festhalten, dass eine $H(N, M, n)$ -Verteilung dann hinreichend genau durch eine $B(n, M/N)$ -Verteilung approximiert werden kann, wenn für den Auswahlsatz n/N

$$\frac{n}{N} \leq 0,1 \quad \rightarrow \quad H(N, M, n) \approx B(n, M / N) \quad (\text{II3: 15})$$

gilt. Ist (II3: 15) erfüllt, ist der Effekt des Nichtzurücklegens vernachlässigbar klein, sodass auch wenn eine Stichprobe ohne Zurücklegen gezogen wird, anstelle der Hypergeometrischen Verteilung die Binomialverteilung zur Wahrscheinlichkeitsberechnung verwendet werden kann.

3.1.3 Poissonverteilung

3.1.3.1 Wahrscheinlichkeits- und Verteilungsfunktion

Binomial- oder hypergeometrisch verteilte Zufallsvariablen zählen, wie oft bei normaligem Ziehen aus einer Grundgesamtheit ein bestimmtes Ereignis A eintritt. Der Wertebereich ist nach oben durch n begrenzt und somit endlich. Die Poissonverteilung eignet sich ebenfalls zur Modellierung von Zählvorgängen.² Dabei werden bestimmte Ereignisse gezählt, die innerhalb eines festen, vorgegebenen Zeitintervall es eintreten können. Die mögliche Anzahl der Ereignisse ist nicht nach oben begrenzt. Typische Zählvorgänge dieser Art sind beispielsweise die Anzahl von Schadensmeldungen bei einer Versicherung innerhalb eines Jahres, die Anzahl von Krankheitsfällen einer seltenen Krankheit, oder die Anzahl von Kunden, die in einem Monat bei einer Bank einen Kredit beantragen.

Eine diskrete Zufallsvariable X heißt poissonverteilt, wenn ihre *Wahrscheinlichkeitsfunktion* $f_{po}(x)$ mit $\lambda > 0$ und e als Eulersche Zahl ($2,71828\dots$) die Gestalt

$$f_{po}(x) = \begin{cases} \frac{\lambda^x}{x!} \cdot e^{-\lambda} & \text{für } x = 0, 1, 2, \dots \\ 0 & \text{sonst} \end{cases} \quad (\text{II3: 16})$$

besitzt. Wir schreiben kurz $X \sim Po(\lambda)$. Das Besondere an dieser Verteilung ist, dass X die Null und alle natürlichen Zahlen annehmen kann. Der Faktor $x!$ im Nenner der Wahrscheinlichkeitsfunktion sorgt dafür, dass die Wahrscheinlichkeiten mit zunehmendem x trotz des λ^x im Zähler schnell sehr klein werden. Egal wie groß x aber auch werden mag, bleiben die $f_{po}(x)$ immer positiv (vgl. Beispiel II 3-13).

Funktionalparameter einer Poissonverteilung ist allein λ . Je größer λ ist, desto mehr verschiebt sich das Zentrum der Verteilung entlang der natürlichen Zahlen gerade nach rechts und desto größer wird die Streuung der Verteilung.

Die *Verteilungsfunktion* $F_{po}(x)$ der Poissonverteilung ist für $x \geq 0$ definiert als

$$F_{po}(x) = e^{-\lambda} \cdot \sum_{v=0}^x \frac{\lambda^v}{v!}. \quad (\text{II3: 17})$$

² Auch die geometrische Verteilung (vgl. Abschnitt II 3.2.2.2) modelliert einen Zählvorgang. Hier zählt man, wie lange man warten muss, bis ein Ereignis A zum erstenmal eintritt.

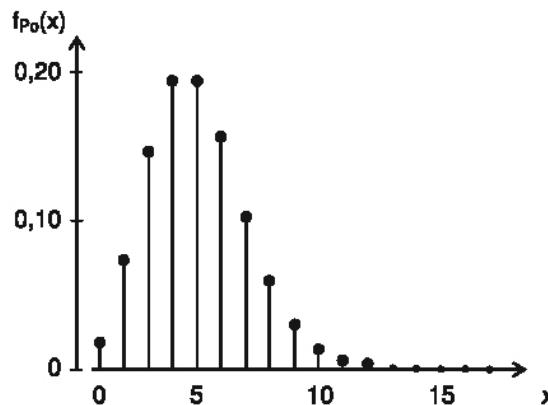
Beispiel II 3-13

Po(4)

Für eine Po(4)-verteilte Zufallsvariable X nehmen Wahrscheinlichkeits- und Verteilungsfunktion folgende Werte an:

x	$f_{Po}(x)$	$F_{Po}(x)$	x	$f_{Po}(x)$	$F_{Po}(x)$	x	$f_{Po}(x)$	$F_{Po}(x)$
0	0,0183	0,0183	5	0,1563	0,7851	10	0,0053	0,9972
1	0,0733	0,0916	6	0,1042	0,8893	11	0,0019	0,9991
2	0,1465	0,2381	7	0,0595	0,9489	12	0,0006	0,9997
3	0,1954	0,4335	8	0,0298	0,9786	13	0,0002	0,9999
4	0,1954	0,6288	9	0,0132	0,9919	14	0,0001	1,0000

Grafisch zeigt die Wahrscheinlichkeitsverteilung damit folgendes Bild:



Eine klassische Anwendung der Poissonverteilung findet sich in sog. *Warteschlangenmodellen*, bei denen die *Anzahl der Ankünfte in einer Warteschlange pro Zeiteinheit* (z.B. vor einem Bankschalter oder einem Grenzübergang) untersucht wird.³ Die Ankünfte werden dabei als stochastisch und die Zahl der Ankünfte als poissonverteilt angesehen. Eine wichtige Voraussetzung für Verwendung der Poissonverteilung ist aber, dass die *Ankünfte unabhängig voneinander* erfolgen. Als λ verwenden wir in solchen Modellen die *mittlere Ankunftsrate*.

Beispiel II 3-14

Warteschlangenmodell

In einer kleinen Bankfiliale auf dem Land erscheinen an einem Bankschalter im Durchschnitt pro Stunde 4 Kunden, um Ihre Bankgeschäfte abzuwickeln. Der Filialdirektor interessiert sich nun dafür, wie groß die Wahrscheinlichkeit dafür ist, dass in einer Stunde mehr als 8 Kunden auftauchen. Unter Verwendung der Tabelle aus Beispiel II 3-13 gilt für diese Wahrscheinlichkeit $P(X > 9) = 1 - P(X \leq 8) = 1 - 0,9786 = 0,0214$.

³ Eine detaillierte Beschreibung des Zusammenhangs der Warteschlangentheorie und der Poisson-Verteilung findet sich z.B. in Fahrmeir, L., Künstler, R., Pigeot, I., Tutz, G. (2007).

Eine weitere Anwendung findet man beim *Gesetz des radioaktiven Zerfalls*. Hier wird die Anzahl der in einem festen Zeitintervall zerfallenden Atome einer radioaktiven Substanz als poissonverteilt angesehen. Auch gelten z.B. die Anzahl der innerhalb von 15 Minuten in einer Telefonzentrale eingehenden Gespräche oder die innerhalb einer Stunde aus einem Lager entnommenen Waren als poissonverteilt.

3.1.3.2 Eigenschaften

Es lässt sich zeigen, dass *Erwartungswert* und *Varianz* der Poissonverteilung stets gleich ihrem Funktionalparameter λ sind. Es gilt also

$$E(X) = \lambda \quad | \quad (\text{II3: 18})$$

$$\text{Var}(X) = \lambda . \quad | \quad (\text{II3: 19})$$

Wie die Binomialverteilung besitzt auch die Poissonverteilung die Eigenschaft der *Reproduktivität*. Dies bedeutet, dass für die beiden unabhängigen Zufallsvariablen $X \sim \text{Po}(\lambda_1)$ und $Y \sim \text{Po}(\lambda_2)$ stets $X + Y \sim \text{Po}(\lambda_1 + \lambda_2)$ gilt.

3.1.3.3 Approximation

Die Poissonverteilung ist vor allem dadurch interessant, dass sie eine Approximationsmöglichkeit für die Binomialverteilung (und damit auch für die Hypergeometrische Verteilung) darstellt. Diese ist konkret gegeben, wenn eine sehr große Zahl n von Versuchen und sehr kleine Wahrscheinlichkeiten p für das Auftreten des Ereignisses A vorliegen (streng genommen $n \rightarrow \infty$, $p \rightarrow 0$ und $n \cdot p = \lambda = \text{konstant}$). Empirisch ist die Poissonverteilung also besonders dann gut geeignet, wenn man registriert, wie oft ein bei einmaliger Durchführung sehr unwahrscheinliches Ereignis bei vielen Wiederholungen eintritt. Man spricht daher auch häufig von einer *Verteilung der seltenen Ereignisse*. So könnte also z.B. die Anzahl der "Sechser" pro Ausspielung im Zahlenlotto oder die Anzahl von einer Versicherung zu regulierenden Schadensfälle pro Zeiteinheit als poissonverteilt angenommen werden.

Eine *Binomialverteilung* oder genauer $B(n,p)$ -Verteilung kann durch eine $\text{Po}(n \cdot p)$ -Verteilung approximiert werden, wenn die beiden Bedingungen

$$\left. \begin{array}{l} n \geq 50 \\ p \leq 0,1 \end{array} \right\} \rightarrow B(n, p) \approx \text{Po}(n \cdot p) \quad | \quad (\text{II3: 20})$$

erfüllt sind. Liegt (II3: 20) vor, verursacht die Approximation meist nur einen vernachlässigbar kleinen Fehler. Wir erkennen, dass bei der Approximation für den Parameter λ der Poissonverteilung der Erwartungswert der Binomialverteilung verwendet wird. Wir sehen auch, dass für sehr kleine p bei der Binomialverteilung $E(X) \approx \text{Var}(X)$ gilt.

Beispiel II 3-15

Approximation der Binomialverteilung

In der Buchhaltung eines Großkonzerns kommt es bei der sehr großen Zahl an täglich anfallenden Buchungen in der Regel zu 0,1 % Fehlbuchungen. Es werden nun 3.000 Buchungen rein zufällig auf Richtigkeit überprüft. Wie groß ist die Wahrscheinlichkeit, dass in dieser Stichprobe 5 fehlerhafte gefunden werden?

Da ohne Zurücklegen gezogen wird, wäre die gesuchte Wahrscheinlichkeit mit Hilfe der Hypergeometrischen Verteilung zu ermitteln. Da die Gesamtzahl der Buchungen jedoch sehr groß ist, ist eine Näherung durch die $B(3.000; 0,001)$ -Verteilung legitim. Aufgrund der sehr geringen Wahrscheinlichkeit p kann aber diese wiederum durch die Poissonverteilung approximiert werden. Es gilt also mit $\lambda = 3.000 \cdot 0,001 = 3$

$$P(X = 5) \approx f_{Po}(5) = \frac{3^5}{5!} \cdot e^{-3} = 0,1008.$$

Beispiel II 3-16

Anwendungsfall: Kreditrisiko

Bei der Kalkulation von Prämien (vgl. auch Beispiel II 3-27) für das von Banken zu tragende Kreditausfallrisiko werden häufig stochastische Modelle zur Unterstützung herangezogen. So kann z.B. die Häufigkeit von Kreditausfällen bei sehr kleinen Ausfallwahrscheinlichkeiten der einzelnen Kredite als näherungsweise $Po(\lambda)$ -verteilt angesehen werden, wenn die Ausfälle unabhängig voneinander erfolgen. Betrachten wir dazu ein Beispiel:

Eine Bankfiliale mit einem Portfolio von 1.000 Krediten schätzt die Ausfallwahrscheinlichkeit eines einzelnen Kredits auf 1 %. Die Wahrscheinlichkeit, dass es innerhalb eines Jahres zu mehr als x Ausfällen kommt, kann nun bei einem Erwartungswert von $\lambda = 1.000 \cdot 0,01 = 10$ wie folgt angegeben werden:

$$P(X > x) = 1 - P(X \leq x) = 1 - e^{-10} \cdot \sum_{v=0}^x \frac{10^v}{v!}$$

Die Wahrscheinlichkeit dafür, dass mehr als 2 % der Kredite ausfallen, liegt demnach unter Verwendung der Tabellen für die Verteilungsfunktion der Poissonverteilung bei

$$P(X > 20) = 1 - P(X \leq 20) = 1 - e^{-10} \cdot \sum_{v=0}^{20} \frac{10^v}{v!} = 1 - 0,9984 = 0,0016.$$

Wir erkennen an der formelmäßigen Zusammensetzung der Wahrscheinlichkeit, dass der Summenterm bei zunehmendem x immer größer und damit die Wahrscheinlichkeit $P(X > x)$ immer kleiner wird. So liegt etwa die Wahrscheinlichkeit dafür, dass mehr als 3 % der Kredite uneinbringlich werden, nur noch bei 0,00000008.

Beispiel II 3-17

Güte der Approximation einer Binomialverteilung



Die folgende Tabelle verdeutlicht die Güte der Annäherung einer Binomialverteilung mit $n = 100$ und $p = 0,03$ durch eine Poissonverteilung mit $\lambda = 100 \cdot 0,03 = 3$.

x	$B(100; 0,03)$	$Po(3)$
0	0,0476	0,0498
1	0,1471	0,1494
2	0,2252	0,2240
3	0,2275	0,2240
4	0,1706	0,1680
5	0,1013	0,1008
6	0,0496	0,0504
7	0,0206	0,0216
8	0,0074	0,0081
9	0,0023	0,0027
10	0,0007	0,0008
$E(X)$	3,00	3,00
$Var(X)$	2,91	3,00

Für die Approximation der *Hypergeometrischen Verteilung* wird zusätzlich zu den Bedingungen in (II3: 20) geprüft, ob $n/N \leq 0,1$ ist, und bei Erfüllung $H(M,N,n)$ durch $P(n \cdot M/N)$ approximiert. Wir sehen auch hier wieder, dass im Rahmen der Approximation für λ der Erwartungswert der zu approximierenden Verteilung herangezogen wird.

3.2 Stetige Verteilungen

3.2.1 Gleichverteilung

3.2.1.1 Dichte- und Verteilungsfunktion

Sind a und b reelle Zahlen mit $a < b$, so heißt eine *stetige* Zufallsvariable X , die folgende *Dichtefunktion* $f_x(x)$ besitzt, *gleichverteilt* im Intervall $[a,b]$, rechteckverteilt oder uniform-verteilt. Wir schreiben kurz $X \sim R(a, b)$.

$$f_x(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \quad (\text{II3: 21})$$

Die Realisationen x der stetigen Gleichverteilung liegen zwischen den beiden endlichen Grenzen a und b . Abbildung II 19 zeigt anschaulich, warum diese Verteilung auch als Rechteckverteilung bezeichnet wird. Da die Breite dieses "Rechtecks" bei $b - a$ und seine Höhe bei $1/(b - a)$ liegt, ergibt sich eine Rechtecksfläche von genau 1. Dies muss auch gelten, da es sich bei $f_x(x)$ sonst nicht um eine Dichtefunktion handeln würde.

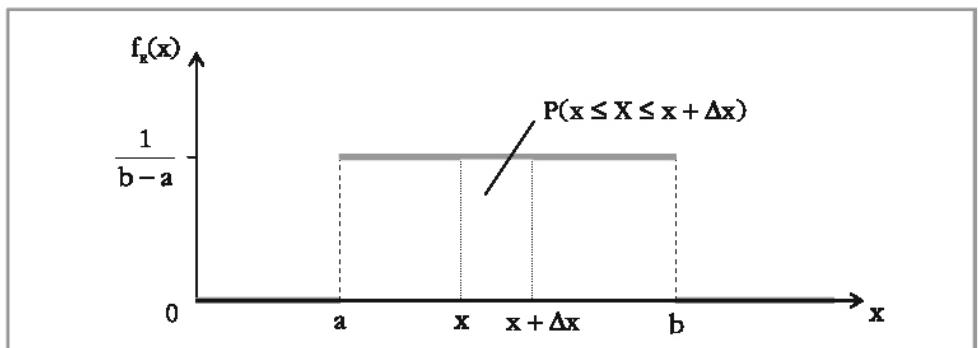


Abbildung II 20: Dichtefunktion der stetigen Gleichverteilung

Abbildung II 20 zeigt auch, dass sich die Wahrscheinlichkeit dafür, dass die gleichverteilte Zufallsvariable X einen Wert zwischen x und $x + \Delta x$ annimmt (allein abhängig von Δx und nicht von x), wie folgt ergibt:

$$P(x \leq X \leq x + \Delta x) = \frac{1}{b-a} \cdot \Delta x \quad (\text{II3: 22})$$

Durch Integration der Dichtefunktion ergibt sich die *Verteilungsfunktion* $F_x(x)$ der stetigen Gleichverteilung als

$$F_R(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 1 & x > b \end{cases} \quad (\text{II3: 23})$$

Die *Funktionalparameter* der Verteilung sind a und b . Der Graf der Verteilungsfunktion $F_R(x)$ ist in Abbildung II 21 dargestellt.

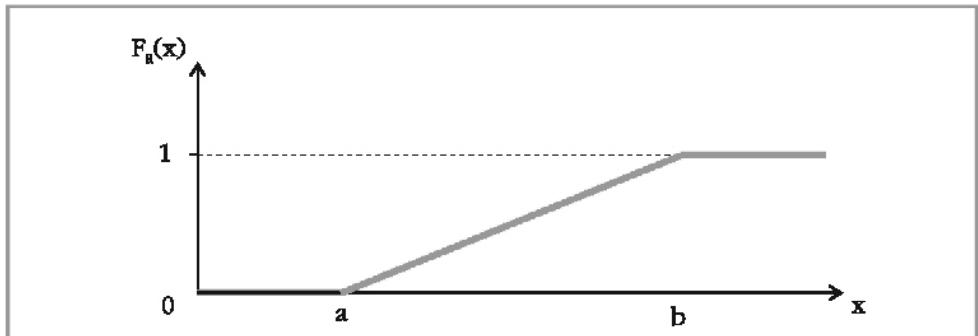


Abbildung II 21: Verteilungsfunktion der stetigen Gleichverteilung

Erwartungswert und *Varianz* einer gleichverteilten stetigen Zufallsvariablen X ergeben sich nach den in Kapitel II 2.4.1 und II 2.4.2 definierten Vorschriften zu

$$E(X) = \frac{a+b}{2} \quad (\text{II3: 24})$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}. \quad (\text{II3: 25})$$

Beispiel II 3-18

Stetige Gleichverteilung

Die Zeit, die ein Arbeiter von seinem Arbeitsplatz zu seinem Zuhause benötigt, sei gleichverteilt zwischen 20 und 30 Minuten. Wieviel Zeit nimmt sein Heimweg im Durchschnitt in Anspruch? Wie groß ist die Wahrscheinlichkeit, dass der Arbeiter zwischen 24 und 26 Minuten benötigt?

Die Zufallsvariable X : "Zeit für den Heimweg" ist $R(20;30)$ -verteilt, sodass Folgendes gilt:

$$E(X) = (20+30)/2 = 25 \text{ Minuten}$$

$$P(24 \leq X \leq 26) = 0,1 \cdot (26-24) = 0,2$$

3.2.1.2 Diskretes Gegenstück

Besitzt eine *diskrete* Zufallsvariable X genau k Ausprägungen x_i mit einer Eintrittswahrscheinlichkeit von jeweils $1/k$, so lautet ihre *Wahrscheinlichkeitsfunktion* $f_U(x)$

$$f_U(x) = \begin{cases} 1/k & x = x_1, x_2, \dots, x_k \\ 0 & \text{sonst} \end{cases}. \quad (\text{II3: 26})$$

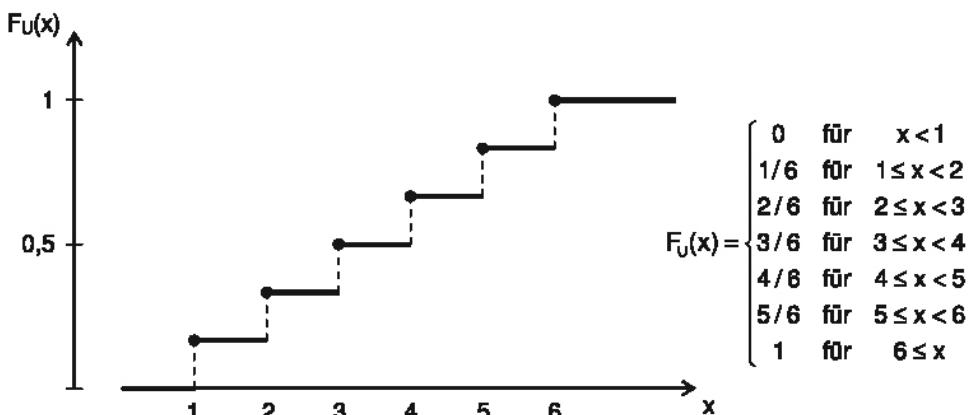
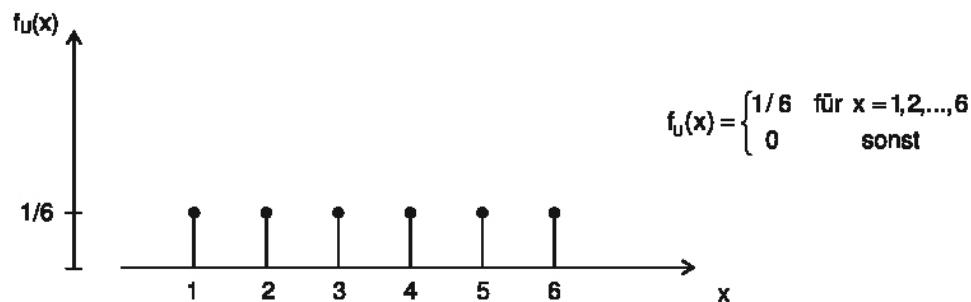
Die **Verteilungsfunktion** $F_U(x)$ ergibt sich, wie bei jeder diskreten Zufallsvariable, durch Kumulation der entsprechenden Werte der Wahrscheinlichkeitsfunktion.

Der **Erwartungswert** und die **Varianz** der Zufallsvariablen X können ganz einfach mit (II2: 21) und (II2: 31) berechnet werden.

Beispiel II 3-19

Diskrete Gleichverteilung

Ein typisches Beispiel für eine diskrete Gleichverteilung bietet die Zufallsvariable X : "Augenzahl beim einmaligen Würfelwurf" (Spezialfall $x_i = i$). Wahrscheinlichkeits- und Verteilungsfunktion zeigen dabei folgendes Bild:



Für die Zufallsvariable X ergibt sich für den Erwartungswert

$$E(X) = \frac{1}{6} \sum_{i=1}^6 i = 3,5$$

und die Varianz

$$\text{Var}(X) = \frac{1}{6} \sum_{i=1}^6 (i - 3,5)^2 = 2,92.$$

Bei einer sehr großen Anzahl wiederholter Würfe, z.B. 1.000, wird also die Summe der gewürfelten Augenzahlen etwa $1.000 \cdot 3,5 = 3.500$ betragen.

3.2.2 Exponentialverteilung

3.2.2.1 Dichte- und Verteilungsfunktion

Besitzt eine *stetige* Zufallsvariable X folgende *Dichtefunktion* $f_{\text{Ex}}(x)$ mit $\lambda > 0$ und e als Eulersche Zahl, so heißt sie *exponentialverteilt*. Wir schreiben kurz $X \sim \text{Ex}(\lambda)$.

$$f_{\text{Ex}}(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{II3: 27})$$

Die Wahrscheinlichkeit $P(X \leq x)$, mit der eine exponentialverteilte Zufallsvariable X höchstens einen Wert x annehmen kann, ist durch ihre *Verteilungsfunktion* $F_{\text{Ex}}(x)$ bestimmt. Diese entsteht durch Integration der Dichtefunktion $f_{\text{Ex}}(x)$ als

$$F_{\text{Ex}}(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-\lambda \cdot x} & \text{für } x \geq 0 \end{cases}. \quad (\text{II3: 28})$$

Für *Erwartungswert* und *Varianz* einer exponentialverteilten Zufallsvariablen X gilt

$$E(X) = \frac{1}{\lambda} \quad (\text{II3: 29})$$

$$\text{Var}(X) = \frac{1}{\lambda^2}. \quad (\text{II3: 30})$$

Die Exponentialverteilung findet häufig bei Zufallsvariablen Anwendung, die bei der *Messung von Zeitspannen* (z.B. in Warteschlangenmodellen und Modellen der Instandhaltungsplanung) auftreten. So können etwa die Lebensdauern von Bauteilen, Servicezeiten (Reparaturdauer, Beladezeit eines Lieferwagens, Kundenabfertigungszeit) oder Zeitlücken zwischen dem Eintreffen zweier Pkws an einem Grenzübergang als exponentialverteilt angenommen werden. Der Parameter λ ist dann als durchschnittliche Abfertigungszeit, Servicezeit etc. zu interpretieren.

Beispiel II 3-20

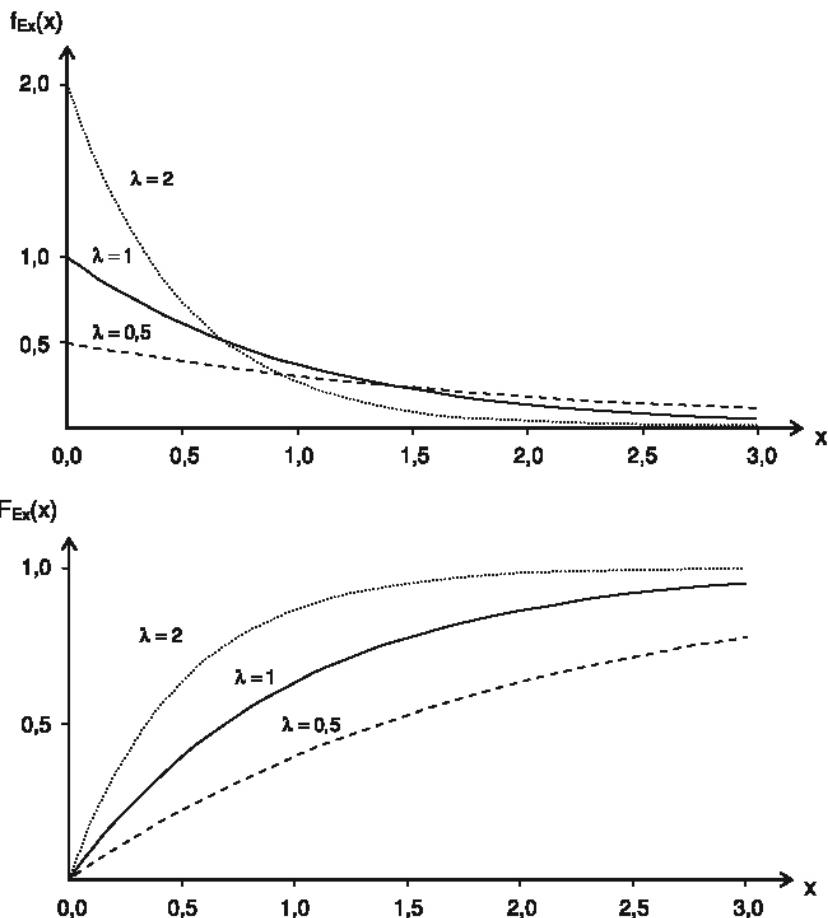
Warteschlangen

Die Zufallsvariable X beschreibe die Zeit zwischen den Ankünften zweier Kunden am Schalter einer Bank. Durchschnittlich vergehen 0,6 Minuten zwischen dem Eintreffen zweier Kunden.

Wie groß ist die Wahrscheinlichkeit dafür, dass zwischen den Ankünften zweier Kunden mehr als 3 Minuten verstreichen, falls wir die Exponentialverteilung unterstellen?

$$P(X > 3) = 1 - P(X \leq 3) = 1 - F_{\text{Ex}}(3) = 1 - (1 - e^{-0,6 \cdot 3}) = 1 - 0,8347 = 0,1653$$

Der einzige *Funktionalparameter* λ bestimmt die exakte Form der Dichtefunktion. Im Beispiel II 3-21 sind die Dichtefunktionen für einige ausgewählte Werte von λ abgebildet. Wir erkennen, dass die Dichtefunktion für negative x den Wert Null annimmt und bei $x = 0$ auf den Wert λ springt, um von da an monoton zu fallen und asymptotisch gegen die x -Achse zu konvergieren. Je größer λ ist, desto "schneller" konvergiert die Dichte $f_{\text{Ex}}(x)$ bei zunehmendem x gegen Null. Bei großem λ konvergiert natürlich auch die Verteilungsfunktion $F_{\text{Ex}}(x)$ "schneller" gegen Eins.

Beispiel II 3-21**Ausgewählte Exponentialverteilungen**

Die Exponentialverteilung wird gelegentlich auch als *Verteilung ohne Gedächtnis* bezeichnet. Dies beruht auf der Tatsache, dass eine stetige Zufallsvariable genau dann exponentialverteilt ist, wenn für alle $a \geq 0$ und $b \geq 0$ die folgende Bedingung (sog. Markov-Eigenschaft) gilt:⁴

$$P(X \leq a + b | X \geq a) = P(X \leq b) \quad (\text{II3: 31})$$

Angenommen, die Zufallsvariable X beschreibt die Lebensdauer einer Glühbirne, so besagt (II3: 31) dabei, dass die bedingte Verteilung der weiteren Lebensdauer b unabhängig von der bereits erreichten Lebensdauer a ist. Anders ausgedrückt, hat die Zeit, die die Glühbirne bisher brannte, keinen Einfluss auf die Zeit, die die Glühbirne in Zukunft noch brennen wird.

Gerade diese Eigenschaft der Exponentialverteilung macht sie aber für gewisse praktische Sachverhalte untauglich. Erfasst eine Zufallsvariable z.B. die Dauer der

⁴ Den Beweis dieser Eigenschaft überlassen wir in Aufgabe V9 dem Leser.

Arbeitslosigkeit einer Person, so zeigt sich in der Empirie, dass die Wahrscheinlichkeit, die Arbeitslosigkeit in einem bestimmten Zeitintervall zu beenden, durchaus von der bisherigen Dauer der Arbeitslosigkeit abhängt, da es für Langzeitarbeitslose in der Regel problematischer ist, wieder ins Berufsleben einzusteigen. Eine Exponentialverteilung wäre daher für die Modellierung der Arbeitslosigkeitsdauer ungeeignet.⁵

3.2.2.2 Diskretes Gegenstück

Führen wir ein Bernoulli-Experiment mit einer Erfolgswahrscheinlichkeit von p (Misserfolgswahrscheinlichkeit $1 - p$) so oft aus, bis zum ersten Mal ein Erfolg eintritt, und verwenden wir die Zufallsvariable X : "Anzahl der Misserfolge bis zum ersten Erfolg" ($x = 0, 1, 2, \dots$), so beträgt die Wahrscheinlichkeit dabei genau x Misserfolge zu erzielen nach dem Multiplikationssatz für *unabhängige Ereignisse*

$$P(X = x) = (1 - p)^x \cdot p. \quad (\text{II3: 32})$$

Eine diskrete Zufallsvariable X mit folgender *Wahrscheinlichkeitsfunktion* $f_G(x)$ bezeichnen wir als *geometrisch verteilt*. Wir schreiben kurz $X \sim G(p)$.

$$f_G(x) = \begin{cases} (1 - p)^x \cdot p & x = 0, 1, 2, \dots \\ 0 & \text{sonst} \end{cases} \quad (\text{II3: 33})$$

Mit zunehmendem x nehmen hier die Wahrscheinlichkeiten beginnend beim Wert $f_G(0) = p$ ab. Wie daran und auch an der grafischen Form der Verteilung (vgl. Beispiel II 3-22) unschwer zu erkennen ist, handelt es sich bei der geometrischen Verteilung um das diskrete Gegenstück zur Exponentialverteilung.⁶

Eine weitere interessante Eigenschaft dieser Verteilung ist

$$\frac{f_G(x+1)}{f_G(x)} = (1 - p). \quad (\text{II3: 34})$$

Die Misserfolgswahrscheinlichkeit $(1 - p)$ kann also aus zwei aufeinanderfolgenden Funktionswerten der Wahrscheinlichkeitsfunktion ermittelt werden. Gleichzeitig können wir sagen, dass die Wahrscheinlichkeiten bei zunehmendem x mit dem Term $1 - p$ abnehmen, da $f_G(x+1) = f_G(x) \cdot (1 - p)$ gilt. Je größer also p ist, desto kleiner ist der Term $1 - p$ und desto schneller nehmen die Wahrscheinlichkeiten ab. Diese Abnahme der Wahrscheinlichkeiten sorgt schließlich dafür, dass ihre Summe den Wert 1 annimmt:

$$\sum_{x=0}^{\infty} f_G(x) = \sum_{x=0}^{\infty} p \cdot (1 - p)^x = p \cdot \underbrace{\sum_{x=0}^{\infty} (1 - p)^x}_{\text{unendl. geom. Reihe}} = p \cdot \frac{1}{1 - (1 - p)} = 1$$

⁵ Die sog. Weibull-Verteilung weiß diese Eigenschaft nicht auf. Für eine detaillierte Behandlung dieser speziellen Verteilung vgl. z.B. Hartung, J., Elpelt, B., Klösener, K.-H. (2005).

⁶ Die Exponentialverteilung steht auch in engem Zusammenhang zur Poissonverteilung. Die Anzahl von Ereignissen in einem Zeitintervall ist nämlich genau dann $\text{Po}(\lambda)$ -verteilt, wenn die Zeitspannen zwischen aufeinanderfolgenden Ereignissen unabhängig und exponentialverteilt mit dem Parameter λ sind.

Die Verteilungsfunktion $F_G(x)$ der geometrischen Verteilung lautet

$$F_G(x) = \begin{cases} 0 & x < 0 \\ 1 - (1-p)^{x+1} & x = 0, 1, 2, \dots \end{cases} \quad (\text{II3: 35})$$

Funktionalparameter der geometrischen Verteilung ist allein die Erfolgswahrscheinlichkeit p . Für Erwartungswert und Varianz der geometrischen Verteilung gilt:

$$E(X) = \frac{1-p}{p} \quad (\text{II3: 36})$$

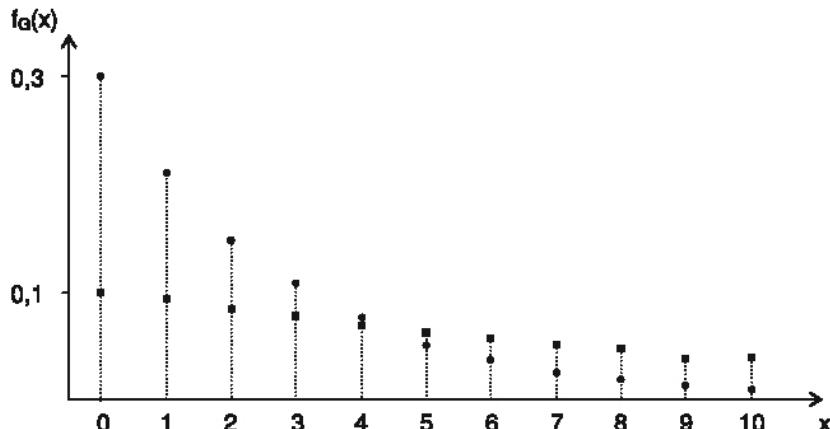
$$V(X) = \frac{1-p}{p^2} \quad (\text{II3: 37})$$

Beispiel II 3-22

Wahrscheinlichkeitsfunktionen von $G(0,3)$ und $G(0,1)$



Für die Werte $p_1 = 0,3$ (Punkte) und $p_2 = 0,1$ (Rechtecke) ergeben sich folgende Wahrscheinlichkeitsfunktionen:



Beispiel II 3-23

Gesamtbeispiel geometrische Verteilung

Sie befinden sich in der Frankfurter Innenstadt und haben Ihr Mobiltelefon vergessen, sodass Sie sich kein Taxi rufen können. Sie sind also darauf angewiesen zu warten, bis das nächste vorbeikommt. Aus Erfahrungswerten wissen Sie, dass die Wahrscheinlichkeit, dass in der nächsten Minute eines kommt, bei 10 % liegt.

Die Wahrscheinlichkeit nicht länger als 2 Minuten warten zu müssen liegt damit bei

$$F_G(1) = 1 - (1 - 0,1)^{1+1} = 0,19.$$

Die Wahrscheinlichkeit länger als 10 Minuten warten zu müssen ergibt sich zu

$$P(X > 10) = 1 - F_G(10) = 1 - [1 - (1 - 0,1)^{10+1}] = 1 - 0,6862 = 0,3138.$$

Im Durchschnitt müssen Sie

$$E(X) = \frac{1-0,1}{0,1} = 9 \text{ Minuten}$$

warten.

Auch die geometrische Verteilung ist eine *Verteilung ohne Gedächtnis*. So können wir z.B. in Beispiel II 3-23 zeigen, dass die Wahrscheinlichkeit weitere 10 Minuten warten zu müssen, nachdem man bereits 20 Minuten gewartet hat, immer noch 0,3138 beträgt.

3.2.3 Normalverteilung

3.2.3.1 Dichte- und Verteilungsfunktion

Die auch als Gauß-Verteilung bezeichnete *Normalverteilung* (nach Abraham De Moivre und Carl Friedrich Gauß) ist die wohl wichtigste statistische Verteilung. Für die zentrale Rolle der Normalverteilung in der Statistik gibt es mehrere Gründe:

- Viele Zufallsvariablen, die in Experimenten oder Stichprobenerhebungen in der *Praxis* auftreten, sind zumindest näherungsweise normalverteilt. Dies gilt insbesondere für biologische, physikalische oder technische Variablen, wie z.B. die Größe, das Gewicht oder den Intelligenzquotienten von Menschen.
- Gewisse nicht normalverteilte Zufallsvariablen kann man so *transformieren*, dass die resultierenden Variablen normalverteilt sind.⁷
- Die Normalverteilung ist eine Grenzverteilung zahlreicher anderer Verteilungen, was unter bestimmten Bedingungen die *Approximation* dieser Verteilungen durch die Normalverteilung ermöglicht.⁸
- In statistischen Schätz- und Testverfahren kommen oft Größen vor, die normalverteilt sind oder sich bei Grenzübergängen einer Normalverteilung nähern. Letzteres findet eine theoretische Begründung durch den sog. *Zentralen Grenzwertsatz* (vgl. Abschnitt II 3.4.1).

Ist X eine mit den Parametern μ und σ^2 normalverteilte Zufallsvariable, $X \sim N(\mu, \sigma^2)$, so lautet die *Dichtefunktion* $f_N(x)$ folgendermaßen:

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{für } -\infty < x < +\infty \quad (\text{II3: 38})$$

Dabei sind $\sigma > 0$, $\pi = 3,14159\dots$ und $e = 2,71828\dots$ Diese Dichtefunktion (vgl. Abbildung II 22) weist folgende Eigenschaften auf:

- Für alle x gilt stets $f_N(x) > 0$ und $f_N(x) \rightarrow 0$ für $x \rightarrow \pm\infty$.
- $f_N(x)$ ist symmetrisch zur Senkrechten $x = \mu$. Es gilt daher

$$f_N(\mu - x) = f_N(\mu + x) \quad \text{für alle } x \in \mathbb{R}. \quad (\text{II3: 39})$$

- Die Fläche unter $f_N(x)$ von $-\infty$ bis μ entspricht der Fläche unter $f_N(x)$ von μ bis $+\infty$. Beide Flächen weisen einen Flächeninhalt von jeweils 0,5 auf.
- $f_N(x)$ besitzt an der Stelle $x = \mu$ ihr absolutes Maximum von $f_N(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$.
- Ihre Wendepunkte liegen bei $x = \mu - \sigma$ und $x = \mu + \sigma$.

⁷ Vgl. z.B. die Potenzierungstransformation nach Chen, W.W., Deo, R.S. (2004).

⁸ Die Approximation ausgewählter diskreter Verteilungen behandeln wir unter II 3.4. Die Approximation sog. Testverteilungen haben wir in die Beschreibung dieser Verteilungen in Abschnitt II 3.3 integriert.

Die Verteilungsfunktion $F_N(x)$ der Normalverteilung ist als

$$F_N(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv \quad (\text{II3: 40})$$

definiert. Sie zeigt einen S-förmigen Verlauf (vgl. Abbildung II 22), besitzt einen Wendepunkt bei $x = \mu$ und es gilt $F_N(x) \rightarrow 0$ für $x \rightarrow -\infty$ sowie $F_N(x) \rightarrow 1$ für $x \rightarrow +\infty$.

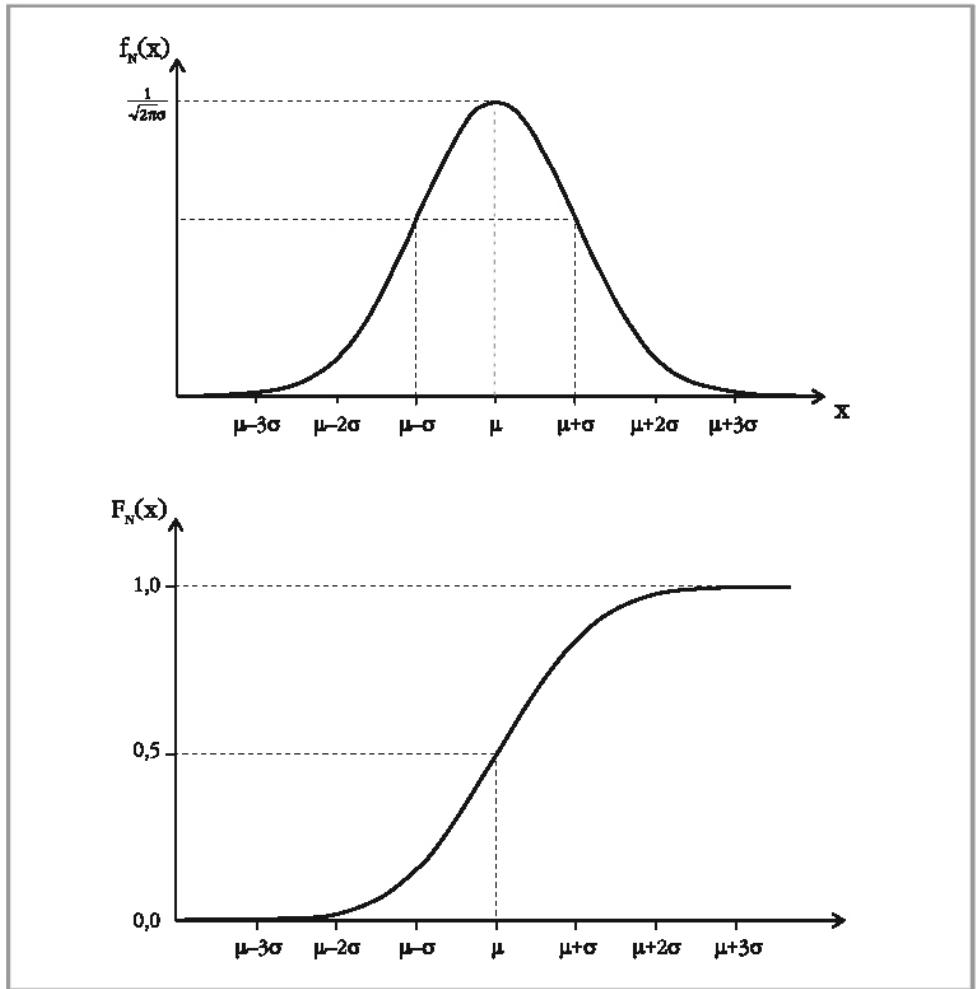


Abbildung II 22: Dichte- und Verteilungsfunktion der Normalverteilung

Für unterschiedliche Ausprägungen der *Funktionalparameter* μ und σ^2 ergeben sich unterschiedliche Normalverteilungsdichten. Halten wir etwa μ konstant und variieren nur σ^2 , so wird die Kurve immer flacher, je größer σ^2 wird. Abbildung II 23 veranschaulicht dies. Verändern wir nur μ bei konstantem σ^2 , so kommt es zu einer Verschiebung der Kurve nach "links" oder "rechts" (vgl. Abbildung II 24). Wir erkennen an den Abbildungen, dass es sich bei μ um einen Lageparameter und bei

σ^2 um einen Streuungsparameter der Verteilung handelt, d.h. für *Erwartungswert* und *Varianz* einer $N(\mu, \sigma^2)$ -verteilten Zufallsvariable X gilt

$$E(X) = \mu \quad (\text{II3: 41})$$

$$\text{Var}(X) = \sigma^2 . \quad (\text{II3: 42})$$

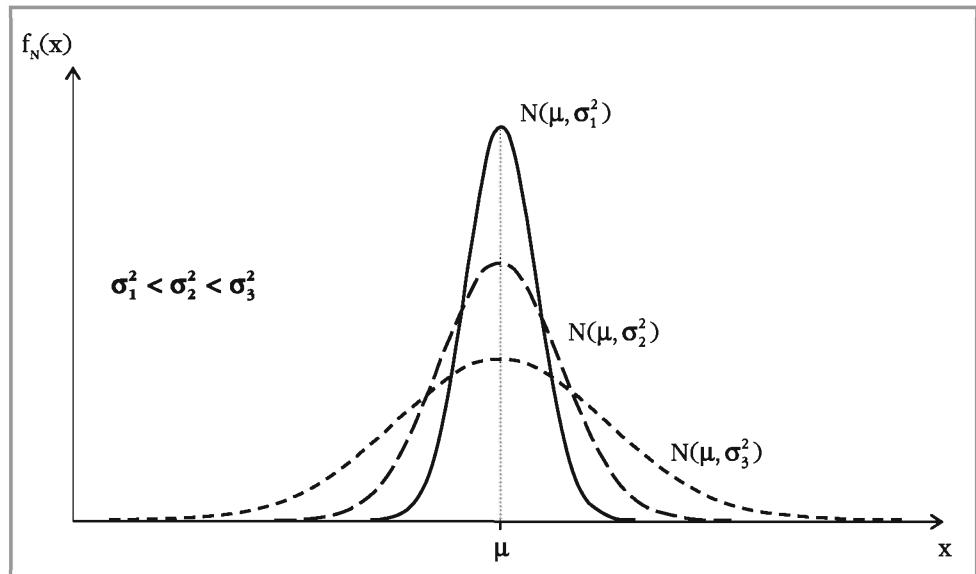


Abbildung II 23: Skizzierte NV-Dichtefunktionen für verschiedene σ

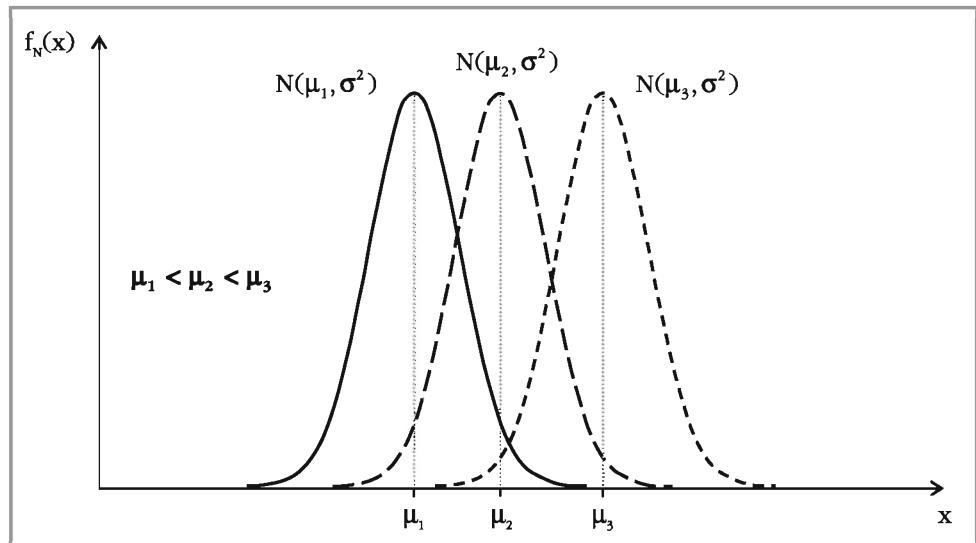


Abbildung II 24: Skizzierte NV-Dichtefunktionen für verschiedene μ

3.2.3.2 Standardnormalverteilung

Eine Normalverteilung mit $\mu = 0$ und $\sigma^2 = 1$, also $N(0;1)$, heißt *Standardnormalverteilung* (normierte Normalverteilung). Da in der statistischen Literatur eine standardnormalverteilte Zufallsvariable meist mit Z und deren Dichte mit φ bezeichnet werden, wollen wir uns ebenso dieser Notation bedienen, sodass sich für die *Dichte*

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (\text{II3: 43})$$

schreiben lässt. Diese Dichtefunktion ist in Abbildung II 25 skizziert. Wir erkennen daran, dass bei einer $N(0;1)$ -Verteilung 68,27 % der x -Werte im Intervall $[-1; 1]$ (entspricht einem Intervall von einer Standardabweichung um den Erwartungswert) liegen. Im Intervall $[-3; 3]$ (entspricht einem Intervall von drei Standardabweichungen um den Erwartungswert) befinden sich bereits 99,73 % der Werte.

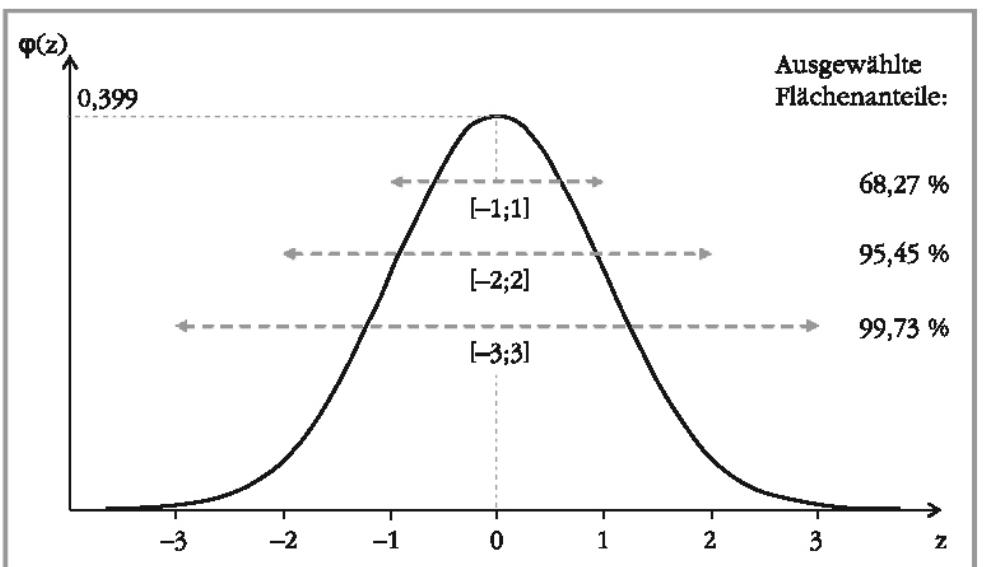


Abbildung II 25: Dichtefunktion der Standardnormalverteilung

Durch Integration der Dichtefunktion $\varphi(z)$ erhalten wir die *Verteilungsfunktion* $\Phi(z)$ der Standardnormalverteilung. Auch sie weist wie die Verteilungsfunktion der Normalverteilung einen S-förmigen Verlauf auf und gibt die Wahrscheinlichkeit $P(Z \leq z) = \Phi(z)$ an.

Ist c eine beliebige Konstante, so können wir aufgrund der *Symmetrieeigenschaft* der Normalverteilung für ihre Verteilungsfunktion den Zusammenhang

$$P(X \leq \mu - c) = P(X \geq \mu + c) \rightarrow \Phi(\mu - c) = 1 - \Phi(\mu + c)$$

festhalten, wonach für die Verteilungsfunktion der Standardnormalverteilung

$$P(Z \leq -c) = P(Z \geq c) \rightarrow \Phi(-c) = 1 - \Phi(c) \quad (\text{II3: 44})$$

gilt.

Aufgrund ihrer enormen Bedeutung für die induktive Statistik wollen wir im Folgenden noch auf weitere *Eigenschaften* der Standardnormalverteilung eingehen:

1. Transformationseigenschaft der Standardnormalverteilung

Ist X eine $N(\mu, \sigma^2)$ -verteilte Zufallsvariable, so kann diese durch eine spezielle Art der linearen Transformation, nämlich die Standardisierung

$$Z = a + bX \quad \text{mit} \quad a = \frac{-\mu}{\sigma}, \quad b = \frac{1}{\sigma} \quad \rightarrow \quad Z = \frac{X - \mu}{\sigma}$$

in eine $N(0;1)$ - bzw. standardnormalverteilte Zufallsvariable Z überführt werden. Durch die Standardisierung wird $E(Z) = 0$ und $\text{Var}(Z) = 1$ erreicht (vgl. Abschnitt II 2.4.2.3), sodass wir zusammenfassend Folgendes festhalten können:

$$X \sim N(\mu, \sigma^2) \quad \rightarrow \quad Z = \frac{X - \mu}{\sigma} \sim N(0; 1) \quad (\text{II3: 45})$$

Aufgrund dieser sog. *Transformationseigenschaft* reicht es für die Anwendung der Normalverteilung aus, die Verteilungsfunktion der Standardnormalverteilung zu tabellieren.⁹ Da außerdem nach (II3: 44) der Zusammenhang $\Phi(-z) = 1 - \Phi(z)$ gilt, ist es ausreichend in Tabellen zur Standardnormalverteilung nur die Wahrscheinlichkeiten für $z \geq 0$ aufzunehmen. Auszugsweise sieht eine solche Tabellierung daher wie folgt aus:

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0,0	0,5000	1,0	0,8413	2,0	0,9772
0,2	0,5793	1,2	0,8849	2,2	0,9861
0,4	0,6554	1,4	0,9192	2,4	0,9918
0,6	0,7257	1,6	0,9452	2,6	0,9953
0,8	0,7881	1,8	0,9641	2,8	0,9974

Interessieren wir uns für die Wahrscheinlichkeit $P(Z \leq 2)$, so können wir diese aus obiger Tabelle als $\Phi(2) = 0,9772$ ablesen. $P(Z \leq -2)$ können wir aufgrund von (II3: 44) als $\Phi(-2) = 1 - \Phi(2) = 1 - 0,9772 = 0,0228$ bestimmen.

Für die Verteilungsfunktionen von Normalverteilung $F_N(x)$ und Standardnormalverteilung $\Phi(x)$ können wir folgenden Zusammenhang zeigen:

$$F_N(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \Phi(z)$$

Die Wahrscheinlichkeit, dass eine mit μ und σ^2 normalverteilte Zufallsvariable X höchstens den Wert x annimmt, kann also direkt aus den Tabellen der Standardnormalverteilung entnommen werden:

$$P(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad (\text{II3: 46})$$

⁹ Eine Tabellierung ist erforderlich, da keine Stammfunktion der Dichte der Normalverteilung existiert. Das Integral $F_N(x)$ kann nur mit Näherungsverfahren der numerischen Mathematik gelöst werden.

2. Intervallwahrscheinlichkeiten

- Ist die Wahrscheinlichkeit dafür gesucht, dass eine $N(\mu, \sigma^2)$ -verteilte Zufallsvariable X einen Wert zwischen x_u und x_o annimmt, so gilt

$$\begin{aligned} P(x_u \leq X \leq x_o) &= P(x_u < X \leq x_o) \\ &= P(X \leq x_o) - P(X \leq x_u) = F_N(x_o) - F_N(x_u). \end{aligned}$$

Unter Verwendung von (II3: 46) können wir auch Folgendes schreiben:

$$P(x_u \leq X \leq x_o) = \Phi\left(\frac{x_o - \mu}{\sigma}\right) - \Phi\left(\frac{x_u - \mu}{\sigma}\right) = \Phi(z_o) - \Phi(z_u) \quad (\text{II3: 47})$$

- Eine weitere in der Praxis interessante Fragestellung ist häufig, wie groß die Wahrscheinlichkeit dafür ist, dass eine $N(\mu, \sigma^2)$ -verteilte Zufallsvariable X um höchstens einen Wert $c > 0$ von μ abweicht. Diese errechnet man unter Berücksichtigung von (II3: 44) wie folgt:

$$\begin{aligned} P(\mu - c \leq X \leq \mu + c) &= F_N(\mu + c) - F_N(\mu - c) = \Phi\left(\frac{\mu + c - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - c - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{c}{\sigma}\right) - \Phi\left(\frac{-c}{\sigma}\right) = \Phi\left(\frac{c}{\sigma}\right) - \left[1 - \Phi\left(\frac{c}{\sigma}\right)\right] \end{aligned}$$

Dies kann man weiter zusammenfassen zu

$$P(\mu - c \leq X \leq \mu + c) = 2 \cdot \Phi\left(\frac{c}{\sigma}\right) - 1. \quad (\text{II3: 48})$$

- Interessieren wir uns für die Wahrscheinlichkeit, dass eine $N(\mu, \sigma^2)$ -verteilte Zufallsvariable X Werte in einem Intervall $\mu - k\sigma \leq X \leq \mu + k\sigma$ mit $k > 0$ (sog. $k\sigma$ -Bereich) annimmt, so ergibt sich diese zu

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) = \Phi(k) - \Phi(-k).$$

Unter Verwendung von (II3: 44) können wir diesen Zusammenhang auch als

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = 2 \cdot \Phi(k) - 1 \quad (\text{II3: 49})$$

ausdrücken. Zu erkennen ist, dass die Fläche von $\mu - k\sigma$ bis $\mu + k\sigma$ unter der Gaußschen Glockenkurve unabhängig von μ und σ ist. Für verschiedene Werte von k erhalten wir folgende Wahrscheinlichkeiten:

k	Intervall	Fläche	k	Intervall	Fläche
1	$\mu \pm 1\sigma$	68,27 %	1,64	$\mu \pm 1,64\sigma$	90 %
2	$\mu \pm 2\sigma$	95,45 %	1,96	$\mu \pm 1,96\sigma$	95 %
3	$\mu \pm 3\sigma$	99,73 %	2,58	$\mu \pm 2,58\sigma$	99 %

Die Wahrscheinlichkeiten für $k = 1$, $k = 2$ und $k = 3$ sind uns dabei bereits in Abbildung II 25 begegnet.

Beispiel II 3-24**Zusammenfassendes Beispiel zur Normalverteilung**

Eine Maschine produziert Metallplättchen mit einer Länge von 30 mm. Da zufallsabhängige Ungenauigkeiten bei der Herstellung auftreten können, kann die Länge X eines Plättchens als Zufallsvariable angesehen werden, die wir als mit $\mu = 30$ und $\sigma^2 = 0,25$ normalverteilt betrachten wollen. Gesucht sind nun folgende Wahrscheinlichkeiten, die wir unter Heranziehung der tabellierten Standardnormalverteilung im Anhang bestimmen wollen:

Ein zufällig aus der Produktion entnommenes Plättchen ist

- a) höchstens 30,5 mm lang:

$$P(X \leq 30,5) = \Phi\left(\frac{30,5 - 30}{0,5}\right) = \Phi(1) = 0,8413$$

- b) mindestens 29,6 mm lang:

$$\begin{aligned} P(X \geq 29,6) &= 1 - P(X \leq 29,6) = 1 - \Phi\left(\frac{29,6 - 30}{0,5}\right) = 1 - \Phi(-0,8) = 1 - [1 - \Phi(0,8)] \\ &= \Phi(0,8) = 0,7881 \end{aligned}$$

- c) zwischen 29,1 mm und 30,7 mm lang:

$$\begin{aligned} P(29,1 \leq X \leq 30,7) &= \Phi\left(\frac{30,7 - 30}{0,5}\right) - \Phi\left(\frac{29,1 - 30}{0,5}\right) = \Phi(1,4) - \Phi(-1,8) \\ &= \Phi(1,4) + \Phi(1,8) - 1 = 0,9192 + 0,9641 - 1 = 0,8833 \end{aligned}$$

- d) maximal 0,2 mm von der Soll-Länge entfernt:

$$P(29,8 \leq X \leq 30,2) = 2 \cdot \Phi\left(\frac{0,2}{0,5}\right) - 1 = 2 \cdot \Phi(0,4) - 1 = 2 \cdot 0,6554 - 1 = 0,3108$$

3. Quantile der Standardnormalverteilung

Für ein α -Quantil (oder Fraktil) z_α nimmt die Wahrscheinlichkeit $P(Z \leq z_\alpha) = \Phi(z_\alpha)$ den Wert α an. Zur Bestimmung des Wertes z_α wird also in der Verteilungstabelle zunächst die Wahrscheinlichkeit α gesucht und das dazugehörige z_α abgelesen.

$$\Phi(z_\alpha) = \alpha \quad (\text{II3: 50})$$

Einige wichtige Quantile, die wir insbesondere in der induktiven Statistik noch des Öfteren benötigen werden, sind in folgender Tabelle aufgeführt:

α	z_α	α	z_α
0,50	0,00	0,975	1,96
0,90	1,28	0,99	2,33
0,95	1,64	0,995	2,58

Demnach liegt also z.B. das 99 %-Quantil $z_{0,99}$ der $N(0;1)$ -Verteilung bei 2,33. Interessieren wir uns für das 1 %-Quantil $z_{0,01}$, so können wir dieses aufgrund der Symmetrie der Dichte bezüglich $z = 0$ über $z_\alpha = -z_{1-\alpha}$, d.h. hier als $z_{0,01} = -z_{0,99} = -2,33$ bestimmen.

Interessieren wir uns für die Wahrscheinlichkeit, dass eine $N(\mu, \sigma^2)$ -verteilte Zufallsvariable X Werte in einem Intervall $x_{\alpha/2} \leq X \leq x_{1-\alpha/2}$ annimmt, so ergibt sich diese unmittelbar aus der uns bekannten Quantilsdefinition. Wie die Abbildung II 26 anschaulich zeigt, ist der Flächeninhalt der Flächen links von $x_{\alpha/2}$ und rechts von $x_{1-\alpha/2}$ jeweils gerade $\alpha/2$, sodass für die gesuchte Intervallwahrscheinlichkeit offensichtlich $1 - (\alpha/2 + \alpha/2) = 1 - \alpha$ gilt.

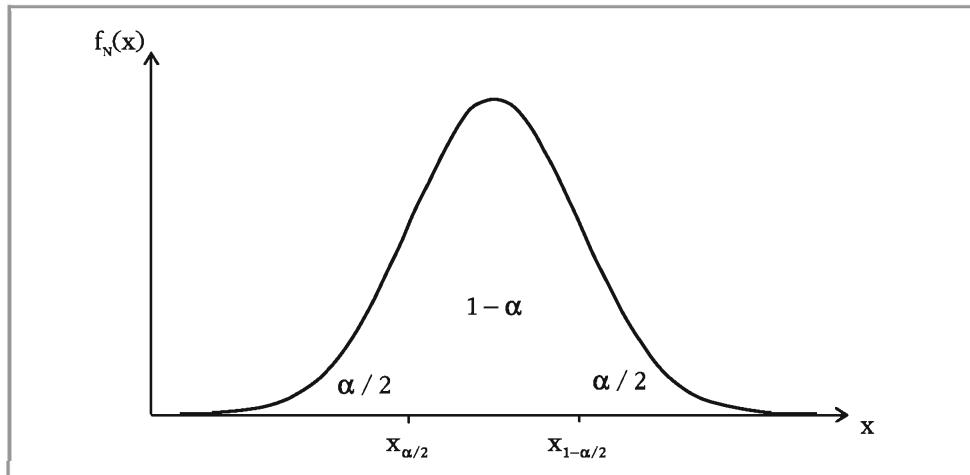


Abbildung II 26: Dichtefunktion, Quantile und Intervallwahrscheinlichkeiten

3.2.3.3 Reproduktionseigenschaft

Auch die Normalverteilung besitzt die bereits mehrfach erwähnte *Reproduktions-eigenschaft*. Diese kommt wie folgt zum Ausdruck:

Sind X_1, \dots, X_n *n unabhängige* normalverteilte Zufallsvariablen, so ist ihre Summe X , also $X = X_1 + \dots + X_n$ ebenfalls normalverteilt.¹⁰ Der Erwartungswert von X , μ_x , ergibt sich dann als Summe der einzelnen Erwartungswerte μ_1, \dots, μ_n der Zufallsvariablen X_1, \dots, X_n :

$$\mu_x = \mu_1 + \dots + \mu_n \quad (\text{II3: 51})$$

Für die Varianz von X , σ_x^2 , ergibt sich aufgrund der *Unabhängigkeit* der X_i

$$\sigma_x^2 = \sigma_1^2 + \dots + \sigma_n^2. \quad (\text{II3: 52})$$

Obwohl sie nicht mit der Reproduktionseigenschaft zusammenhängt, wollen wir an dieser Stelle auf eine weitere bedeutende Eigenschaft der Normalverteilung hinweisen: Entsteht eine Zufallsvariable Y durch eine Transformation $Y = a + bX$ mit $b \neq 0$ aus einer $N(\mu, \sigma^2)$ -verteilten Zufallsvariable X , so ist $Y \sim N(a + b\mu, b^2\sigma^2)$ -verteilt.

¹⁰ Genau genommen ist die Summe ΣX_i *gemeinsam* normalverteilter Zufallsvariablen X_1, \dots, X_n wieder normalverteilt (oder eine Konstante). Zur Definition gemeinsam normalverteilter Zufallsvariablen vgl. z.B. Schmid, F., Trede, M. (2005). Insbesondere sind n unabhängige normalverteilte Zufallsvariablen auch gemeinsam normalverteilt.

Beispiel II 3-25

Reproduktionseigenschaft

In einer Winzerei werden Weinflaschen mit einem Sollinhalt von 1,5 l bei einer Varianz von 0,0025 l² abgefüllt. Wie groß ist unter Annahme der Normalverteilung und Unabhängigkeit der Flascheninhalte die Wahrscheinlichkeit dafür, dass ein Kunde, der 4 Flaschen kauft, wenigstens 5,95 l erhält?

Die Gesamtliterzahl X ergibt sich hier aus den (unabhängigen) Literzahlen X_1, \dots, X_4 der einzelnen Flaschen und ist damit auch normalverteilt. Es gilt dann:

$$X = X_1 + X_2 + X_3 + X_4 \quad \rightarrow \quad \mu_X = 4 \cdot 1,5 = 6 \text{ l}$$

$$\sigma_X^2 = 4 \cdot 0,0025 = 0,01 \text{ l}^2$$

Für die gesuchte Wahrscheinlichkeit ergibt sich demnach

$$P(X \geq 5,95) = 1 - \Phi\left(\frac{5,95 - 6}{0,1}\right) = 1 - \Phi(-0,5) = 1 - [1 - \Phi(0,5)] = 0,6915.$$

3.2.4 Logarithmische Normalverteilung

Eine nichtnegative Zufallsvariable X heißt logarithmisch normalverteilt oder *log-normalverteilt* mit den Parametern μ und σ^2 , kurz $X \sim L(\mu, \sigma^2)$, wenn nicht die Zufallsvariable X selbst, sondern ihr Logarithmus normalverteilt bzw. konkret $N(\mu, \sigma^2)$ -verteilt ist. Ist also $Y = \ln X$ normalverteilt, dann ist $X = e^Y$ lognormalverteilt. Für die Verteilungsfunktion $F_L(x)$ gilt dann

$$F_L(x) = F_N(y) = F_N(\ln x).$$

Durch Ableitung nach x erhalten wir unter Berücksichtigung der Kettenregel

$$f_L(x) = \frac{d F_L(x)}{d x} = \frac{d F_N(\ln x)}{d \ln x} \cdot \frac{d (\ln x)}{d x} = f_N(\ln x) \cdot \frac{1}{x},$$

was die *Dichtefunktion* $f_L(x)$ der *nicht symmetrischen* Lognormalverteilung als

$$f_L(x) = \begin{cases} \frac{1}{x \cdot \sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left(\frac{\ln x - \mu}{\sigma} \right)^2} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \quad (\text{II3: 53})$$

liefert. Da $\ln x$ nur für $x > 0$ definiert ist, ist $f_L(x)$ für $x \leq 0$ gleich Null. Die Dichtefunktion beginnt bei $x = 0$ mit dem Wert $f_L(0) = 0$, steigt bis zu einem Maximum an, fällt dann wieder und nähert sich asymptotisch der x -Achse (vgl. Beispiel II 3-26). Die Lognormalverteilung besitzt die zwei Funktionalparameter $\mu = E(\ln X)$ und $\sigma^2 = \text{Var}(\ln X)$. Ihre α -Quantile können wir über

$$x_\alpha = e^{\mu + \sigma z_\alpha} \quad (\text{II3: 54})$$

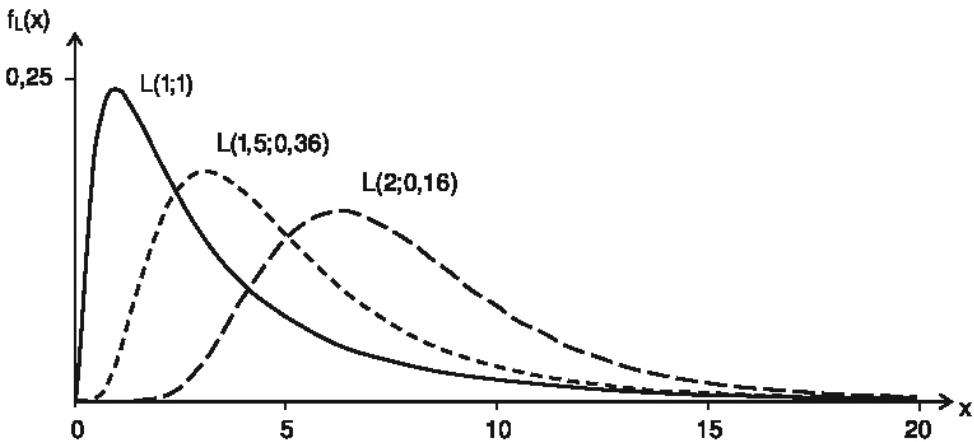
berechnen, wobei z_α das entsprechende Quantil der Standardnormalverteilung ist.

Die rechtsschiefe bzw. linkssteile $L(\mu, \sigma^2)$ -Verteilung wird in der Praxis besonders gerne verwendet, wenn keine negativen Merkmalsausprägungen vorkommen können und eine linkssteile Verteilung gebraucht wird. Dies ist etwa bei Einkommens-

oder Vermögensverteilungen gegeben. Häufig werden in der Finanzmarkttheorie auch Aktienkurse oder -renditen als lognormalverteilt unterstellt.¹¹

Beispiel II 3-26

Verschiedene Lognormalverteilungen



Die *Funktionalparameter* μ und σ^2 der $L(\mu, \sigma^2)$ -Verteilung können anders als bei der Normalverteilung nicht als Erwartungswert und Varianz der Verteilung interpretiert werden, da sie Erwartungswert und Varianz der logarithmierten Variablen darstellen. Stattdessen ergeben sich Erwartungswert und Varianz als

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} \quad (\text{II3: 55})$$

$$\text{Var}(X) = e^{2\mu + \sigma^2} \cdot (e^{\sigma^2} - 1). \quad (\text{II3: 56})$$

Beispiel II 3-27

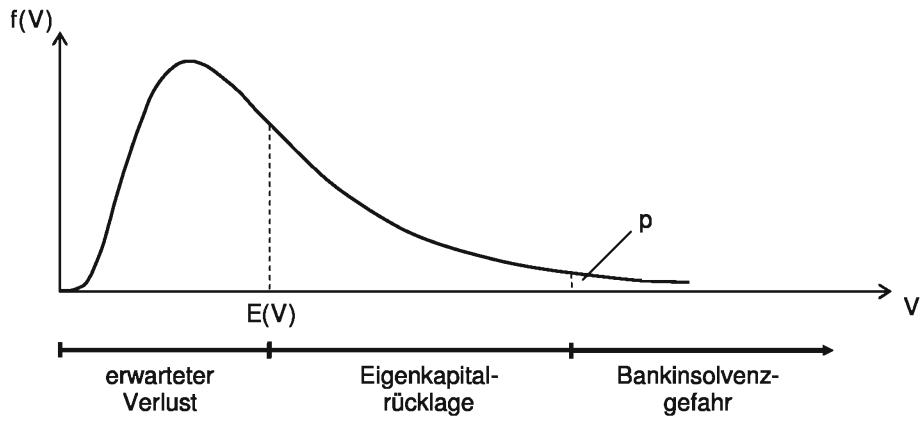
Anwendungsfall: Kreditrisiko und Bank-Rating

In Beispiel II 3-16 wurde behandelt, dass die Häufigkeit von Kreditausfällen als annähernd $P(\lambda)$ -verteilt betrachtet werden kann. Da nun aber jeder Kreditausfall für die Bank einen anderen Schaden bedeutet, ist es letztendlich nicht die Anzahl der Kreditausfälle, sondern die Höhe des damit verbundenen Verlustes die interessiert. In der Praxis hat sich gezeigt, dass die *Wahrscheinlichkeit der aggregierten Ausfallverluste* durch eine *Lognormalverteilung* gut modelliert werden kann.¹²

Bei der Bestimmung des Kreditzinses (bzw. der darin einzubeziehenden Risikoprämie) durch die Banken sollten diese beachten, dass er die erwarteten Ausfallverluste $E(V)$ abdeckt. Für den Fall, dass die tatsächlich eintretenden Verluste diesen Erwartungswert übersteigen, besitzen die Banken immer noch Eigenkapitalrücklagen, die sie vor Insolvenz bewahren. Erst wenn die Verluste die erwarteten Verluste (abgedeckt durch die Risikoprämie im Kreditzins) und unerwartete Verluste (abgedeckt durch die Eigenkapitalrücklage) übersteigen, ist die Bank bedroht. Es wird daher versucht, die Wahrscheinlichkeit für eine Insolvenz der Bank durch entsprechende Festsetzung der Eigenkapitalrücklage auf einem niedrigen Niveau p zu halten. Je niedriger p , desto besser ist das Rating der Bank.

¹¹ Vgl. z.B. Campbell, J.Y., Lo, A.W., MacKinlay, A.C. (1997), Kapitel 1 und 9.

¹² Vgl. z.B. Rau-Bredow, H. (2002) und Rolfs, B. (2008).



3.3 Test-Verteilungen

Die folgenden *stetigen* Verteilungen, die auch als *Testverteilungen* bezeichnet werden, finden besonders bei bestimmten Testverfahren und bei der Bestimmung von Konfidenzintervallen in der induktiven Statistik Anwendung. Wir wollen sie daher zunächst nur theoretisch betrachten.¹³ Auf ihre konkrete Anwendung werden wir schließlich im Kapitel III eingehen.

3.3.1 Chi-Quadrat-Verteilung

Sind Z_1, Z_2, \dots, Z_w unabhängig $N(0; 1)$ -verteilte Zufallsvariablen, so folgt die Quadratesumme

$$\chi^2 = Z_1^2 + Z_2^2 + \dots + Z_w^2 \quad (\text{II3: 57})$$

einer *Chi-Quadrat-Verteilung* mit w Freiheitsgraden oder kurz $\chi^2(w)$ -Verteilung. Der *Funktionalparameter* w der Chi-Quadrat-Verteilung, welcher die Anzahl der unabhängigen Zufallsvariablen in χ^2 beschreibt, bestimmt die Gestalt ihrer *Dichtefunktion* (vgl. Beispiel II 3-28). Da χ^2 nur positive Realisationen annehmen kann, ist sie nur für solche definiert und ihre Dichtefunktion strebt mit $\chi^2 \rightarrow \infty$ gegen Null.¹⁴

Eine Chi-Quadrat-verteilte Zufallsvariable χ^2 besitzt den *Erwartungswert*

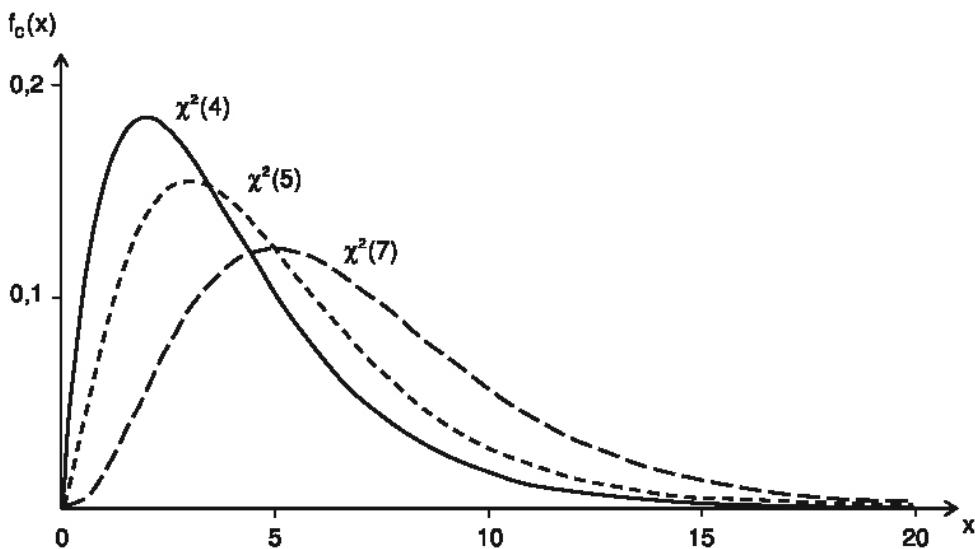
$$E(\chi^2) = w \quad (\text{II3: 58})$$

und die *Varianz*

$$\text{Var}(\chi^2) = 2w. \quad (\text{II3: 59})$$

¹³ Wir verzichten bei der Behandlung der Verteilungen auf die konkrete Angabe der Dichtefunktionen, da wir sie im weiteren Verlauf nicht benötigen. Der interessierte Leser findet die Funktionen jedoch in der im Online-Service des Buches enthaltenen Formelsammlung.

¹⁴ Um Konflikte mit der bisherigen Notation zu vermeiden, bezeichnen wir Realisationen von Chi-Quadrat-, t- und F-verteilten Zufallsvariablen in Grafiken im Folgenden mit x .

Beispiel II 3-28**Verschiedene Chi-Quadrat-Verteilungsdichten**

Beispiel II 3-28 zeigt, dass die Dichten für kleine w deutlich linkssteil sind. Für wachsendes w nähern sie sich der Gaußschen Glockenkurve. Dies ist eine Folge des Zentralen Grenzwertsatzes (vgl. Abschnitt II 3.4.1), da die Summe der Zufallsvariablen $Z_1^2, Z_2^2, \dots, Z_w^2$ dessen Voraussetzungen erfüllt.

Im Anhang sind für eine Reihe von α -Werten und $w \leq 30$ die α -Fraktile der Chi-Quadrat-Verteilung vertafelt. Es lässt sich zeigen, dass sich mit wachsender Anzahl von Freiheitsgraden w die Größe $\sqrt{2\chi^2}$ immer mehr einer Normalverteilung mit $\mu = \sqrt{2w - 1}$ und $\sigma^2 = 1$ annähert. Die standardisierte Größe

$$Z = \sqrt{2\chi^2} - \sqrt{2w - 1}$$

ist daher annähernd standardnormalverteilt. Diese *Näherung* erlaubt uns, für $w > 30$ die α -Fraktile $\chi_{\alpha,w}^2$ einer Chi-Quadrat-Verteilung mit w Freiheitsgraden durch

$$\chi_{\alpha,w}^2 \approx \frac{1}{2} \cdot (z_\alpha + \sqrt{2w - 1})^2 \quad (\text{II3: 60})$$

zu approximieren, wobei z_α das α -Fraktil der $N(0;1)$ -Verteilung ist.¹⁵

Eine *Reproduktionseigenschaft*, wie wir sie bei der Normalverteilung festgestellt haben, weist auch die Chi-Quadrat-Verteilung auf. So ist die Summe von zwei oder mehr Chi-Quadrat-verteilten Zufallsvariablen wiederum Chi-Quadrat-verteilt. Die Freiheitsgrade der neuen Zufallsvariable ergeben sich dabei als Summe der Freiheitsgrade der einzelnen Summanden.

¹⁵ Diese Approximation ist besser als die direkte Approximation der $\chi^2(w)$ -Verteilung durch eine $N(w, 2w)$ -Verteilung. Vgl. hierzu z.B. Bamberg, G., Baur, F. (2002), Kapitel 11.

Beispiel II 3-29**Fraktilapproximation**

Liegt das 10 %-Fraktil einer Chi-Quadrat-Verteilung mit $w = 5$ bei $\chi^2_{0,10} = 1,61$, bedeutet dies, dass die Wahrscheinlichkeit, dass χ^2 höchstens den Wert 1,61 annimmt, bei 10 % liegt. Würden wir uns nun für das 10 %-Fraktil für $w = 50$ interessieren, könnten wir dieses näherungsweise wie folgt bestimmen:

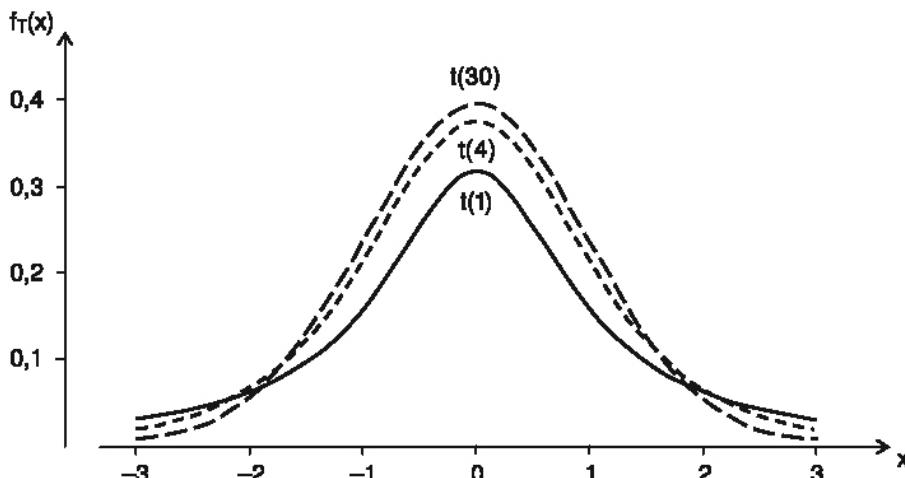
$$\chi^2_{0,10} \approx \frac{1}{2} (-1,282 + \sqrt{99})^2 = 37,57$$

3.3.2 t-Verteilung

Betrachten wir zwei *unabhängige* Zufallsvariablen Z und χ^2 , wobei Z *standardnormalverteilt* und χ^2 mit w Freiheitsgraden *Chi-Quadrat-verteilt* ist, dann gehorcht die Zufallsvariable

$$T = \frac{Z}{\sqrt{\frac{\chi^2}{w}}} \quad (\text{II3: 61})$$

einer *t-Verteilung* (oder *Studentverteilung* nach W.S. Gosset) mit w Freiheitsgraden oder kurz einer *t(w)-Verteilung*. Ihre *Dichtefunktion* besitzt eine ähnliche symmetrische Glockenform wie die Normalverteilung. Ihre Funktionsgestalt ist vom *Funktionalparameter* w abhängig (vgl. Beispiel II 3-30). Sie ist *symmetrisch um Null* und besitzt für kleine w *dickere (höhere) Enden* (engl. *heavy tail distribution*) als die Standardnormalverteilung. Dies bedeutet, dass weniger Wahrscheinlichkeitsmasse im Zentrum um Null und mehr an den Rändern verteilt ist. Gerade daher eignet sich die t-Verteilung zur Modellierung von Daten, die im Vergleich zur Normalverteilung einen größeren Anteil an extremen Werten enthalten.

Beispiel II 3-30**Verschiedene t-Verteilungen**

Wie Beispiel II 3-30 erkennen lässt, geht die t-Verteilung für $w \rightarrow \infty$ in die Standardnormalverteilung über. Als Faustregel können wir festhalten, dass ab $w > 30$ die t-Verteilung in guter Näherung durch die Standardnormalverteilung *approximiert* werden kann. Dies ist auch der Grund, warum meist nur für $w \leq 30$ ausgewählte α -Fraktilen tabelliert sind. Für $w > 30$ können wir als Näherung des α -Fraktils $t_{\alpha,w}$ der t(w)-Verteilung das α -Fraktile z_α der N(0;1)-Verteilung heranziehen. Zudem beschränkt sich die Tabellierung auf α -Werte größer als 0,5. Wegen der Symmetrie der Dichtefunktion der t-Verteilung gilt nämlich $t_{\alpha,w} = -t_{1-\alpha,w}$.

Für $w > 1$ ergibt sich der *Erwartungswert* der t-Verteilung als

$$E(T) = 0. \quad (\text{II3: 62})$$

Liegt die Anzahl der Freiheitsgrade w bei 1, so existiert kein Erwartungswert. In allen anderen Fällen ist bereits zu erkennen, dass (II3: 62) mit dem Erwartungswert der Standardnormalverteilung identisch ist.

Für $w > 2$ erhalten wir die *Varianz*

$$\text{Var}(T) = \frac{w}{w-2}. \quad (\text{II3: 63})$$

Ist $w \leq 2$ existiert keine Varianz. Mit zunehmendem w ($w \rightarrow \infty$) nähert sich die Varianz dem Wert 1, wie ihn auch die Standardnormalverteilung besitzt.

3.3.3 F-Verteilung

Sind eine Zufallsvariable X mit u und eine Zufallsvariable Y mit w Freiheitsgraden *Chi-Quadrat-verteilt* und sind X und Y *unabhängig*, so wird die Verteilung der Zufallsvariablen

$$F = \frac{\frac{1}{u} X}{\frac{1}{w} Y} \quad (\text{II3: 64})$$

als *F-Verteilung* mit den Freiheitsgraden u und w oder kurz *F(u,w)-Verteilung* bezeichnet. Wie eine Chi-Quadrat-verteilte Zufallsvariable, kann auch F nur positive Werte annehmen. Die Gestalt der *Dichtefunktion* ist wiederum von den *Funktionalparametern* u und w abhängig (vgl. Beispiel II 3-32). Für große w lässt sich die F(u,w)-Verteilung durch die $\chi^2(u)$ -Verteilung approximieren.

Im Anhang sind Tabellen für ausgewählte *Fraktile* der F-Verteilung zu finden. Beim Ablesen der Fraktile aus diesen Tabellen ist die Beziehung (II3: 65) zwischen dem α -Fraktile $f_{\alpha,u,w}$ der F(u,w)-Verteilung und dem $(1-\alpha)$ -Fraktile $f_{1-\alpha,w,u}$ der F(w,u)-Verteilung wertvoll. Sie lässt sich darauf zurückführen, dass wegen (II3: 64) stets $F \sim F(u,w) \leftrightarrow 1/F \sim F(w,u)$ gilt.

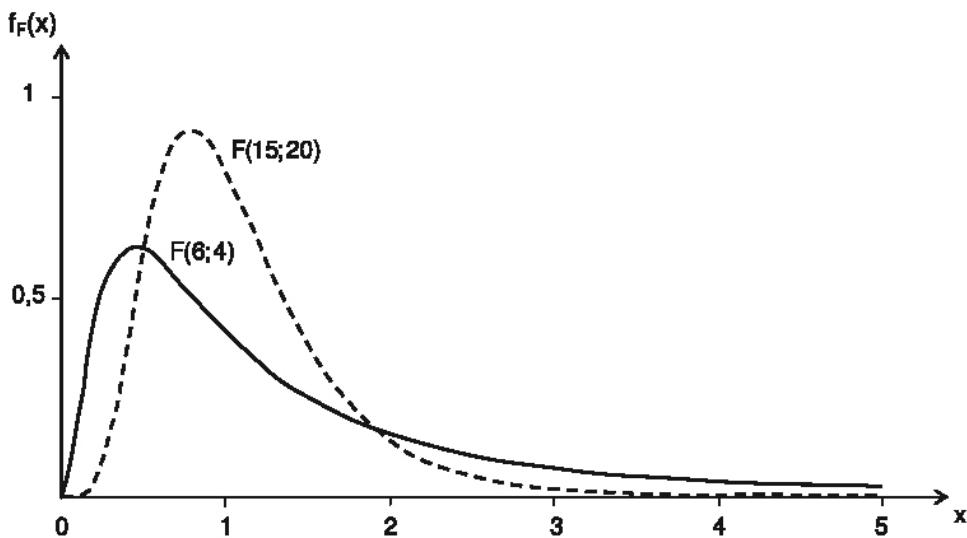
$$f_{\alpha,u,w} = \frac{1}{f_{1-\alpha,w,u}} \quad (\text{II3: 65})$$

Beispiel II 3-31

F-Verteilungs-Fraktile

Das 5 %-Fraktile der F(3;4)-Verteilung können wir finden, indem wir das 95 %-Fraktile der F(4;3)-Verteilung nachschlagen und den Kehrwert davon bilden:

$$f_{0,05;3;4} = 1/9,12 = 0,11$$

Beispiel II 3-32**Verschiedene F-Verteilungen**

Für den *Erwartungswert*, der nur vom Freiheitsgrad der Zufallsvariable im Nenner abhängt, und die *Varianz* der F-Verteilung kann Folgendes festgehalten werden:

$$E(F) = \frac{w}{w-2} \quad \text{für } w > 2 \quad (\text{II3: 66})$$

$$\text{Var}(F) = \frac{2w^2(u+w-2)}{u(w-2)^2(w-4)} \quad \text{für } w > 4 \quad (\text{II3: 67})$$

Mit zunehmendem Freiheitsgrad w konvergiert der Erwartungswert gegen Eins. Da für große Freiheitsgrade die beiden Subtrahenden (2 und 4) im Nenner von (II3: 67) immer mehr an Bedeutung verlieren, konvergiert die Varianz gegen Null.

Zusammenfassend zeigen die folgenden Übersichten die Funktionalparameter und die wichtigsten Momente (Erwartungswert und Varianz) der wesentlichsten diskreten und stetigen Verteilungen noch einmal im Überblick.

Verteilung	Parameter	Erwartungswert	Varianz
Binomial...	$0 \leq p \leq 1, n$	np	$np(1-p)$
Hypergeo- metrische ...	$N = 1, 2, \dots$ $M = 0, 1, \dots, N$ $n = 1, 2, \dots, N$	$n \frac{M}{N}$	$n \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1}$
Poisson...	$\lambda > 0$	λ	λ

Abbildung II 27: Diskrete Verteilungen im Überblick

Verteilung	Parameter	Erwartungswert	Varianz
Gleich...	$-\infty < a < b < +\infty$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential...	$\lambda > 1$	$1/\lambda$	$1/\lambda^2$
Normal...	$-\infty < \mu < +\infty$ $\sigma > 0$	μ	σ^2
Chi-Quadrat...	$w = 1, 2, \dots$	w	$2w$
t...	$w = 1, 2, \dots$	0 für $w > 1$	$\frac{w}{w-2}$ für $w > 2$
F...	$u = 1, 2, \dots$ $w = 1, 2, \dots$	$\frac{w}{w-2}$ für $w > 2$	$\frac{2w^2(u+w-2)}{u(w-2)^2(w-4)}$ für $w > 4$

Abbildung II 28: Stetige Verteilungen im Überblick

3.4 Bedeutung der Normalverteilung

Unter gewissen Voraussetzungen können andere Verteilungen durch die Normalverteilung approximiert werden. Diese Eigenschaft der Normalverteilung wurde bereits angesprochen und ist für die oben genannten Testverteilungen besonders nützlich. In diesem Abschnitt wollen wir uns nun näher mit dem Zentralen Grenzwertsatz auseinandersetzen, der die theoretische Grundlage für diese Approximationsmöglichkeit liefert und außerdem die konkrete Approximation ausgewählter diskreter Verteilungen durch die Normalverteilung näher betrachten.

3.4.1 Zentraler Grenzwertsatz

Sind X_1, X_2, \dots, X_n identisch verteilte¹⁶ und unabhängige Zufallsvariablen (egal ob stetig oder diskret) mit jeweils dem Erwartungswert μ und der Varianz σ^2 , dann passt sich die Verteilung der Summe ΣX_i für wachsende Anzahl n der Summanden immer besser der Normalverteilung

$$N(n\mu, n\sigma^2) \quad (\text{II3: 68})$$

an. Dies ist grob gesprochen die wesentlichste Aussage des *Zentralen Grenzwertsatzes*. Durch Standardisierung der Summe ΣX_i erhalten wir die Zufallsvariable

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n} \cdot \sigma} = \frac{n \cdot \bar{X}_n - n\mu}{\sqrt{n} \cdot \sigma} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (\text{II3: 69})$$

¹⁶ Dies bedeutet konkret, dass $F_{X_1}(x) = \dots = F_{X_n}(x)$ für alle reellen x gilt.

mit $E(Z_n) = 0$ und $\text{Var}(Z_n) = 1$. Der Zentrale Grenzwertsatz besagt also auch, dass die Verteilungsfunktion $F_{Z_n}(z) = P(Z_n \leq z)$ der Zufallsvariablen Z_n , mit wachsender Summandenzahl n gegen die Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung konvergiert:

$$P(Z_n \leq z) \xrightarrow{n \rightarrow \infty} \Phi(z) \quad (\text{II3: 70})$$

Die *praktische Bedeutung* des Satzes liegt darin, dass die Verteilung der Zufallsvariable Z_n bereits für endliches n durch die $N(0;1)$ -Verteilung angenähert werden kann. Entsprechend kann die Verteilung der nicht standardisierten Summe ΣX_i für endliches n durch die $N(n\mu, n\sigma^2)$ -Verteilung approximiert werden. Die Güte der Näherung hängt dabei natürlich von der Verteilung der Zufallsvariablen X_i ab. Die Konvergenz gegen die Normalverteilung ist besonders schnell, wenn die X_i symmetrisch verteilt sind, und umso langsamer, je schiefer die Verteilung ist.

Im Zusammenhang mit diesem fundamentalen statistischen Satz sind folgende Aussagen von außerordentlicher *praktischer Relevanz*:

- Für $n > 30$ wird die Normalverteilung meist als hinreichend genau für die Approximation stetiger Verteilungen angesehen.
- Der zentrale Grenzwertsatz trifft auf viele Verteilungen zu und eignet sich deshalb zu deren *Approximation*. So kann z.B. die $B(n,p)$ -Verteilung durch die $N(0;1)$ -Verteilung approximiert werden (vgl. Abschnitt II 3.4.2.1), da eine $B(n,p)$ -verteilte Zufallsvariable X als Summe $X = \Sigma X_i$ unabhängiger $B(1;p)$ -verteilter X_i dargestellt werden kann. Da $E(X) = np$ und $\text{Var}(X) = np(1-p)$ sind, konvergiert also als Spezialfall von (II3: 69) die Verteilungsfunktion der Zufallsvariable

$$Z_n = \frac{X - np}{\sqrt{np(1-p)}} \quad (\text{II3: 71})$$

für $n \rightarrow \infty$ gegen die Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung.

- Wie die Summe ist auch das *arithmetische Mittel* der X_i in einer Stichprobe, also $\bar{X} = \Sigma X_i/n$, bei hinreichend großem n annähernd normalverteilt.
- Selbst für den Fall, dass die einzelnen unabhängigen Zufallsvariablen X_i *nicht identisch verteilt* sind, ist die Summe ΣX_i für ein großes n hinreichend genau normalverteilt.
- Die entscheidende *Bedeutung der Normalverteilung* in der Wahrscheinlichkeitsrechnung und der induktiven Statistik beruht neben der Tatsache, dass die Summe und der Durchschnitt von unabhängigen Zufallsvariablen approximativ normalverteilt sind, auch darauf, dass Zufallsvariablen, die aus dem Zusammenwirken vieler unabhängiger Einzeleinflüsse resultieren (z.B. Messfehler bei einem Messvorgang), unter Berufung auf den Zentralen Grenzwertsatz als näherungsweise normalverteilt angenommen werden können.

3.4.2 Approximation diskreter Verteilungen

3.4.2.1 Binomialverteilung

Mit wachsendem n nähert sich die Form der unter II 3.1.1 behandelten $B(n,p)$ -Verteilung immer mehr der Glockenform der Normalverteilung an. Diese Annäherung erfolgt umso schneller, je näher p bei 0,5 liegt, d.h. je symmetrischer die Binomialverteilung ist. Als Faustregel können wir festhalten, dass eine Approximation dann zulässig ist, wenn die Varianz der Binomialverteilung größer oder gleich 9 ist, also $np(1-p) \geq 9$ gilt. Für *Erwartungswert* und *Varianz* der approximierenden Normalverteilung verwenden wir Erwartungswert und Varianz der Binomialverteilung, also $\mu = np$ und $\sigma^2 = np(1-p)$.

Zusammenfassend können wir also als *Approximationsregel* für eine binomialverteilte Zufallsvariable X

$$\left. \begin{array}{l} X \sim B(n,p) \\ np(1-p) \geq 9 \end{array} \right\} \rightarrow X \stackrel{\text{d}}{\sim} N(np, np(1-p)) \quad (\text{II3: 72})$$

angeben. Da hier eine Approximation einer diskreten durch eine stetige Verteilung erfolgt, ist eine sog. *Stetigkeitskorrektur* (vgl. Beispiel II 3-33) erforderlich. Andernfalls würden wir die Wahrscheinlichkeiten unterschätzen. Bei *größeren Werten von $np(1-p)$* machen wir jedoch keinen allzu großen Fehler, wenn wir die *Stetigkeitskorrektur unterlassen*.

Beispiel II 3-33

Approximation der Binomialverteilung



In einem Produktionsverfahren genügen i.d.R. 30 % der gefertigten Stücke nicht den Qualitätsanforderungen und müssen daher ausgebessert werden oder sind Ausschuss. Wir interessieren uns für die Wahrscheinlichkeit, dass sich in einer Produktionsserie von 50 Stück

- a) zwischen 14 und 16
- b) mehr als 22

fehlerhafte Stücke X befinden. Wir gehen dabei davon aus, dass der Produktionsvorgang als Bernoulli-Experiment mit $p = 0,3$ angesehen werden kann.

Da X eine diskrete Zufallsvariable ist, gilt für die unter a) gesuchte Wahrscheinlichkeit $P(14 \leq X \leq 16) = P(X \leq 16) - P(X \leq 13)$. Nach der Formel für die Binomialverteilung bzw. den Werten der tabellierten Verteilungsfunktion ergibt sich diese Wahrscheinlichkeit zu

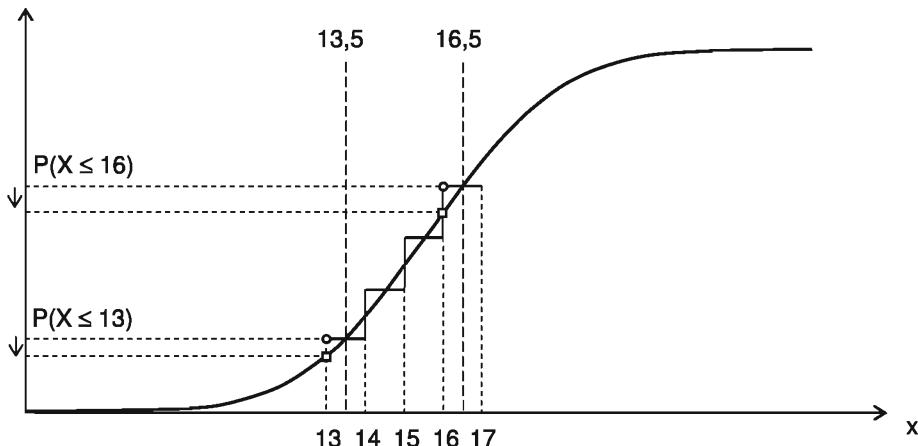
$$P(14 \leq X \leq 16) = F_B(16) - F_B(13) = 0,6839 - 0,3279 = 0,3560.$$

Da die Varianz der hier vorliegenden $B(50;0,3)$ -Verteilung gleich 10,5 ist und somit die Bedingungen aus (II3: 72) erfüllt sind, kann diese auch durch eine $N(\mu, \sigma^2)$ -Verteilung mit $\mu = 50 \cdot 0,3 = 15$ und $\sigma^2 = 50 \cdot 0,3 \cdot 0,7 = 10,5$ approximiert werden. Die gesuchte Wahrscheinlichkeit kann also alternativ auch in guter Näherung über die approximierende Normalverteilung bestimmt werden. Die approximative Wahrscheinlichkeit ergibt sich unter Berücksichtigung einer Stetigkeitskorrektur wie folgt:

$$\begin{aligned} P(14 \leq X \leq 16) &\approx F_N(16 + 0,5) - F_N(13 + 0,5) = \Phi\left(\frac{16,5 - 15}{\sqrt{10,5}}\right) - \Phi\left(\frac{13,5 - 15}{\sqrt{10,5}}\right) \\ &= \Phi(0,46) - \Phi(-0,46) = 2 \cdot \Phi(0,46) - 1 = 2 \cdot 0,6772 - 1 = 0,3544. \end{aligned}$$

Zum besseren Verständnis der hier angewendeten Stetigkeitskorrektur (bzw. der Addition des Wertes 0,5) betrachten wir die nachfolgende Grafik. Sie skizziert grob die Verteilungsfunktion der approximierenden Normalverteilung und abschnittsweise die Verteilungsfunktion der vorliegenden Binomialverteilung. Interessieren wir uns für die Wahrscheinlichkeiten $P(X \leq 13)$ und $P(X \leq 16)$, so wird schnell klar, dass wir bei einer Approximation über die Werte $F_N(13)$ und $F_N(16)$, die in der Grafik durch Rechtecke gekennzeichnet sind, die wahren Wahrscheinlichkeiten unterschätzen würden. Wir verwenden daher $F_N(13,5)$ und $F_N(16,5)$, um auf dem "richtigen Wahrscheinlichkeitslevel" zu bleiben bzw. eine bessere Approximation zu erhalten.

$F_B(x), F_N(x)$



Die unter b) zu bestimmende Wahrscheinlichkeit ergibt sich unter Verwendung der tabellierten Binomialverteilung als

$$P(X > 22) = 1 - P(X \leq 22) = 1 - 0,9877 = 0,0123.$$

Analog zu den Überlegungen aus obiger Grafik erhalten wir mittels der approximierenden Normalverteilung den Wert

$$P(X > 22) \approx 1 - \underbrace{F_N(22,5)}_{P(X \leq 22)} = 1 - \Phi\left(\frac{22,5 - 15}{\sqrt{10,5}}\right) = 1 - \Phi(2,32) = 1 - 0,9898 = 0,0102.^{17}$$

3.4.2.2 Hypergeometrische Verteilung

Wie wir unter II 3.1.2.3 gesehen haben, kann bei Vorliegen bestimmter Voraussetzungen eine $H(N,M,n)$ - durch eine $B(n,M/N)$ -Verteilung approximiert werden. Die Approximation einer $H(N,M,n)$ -Verteilung durch eine Normalverteilung mit

$$\mu = n \frac{M}{N} = np$$

$$\sigma^2 = n \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1} = np(1-p) \frac{N-n}{N-1}$$

¹⁷ Hinweis: Auch die Wahrscheinlichkeit $P(X = 22)$ ließe sich über $P(X \leq 22) - P(X \leq 21) \approx F_N(22,5) - F_N(21,5)$ approximieren. Der echte Zeitvorteil bei der Approximationsrechnung ergibt sich aber vielmehr in Fällen wie $X < 22$ oder $X > 22$.

ist immer dann zulässig, wenn $np(1-p) \geq 9$ erfüllt ist und n im Verhältnis zu N nicht allzu groß ist. Da der Faktor $(N-n)/(N-1)$ nach Ansicht einiger Autoren bereits bei $n \leq 0,1N$ vernachlässigt werden kann,¹⁸ können wir folgende allgemeine *Approximationsregel* für eine hypergeometrisch verteilte Zufallsvariable X angeben:

$$\left. \begin{array}{l} X \sim H(N, M, n) \\ np(1-p) \geq 9 \\ n \leq 0,1N \end{array} \right\} \rightarrow X \stackrel{a}{\sim} N(np, np(1-p)) \quad (\text{II3: 73})$$

Die rechnerische Bestimmung von Wahrscheinlichkeiten erfolgt ebenfalls unter Berücksichtigung einer *Stetigkeitskorrektur*.

Beispiel II 3-34

Approximation der Hypergeometrischen Verteilung

Von den 500 erfolgreichsten Investmentbankern Deutschlands fahren 200 eine bestimmte Luxusautomarke. Wie groß ist die Wahrscheinlichkeit, dass unter 50 zufällig befragten Bankern 20 bis 30 ein solches Fahrzeug besitzen?

Da hier die Bedingungen (II3: 73) erfüllt sind, kann die vorliegende $H(500;200;50)$ -Verteilung durch eine $N(\mu, \sigma^2)$ -Verteilung mit den Parametern

$$\mu = 50 \cdot \frac{200}{500} = 50 \cdot 0,4 = 20 \quad \text{und} \quad \sigma^2 = 50 \cdot 0,4 \cdot 0,6 = 12$$

approximiert werden. Die gesuchte Wahrscheinlichkeit ergibt sich daher als

$$\begin{aligned} P(20 \leq X \leq 30) &\approx F_N(30,5) - F_N(19,5) = \Phi\left(\frac{30,5-20}{\sqrt{12}}\right) - \Phi\left(\frac{19,5-20}{\sqrt{12}}\right) \\ &= \Phi(3,03) - \Phi(-0,14) = 0,9988 - 0,4443 = 0,5545. \end{aligned}$$

Berücksichtigen wir den Faktor $(N-n)(N-1)$ bzw. konkret $450/499$ dennoch, so erhalten wir

$$\begin{aligned} P(20 \leq X \leq 30) &\approx F_N(30,5) - F_N(19,5) = \Phi\left(\frac{30,5-20}{\sqrt{10,8216}}\right) - \Phi\left(\frac{19,5-20}{\sqrt{10,8216}}\right) \\ &= \Phi(3,19) - \Phi(-0,15) = 0,9993 - 0,4404 = 0,5589. \end{aligned}$$

Der Fehler, den wir durch Vernachlässigung des Faktors machen, ist also nicht all zu groß.

3.4.2.3 Poissonverteilung

Genau dann, wenn der Parameter λ der Poissonverteilung gegen ∞ strebt, geht diese in die Normalverteilung über. Als *Faustregel* für die Approximation gilt

$$\left. \begin{array}{l} X \sim Po(\lambda) \\ \lambda \geq 9 \end{array} \right\} \rightarrow X \stackrel{a}{\sim} N(\lambda, \lambda). \quad (\text{II3: 74})$$

Da der Funktionalparameter λ der Poissonverteilung gleichzeitig ihr Erwartungswert und ihre Varianz ist, ist er, wie man sieht, auch Erwartungswert und Varianz der approximierenden Normalverteilung. Die Wahrscheinlichkeitsberechnung erfolgt auch hier wieder unter Berücksichtigung einer *Stetigkeitskorrektur*.

¹⁸ Vgl. dazu z.B. Bamberg, G., Baur, F. (2004).

Beispiel II 3-35**Approximation der Poissonverteilung**

Betrachten wir die Anzahl der pro 15 Minuten an einer Supermarktkasse eintreffenden Kunden als $\text{Po}(10)$ -verteilt, so können wir die Wahrscheinlichkeit dafür, dass innerhalb einer bestimmten Viertelstunde mindestens 12 Kunden eintreffen, wie folgt approximieren:

$$\begin{aligned} P(X \geq 12) &= 1 - P(X \leq 11) \approx 1 - F_N(11,5) = 1 - \Phi\left(\frac{11,5 - 10}{\sqrt{10}}\right) = 1 - \Phi(0,47) \\ &= 1 - 0,6808 = 0,3192 \end{aligned}$$

Der exakte Wert läge bei

$$P(X \geq 12) = 1 - P(X \leq 11) = 1 - F_{\text{Po}}(11) = 1 - e^{-10} \cdot \sum_{v=0}^{11} \frac{10^v}{10!} = 1 - 0,6968 = 0,3032.$$

3.4.2.4 Überblick zur Approximation eindimensionaler Verteilungen

Die folgende Abbildung stellt abschließend die wichtigsten Beziehungen zwischen ausgewählten statistischen Verteilungen in schematisierter Form dar (Approximationsregeln im Fettdruck in weißen Kästen).¹⁹

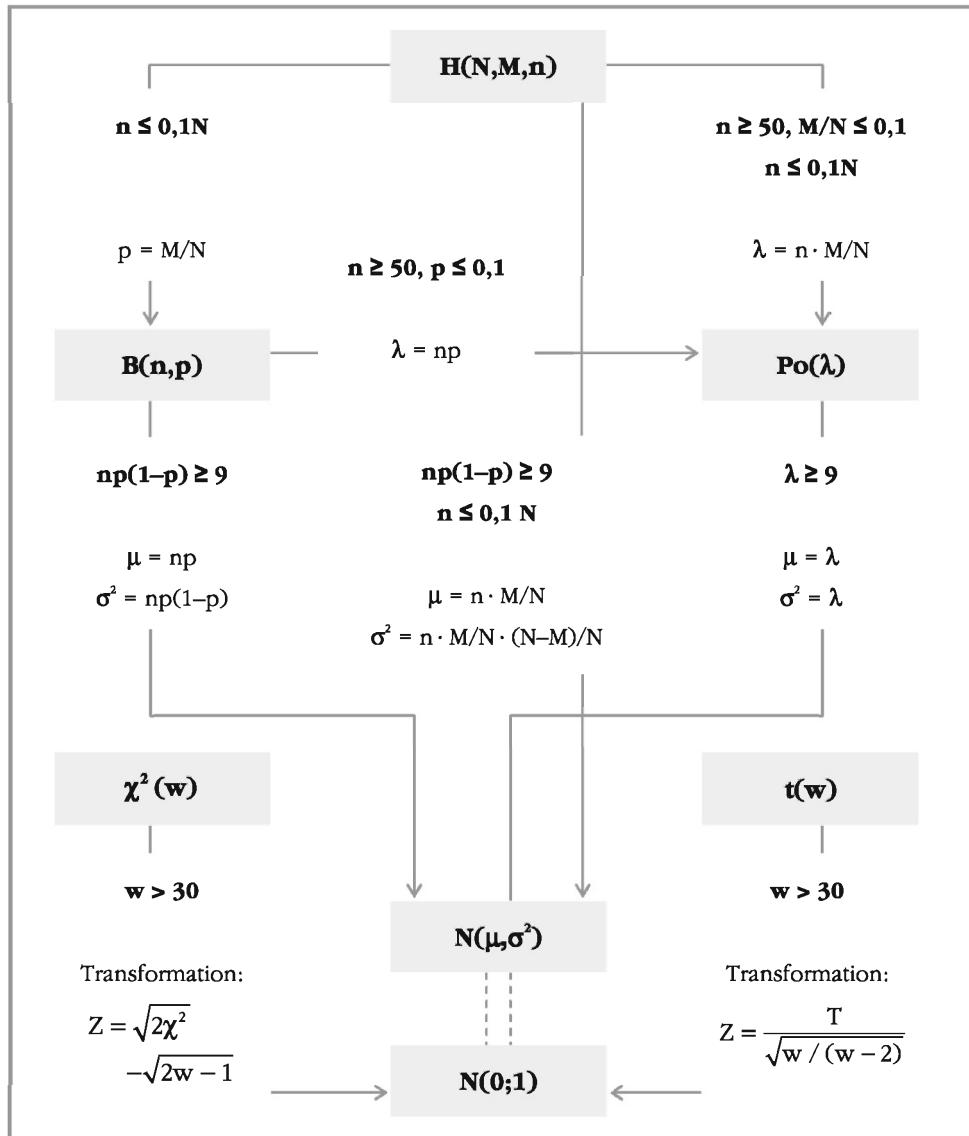


Abbildung II 29: Approximationsregeln ausgewählter Verteilungen

¹⁹ Einen umfassenden Überblick über die Vielfalt stetiger und diskreter Verteilungen geben z.B. Evans, M., Hastings, N., Peacock, B. (2000), Johnson, N.L., Kotz, S., Balakrishnan, N. (1993, 1994) und Johnson, N.L., Kotz, S., Kemp, A.W. (1993).

3.4.2.5 Empirische Verteilungen

Nicht nur theoretischen Verteilungen können sich der Normalverteilung annähern. In der Praxis stellt man immer wieder fest, dass auch empirische Häufigkeitsverteilungen eine solche Eigenschaft aufweisen. Es wird sogar gezielt untersucht, ob solche Verteilungen durch Normalverteilungen beschrieben werden können.

Erwartungswert und Varianz der approximierenden Normalverteilung sind Mittelwert \bar{x} und Varianz s^2 der *empirisch beobachteten Verteilung*.²⁰ Zum Vergleich von empirischer und approximierender Normalverteilung empfiehlt sich folgendes Vorgehen:

1. Bestimmung der nach der Normalverteilung zu erwartenden Werte der Verteilungsfunktion anhand der Klassengrenzen a_j^o ($j = 1, 2, \dots, k$) der empirischen Verteilung:

$$F_N(a_j^o) = \Phi\left(\frac{a_j^o - \bar{x}}{s}\right) = \Phi(a_j^o)$$

Für die offene Randklasse $a_k^u < x \leq \infty$ nimmt die Wahrscheinlichkeit an der Klassenobergrenze den Wert 1 an, d.h. es gilt $\Phi(\infty) = 1$.

2. Errechnung der nach der Normalverteilung theoretisch zu erwartenden relativen Häufigkeiten:

$$f_j^e = \Phi(z_{j+1}^o) - \Phi(z_j^o)$$

Sofern Schritt 3 mit Hilfe absoluter Häufigkeiten umgesetzt werden soll, können die erwarteten absoluten Häufigkeiten $h_j^e = n \cdot f_j^e$ zusätzlich ermittelt werden.

3. Grafische Abbildung

Beispiel II 3-36

Approximation empirischer Verteilungen

Die mit einer Stichprobe ermittelte Häufigkeitsverteilung der Lebensdauer eines Gerätes (mit $\bar{x} = 5$ Jahre und $s^2 = 2,8508$ Jahre²) ist in folgender Tabelle gegeben:

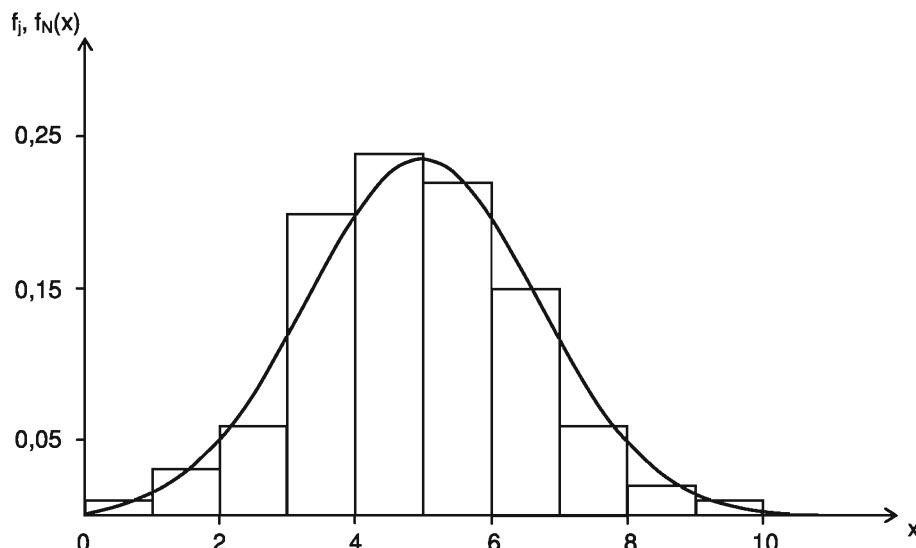
Klasse j	$> a_j^u$	$\leq a_j^o$	h_j	f_j	F_j
1	0	1	1	0,01	0,01
2	1	2	3	0,03	0,04
3	2	3	6	0,06	0,10
4	3	4	20	0,20	0,30
5	4	5	24	0,24	0,54
6	5	6	22	0,22	0,76
7	6	7	15	0,15	0,91
8	7	8	6	0,06	0,97
9	8	9	2	0,02	0,99
10	9	10	1	0,01	1,00
11	10	.	0	0,00	1,00

²⁰ Genau genommen sind die Parameter der Normalverteilung durch geeignete Schätzfunktionen (vgl. Abschnitt III 1) aus dem vorliegenden Datenmaterial zu schätzen. Der Fehler, den wir durch die hier beschriebene Vorgehensweise begehen, ist allerdings in ausreichend großen Stichproben vernachlässigbar gering.

Die nachfolgende Tabelle enthält nun für jede Klasse die Klassenobergrenze a_j^o , die standardisierte Klassenobergrenze z_j^o und den dazugehörigen Wert $\Phi(z_j^o)$ der theoretischen Verteilungsfunktion. Außerdem sind bereits die erwarteten theoretischen relativen Häufigkeiten f_j^e aufgenommen und den empirischen relativen Häufigkeiten f_j gegenübergestellt.

j	a_j^o	z_j^o	$\Phi(z_j^o)$	→	f_j^e	f_j
1	1	-2,3691	0,0089		0,0044	0,01
2	2	-1,7768	0,0378		0,0289	0,03
3	3	-1,1845	0,1181		0,0803	0,06
4	4	-0,5923	0,2768		0,1587	0,20
5	5	0,0000	0,5000		0,2232	0,24
6	6	0,5923	0,7232		0,2232	0,22
7	7	1,1845	0,8819		0,1587	0,15
8	8	1,7768	0,9622		0,0803	0,06
9	9	2,3691	0,9911		0,0289	0,02
10	10	2,9613	0,9985		0,0074	0,01
11	.	.	1,0000		0,0015	0,00

Wir erkennen an den letzten beiden Spalten der Tabelle und auch an der nachfolgenden Grafik, dass zwischen empirischer und theoretischer Verteilung eine recht gute Übereinstimmung besteht.



4.

Aufgaben

Wahrscheinlichkeitsrechnung

Aufgabe W1

Ein Unfallzeuge hat beobachtet, wie der Unfallverursacher Fahrerflucht begangen hat. An das Kennzeichen des Flüchtenden kann er sich allerdings nur ungenau erinnern. Er kann lediglich folgende Aussagen machen:

- Ortskennbuchstaben: WEN
- Buchstabengruppe AU, AV oder AY
- Drei Ziffern, von denen die erste die 1 und noch mindestens eine 2 dabei war

Wie viele Kfz-Kennzeichen sind noch möglich bzw. hat die Polizei zu untersuchen, wenn man dem Zeugen Glauben schenkt? Der Zeuge sieht die auf Grund seiner Wahrnehmung noch möglichen Kennzeichen als gleichwahrscheinlich an. Berechnen Sie damit die Wahrscheinlichkeit dafür, dass

- a) die Buchstabengruppe AY vorliegt!
- b) die ersten beiden Ziffern 12 lauten!
- c) die letzten beiden Ziffern 29 lauten!
- d) die letzte Ziffer 2 ist!
- e) unter den drei Ziffern die 0 vorkommt!
- f) die letzte Ziffer größer als die beiden anderen ist!

Aufgabe W2

Eine Münze und ein Würfel werden gemeinsam geworfen. Wie groß ist die Wahrscheinlichkeit für das Ereignis A, dass dabei Kopf und eine Augenzahl größer als 4 erscheint?

Aufgabe W3

Die Ermittlung der Gewinnzahlen beim europäischen Roulette stellt einen Zufallsvorgang mit den 37 möglichen und gleichwahrscheinlichen Elementarereignissen $0, 1, 2, 3, \dots, 36$ dar. Die Zahlen, welche von 0 verschieden sind, sind je zur Hälfte in rote und schwarze Felder eingeteilt. Berechnen Sie die jeweiligen Gewinnwahrscheinlichkeiten, die für einen Spieler gelten, der sich für die alternativen Taktiken

A: Setzt auf "rot"

B: Setzt auf "gerade Zahl" (ohne 0) und "19"

C: Setzt auf "ungerade Zahl" und die Querreihe 34, 35, 36

entscheiden kann.

Aufgabe W4

Bei einem Automobilhersteller wurden 400 aus der Produktion kommende Pkws in einer Kontrolle K_1 auf Lackmängel und einer Kontrolle K_2 auf Mängel im Innenraum untersucht. 50-mal trat bei K_1 ein positives Ergebnis auf, 40-mal bei K_2 und 20-mal bei beiden Kontrollen zugleich.

- Berechnen Sie unter der Annahme, dass die ermittelten relativen Häufigkeiten als Wahrscheinlichkeiten für die Ereignisse A_i : "das Ergebnis bei K_i ist positiv" anzusehen sind, die Wahrscheinlichkeiten $P(A_1)$, $P(A_2)$ und $P(A_1 \cap A_2)$!
- Berechnen Sie die Wahrscheinlichkeit dafür, dass ein Pkw bei K_1 ein positives Ergebnis liefert und bei K_2 nicht!
- Wie wahrscheinlich ist es, dass bei einem kontrollierten Pkw höchstens (mindestens oder genau) eine der beiden Kontrollen Mängel entdeckt?
- Wie groß ist die Wahrscheinlichkeit, dass K_2 Mängel entdeckt, wenn dies bereits bei K_1 der Fall ist?
- Berechnen Sie des Weiteren die Wahrscheinlichkeiten dafür,
 - K_2 Mängel entdeckt, wenn K_1 keine entdeckt!
 - falls genau eine der beiden Kontrollen Mängel findet, es sich dabei um K_2 handelt!
- Prüfen Sie, ob die beiden Ereignisse A_1 und A_2 unabhängig sind!

Aufgabe W5

Ein Unternehmen beabsichtigt 2 Stellen neu zu besetzen. Dazu hat es die Möglichkeit unter 4 Bewerbern auszuwählen. Von diesen werden nun 2 zufällig ausgewählt, da das Einstellungsverfahren eine nahezu identische Qualifikation dieser Bewerber ergeben hat. Tatsächlich ist es aber so, dass die Bewerber A und B im Vergleich zu C und D die Besseren sind.

Wie groß ist nun die Wahrscheinlichkeit dafür, dass die beiden besten Bewerber "erwischt" werden?

Aufgabe W6

In einer Urne befinden sich insgesamt fünf Kugeln, die von 1 bis 5 durchnummieriert sind. Die Kugeln 1 bis 3 sind dabei rot gefärbt. Es wird eine Stichprobe von 2 Kugeln entnommen.

- Gesucht ist nun die Wahrscheinlichkeit $P(A)$ des Ereignisses A: "Es befinden sich zwei rote Kugeln in der Stichprobe". Gehen Sie dabei einmal von der behandelten Situation III und einmal von Situation I aus!
- Geben Sie unter Beachtung der für Situation III geltenden Bedingungen des Weiteren die Wahrscheinlichkeit dafür an, beim ersten Zug eine rote Kugel zu erwischen und die Wahrscheinlichkeit dafür, beim zweiten Zug eine rote Kugel zu ziehen, wenn der erste Zug bereits eine rote Kugel ergab!

Aufgabe W7

In einer Schublade befinden sich 5 Kugelschreiber, denen äußerlich nicht anzusehen ist, dass 2 von ihnen mit einer schwarzen und 3 mit einer blauen Mine ausgestattet sind. Eine Bürokraft greift nun drei Tage in Folge in diese Schublade und nimmt zufällig einen dieser Kugelschreiber zum Gebrauch heraus.

- Mit welcher Wahrscheinlichkeit wurde am ersten Tag ein Stift mit schwarzer Mine erwischt?
- Wie groß ist die Wahrscheinlichkeit, dass an keinem der 3 Tage mit schwarzer Mine geschrieben wurde? Spielt es dabei für den Wert der Wahrscheinlichkeit eine Rolle, ob der Schreiber jeweils am Tagesende wieder in die Schublade gelegt (Ziehen mit Zurücklegen) oder nicht mehr zurückgelegt wurde (Ziehen ohne Zurücklegen)?

Aufgabe W8

2 von 10 hergestellten Porzellanfiguren seien mangelhaft. Es werden nun zufällig zwei Exemplare entnommen. Bestimmen Sie für die folgenden Ereignisse sowohl im Falle mit Zurücklegen als auch im Falle ohne Zurücklegen die Eintrittswahrscheinlichkeiten!

- "Die erste ausgewählte Porzellanfigur ist mangelhaft."
- "Die zweite ausgewählte Porzellanfigur ist mangelhaft."
- "Mindestens eine der beiden ausgewählten Figuren ist mangelhaft."

Aufgabe W9

Berechnen Sie jeweils folgende Wahrscheinlichkeiten beim zweimaligen Werfen eines Würfels:

- Gegeben seien die Ereignisse A_1 : "Augensumme mindestens 9" und A_2 : "Erster Wurf = 6". Berechnen Sie $P(A_1 \mid A_2)$!
- Wie groß ist die Wahrscheinlichkeit, dass der erste Würfel eine Sechs zeigt, unter der Bedingung, dass die Augensumme mindestens 9 ist? Versuchen Sie den Sachverhalt auch grafisch zu veranschaulichen!

Aufgabe W10

In einer Schachtel befinden sich insgesamt 6 Zündkerzen, wovon genau 2 Stück defekt sind. Es werden zwei Zündkerzen nacheinander entnommen, ohne dass die erste vor der Entnahme der zweiten wieder in die Schachtel zurückgelegt wird.

Gesucht ist die Wahrscheinlichkeit dafür, dass beim zweiten Zug eine brauchbare Zündkerze erwischt wird, unter der Voraussetzung, dass bereits die erste gezogene Zündkerze brauchbar war.

Aufgabe W11

Zwei Studenten versuchen unabhängig voneinander eine Klausuraufgabe zu lösen, wobei jeder von ihnen mit einer Lösungswahrscheinlichkeit von 0,70 arbeitet. Wie groß ist die Wahrscheinlichkeit, dass wenigstens einer der Studenten die korrekte Lösung findet?

Aufgabe W12

Zu Beginn seines Studiums rechnet ein Student damit, dieses mit einer (subjektiven) Wahrscheinlichkeit von 60 % abzuschließen. Mit erfolgreich abgeschlossenem Studium beträgt die Wahrscheinlichkeit, die gewünschte Position zu erhalten, 70 %. Ohne Studienabschluss besetzt er diese nur mit 20 % Wahrscheinlichkeit.

Wie groß ist die Wahrscheinlichkeit, dass der Student die Position erhalten wird?

Aufgabe W13

Ein Student bewirbt sich vor Abschluss seines Studiums bei zwei Firmen X und Y. Er rechnet mit einer Wahrscheinlichkeit von 0,4 bei X und mit einer Wahrscheinlichkeit von 0,70 bei Y eingestellt zu werden. Zudem geht er mit einer Wahrscheinlichkeit von 0,20 davon aus, von beiden Firmen angenommen zu werden.

Wie wahrscheinlich ist es nun, dass der besagte Student von mindestens einer der beiden Firmen eine Zusage erhält?

Aufgabe W14

Ein Radargerät auf einer Autobahn wird innerhalb einer Stunde mit einer Wahrscheinlichkeit von 90 % von mindestens 500 Fahrzeugen passiert und mit einer Wahrscheinlichkeit von 50 % von höchstens 1000 Fahrzeugen.

Wie groß ist die Wahrscheinlichkeit, dass innerhalb der nächsten Stunde 500, 501, 502, ... oder 1000 Fahrzeuge gezählt werden können?

Aufgabe W15

Sie fragen in der Mensa zufällig drei Kommilitonen nach ihrem Geburtsdatum. Wie groß ist die Wahrscheinlichkeit dafür, dass

- alle drei am gleichen Wochentag geboren sind?
- alle drei an verschiedenen Wochentagen geboren sind?
- alle drei an einem Sonntag geboren sind?
- genau zwei am gleichen Wochentag geboren sind?

Aufgabe W16

Angenommen es gäbe weltweit nur drei Fluggesellschaften A, B und C. 30 % aller Flugreisen werden bei Gesellschaft A, 20 % bei B und 50 % bei C gebucht. Eine Analyse von Kundenbeschwerden ergab, dass bei A 0,1 % des Gepäcks bei Flügen an den falschen Zielflughafen geschickt wird. Bei B sind dies 0,5 % und bei C genau 0,12 %.

- Mit welcher Wahrscheinlichkeit landet das Gepäck eines Passagiers, der zufällig ausgewählt wird, nicht an seinem Zielflughafen?
- Ein Passagier wartet an seinem Zielflughafen vergeblich auf seinen Koffer. Dieser scheint wohl verloren gegangen oder fehlgeleitet worden zu sein. Mit welcher Wahrscheinlichkeit ist dieser Passagier mit Fluggesellschaft A geflogen?
- Prüfen Sie, ob die beiden Ereignisse E_1 : "Ein Passagier fliegt mit Gesellschaft B" und E_2 : "Sein Gepäck wird fehlgeleitet" voneinander abhängig sind!

Aufgabe W17

In einer Firma wird ein Produkt in zwei unabhängig voneinander und nacheinander ablaufenden Arbeitsgängen hergestellt. Es gilt als Ausschuss, wenn bei einem der beiden Arbeitsgänge ein Fehler passiert. Die Wahrscheinlichkeit für das Entstehen eines Ausschusstückes liegt nach Erfahrungswerten bei 0,06. Dabei geschieht mit einer Wahrscheinlichkeit von 0,04 ein Fehler im ersten Arbeitsgang.

Wie groß ist nun die Wahrscheinlichkeit für einen Fehler im zweiten Arbeitsgang?

Aufgabe W18

Um teuere Rückrufaktionen von produzierten Laptops zu vermeiden, werden von einem Hersteller eingehende Akkus, welche von Zulieferern hergestellt werden, einem zweistufigen Inspektionsplan unterzogen.

Es werden 10 Akkus aus einer eingegangenen Lieferung entnommen und auf Funktionsfähigkeit getestet. Sind alle 10 Akkus in Ordnung (Ereignis A_0), wird die Lieferung angenommen. Sind zwei oder mehr defekt (Ereignis A_2), wird die gesamte Lieferung zurückgeschickt. Ist jedoch nur ein Stück fehlerhaft (Ereignis A_1), wird eine weitere Stichprobe von 20 Stück entnommen. Nur dann, wenn in dieser zweiten Stichprobe alle Akkus einwandfrei sind, wird die Lieferung behalten, andernfalls zurückgeschickt.

Wie groß ist die Wahrscheinlichkeit, dass eine Lieferung angenommen wird (Ereignis B), in welcher 30% der Akkus fehlerhaft sind? Gehen Sie dabei der Einfachheit halber davon aus, dass die Stichproben mit Zurücklegen durchgeführt werden. In der Praxis der statistischen Qualitätskontrolle werden Stichproben natürlich ohne Zurücklegen gezogen, so dass die Wahrscheinlichkeit, einen defekten Akku zu finden, davon abhängig ist, ob der vorher entnommene Akku defekt war oder nicht.

Aufgabe W19

Eine Fußballmannschaft bestehe aus 25 Spielern. Davon nehmen in jeder Begegnung nur 11 aktiv am Spiel teil. Die verbleibenden 14 Spieler dienen als Reserve.

Die traditionelle Aufstellung dieser Mannschaft sei:

Sturmkette	• •
Mittelfeldkette	• • • • •
Abwehrkette	• • •
Torwart	•

Vor einem wichtigen Entscheidungsspiel stellt sich die Situation der Mannschaft nun wie folgt dar:

- Genau 3 der 25 Spieler sind als Torhüter spezialisiert und können nicht als Feldspieler eingesetzt werden. Alle anderen Spieler können nicht im Tor, sondern nur auf dem Feld spielen.
- Verletzungsbedingt können 3 Feldspieler nicht am Entscheidungsspiel teilnehmen.
- Ein weiterer Feldspieler hat im vorhergehenden Spiel eine rote Karte kassiert, sodass er ebenfalls ausfällt.

Es stehen damit für das Entscheidungsspiel noch 5 Stürmer, sieben Mittelfeldspieler und x Verteidiger zur Verfügung.

- Ermitteln Sie zunächst die Anzahl x der Verteidiger, die noch verfügbar sind!
- Wie viele Aufstellungen sind beim Spiel prinzipiell möglich? Beachten Sie dabei, dass die Spieler nur entsprechend ihrer Spezialisierung eingesetzt werden können. Mittelfeldspieler können also z.B. nur im Mittelfeld und nicht als Stürmer spielen. Welcher Mittelfeldspieler an welcher Position im Mittelfeld spielt ist jedoch unerheblich.

Aufgabe W20

Gegeben seien die Wahrscheinlichkeiten $P(A) = 0,6$, $P(B) = 0,2$ und $P(A \cap B) = 0,1$. Berechnen Sie mittels dieser Werte folgende Wahrscheinlichkeiten zahlenmäßig:

- | | | |
|-----------------------------|---------------|------------------------------|
| a) $P(A \cup B)$ | d) $P(A B)$ | f) $P(\bar{A} \cap \bar{B})$ |
| b) $P(\overline{A \cup B})$ | e) $P(B A)$ | g) $P(\bar{A} \cup \bar{B})$ |
| c) $P(A \cap \bar{B})$ | | |

Aufgabe W21

Beim Bingo dürfen Sie 10 verschiedene Zahlen zwischen 1 und 100 wählen. Dann werden 60 Zahlen zwischen 1 und 100 gezogen.

Wie groß ist die Wahrscheinlichkeit, dass dabei alle Ihre gewählten Zahlen dabei sind und Sie somit gewonnen haben?

Aufgabe W22

Ein Unternehmen erhält regelmäßig eine Lieferung von 50 Lithium-Ionen-Akkus. Die Annahmekontrolle geschieht nach folgendem Inspektionsplan:

Es wird ein Akku entnommen und auf Mängel untersucht. Ist dieser in Ordnung, wird ein zweiter Akku ohne Zurücklegen des vorher geprüften Akkus entnommen. Dieser Vorgang wird maximal 7-mal durchgeführt. Sobald ein Akku als defekt eingestuft wird, weist man die gesamte Lieferung zurück. Sind die ersten 7 Akkus allerdings einwandfrei, wird die Lieferung angenommen (Ereignis X).

Wie groß ist die Annahmewahrscheinlichkeit $P(X)$ für eine Lieferung, wenn 0, 5 oder 10 Stücke der Sendung schadhaft sind?

Aufgabe W23

Wie groß ist die Wahrscheinlichkeit, dass in einer Familie mit 4 Kindern

- alle 4 Jungen sind,
- genau 2 Jungen und 2 Mädchen sind,
- das älteste Kind ein Mädchen und die folgenden Jungen sind?

Gehen Sie dabei davon aus, dass das Auftreten eines Jungen oder eines Mädchens gleich wahrscheinlich ist!

Aufgabe W24

Das sog. Sterling-Silber ist eine Legierung aus Kupfer und Silber. Beim Legierungs-vorgang tritt in einer Silbermanufaktur mit einer Wahrscheinlichkeit von $P(L)=0,01$ ein Fehler auf. Verunreinigungen im Silber bzw. Kupfer kommen mit $P(S) = 0,02$ bzw. $P(K) = 0,03$ vor. Das entstehende Sterling-Silber gilt nur dann als einwandfrei, wenn die verwendeten Materialien makellos sind und es im Produktionsverfahren zu keinen Fehlern kommt.

Berechnen Sie unter der Annahme, dass sich die möglicherweise auftretenden Mängel bzw. Produktionsfehler nicht gegenseitig beeinflussen, die Wahrscheinlichkeit für die Produktion von hochwertigem Sterling-Silber!

Aufgabe W25

Ein Neuling beim Pferderennen, der über die Stärken und Schwächen der einzelnen Pferde keine Kenntnis hat, versucht sein Glück durch zufälliges Tippen bei einem Rennen, an dem 10 Pferde teilnehmen.

- a) Wie groß ist jeweils die Wahrscheinlichkeit, dass er die folgenden Ereignisse A, B, C richtig tippt?
 - A: "siegenderes Pferd"
 - B: "die ersten drei Pferde in der Reihenfolge ihres Einlaufes"
 - C: "die ersten drei Pferde ohne Berücksichtigung der Einlaufreihenfolge"
- b) Angenommen, der Neuling hat einen todsicheren Wetttipp erhalten, der ihn über den Erstplatzierten des Rennens informiert. Kommt es dadurch zu einer Änderung der Wahrscheinlichkeiten obiger Ereignisse?

Aufgabe W26

Gegeben seien drei Urnen U_i ($i = 1,2,3$). Jede Urne enthält insgesamt 10 Kugeln. Der Wert i gibt dabei jeweils an, wie viele rote Kugeln sich in dieser Urne befinden. Die restlichen Kugeln sind schwarz.

Es wird nun gewürfelt, wobei der Ausgang des Wurfs darüber entscheidet, aus welcher Urne eine einzige Kugel entnommen wird. Würfelt man 1 oder 2, so wird die Kugel aus der Urne U_1 gezogen, bei 3 oder 4 aus der Urne U_2 und bei 5 oder 6 aus der Urne U_3 .

Wie groß ist die Wahrscheinlichkeit, dass die eine gezogene Kugel rot ist?

Zufallsvariablen

Aufgabe Z1

Eine Münze wird dreimal geworfen. Geben Sie für die Zufallsvariable X: "Anzahl von Kopf" sowohl die Wahrscheinlichkeits- als auch die Verteilungsfunktion in Tabellenform an!

Aufgabe Z2

Nehmen Sie an, Sie schreiben zusammen mit Ihrem Professor ein Statistik-Lehrbuch. Sie wollen eine Übungsaufgabe stellen, in der die zukünftigen Leser prüfen sollen, ob es sich bei der gegebenen Funktion um eine Dichtefunktion handelt. Sie beabsichtigen, dass es sich bei der Funktion tatsächlich um eine Dichtefunktion handelt und diese im Bereich der möglichen Werte die Gestalt einer fallenden Gerade besitzt, welche bei $x = 6$ eine Nullstelle aufweist. Die Zufallsvariable X, die durch diese Dichtefunktion beschrieben wird, soll nur Werte im Intervall $[0;6]$ annehmen können. Welche Dichtefunktion liefern Sie Ihrem Professor?

Aufgabe Z3

Eine diskrete Zufallsvariable X habe die folgende Wahrscheinlichkeitsfunktion:

$$f(x) = \begin{cases} 0,2 & \text{für } x = 1 \\ 0,8 & \text{für } x = 2 \\ 0 & \text{sonst.} \end{cases}$$

Berechnen Sie den Erwartungswert von X!

Aufgabe Z4

Begründen Sie, warum für den Erwartungswert einer konstanten Größe a stets der Zusammenhang $E(a) = a$ gilt!

Aufgabe Z5

Bestimmen Sie nicht durch Berechnung, sondern rein durch Argumentation, den Erwartungswert einer Zufallsvariablen X mit folgender Wahrscheinlichkeitsfunktion:

x_i	4	5	6
$f(x_i)$	0,2	0,6	0,2

Aufgabe Z6

Ein Händler hat zwei gleichartige High-End-Laptops auf Lager. Er möchte beide innerhalb einer Periode zu je 1.500 Euro verkaufen. Dem Händler sind folgende Verkaufswahrscheinlichkeiten bekannt: Beide Geräte werden mit 70 %, nur ein Gerät mit 20 % und kein Gerät mit 10 % Wahrscheinlichkeit innerhalb der Periode verkauft. Wie hoch ist damit in der Periode der erwartete Umsatz mit den Laptops?

Aufgabe Z7

Ein Spekulant kauft 1.000 US-Dollar für 800 Euro, beabsichtigt diese aber nach einem Monat wieder zu verkaufen. Zum Verkaufszeitpunkt wird der Dollar-Kurs wie folgt eingeschätzt:

Wert des Dollar in Euro	0,50	0,70	0,80	1,00	1,20
Wahrscheinlichkeit	10 %	10 %	40 %	20 %	20 %

Wie hoch ist der erwartete Gewinn bzw. Verlust aus der Spekulation? Berechnen Sie ein geeignetes Maß zur Messung des Risikos dieser Spekulation!

Aufgabe Z8

In einer Studiengruppe befinden sich 6 sehr gute und 6 mittelmäßige Studenten. Ihr Professor benötigt nun einen dieser Studenten für die Durchführung eines Mathematiktutoriums. Da er aber ein schlechtes Personengedächtnis hat, kann er nicht unterscheiden, welche die 6 ausgezeichneten Studenten sind. Für die Auswahl eines Kandidaten geht er nun folgendermaßen vor: Er greift zufällig einen Studenten heraus und stellt ihm eine Rechenaufgabe, von der er weiß, dass diese nur von den sehr guten Studenten gelöst werden kann. Kann der befragte Student keine korrekte Antwort geben, wählt der Professor zufällig einen weiteren Studenten aus und stellt eine neue Aufgabe. Dies macht er so lange, bis ein Student die gestellte Aufgabe lösen kann. Dieser wird dann (seine Bereitschaft vorausgesetzt) Tutor.

Die Zufallsvariable X stehe nun für die Anzahl der befragten Studenten aus der Studiengruppe. Berechnen Sie $E(X)$! Was sagt $E(X)$ hier aus?

Aufgabe Z9

Es sei folgende Funktion $f(x)$ der stetigen Zufallsvariable X gegeben:

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

- a) Zeigen Sie, dass $f(x)$ eine Dichtefunktion ist!
- b) Bestimmen Sie die Verteilungsfunktion $F(x)$!
- c) Wie groß sind die Wahrscheinlichkeiten $P(0,3 \leq X \leq 0,7)$ und $P(X > 0,8)$?

Aufgabe Z10

Berechnen Sie Erwartungswert und Varianz einer stetigen Zufallsvariablen X , die im Intervall $0 < x < 3$ gleichmäßig mit der konstanten Dichte von $1/3$ verteilt ist! Bestimmen Sie außerdem die Wahrscheinlichkeiten $P(X = 1)$, $P(X \leq 2)$, $P(X < 2)$ sowie $P(1 < X < 2)$ und $P(1 \leq X < 2)$!

Aufgabe Z11

Nehmen Sie an, eine Zufallsvariable X besitzt den Erwartungswert $\mu_x = 8$ und die Varianz $\sigma_x^2 = 4$. Schätzen Sie ab, wie groß die Wahrscheinlichkeiten

- a) $P(4 < X < 12)$
 - b) $P(X \leq 2 \cup X \geq 14)$
- mindestens bzw. höchstens sind!

Aufgabe Z12

Eine diskrete Zufallsvariable Y mit den drei Realisationen 0, 1 und 2 besitzt die folgende Wahrscheinlichkeitsfunktion:

$$f(y) = \begin{cases} 0,25\sqrt{a} - 2^{-y-1} & \text{für } y = 0, 1, 2 \\ 0 & \text{sonst} \end{cases}$$

- Welchen Wert muss die Konstante a annehmen?
- Berechnen Sie $E(Y)$ und $\text{Var}(Y)$!
- Wie groß sind die Wahrscheinlichkeiten $P(Y > 1)$, $P(Y \geq 1)$ und $P(0 \leq Y < 5)$?

Aufgabe Z13

Ein idealer Würfel wird n-mal geworfen. Es sei X: "Augensumme" und Y: "Augenprodukt". Berechnen Sie $E(X)$ und $E(Y)$!

Aufgabe Z14

Drei Anleger A, B und C wollen sich ein Portfolio aus Aktien der Firmen G und H zusammenstellen. Über die Rendite X von Firma G und Y von H ist bekannt, dass

$$\begin{array}{ll} E(X) = 0,07 & \text{Var}(X) = 0,16 \\ E(Y) = 0,12 & \text{Var}(Y) = 0,25 \\ & \rho_{XY} = 0,20. \end{array}$$

Beantworten Sie dazu folgende Fragen:

- Welchen Wert besitzt die Kovarianz der Aktienrenditen?
- Anleger A entschließt sich sein gesamtes Vermögen in Aktie G zu investieren, da diese (gemessen an der Varianz) ein geringeres Risiko hat. Anleger B entscheidet sich hingegen zu einer 50:50-Investition. Welcher der beiden Anleger hat nun die Anlageform mit dem geringeren Risiko gewählt?
- Anleger C behauptet, er kenne eine optimale Strategie im Sinne eines minimalen Risikos. In welchen Relationen wird er in die Aktien G und H investieren?

Aufgabe Z15

Für die Aktien zweier Unternehmen A und B ergeben sich zu verschiedenen Situationen 1, 2 und 3, die mit gewissen Wahrscheinlichkeiten auftreten können, folgende Renditen:

Situation i	Wahrscheinlichkeit p_i	Rendite r_i^A	Rendite r_i^B
1	1/3	0,385	-0,225
2	1/3	0,140	0,020
3	1/3	-0,105	0,265

Untersuchen Sie die Korrelation der Renditen von A und B und geben Sie an, wie hoch die erwartete Rendite und die Standardabweichung eines Portfolios ist, welches je zu 50 % in A und B investiert!

Aufgabe Z16

Aus einer Gruppe von berufstätigen Ehepaaren, die jeweils in einem Haushalt zusammenleben, wird eines zufällig ausgewählt. Es ist bekannt, dass das erwartete Einkommen des Mannes bei einer Varianz von $\text{Var}(X) = 300 \text{ €}^2$ bei $E(X) = 3.000 \text{ €}$ liegt. Für die Frau gilt $E(Y) = 2.500 \text{ €}$ und $\text{Var}(Y) = 200 \text{ €}^2$. Ermitteln Sie unter der Annahme eines Korrelationskoeffizienten von 0,50 den Erwartungswert und die Varianz des Haushaltseinkommens Z!

Aufgabe Z17

Bei einem Wettbewerb werden die besten Doktorarbeiten Deutschlands von einer Jury hinsichtlich formaler Korrektheit X in zwei und bezüglich der praktischen Verwendbarkeit Y in drei Güteklassen eingeteilt. Ein Doktorand rechnet mit folgenden Wahrscheinlichkeiten, dass seine Arbeit bezüglich der beiden Merkmale X und Y in die einzelnen Güteklassen fällt:

x	y	1	2	3
1	?	0,20	0,05	
2	0,15	0,10	0,05	

Ermitteln und interpretieren Sie

- den fehlenden Tabellenwert im Hinblick darauf, wie der Doktorand seine Leistung einschätzt,
- $f_Y(1)$,
- $f(X = 2 | Y = 3)$,
- $E(X)$, $E(Y)$, $\text{Var}(X)$ und $\text{Var}(Z)$,
- $E(X | 1)$ und $\text{Var}(X | 1)$ sowie
- ρ_{XY} !

Prüfen Sie des Weiteren, ob

- X und Y voneinander stochastisch unabhängig sind!

Aufgabe Z18

Eine Aktie der AW-Clan AG werfe einen mittleren Jahresgewinn von $\mu_G = 10 \text{ €}$ ab. Die Unsicherheit im tatsächlichen Gewinn kommt in seiner Standardabweichung von $\sigma_G = 30 \text{ €}$ zum Ausdruck.

Berechnen Sie Erwartungswert und Standardabweichung des Gewinns eines Portfolios P, das sich

- aus drei Aktien des AW-Clans
- aus einer Aktie des AW-Clans und zwei weiteren Gesellschaften, die jeweils denselben Erwartungswert und dieselbe Standardabweichung aufweisen,
zusammensetzt! Unterstellen Sie bei b), dass sich die Gewinne der einzelnen Gesellschaften völlig unabhängig voneinander entwickeln!

Verteilungen

Aufgabe V1

20 Personen gleichen Alters und gleicher physischer Konstitution, die bei einer Versicherungsgesellschaft Lebensversicherungsverträge abgeschlossen haben, sind an einer schweren Grippe erkrankt. Nach Meinung der Ärzte beträgt die Wahrscheinlichkeit für jeden der Erkrankten, das nächste Jahr nicht zu überleben, 10 %.

Wie wahrscheinlich ist es, dass nach einem Jahr wenigstens a) 18 bzw. b) 1 Erkrankte(r) noch am Leben sind (ist)? Bestimmen Sie die gesuchten Wahrscheinlichkeiten ohne Nutzung der Verteilungstabellen aus dem Anhang!

Aufgabe V2

Ein Diktator hat eine Liste mit 12 Offizieren einer Elitetruppe vor sich. Der Diktator würfelt 6 der 12 Offiziere aus. Diese 6 bilden seine Leibwache. Von seinem Geheimdienst erfährt der Diktator, dass unter den 12 Offizieren 2 Attentäter sind.

Wie wahrscheinlich ist a) kein Attentäter bzw. b) mindestens ein Attentäter unter der Leibwache?

Aufgabe V3

Nehmen Sie an, in einem Wahlkreis beabsichtigen 45 % der Wahlberechtigten eine Partei B zu wählen, die sich dafür einsetzt, dass Bayrisch als eigenständige Fremdsprache anerkannt wird. Wie groß ist näherungsweise die Wahrscheinlichkeit, dass die Wähler dieser Partei in einer Zufallsstichprobe von $n = 10$ (ohne Zurücklegen) die Mehrheit (mehr als 50 % der Stimmen) haben?

Aufgabe V4

Eine Schachtel enthält 500 elektrische Bauteile, von denen 100 defekt sind. Wie groß ist näherungsweise die Wahrscheinlichkeit, dass höchstens 2 von 10 (ohne Zurücklegen) entnommenen Bauteilen defekt sind? Bestimmen Sie die gesuchten Wahrscheinlichkeiten ohne Nutzung der Verteilungstabellen aus dem Anhang!

Aufgabe V5

In einem Telefonnetz mit 10.000 Telefonanschlüssen tritt mit einer Wahrscheinlichkeit von 0,0005 pro Tag pro Anschluss eine Funktionsstörung auf. Störungen treten dabei unabhängig voneinander auf. Die Wahrscheinlichkeit, dass pro Tag und Anschluss mehr als eine Störung auftritt, kann vernachlässigt werden.

Wie groß ist die Wahrscheinlichkeit, dass genau 5 (a), mindestens 10 (b) und höchstens 20 (c) Störungen auftreten?

Aufgabe V6

In einem Callcenter gehen im Durchschnitt pro Minute 3 Anrufe ein. Wie groß ist die Wahrscheinlichkeit, dass die Callcentermitarbeiter in einer bestimmten Minute "arbeitslos" sind?

Aufgabe V7

Von 10.000 Einwohnern einer Stadt sind nur 500 daran interessiert, dass der aktuelle Bürgermeister im Amt bleibt. Dieser Bürgermeister möchte nun vor der Wahl mit 50 rein zufällig ausgewählten Bürgern sprechen, um sich einen Eindruck von seinen Wiederwahlchancen zu machen.

Wie groß ist näherungsweise die Wahrscheinlichkeit, dass ihm von mindestens 7, jedoch höchstens 10 Bürgern eine positive Resonanz gegeben wird?

Aufgabe V8

Nach Herstellerangaben liegt die durchschnittliche Lebensdauer einer Energiesparlampe bei 4.000 Stunden. Wie groß ist die Wahrscheinlichkeit dafür, dass die Glühbirne weniger als halb so lange (a) oder mehr als doppelt so lange (b) brennt, wenn als geeignetes Modell für die Verteilung der Lebensdauer X die Exponentialverteilung verwendet wird? Skizzieren Sie die gesuchten Wahrscheinlichkeiten unter Verwendung dieser Dichtefunktion!

Aufgabe V9

Die Zufallsvariable X beschreibe die Wartezeit eines Kunden an einem Bankschalter. Zeigen Sie allgemein, dass die Wahrscheinlichkeit, endlich an die Reihe zu kommen, mit zunehmender Wartezeit w nicht zunimmt, wenn die Zeitdauer einer Exponentialverteilung gehorcht!

Aufgabe V10

Berechnen Sie für eine standardnormalverteilte Zufallsvariable Z die Wahrscheinlichkeiten $P(0 \leq Z \leq 1,6)$, $P(-1,6 < Z < 0)$, $P(Z > 1,6)$ und $P(-2,5 \leq Z \leq 2,5)$!

Erklären Sie, warum es hier keine Rolle spielt bei der Berechnung der Wahrscheinlichkeiten zwischen " $<$ " und " \leq " bzw. " $>$ " und " \geq " zu differenzieren!

Aufgabe V11

Die Brenndauer von Glühbirnen einer speziellen Marke sei normalverteilt mit dem Mittelwert 900 Stunden und der Standardabweichung 100 Stunden. Wie groß ist die Wahrscheinlichkeit, dass eine zufällig aus der Produktion entnommene Glühbirne

- a) mindestens 1.200 Stunden
 - b) höchstens 650 Stunden
 - c) zwischen 750 und 1.050 Stunden
- brennt?

Aufgabe V12

Das Gewicht (in Gramm) von Zuckerpaketen sei $N(1.000; 25)$ -verteilt. Ermitteln Sie

- a) das Gewicht, das ein zufällig entnommenes Zuckerpaket mit einer Wahrscheinlichkeit von 90 % höchstens wiegt,
- b) das Gewicht, das ein zufällig entnommenes Zuckerpaket mit einer Wahrscheinlichkeit von 95 % mindestens wiegt,
- c) den zum Erwartungswert symmetrischen Bereich, der das Gewicht eines zufällig entnommenen Pakets mit einer Wahrscheinlichkeit von 98 % enthält.

Aufgabe V13

Eine Molkerei produziert unter anderem Butterriegel, deren Gewichte als normalverteilt angesehen werden. Aus Erfahrung ist nun bekannt, dass die Wahrscheinlichkeit dafür, dass ein zufällig entnommener Riegel maximal 1 g vom Sollwert $\mu = 250$ g abweicht, bei 98 % liegt. Welche Streuung σ implizieren diese Erfahrungswerte?

Aufgabe V14

- Bestimmen Sie das 0,90-Fraktil und das 0,10-Fraktil einer $\chi^2(10)$ -Verteilung!
- Bestimmen Sie das 0,99-Fraktil und das 0,01-Fraktil einer $t(20)$ -Verteilung!
- Bestimmen Sie das 0,95- und das 0,01-Fraktil einer $F(8; 10)$ -Verteilung!

Aufgabe V15

An einer Bushaltestelle in Frankfurt fährt exakt alle 10 Minuten ein Bus ab. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig eintreffender Fahrgäst länger als 5 Minuten warten muss?

Aufgabe V16

Eine Brauerei füllt Weißbier in extra große Flaschen ab. Umfangreiche Messungen haben ergeben, dass der Flascheninhalt mit $\mu = 751,5$ ml und $\sigma = 0,6$ ml normalverteilt ist. Der Mindestinhalt einer Flasche beträgt 750 ml.

- Wie groß ist die Wahrscheinlichkeit, dass eine Flasche unterfüllt ist?
- Wie groß ist die Wahrscheinlichkeit, dass in einer Flasche zwischen 750 und 753 ml enthalten sind?
- Wie groß ist bei geeigneter Unabhängigkeitsannahme die Wahrscheinlichkeit, dass beim Kauf von sechs Flaschen insgesamt eine Unterfüllung vorliegt?
- Die Abfüllanlage war bisher auf 751,5 ml justiert. Wie ist die Maschine bei $\sigma = 0,6$ ml einzustellen, wenn höchstens 2,5 % der Flaschen unterfüllt sein sollen?

Aufgabe V17

Von den Mitarbeitern einer Bankfiliale werden im Jahresdurchschnitt 1.000.000 Überweisungen verbucht. Die Fehlerquote bei der Archivierung der Überweisungsbelege liege bei 0,5 %. Für diese Filiale werden im Rahmen der Innenrevision nun zufällig 3.000 Belege überprüft. Ein Fehleranteil von 1 % wird dabei als tolerierbar angesehen. Andernfalls hat sich der Filialdirektor zu verantworten. Wie wahrscheinlich sind nun negative Folgen für den Filialdirektor?

Aufgabe V18

Eine Maschine füllt Mörschen in Blechdosen ab. Die Gewichte des Inhalts und der leeren Dosen seien unabhängig normalverteilt mit dem Mittelwert 750 g bzw. 50 g und der Varianz 100 g^2 bzw. 5 g^2 . Genau 1.000 dieser gefüllten Dosen werden nun rein zufällig in einen Lieferwagen verladen.

- Wie wahrscheinlich ist es, dass eine einzelne Dose weniger als 790 g wiegt?
- Wie groß ist die Wahrscheinlichkeit, dass die in den Lieferwagen verladenen Dosen ein Gewicht von 801 kg überschreiten?

III INDUKTIVE STATISTIK

Häufig stellt sich in der Statistik die Aufgabe, Informationen über bestimmte Eigenschaften einer statistischen Grundgesamtheit zu beschaffen (z.B. Durchschnittseinkommen bayerischer Studenten oder der Zusammenhang zwischen Arbeitseinkommen und Bildung bei Erwerbspersonen). Da in vielen Anwendungsfällen eine vollständige Datenerhebung (Einbeziehung sämtlicher Elemente der Grundgesamtheit) aber unmöglich oder nicht zweckmäßig (zu kosten- und zeitintensiv) ist, werden durch die in einer Teilerhebung (Stichprobe; Teilmenge der Grundgesamtheit) gewonnenen Daten Rückschlüsse auf die Gesamtsituation gezogen. So können beispielsweise Mittel- oder Anteilswerte in der Grundgesamtheit aufgrund der Ergebnisse der Teilerhebung mehr oder weniger genau geschätzt werden.

Zur Schätzung können sog. **Punktschätzverfahren** verwendet werden, die einen einzigen Schätzwert liefern. Es kann außerdem ein **Intervallschätzverfahren** herangezogen werden. Hier wird als Schätzergebnis ein Intervall bestimmt, in das der unbekannte Parameter mit einer vorgegebenen hohen Wahrscheinlichkeit fällt.

Soll eine bezüglich der Grundgesamtheit bestehende Hypothese anhand der Stichprobenergebnisse auf ihre Richtigkeit überprüft werden, so geschieht dies mit sog. **Hypothesentests**. Sie sind so konstruiert, dass eine korrekte Hypothese nur mit einer geringen Wahrscheinlichkeit abgelehnt wird.

1.

Punktschätzung

Häufig ist es wünschenswert, einen unbekannten charakteristischen Wert der Verteilung einer Grundgesamtheit möglichst genau zu kennen. So kann sich z.B. eine Partei für ihren Stimmanteil bei der nächsten Wahl (Anteilswert) oder eine Versicherungsgesellschaft für die Lebenserwartung von Rauchern (Mittelwert) interessieren. Wird bei solchen Fragestellungen für einen unbekannten Parameter der Verteilung der Grundgesamtheit aufgrund des Ergebnisses einer Stichprobe ein einziger numerischer Schätzwert festgelegt, sprechen wir von einer Punktschätzung. Die Stichprobenergebnisse werden dabei in sog. Schätzfunktionen verarbeitet, die dem Stichprobendatenmaterial einen charakteristischen Zahlenwert zuordnen. Dieser Zahlenwert dient dann als Schätzung.

1.1 Stichproben

Allgemein sind bei Teilerhebungen zwei Arten von Auswahlverfahren zu unterscheiden. Dies sind Zufallsstichproben und Verfahren der bewussten Auswahl.

Bei **Zufallsstichproben** besitzen sämtliche Elemente der Grundgesamtheit bestimmte, von Null verschiedene Wahrscheinlichkeiten, in die Stichprobe zu gelangen. Das einfachste Verfahren dieser Gruppe ist die *reine (uneingeschränkte) Zufallsauswahl*, bei der jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit besitzt, gezogen zu werden.

Eine Zufallsstichprobe kann durch ein *Urnenmodell* vollständig beschrieben werden. Die Urne enthält dabei genau N Kugeln (= Anzahl statistischer Einheiten in der Grundgesamtheit), die nummeriert sind und von denen jede genau einer statistischen Einheit zugeordnet ist. Für eine reine Zufallsstichprobe vom Umfang n ziehen wir n Kugeln, notieren die Nummern und haben damit die Auswahl der statistischen Einheiten vorgenommen. Das Ziehen kann dabei mit oder ohne Zurücklegen vorgenommen werden.

Untersuchen wir alle Objekte der Grundgesamtheit hinsichtlich eines Merkmals X, so können die festgestellten Merkmalsausprägungen als Realisationen einer Zufallsvariablen X betrachtet werden. Die sich daraus ergebende Verteilung von X wird auch als *Verteilung der Grundgesamtheit* bezüglich des Merkmals X bezeichnet. Daraus können schließlich *Mittelwert und Varianz der Grundgesamtheit* (bezüglich des Merkmals X) bestimmt werden. In der induktiven Statistik gehen wir nun aber gerade einen anderen Weg. Aus einer Stichprobenerhebung sollen Verteilung, Mittelwert, Varianz und andere Charakteristika der Grundgesamtheit geschätzt werden. Es wird daher nach bestimmten Ziehungsvorschriften eine bestimmte Anzahl von

Objekten aus der Grundgesamtheit ausgewählt, wobei bei jedem Objekt die Ausprägung des Merkmals X registriert wird. Die Anzahl n der so durchgeführten Beobachtungsvorgänge wird als *Stichprobenumfang* bezeichnet. Jeden Beobachtungswert x_i können wir als Realisierung einer Zufallsvariablen, der sog. i-ten *Stichprobenvariablen* X_i mit $i = 1, 2, \dots, n$, ansehen.

Alle Stichprobenvariablen X_i sind *unabhängig* und bezüglich der Grundgesamtheit *identisch verteilt* (d.h. genau wie X),¹ wenn die Entnahme der Objekte aus der Grundgesamtheit als reine Zufallsauswahl und unabhängig voneinander erfolgt. Beim derartigen Ziehen eines Objekts aus der Grundgesamtheit handelt es sich beim interessierenden Merkmal also um eine Zufallsvariable, die genauso verteilt ist wie die Grundgesamtheit selbst. Wir bezeichnen eine solche unabhängige Zufallsauswahl als **einfache Stichprobe** vom Umfang n. Das *Stichprobenergebnis* ergibt sich durch die Realisierungen x_1, x_2, \dots, x_n der n Stichprobenvariablen X_1, X_2, \dots, X_n .

Einer einfachen Stichprobe liegt die Vorstellung eines zufälligen *Ziebens mit Zurücklegen* zugrunde. Nur in einem solchen Fall ist nämlich gewährleistet, dass die Züge und damit auch die Stichprobenvariablen *unabhängig* sind. Einfache Stichproben sind jedoch in der Praxis nicht immer zu realisieren. Beim zufälligen *Ziehen ohne Zurücklegen* sind die Züge *abhängig* voneinander. Im Folgenden werden wir dennoch vor allem einfache Stichproben behandeln. Wenn wir von einer Stichprobe sprechen, meinen wir daher i.d.R. eine einfache Stichprobe. Die Aussagen, die entwickelt werden, gelten beim Ziehen ohne Zurücklegen zumindest näherungsweise, wenn die Anzahl N der Elemente der Grundgesamtheit sehr viel größer als der Stichprobenumfang n ist. Als Faustregel sollte $n \leq 0,1N$ vorliegen.

Verfahren der bewussten Auswahl beruhen nicht auf dem Zufallsprinzip. Ein häufig in der Markt- und Meinungsforschung eingesetztes Verfahren dieser Richtung ist die sog. *Quotenauswahl*. Dabei wird versucht, die in die Stichprobe aufzunehmenden Elemente so zu wählen, dass die Teilgesamtheit einen modellgerechten Miniaturquerschnitt der zu untersuchenden Grundgesamtheit darstellt (z.B. bestimmte Frauenquote). Da die Interviewer allerdings mehr oder weniger frei bei der Auswahl der zu befragenden Personen sind (subjektive Auswahl), können keine wahrscheinlichkeitstheoretisch fundierten Aussagen über die Zuverlässigkeit der gewonnenen Ergebnisse gemacht werden.

1.2 Schätzer und ihre Stichprobenverteilungen

1.2.1 Grundlagen der Punktschätzung

Bei einer Punktschätzung wird für einen unbekannten Parameter ϑ der Verteilung einer Grundgesamtheit aufgrund des Ergebnisses x_1, x_2, \dots, x_n einer Stichprobe ein numerischer Schätzwert $\hat{\vartheta}$ festgelegt.² Ein solcher Schätzwert kann als Realisierung einer Stichprobenfunktion

$$\hat{\Theta} = g(X_1, X_2, \dots, X_n) \quad (\text{III1: 1})$$

¹ Diese Eigenschaften der Stichprobenvariablen bezeichnet man als i.i.d. (engl. independent and identically distributed). Aus ihnen folgt $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$.

² Hinweis: ϑ ist der Kleinbuchstabe zu Θ .

betrachtet werden, die der Verarbeitung des eintretenden Stichprobenergebnisses zu aussagekräftigen Größen dient und als **Schätzfunktion** oder **Schätzer** bezeichnet wird. Da die n StichprobenvARIABLEN X_1, \dots, X_n Zufallsvariablen sind, ist auch $\hat{\Theta}$ als Funktion der n Zufallsvariablen eine Zufallsvariable. Die Wahrscheinlichkeitsverteilung von $\hat{\Theta}$ heißt *Stichprobenverteilung* von $\hat{\Theta}$. Für eine konkrete Stichprobe mit den Ausprägungen x_1, x_2, \dots, x_n liefert der Schätzer $\hat{\Theta}$ die *Schätzung* $\hat{\vartheta} = g(x_1, \dots, x_n)$. $\hat{\vartheta}$ ist also die konkrete Realisation der Zufallsvariable $\hat{\Theta}$. Der Schätzer gibt somit an, wie aus den Ergebnissen einer Stichprobe ein Schätzwert für ϑ zu bestimmen ist.

Unser Ziel ist es, eine konkrete Stichprobenrealisation x_1, x_2, \dots, x_n durch geeignete gewählte Funktionen $g(X_1, \dots, X_n)$ so zu verdichten, dass wir auf uns interessierende Kennzahlen bzw. Parameter der Grundgesamtheit bzw. der Verteilung von X schließen können. Nach welchen Kriterien derartige Funktionen auszuwählen sind, werden wir im Abschnitt III 1.3 vertiefen. Die Schätzer für das *arithmetische Mittel*, den *Anteilswert* und die *Varianz* einer Grundgesamtheit sind in der folgenden Tabelle enthalten und im Anschluss daran erklärt:

Parameter	Schätzer	
Arithm. Mittel	Stichprobenmittel	
μ	$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	(III1: 2)
Anteilswert	Stichprobenanteilswert	
p	$\hat{p} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{für } X_i = \begin{cases} 1 & A \text{ tritt ein} \\ 0 & A \text{ tritt nicht ein} \end{cases}$	(III1: 3)
Varianz	Stichprobenvarianz	
σ^2	$\hat{\sigma}^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$	(III1: 4)

Hinsichtlich der *Notation* ist zunächst zu beachten, dass wir Schätzer im Folgenden dadurch kennzeichnen wollen, dass wir die zu schätzenden Parameter mit "Dächern" versehen. Für die Realisationen der obigen Schätzer werden wir aus Vereinfachungsgründen außerdem keine eigene Symbolik einführen, da stets aus dem Kontext hervorgeht, worum es sich genau handelt.³

Die obige Übersicht zeigt mit (III1: 2), dass wir als Schätzer $\hat{\mu}$ des arithmetischen Mittels μ einer Grundgesamtheit, das arithmetische Mittel \bar{X} der StichprobenvARIABLEN X_1, \dots, X_n verwenden können. Die Schätzung für μ aus einer konkreten Stichprobe ist also gerade das arithmetische Mittel der Realisationen x_1, \dots, x_n der n StichprobenvARIABLEN.

³ Die Verwendung von Großbuchstaben für Schätzer und Kleinbuchstaben für Schätzungen (Realisationen) ergäbe z.B. das Problem, dass wir den Schätzer für σ^2 mit $\hat{\Sigma}^2$ und seine Realisation mit $\hat{\sigma}^2$ benennen müssten. Diese Symbolik ist aber bereits für das Summenzeichen vergeben.

Beispiel III 1-1**Schätzung des Erwartungswerts**

Von den Studenten einer Vorlesung (Grundgesamtheit) werden 5 zufällig ausgewählt und ihre Körpergröße X in cm festgehalten. Die Stichprobe ergab die x -Werte 182, 178, 186, 180, 177. Wir können damit die durchschnittliche Körpergröße μ aller Studenten in der Vorlesung auf

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{5} \cdot (182 + \dots + 177) = 180,6 \text{ cm}$$

schätzen.

Beim Anteilswert p handelt es sich um nichts anderes als den Mittelwert einer dichotomen Grundgesamtheit. In einer solchen Grundgesamtheit gilt für die Stichprobenvariablen X_1, \dots, X_n hinsichtlich des Eintritts eines Ereignisses A

$$X_i = \begin{cases} 1 & \text{falls A eintritt} \\ 0 & \text{falls A nicht eintritt.} \end{cases}$$

Das Stichprobenmittel beschreibt also im Fall einer dichotomen Grundgesamtheit gerade die relative Häufigkeit der Ausprägung 1 in der Stichprobe und wird daher in diesem Kontext als Stichprobenanteilswert bezeichnet.

Beispiel III 1-2**Schätzung des Anteilswerts**

Aus einer Gruppe von 5.000 Personen werden zufällig 100 befragt, ob sie einer Partei A bei den nächsten Wahlen ihre Stimme geben werden. Nur 5 davon äußern sich diesbezüglich positiv. Genau 5 der X_i nehmen daher die Realisation 1 an, sodass wir den unbekannten Anteil p potenzieller Wähler unter den 5.000 Personen auf

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{100} \cdot 5 = 0,05$$

bzw. 5 % schätzen können.

Wie (III1: 4) zeigt, verwenden wir (beim Ziehen mit Zurücklegen) als Schätzer $\hat{\sigma}^2$ für die Varianz σ^2 nicht – wie man vielleicht vermuten würde – die mittlere quadratische Abweichung

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad (\text{III1: 5})$$

sondern modifizieren diese. Die so entstehende Stichprobenvarianz $\hat{\sigma}^2$ besitzt als Schätzer bessere Eigenschaften als S^2 . Zu diesen zählt etwa die Erwartungstreue, die wir im Abschnitt III 1.3.1 noch näher betrachten werden.

Beispiel III 1-3**Schätzung der Varianz**

Für Beispiel III 1-1 erhalten wir (beim Ziehen mit Zurücklegen) als Schätzung die Varianz

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{5-1} \cdot ((182 - 180,6)^2 + \dots + (177 - 180,6)^2) = 12,8 \text{ cm}^2.$$

In den folgenden Abschnitten wollen wir uns nun den Verteilungen der bisher behandelten Schätzer zuwenden. Diese erlauben uns, *Wahrscheinlichkeiten* dafür zu bestimmen, dass die *Stichprobengrößen in einer gezogenen Stichprobe einen bestimmten Wert annehmen*. Sie werden außerdem noch im Rahmen von Intervallschätzungen (Kapitel III 2) und Testverfahren (Kapitel III 3) von Bedeutung sein.

1.2.2 Verteilung des Stichprobenmittels

Zur Herleitung der Verteilung des arithmetischen Mittels einer Zufallsstichprobe bedienen wir uns des *Urnenmodells für ein metrisch skaliertes Merkmal*. Es werden hier die N Elemente der Grundgesamtheit durch N Kugeln repräsentiert. Auf diesen Kugeln sind die entsprechenden Merkmalswerte x_i ($i = 1, \dots, N$) aufgedruckt. Der Mittelwert μ der Grundgesamtheit ergibt sich als arithmetisches Mittel dieser Merkmalswerte. Wird nun eine Stichprobe im Umfang n entnommen, so hängt das Ergebnis eines jeden Zuges vom Zufall ab. Das Ergebnis des i -ten Zuges, welches wir mit X_i bezeichnen wollen, ist daher eine Zufallsvariable, die bei einer bestimmten Ausführung des Zufallsexperiments die konkrete Realisation x_i annimmt.

Ziehung Nr.	1	2	...	i	...	n
Stichprobenvariable	X_1	X_2	...	X_i	...	X_n
Realisation	x_1	x_2	...	x_i	...	x_n

Als Funktion von n Zufallsvariablen ist das arithmetische Mittel der Stichprobe, wie wir bereits gesehen haben, ebenfalls eine Zufallsvariable, und zwar

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i .$$

Die Verteilung von \bar{X} hängt davon ab, ob die Ziehung mit oder ohne Zurücklegen erfolgt. Wir wollen daher im Folgenden beide Entnahmetechniken betrachten.

1.2.2.1 Ziehen mit Zurücklegen

Bezeichnen wir das arithmetische Mittel bzw. die Varianz der Grundgesamtheit mit μ und σ^2 . Da die n Zufallsvariablen X_i wie die Grundgesamtheit verteilt sind, d.h. $E(X_1) = \dots = E(X_n) = \mu$ und $Var(X_1) = \dots = Var(X_n) = \sigma^2$ gilt, erhalten wir unter Anwendung der unter II 2.7.5 behandelten Rechenregeln für Linearkombinationen von Zufallsvariablen für den **Erwartungswert des Stichprobenmittels**

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu$$

d.h.

$$E(\bar{X}) = \mu . \quad (\text{III1: 6})$$

Der Erwartungswert von \bar{X} ist somit das unbekannte Mittel μ der Grundgesamtheit. Beachten wir zudem, dass aufgrund der Unabhängigkeit der einzelnen Züge beim Ziehen mit Zurücklegen auch die n Zufallsvariablen X_i voneinander *unabhängig* sind, so erhalten wir die **Varianz des Stichprobenmittels** als

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} n \sigma^2.$$

Das vorletzte Gleichheitszeichen resultiert darin aufgrund der getroffenen Unabhängigkeitssannahme. Eine weitere Vereinfachung liefert schließlich

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}. \quad (\text{III1: 7})$$

Für die auch als *Standardfehler des Stichprobenmittels* bezeichnete Standardabweichung des arithmetischen Mittels der Stichprobe gilt daher

$$\sigma_{\bar{x}} = \sqrt{\text{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}. \quad (\text{III1: 8})$$

Die Streuung oder der Standardfehler des Schätzers hängt also von der Standardabweichung σ der Grundgesamtheit und dem Stichprobenumfang n ab. Sie wird offensichtlich immer geringer, je größer der Stichprobenumfang n gewählt wird. Die Reduktion vollzieht sich jedoch nicht linear, sondern mit dem Faktor $1/\sqrt{n}$. So ist z.B. ein 4-facher Stichprobenumfang nötig, um die Streuung des Schätzers zu halbieren. Dieser Sachverhalt wird auch als \sqrt{n} -Gesetz bezeichnet.

Mit wachsendem Stichprobenumfang n werden sich also die Realisierungen von \bar{X} immer mehr um μ konzentrieren. Dies ist die wesentliche Aussage des **Gesetzes der großen Zahlen**. Konkret besagt es, dass für beliebig kleine $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0 \quad (\text{III1: 9})$$

gilt. Die Wahrscheinlichkeit, dass \bar{X} um einen beliebig kleinen positiven Wert ϵ von μ abweicht, geht also mit zunehmendem Stichprobenumfang n gegen Null.⁴

Von der Wahrscheinlichkeitsverteilung des Stichprobenmittels kennen wir nun Erwartungswert und Varianz. Um Aussagen über die konkrete Verteilungsform machen zu können, müssen wir die drei folgenden Fälle unterscheiden:

1. Normalverteilte Grundgesamtheit mit bekannter Varianz σ^2

Ist die Grundgesamtheit bzw. das Merkmal X in der Grundgesamtheit mit bekannter Varianz σ^2 normalverteilt, so ist die Summe $X_1 + X_2 + \dots + X_n$ der n unabhängigen Zufallsvariablen X_1, X_2, \dots, X_n aufgrund der *Reproduktionseigenschaft* der Normalverteilung (vgl. Abschnitt II 3.2.3.3) ebenfalls normalverteilt. Es gilt konkret

$$X \sim N(\mu, \sigma^2) \rightarrow X_i \sim N(\mu, \sigma^2) \rightarrow \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2). \quad (\text{III1: 10})$$

⁴ Das Gesetz der großen Zahlen besagt beim Ziehen mit Zurücklegen nicht, dass $\lim_{n \rightarrow \infty} \bar{x} = \mu$, also die Folge der beobachteten arithmetischen Mittelwerte gegen μ konvergiert. Es lässt sich nämlich nicht zu jedem kleinen $\epsilon > 0$ ein n_ϵ finden, ab welchem eine Abweichung $|\bar{x} - \mu| \geq \epsilon$ gar nicht mehr vorkäme. Vielmehr kann bei jedem noch so großen n eine Abweichung um ϵ oder mehr vorkommen. Lediglich die Wahrscheinlichkeit dafür wird immer kleiner und geht schließlich gegen Null.

Damit ist auch das Stichprobenmittel \bar{X} *normalverteilt*, da es sich von dieser Summe nur um den Faktor $1/n$ unterscheidet.

$$X \sim N(\mu, \sigma^2) \rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (\text{III1: 11})$$

2. Normalverteilte Grundgesamtheit mit unbekannter Varianz σ^2

Ist die Grundgesamtheit normalverteilt, muss ihre Varianz jedoch mittels $\hat{\sigma}^2$ bzw. (III1: 4) geschätzt werden, so ergibt sich das mit seinem Erwartungswert μ und der geschätzten Standardabweichung $\hat{\sigma}_{\bar{X}} = \hat{\sigma} / \sqrt{n}$ transformierte Stichprobenmittel als

$$T = \frac{\bar{X} - \mu}{\hat{\sigma}_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}.$$

Wir sprechen bei diesem transformierten Stichprobenmittel T häufig auch von der sog. t-Statistik. Durch diverse Umformungen erhalten wir daraus

$$T = \frac{\frac{\bar{X} - \mu}{\sigma} \sqrt{n}}{\sqrt{\frac{(n-1)\hat{\sigma}^2}{(n-1)\sigma^2}}} \quad \text{mit} \quad \frac{\bar{X} - \mu}{\sigma} = Z, \frac{(n-1)\hat{\sigma}^2}{\sigma^2} = B, n-1 = w \rightarrow T = \frac{Z}{\sqrt{\frac{B}{w}}}.$$

Z ist dabei offensichtlich standardnormalverteilt. B ist, wie wir unter III 1.2.4 sehen werden, Chi-Quadrat-verteilt mit $w = n - 1$ Freiheitsgraden. T folgt somit nach den Vorschriften aus Abschnitt II 3.3.2 einer *t-Verteilung* mit $w = n - 1$ Freiheitsgraden.

$$X \sim N(\mu, \hat{\sigma}^2) \rightarrow T = \frac{\bar{X} - \mu}{\hat{\sigma}_{\bar{X}}} \sim t(n-1) \quad \text{mit} \quad \hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{n}} \quad (\text{III1: 12})$$

3. Beliebig verteilte Grundgesamtheit

Ist die Grundgesamtheit mit μ und σ^2 beliebig verteilt, so können wir für große Stichproben auf den *Zentralen Grenzwertsatz* (vgl. Abschnitt II 3.4.1) zurückgreifen, um die Verteilung von \bar{X} abzuleiten. Bis auf den Faktor $1/n$ ist das Stichprobenmittel \bar{X} nämlich eine Summe von n unabhängigen und identisch verteilten Zufallsvariablen X_i . Mit wachsendem Stichprobenumfang n strebt somit die Verteilung von \bar{X} gegen eine Normalverteilung mit dem Erwartungswert $E(\bar{X}) = \mu$ und der Varianz $\text{Var}(\bar{X}) = \sigma^2 / n$. Man sagt, dass \bar{X} *asymptotisch normalverteilt* ist.

$$X \sim (\mu, \sigma^2) \xrightarrow{n \rightarrow \infty} \bar{X} \stackrel{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \quad (\text{III1: 13})$$

Als *Faustregel* kann angegeben werden, dass für $n > 30$ das Stichprobenmittel in guter Näherung normalverteilt ist. Abbildung III 1 verdeutlicht für vier ausgewählte Grundgesamtheiten die beschriebene Konvergenz der Verteilung des Stichprobenmittels gegen die Normalverteilung.⁵ Dabei sind in der ersten Zeile die verschiedenen Grundgesamtheiten dargestellt. Die zweite, dritte und vierte Zeile zeigt die

⁵ Die Grafik ist entnommen aus Bleymüller, J., Gehlert, G., Göllicher, H. (2004), Kapitel 13.

resultierenden Verteilungen der Stichprobenmittelwerte bei Stichprobenumfängen von $n = 2$, $n = 5$ und $n = 30$.

1.2.2.2 Ziehen ohne Zurücklegen

Wird ohne Zurücklegen gezogen, gilt für den Erwartungswert und die Varianz des Stichprobenmittels

$$E(\bar{X}) = \mu \quad (\text{III1: 14})$$

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}. \quad (\text{III1: 15})$$

Wie zu erkennen ist, unterscheiden sich also beim Ziehen mit und ohne Zurücklegen die Varianzen durch den Faktor $(N - n)/(N - 1)$, den wir bereits aus Abschnitt II 3.1.2.2 kennen und der in der induktiven Statistik auch als Korrekturfaktor für endliche Gesamtheiten oder kurz *Endlichkeitskorrekturfaktor* bezeichnet wird. Für $n \leq 0,1N$ können wir diesen wieder vernachlässigen (vgl. Abschnitt II 3.1.2.3), sodass sich die Varianzen entsprechen.

Trotz der Tatsache, dass die einzelnen Zufallsvariablen X_i beim Ziehen ohne Zurücklegen nicht unabhängig sind, wird auch hier mit wachsendem Stichprobenumfang der *Zentrale Grenzwertsatz* wirksam, sofern N im Verhältnis zu n ausreichend groß ist. Als Faustregel können wir die näherungsweise *Normalverteilung* des Stichprobenmittels für $n > 30$ und $n \leq 0,5N$ annehmen.

$$\left. \begin{array}{l} X \sim (\mu, \sigma^2) \\ n > 30 \\ n \leq 0,5N \end{array} \right\} \rightarrow \bar{X} \stackrel{a}{\sim} N\left(\mu, \frac{\sigma^2}{n} \frac{N-n}{N-1}\right) \quad (\text{III1: 16})$$

Beispiel III 1-4

Verteilung des Stichprobenmittelwerts

Die von einer Unternehmung hergestellten Energiesparlampen haben eine durchschnittliche Lebensdauer von $\mu = 900$ Stunden bei einer Standardabweichung von $\sigma = 50$ Stunden. Aus der laufenden Massenproduktion wird eine Stichprobe im Umfang von $n = 35$ Stück entnommen. Wie groß ist die Wahrscheinlichkeit dafür, dass die Stichprobe eine durchschnittliche Lebensdauer \bar{x} von weniger als 880 Stunden liefert?

Das Stichprobenmittel \bar{X} ist annähernd normalverteilt, da (III1: 16) erfüllt ist. Es ist $n > 30$ und außerdem kann $n \leq 0,5N$ als erfüllt angesehen werden, da bei einer Massenproduktion von einem sehr großen N ausgegangen werden kann. Aufgrund des großen N ist auch die Bedingung $n \leq 0,1N$ für die Vernachlässigung des Endlichkeitskorrekturfaktors erfüllt und wir können wir auf diesen verzichten, sodass

$$\bar{X} \stackrel{a}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(900; \frac{50^2}{35}\right) = N\left(900; \frac{50^2}{35}\right) = N(900; 71,43)$$

gilt. Die gesuchte Wahrscheinlichkeit erhalten wir daher als

$$P(\bar{X} < 880) \approx F_N(880) = \Phi\left(\frac{880 - 900}{\sqrt{71,43}}\right) = \Phi(-2,37) = 1 - 0,9911 = 0,0089.$$



Verteilung der Grundgesamtheit

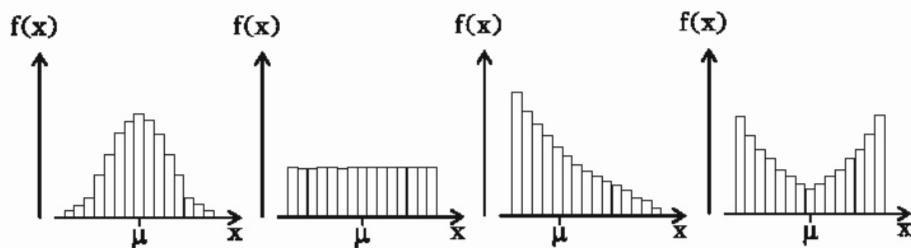
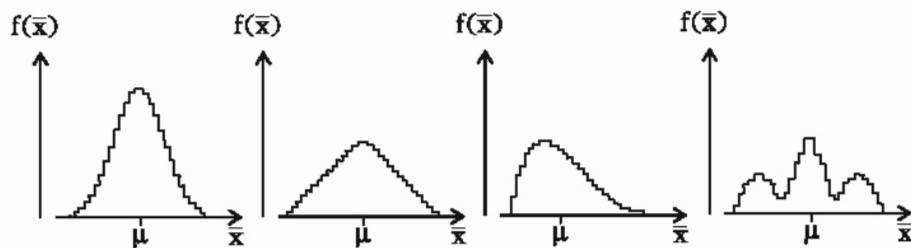
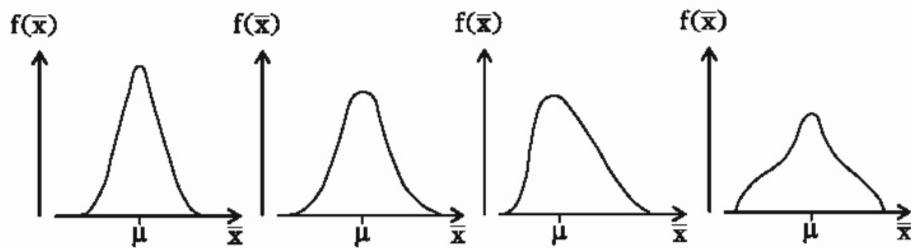
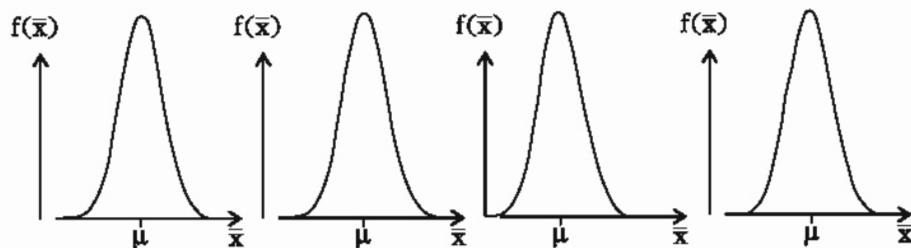
Verteilung des Stichprobenmittels für $n = 2$ Verteilung des Stichprobenmittels für $n = 5$ Verteilung des Stichprobenmittels für $n = 30$ 

Abbildung III 1: Verteilung des Stichprobenmittels

1.2.3 Verteilung des Stichprobenanteilswerts

Zur Herleitung der Stichprobenverteilung des Anteilswerts bedienen wir uns eines *Urnenmodells für ein dichotomes Merkmal*. Nehmen wir an, in einer Urne (Grundgesamtheit) befinden sich N Kugeln, von denen M eine bestimmte Eigenschaft A (z.B. Farbe) aufweisen. Die restlichen $N - M$ Kugeln besitzen diese Eigenschaft nicht. Der Anteil der Kugeln mit der Eigenschaft A ist mit $p = M/N$ bekannt. Es werden nun mehrere Stichproben mit einem Umfang von n gezogen. Enthält die Stichprobe x Kugeln mit der Eigenschaft A, ist der Stichprobenanteilswert $\hat{p} = x/n$. Jede weitere Stichprobe im Umfang n kann einen anderen (zufallsabhängigen) Stichprobenanteilswert liefern, sodass \hat{p} als Realisation einer Zufallsvariablen $\hat{p} = X/n$: "Stichprobenanteilswert" aufgefasst werden kann. \hat{p} gibt also den Anteil der Kugeln mit Eigenschaft A in der Stichprobe an. Je nach Technik der Entnahme ist die Wahrscheinlichkeitsverteilung dieser Zufallsvariablen \hat{p} unterschiedlich.

1.2.3.1 Ziehen mit Zurücklegen

Für den Fall der Stichprobenentnahme *mit Zurücklegen* ist die Zufallsvariable X : "Anzahl der Kugeln mit Eigenschaft A in der Stichprobe", also

$$X = \sum_{i=1}^n X_i \quad \text{mit} \quad X_i = \begin{cases} 1 & \text{falls A eintritt} \\ 0 & \text{falls A nicht eintritt,} \end{cases}$$

binomialverteilt. Für Erwartungswert und Varianz dieser Zufallsvariablen X bzw. der Merkmalssumme können wir daher

$$E(X) = np$$

$$\text{Var}(X) = np(1-p)$$

festhalten. Da sich der Anteilswert \hat{p} aus dem Zusammenhang $\hat{p} = X/n$ ergibt, lassen sich **Erwartungswert und Varianz** der Zufallsvariablen \hat{p} : "Anteil der Kugeln mit Eigenschaft A in der Stichprobe", d.h. des **Sichprobenanteilswerts**, wie folgt bestimmen:

$$E(\hat{p}) = E\left(\frac{1}{n} X\right) = \frac{1}{n} E(X) = p \quad (\text{III1: 17})$$

$$\text{Var}(\hat{p}) = \text{Var}\left(\frac{1}{n} X\right) = \frac{1}{n^2} \text{Var}(X) = \frac{p(1-p)}{n} \quad (\text{III1: 18})$$

Die Standardabweichung von \hat{p} wird als *Standardfehler des Stichprobenanteilswertes* bezeichnet und ergibt sich zu

$$\sigma_{\hat{p}} = \sqrt{\text{Var}(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}. \quad (\text{III1: 19})$$

Sie folgt wie jene des Stichprobenmittels dem \sqrt{n} -Gesetz. Als Spezialfall des *Ge setzes der großen Zahlen* (vgl. Abschnitt III 1.2.2.1) erhalten wir für $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\hat{p} - p| \geq \epsilon) = 0.$$

Dies bedeutet, dass auch für den Stichprobenanteilswert die Wahrscheinlichkeit, dass der Schätzer \hat{p} um ein beliebig kleines positives ϵ vom wahren Wert p ab-

weicht, mit zunehmendem Stichprobenumfang n gegen Null konvergiert. Konsequenz daraus ist, dass die unbekannte Wahrscheinlichkeit $P(A) = p$ eines Ereignisses A durch die relative Häufigkeit \hat{p} in einfachen Zufallsstichproben genügend großen Umfangs näherungsweise bestimmt werden kann. Dieser Spezialfall des Gesetzes der großen Zahlen ist uns bereits im Abschnitt II 1.3.3 begegnet und war die Rechtfertigung für die Verwendung statistischer Wahrscheinlichkeiten.

Ist die im Abschnitt II 3.4.2.1 behandelte Approximationsbedingung $np(1-p) \geq 9$ erfüllt, so kann die Verteilung von \hat{p} auf Basis des *Zentralen Grenzwertsatzes* (vgl. Abschnitt II 3.4.1) durch eine *Normalverteilung* mit den Parametern

$$\mu = E(\hat{p}) = p$$

$$\sigma^2 = \text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

approximiert werden. Kurz können wir daher

$$\left. \begin{array}{l} \hat{p} = \frac{x}{n} \text{ mit } X \sim \text{binomialverteilt} \\ np(1-p) \geq 9 \end{array} \right\} \rightarrow \hat{p} \stackrel{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \quad (\text{III1: 20})$$

festhalten. Da hier eine diskrete durch eine stetige Verteilung approximiert wird, schlagen manche Autoren bei der Wahrscheinlichkeitsberechnung eine *Stetigkeitskorrektur* vor.⁶ Da diese Stetigkeitskorrektur aufgrund der Anwendung des Zentralen Grenzwertsatzes in den meisten Fällen nur einen sehr geringen Wert aufweist, können wir sie im Folgenden vernachlässigen. Der Anteilswert wird also wie eine stetige Variable behandelt.

Beispiel III 1-5

Stichprobenanteilswert (Ziehen mit Zurücklegen)

Im Land Amazonia beträgt der Anteil p weiblicher Einwohner 90 %. Wie groß ist die Wahrscheinlichkeit, dass der Frauenanteil \hat{p} in einer Stichprobe von $n = 500$ größer als 92 % ist?

Da hier die Approximationsbedingung $np(1-p) \geq 9$ erfüllt ist, ist \hat{p} wie folgt verteilt:

$$\hat{p} \stackrel{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right) = N(0,9; 0,00018)$$

Unter Vernachlässigung der hier sehr geringen Stetigkeitskorrektur erhalten wir so für die gesuchte Wahrscheinlichkeit

$$P(\hat{p} > 0,92) \approx 1 - F_N(0,92) = 1 - \Phi\left(\frac{0,92 - 0,9}{\sqrt{0,00018}}\right) = 1 - \Phi(1,49) = 1 - 0,9319 = 0,0681.$$

1.2.3.2 Ziehen ohne Zurücklegen

Bei Stichprobenentnahme *ohne Zurücklegen* liegt das Modell der *Hypergeometrischen Verteilung* vor. Beschreibt die Zufallsvariable X wieder die Anzahl der Kugeln mit Eigenschaft A in der Stichprobe, gilt für Erwartungswert und Varianz

⁶ Diese Stetigkeitskorrektur hat z.B. die Form $P(\hat{p} \leq b) \approx F_N(b + \frac{1}{2n})$. Für eine nähere Behandlung vgl. z.B. Bleymüller, J., Gehlert, G., Göllicher, H. (2004), Kapitel 12.

$$E(X) = n \cdot \frac{M}{N} = np$$

$$\text{Var}(X) = np(1-p) \cdot \frac{N-n}{N-1}.$$

Da sich \hat{p} wieder aus dem Zusammenhang $\hat{p} = X / n$ ergibt, errechnen sich **Erwartungswert und Varianz des Stichprobenanteilswertes \hat{p}** wie folgt:

$$E(\hat{p}) = \frac{1}{n} E(X) = p \quad (\text{III1: 21})$$

$$\text{Var}(\hat{p}) = \frac{1}{n^2} \text{Var}(X) = \frac{p(1-p)}{n} \frac{N-n}{N-1} \quad (\text{III1: 22})$$

Wie bereits beim Stichprobenmittel unterscheiden sich die Varianzen beim Ziehen mit und ohne Zurücklegen durch den *Endlichkeitskorrekturfaktor*, den wir wieder für $n \leq 0,1N$ vernachlässigen können. Die Varianz des Stichprobenanteilswertes reduziert sich dann zu (III1: 18). Wie wir in Abschnitt II 3.4.2.3 außerdem gesehen haben, kann die Hypergeometrische Verteilung bei $np(1-p) \geq 9$ und $n \leq 0,1N$ durch eine *Normalverteilung* approximiert werden. Daher lässt sich bei diesen erfüllten Bedingungen auch die Stichprobenverteilung des Anteilswertes \hat{p} durch eine Normalverteilung mit folgenden Parametern *approximieren*:

$$\mu = E(\hat{p}) = p$$

$$\sigma^2 = \text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

Da unsere Approximationsbedingung auch die Bedingung für die Vernachlässigung des Endlichkeitskorrekturfaktors beinhaltet, kann er vernachlässigt werden. *Der Fall des Ziehens ohne Zurücklegen kann also durch den des Ziehens mit Zurücklegen approximiert werden.*

Beispiel III 1-6

Stichprobenanteilswert (Ziehen ohne Zurücklegen)

In einem Fertigungslos von $N = 10.000$ Stück liegt der Anteil fehlerhafter Produkte bei $p = 20\%$. Es wird eine Stichprobe von $n = 500$ Stück ohne Zurücklegen gezogen. Wie groß ist nun die Wahrscheinlichkeit dafür, dass der Anteil der mangelhaften Produkte in der Stichprobe zwischen 19,8 % und 20,5 % liegt?

Da die Voraussetzungen $np(1-p) \geq 9$ und $n \leq 0,1N$ erfüllt sind, kann die Verteilung des Stichprobenanteilswertes unter Vernachlässigung der Endlichkeitskorrektur durch die Normalverteilung

$$N(E(\hat{p}), \text{Var}(\hat{p})) = N\left(p, \frac{p(1-p)}{n}\right) = N\left(0,2, \frac{0,2 \cdot 0,8}{500}\right) = N(0,2; 0,00032)$$

approximiert werden. Für die gesuchte Wahrscheinlichkeit gilt unter Vernachlässigung der Stetigkeitskorrektur

$$\begin{aligned} P(0,198 \leq \hat{p} \leq 0,205) &\approx F_N(0,205) - F_N(0,198) = \Phi\left(\frac{0,205 - 0,2}{\sqrt{0,00032}}\right) - \Phi\left(\frac{0,198 - 0,2}{\sqrt{0,00032}}\right) \\ &= \Phi(0,28) - \Phi(-0,11) = 0,6103 - (1 - 0,5438) = 0,1541. \end{aligned}$$

1.2.4 Verteilung der Stichprobenvarianz

Zur Bestimmung der Verteilung der Stichprobenvarianz $\hat{\sigma}^2$ unterstellen wir im Folgenden die *Unabhängigkeit* der Ergebnisse der einzelnen Stichprobenziehungen (d.h. *Ziehen mit Zurücklegen*). Zudem nehmen wir an, dass die Grundgesamtheit mit dem Mittelwert μ und der Varianz σ^2 *annähernd normalverteilt* ist. Unter diesen Voraussetzungen sind die n Zufallsvariablen X_i (Ergebnis des i -ten Zugs) alle unabhängig normalverteilt mit $E(X_i) = \mu$ ($i = 1, \dots, n$) und $\text{Var}(X_i) = \sigma^2$ ($i = 1, \dots, n$). Als Funktion von n Zufallsvariablen X_i ist die Stichprobenvarianz

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

ebenfalls eine Zufallsvariable.

Zur Herleitung der Verteilung der Stichprobenvarianz betrachten wir zunächst die n unabhängigen standardnormalverteilten Zufallsvariablen

$$Z_i = \frac{X_i - \mu}{\sigma},$$

die wir quadrieren und aufsummieren, sodass sich

$$\chi^2 = \sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}$$

ergibt. In Abschnitt II 3.3.1 haben wir gesehen, dass eine solche Summe quadrierter unabhängiger Zufallsvariablen Z_i , die mit dem Erwartungswert 0 und der Varianz 1 normalverteilt sind, einer Chi-Quadrat-Verteilung mit $w = n$ Freiheitsgraden gehorcht.

Für eine Summe B , die aus χ^2 hervorgeht, indem wir μ durch \bar{X} ersetzen, also

$$B = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2},$$

kann gezeigt werden, dass sie ebenfalls einer Chi-Quadrat-Verteilung folgt, allerdings mit nur $w = n - 1$ Freiheitsgraden. Formen wir die Definition der Stichprobenvarianz in der Form

$$\sum_{i=1}^n (X_i - \bar{X})^2 = (n-1) \cdot \hat{\sigma}^2$$

um und setzen diesen Term in den Ausdruck für B ein, so erhalten wir

$$B = \frac{(n-1) \cdot \hat{\sigma}^2}{\sigma^2}. \quad (\text{III1: 23})$$

Diese Zufallsvariable B gehorcht einer *Chi-Quadrat-Verteilung* mit $w = n - 1$ Freiheitsgraden. Auf ein konkretes Anwendungsbeispiel wollen wir an dieser Stelle verzichten, da sich (III1: 23) erst im Rahmen von Konfidenzintervallen und Hypothesentests als besonders hilfreich herausstellt.

1.2.5 Verteilung weiterer Stichprobengrößen

Neben den Stichprobenverteilungen der Schätzer $\hat{\mu}$, \hat{p} und $\hat{\sigma}^2$ sind vor allem für die Durchführung von Hypothesentests die Verteilungen der Differenz von Stichprobenmitteln bzw. von Stichprobenanteilswerten sowie der Quotienten von Stichprobenvarianzen von Bedeutung. Wir werden sie daher im Folgenden herleiten.

1.2.5.1 Differenz zweier Stichprobenmittel

Werden zwei *unabhängige* Zufallsstichproben des Umfangs n_1 und n_2 gezogen und die Stichprobenmittel \bar{X}_1 und \bar{X}_2 erfasst, so kann deren Differenz $D = \bar{X}_1 - \bar{X}_2$ als Realisation einer Zufallsvariablen $D = \bar{X}_1 - \bar{X}_2$ interpretiert werden. Zur Herleitung der Verteilung dieser Differenz D zweier Stichprobenmittel rufen wir uns zunächst ins Gedächtnis, dass die Differenz zweier unabhängiger Zufallsvariablen X und Y nach den in Kapitel II 2.7.5 behandelten Vorschriften für Linearkombinationen von Zufallsvariablen einen Erwartungswert von $E(X) - E(Y)$ und eine Varianz von $Var(X) + Var(Y)$ aufweist. Da die Stichproben aus beiden Grundgesamtheiten hier voneinander unabhängig sind, handelt es sich bei der Differenz $D = \bar{X}_1 - \bar{X}_2$ der beiden Stichprobenmittelwerte \bar{X}_1 und \bar{X}_2 , genau um eine solche Form von Linearkombination. Wir erhalten daher unter Berücksichtigung der Erwartungswerte und Varianzen beider Stichprobenmittelwerte

$$E(\bar{X}_1) = \mu_1 \quad \text{Var}(\bar{X}_1) = \frac{\sigma_1^2}{n_1}$$

$$E(\bar{X}_2) = \mu_2 \quad \text{Var}(\bar{X}_2) = \frac{\sigma_2^2}{n_2}$$

den **Erwartungswert** und die **Varianz** der **Differenz zweier Stichprobenmittelwerte** $D = \bar{X}_1 - \bar{X}_2$ folgendermaßen:

$$E(D) = E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2 \quad (\text{III1: 24})$$

$$\text{Var}(D) = \text{Var}(\bar{X}_1 - \bar{X}_2) = \text{Var}(\bar{X}_1) + \text{Var}(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \quad (\text{III1: 25})$$

Die Standardabweichung bzw. der **Standardfehler der Differenz der Stichprobenmittelwerte** beträgt damit

$$\sigma_D = \sqrt{\text{Var}(D)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \quad (\text{III1: 26})$$

Treffen wir nun noch die Annahmen, dass beide Grundgesamtheiten mit den Mittelwerten μ_1 und μ_2 und den Varianzen σ_1^2 und σ_2^2 *annähernd normalverteilt* sind und dass sofern dies nicht gilt, n_1 und n_2 so groß sind, dass jeweils der zentrale Grenzwertsatz greift, so gilt für die Verteilung von D näherungsweise

$$D \xrightarrow{a} N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right). \quad (\text{III1: 27})$$

Beispiel III 1-7**Differenz von Stichprobenmitteln**

Eine bestimmte Reifensorte 1 hat eine durchschnittliche Laufleistung von $\mu_1 = 50.000$ km bei einer Standardabweichung von $\sigma_1 = 5.000$ km. Die durchschnittliche Laufleistung einer Reifensorte 2 liegt bei $\mu_2 = 48.000$ km bei einer Standardabweichung von $\sigma_2 = 4.500$ km. Aus der laufenden Produktion werden nun zufällig $n_1 = 40$ und $n_2 = 50$ Reifen entnommen und einem Laufleistungstest (Test der Materialbelastbarkeit) unterzogen.

Wie groß ist die Wahrscheinlichkeit dafür, dass das arithmetische Mittel der Laufleistung der 40 Reifen der Sorte 1 um mehr als 3.000 km größer ist als das arithmetische Mittel der 50 Reifen der Sorte 2?

Für die Verteilung der Differenz der arithmetischen Mittel $D = \bar{X}_1 - \bar{X}_2$ gilt hier

$$D \stackrel{a}{=} N(E(D), \text{Var}(D)) = N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) = N(2.000; 1.030.000).$$

Die gesuchte Wahrscheinlichkeit ergibt sich zu

$$P(D > 3.000) \approx 1 - F_N(3.000) = 1 - \Phi\left(\frac{3.000 - 2.000}{\sqrt{1.030.000}}\right) = 1 - \Phi(0,99) = 1 - 0,8389 = 0,1611.$$

1.2.5.2 Differenz zweier Stichprobenanteilsätze

Zur Untersuchung der Stichprobenverteilung der Differenz $D = \hat{p}_1 - \hat{p}_2$ zweier Stichprobenanteilsätze \hat{p}_1 und \hat{p}_2 , können wir analog zum Abschnitt III 1.2.5.1 vorgehen. Gehen wir von zwei *unabhängigen* Zufallsstichproben der Umfänge n_1 und n_2 aus, so gilt unter Verwendung der Werte

$$E(\hat{p}_1) = p_1 \quad \text{Var}(\hat{p}_1) = \frac{p_1(1-p_1)}{n_1}$$

$$E(\hat{p}_2) = p_2 \quad \text{Var}(\hat{p}_2) = \frac{p_2(1-p_2)}{n_2},$$

die wir aus Abschnitt III 1.2.3.1 kennen, für den **Erwartungswert** und die **Varianz** der **Differenz zweier Stichprobenanteilsätze** $D = \hat{p}_1 - \hat{p}_2$

$$E(D) = E(\hat{p}_1 - \hat{p}_2) = E(\hat{p}_1) - E(\hat{p}_2) = p_1 - p_2 \quad (\text{III1: 28})$$

$$\text{Var}(D) = \text{Var}(\hat{p}_1 - \hat{p}_2) = \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}. \quad (\text{III1: 29})$$

Den *Standardfehler der Differenz der Stichprobenanteilsätze* können wir damit wie folgt festhalten:

$$\sigma_D = \sqrt{\text{Var}(D)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad (\text{III1: 30})$$

Nehmen wir nun wieder an, dass die zugehörigen Stichprobenumfänge n_1 und n_2 so groß sind, dass die Stichprobenanteilsätze \hat{p}_1 und \hat{p}_2 als *annähernd normalverteilt* angesehen werden können, so ist D näherungsweise normalverteilt, d.h.

$$D \stackrel{a}{\sim} N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right). \quad (\text{III1: 31})$$

Als *Faustregel* für eine Rechtfertigung dieser Approximation sollte $n_1 p_1 (1-p_1) \geq 9$ und $n_2 p_2 (1-p_2) \geq 9$ erfüllt sein.

Beispiel III 1-8

Differenz von Stichprobenanteilswerten

Der Anteil der Stipendiaten beträgt an einer Hochschule $p_1 = 0,15$ und an einer anderen $p_2 = 0,10$. Wie hoch ist die Wahrscheinlichkeit dafür, dass die Differenz $D = \hat{p}_1 - \hat{p}_2$ in zwei unabhängigen Stichproben vom Umfang $n_1 = n_2 = 100$ zwischen 0,05 und 0,15 liegt?

Für die Verteilung von D gilt

$$D \stackrel{a}{\sim} N(E(D), \text{Var}(D)) = N\left(0,15 - 0,10; \frac{0,15 \cdot 0,85}{100} + \frac{0,10 \cdot 0,90}{100}\right) = N(0,05; 0,0022).$$

Die gesuchte Wahrscheinlichkeit kann damit wie folgt bestimmt werden:

$$\begin{aligned} P(0,05 < D < 0,15) &\approx F_N(0,15) - F_N(0,05) = \Phi\left(\frac{0,15 - 0,05}{\sqrt{0,0022}}\right) - \Phi\left(\frac{0,05 - 0,05}{\sqrt{0,0022}}\right) \\ &= \Phi(2,13) - \Phi(0) = 0,9834 - 0,5 = 0,4834 \end{aligned}$$

1.2.5.3 Quotient zweier Stichprobenvarianzen

Werden zwei *unabhängige* Stichproben des Umfangs n_1 und n_2 aus zwei *normalverteilten* Grundgesamtheiten mit den Varianzen σ_1^2 und σ_2^2 gezogen, so sind die beiden Zufallsvariablen

$$B_1 = \frac{(n_1 - 1)\hat{\sigma}_1^2}{\sigma_1^2} \quad B_2 = \frac{(n_2 - 1)\hat{\sigma}_2^2}{\sigma_2^2},$$

wie wir unter III 1.2.4 gesehen haben, mit $w_1 = n_1 - 1$ bzw. $w_2 = n_2 - 1$ Freiheitsgraden Chi-Quadrat-verteilt. Es lässt sich zeigen, dass ein *Quotient F*, der sich aus diesen beiden unabhängigen, Chi-Quadrat-verteilten Zufallsvariablen B_1 und B_2 mit w_1 bzw. w_2 ergibt und die Form

$$F = \frac{B_1 / w_1}{B_2 / w_2}$$

aufweist, einer $F(w_1, w_2)$ -Verteilung gehorcht (vgl. Abschnitt II 3.3.3). Setzen wir in die Formel für F die Ausdrücke für B_1 und B_2 ein, so erhalten wir als andere Darstellung für F den Ausdruck

$$F = \frac{\frac{(n_1 - 1)\hat{\sigma}_1^2}{\sigma_1^2 w_1}}{\frac{(n_2 - 1)\hat{\sigma}_2^2}{\sigma_2^2 w_2}} = \frac{\frac{(n_1 - 1)\hat{\sigma}_1^2}{\sigma_1^2(n_1 - 1)}}{\frac{(n_2 - 1)\hat{\sigma}_2^2}{\sigma_2^2(n_2 - 1)}} = \frac{\hat{\sigma}_1^2 / \sigma_1^2}{\hat{\sigma}_2^2 / \sigma_2^2}, \quad (\text{III1: 32})$$

den wir im Zusammenhang mit den Hypothesentests im Abschnitt III 3 noch benötigen werden. Abbildung III 2 gibt einen abschließenden Überblick über die wichtigsten Stichprobengrößen und deren Verteilungen.

ZV	Stichprobenverteilung	Parameter
\bar{X}	$N(E(\bar{X}), \text{Var}(\bar{X}))$ <i>wenn:</i> Grundgesamtheit normalverteilt oder $n > 30$	$E(\bar{X}) = \mu$ Z.m.Z.: $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$ Z.o.Z.: $\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{n-1}$
$T = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$ mit $\sigma_{\bar{X}} = \frac{\delta}{\sqrt{n}}$	$t(w)$ <i>wenn:</i> Grundgesamtheit normalverteilt $N(0; 1)$ <i>wenn:</i> $n > 30$ $N(E(\hat{p}), \text{Var}(\hat{p}))$ <i>wenn:</i> $np(1-p) \geq 9$	$w = n - 1$ $E(\hat{p}) = p$ Z.m.Z.: $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$ Z.o.Z.: $\text{Var}(\hat{p}) = \frac{p(1-p)}{n} \cdot \frac{N-n}{n-1}$
\hat{p}		$w = n - 1$
$B = \frac{(n-1)\hat{\sigma}^2}{\sigma^2}$	$\chi^2(w)$ <i>wenn:</i> Grundgesamtheit normalverteilt	$E(D) = \mu_1 - \mu_2$ Var(D) = $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
$D = \bar{X}_1 - \bar{X}_2$	$N(E(D), \text{Var}(D))$ <i>wenn:</i> Grundgesamtheiten normalverteilt oder $n_1 > 30$ und $n_2 > 30$	für Z.m.Z. und auch Z.o.Z. wenn $n_1 \leq 0,1N_1$ und $n_2 \leq 0,1N_2$ $E(D) = p_1 - p_2$ Var(D) = $\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$
$D = \hat{p}_1 - \hat{p}_2$	$N(E(D), \text{Var}(D))$ <i>wenn:</i> $n_1 p_1 (1-p_1) \geq 9$ und $n_2 p_2 (1-p_2) \geq 9$	für Z.m.Z. und auch Z.o.Z. wenn $n_1 \leq 0,1N_1$ und $n_2 \leq 0,1N_2$ $w_1 = n_1 - 1; w_2 = n_2 - 1$
$F = \frac{\hat{\sigma}_1^2 / \sigma_1^2}{\hat{\sigma}_2^2 / \sigma_2^2}$	$F(w_1, w_2)$ <i>wenn:</i> Grundgesamtheiten normalverteilt	

Hinweise: 1. Der Endlichkeitskorrekturfaktor kann für $n \leq 0,1N$ vernachlässigt werden.
 2. Z.m.Z. = Ziehen mit Zurücklegen, Z.o.Z. = Ziehen ohne Zurücklegen

Abbildung III 2: Wichtige Stichprobenverteilungen im Überblick

1.3 Güte von Schätzern

Bei der praktischen Durchführung einer Punktschätzung müssen wir damit rechnen, dass der Schätzwert $\hat{\theta}$ vom wahren Wert ϑ mehr oder weniger stark abweicht. Meistens wird es sogar so sein, dass $\hat{\theta} \neq \vartheta$ gilt und nur selten $\hat{\theta} = \vartheta$. Der sog. **Schätzfehler** ist in diesem Zusammenhang definiert als $\hat{\theta} - \vartheta$. Da der Schätzer $\hat{\theta}$ eine Zufallsvariable ist, ist auch der Schätzfehler eine Zufallsgröße. Er zählt zu den wesentlichen Kriterien zur Beurteilung der Qualität eines Schätzers $\hat{\theta}$.

Konkret verwenden wir zur Einstufung der Güte eines Schätzers (bzw. einer Schätzfunktion) einen Katalog von vier wünschenswerten Eigenschaften. Dazu zählen Erwartungstreue (oder asymptotische Erwartungstreue), Effizienz, Konsistenz und ein geringer mittlerer quadratischer Fehler.

1.3.1 Erwartungstreue

Ein Schätzer $\hat{\theta}$ wird dann als **erwartungstreu** (unverzerrt, engl. unbiased) bezeichnet, wenn sein Erwartungswert $E(\hat{\theta})$ bei jedem beliebigen Stichprobenumfang mit dem wahren Parameter ϑ übereinstimmt.

$$E(\hat{\theta}) = \vartheta \quad (\text{III1: 33})$$

Eine solche Eigenschaft ist wünschenswert, da eine Schätzformel im Durchschnitt ihrer Anwendungen den gesuchten Wert treffen sollte. Eine systematische Über- oder Unterschätzung ist zu vermeiden.

Zur Veranschaulichung dieser wesentlichen Eigenschaft von Schätzern betrachten wir Abbildung III 3. Sie enthält die Verteilungen dreier Schätzer $\hat{\theta}_1$, $\hat{\theta}_2$ und $\hat{\theta}_3$. Von diesen Schätzern können nur die Schätzer $\hat{\theta}_1$ und $\hat{\theta}_2$ als erwartungstreu bezeichnet werden, da sich ihre Erwartungswerte $E(\hat{\theta}_1)$ und $E(\hat{\theta}_2)$ mit dem Parameter ϑ der Grundgesamtheit decken. Der Schätzer $\hat{\theta}_3$ liefert im Durchschnitt einen Wert größer als ϑ und gilt daher nicht als erwartungstreu sondern als verzerrt.

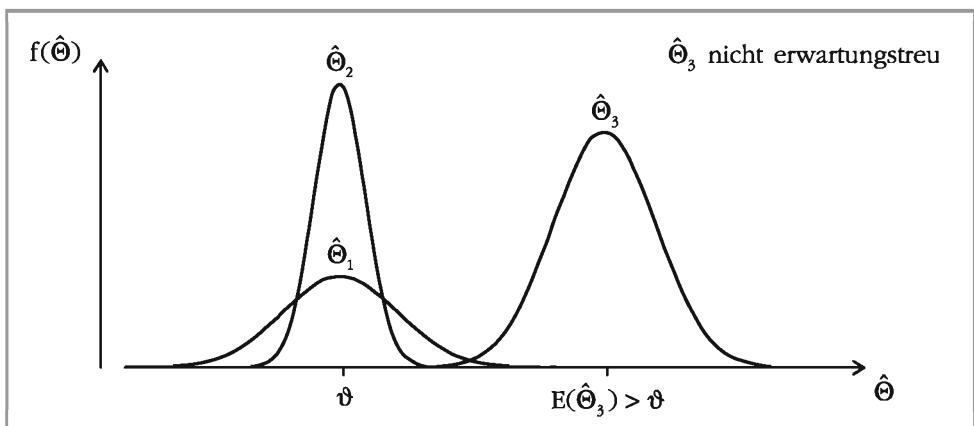


Abbildung III 3: Erwartungstreue und verzerrte Schätzer

Wir wollen nun im Folgenden drei ausgewählte Schätzer hinsichtlich ihrer Erwartungstreue näher betrachten.

- Das Stichprobenmittel $\hat{\mu} = \bar{X}$ ist ein erwartungstreuer Schätzer für das arithmetische Mittel μ der Grundgesamtheit, da

$$E(\hat{\mu}) = \mu$$

gilt. Dies haben wir bereits unter III 1.2.2.1 gezeigt.

- Als Spezialfall des Stichprobenmittels ist natürlich auch der Stichprobenanteilswert \hat{p} eine erwartungstreue Schätzfunktion für den Anteilswert p der Grundgesamtheit. Es gilt nämlich, wie wir bereits unter III 1.2.3.1 sahen

$$E(\hat{p}) = p.$$

- Die Stichprobenvarianz $\hat{\sigma}^2$ bzw. (III1: 4) ist (beim Ziehen mit Zurücklegen) ein *erwartungstreuer* Schätzer für die Varianz σ^2 der Grundgesamtheit,⁷ da

$$E(\hat{\sigma}^2) = \sigma^2.$$

Die mittlere quadratische Abweichung bezüglich \bar{X} , d.h. S^2 bzw. (III1: 5), ist hingegen *nicht erwartungstreu* (erfüllt aber die unter III 1.3.2 beschriebene Eigenschaft der *asymptotischen Erwartungstreue*).⁸ Es lässt sich nämlich

$$E(S^2) = \frac{n-1}{n} \sigma^2$$

zeigen. Man sagt in diesem Fall, dass die Punktschätzung um den Faktor $(n-1)/n$ *verzerrt* ist. Diese Schätzfunktion würde nämlich die Größe σ^2 systematisch unterschätzen. Da die Verzerrung hiermit bekannt ist, können wir sie leicht korrigieren und zwar durch Multiplikation von S^2 mit dem Kehrwert der Verzerrung, also $n/(n-1)$. Wir erhalten so den erwartungstreuen Schätzer $\hat{\sigma}^2$. Abschließend ist festzuhalten, dass $\hat{\sigma}^2$ zwar ein erwartungstreuer Schätzer für σ^2 ist, jedoch $\hat{\sigma}$ keinesfalls ein erwartungstreuer Schätzer für σ ist. Vielmehr kann gezeigt werden, dass $E(\hat{\sigma}) < \sigma$, $\hat{\sigma}$ also σ systematisch unterschätzt.⁹

1.3.2 Asymptotische Erwartungstreue

Ist ein Schätzer nicht erwartungstreu, könnte es zumindest sein, dass die Verzerrung, die er aufweist, mit zunehmendem Stichprobenumfang n immer geringer wird. Ein Schätzer wird also als **asymptotisch erwartungstreu** bezeichnet, wenn

$$\lim_{n \rightarrow \infty} E(\hat{\Theta}) = \vartheta. \quad (\text{III1: 34})$$

In praktischen Anwendungen ist es von Bedeutung, wie schnell diese Konvergenz erfolgt. Sind sehr hohe Stichprobenumfänge notwendig, um sie zu erreichen, ist der Schätzer weniger nützlich.

⁷ Den Beweis hierfür überlassen wir in Aufgabe PS1 dem Leser.

⁸ Die mittlere quadratische Abweichung bezüglich μ wäre hingegen bei bekanntem arithmetischen Mittel μ der Grundgesamtheit ein erwartungstreuer Schätzer für σ^2 .

⁹ Auer, B.R., Seitz, F. (2008) gehen auf dieses Problem im Kontext von Investmentfondsperformance ein und zeigen außerdem einen alternativen, erwartungstreuen Schätzer.

1.3.3 Effizienz

Ein erwartungstreuer Schätzer ist dann *effizienter* als ein anderer ebenfalls erwartungstreuer Schätzer, wenn er eine kleinere Varianz besitzt. So ist etwa in Abbildung III 3 der Schätzer $\hat{\Theta}_2$ effizienter als $\hat{\Theta}_1$.¹⁰ Unter sämtlichen erwartungstreuen Schätzern ist derjenige *absolut effizient* oder kurz **effizient** (oder wirksamster), der die kleinste Varianz aufweist. Ein Schätzer $\hat{\Theta}$ ist also dann effizient, wenn sowohl

$$E(\hat{\Theta}) = \vartheta \quad (\text{III1: 35})$$

als auch

$$\text{Var}(\hat{\Theta}) \leq \text{Var}(\hat{\Theta}^e) \quad (\text{III1: 36})$$

erfüllt sind, wobei in (III1: 36) der Ausdruck $\hat{\Theta}^e$ für jeden beliebigen anderen erwartungstreuen Schätzer steht.

Die Suche des effizienten Schätzers unter allen erwartungstreuen Schätzern ist ein relativ schwieriges Problem, mit dem wir uns hier nicht näher beschäftigen. Wir führen stattdessen an dieser Stelle nur einige einschlägige Erkenntnisse auf:

- Das Stichprobenmittel $\hat{\mu} = \bar{X}$ ist für den Erwartungswert μ der Grundgesamtheit der effiziente Schätzer.
- Der Stichprobenanteilswert \hat{p} ist der effiziente Schätzer für den Anteilswert p einer dichotomen Grundgesamtheit.
- Die mittlere quadratische Abweichung bezüglich μ bzw. die Stichprobenvarianz S^2 , ist der effiziente Schätzer für die Varianz σ^2 einer $N(\mu, \sigma^2)$ -verteilten Grundgesamtheit, wenn μ bekannt bzw. μ unbekannt ist.

1.3.4 Konsistenz

Eine Folge von Schätzern $\hat{\Theta}_n$ gemäß

$$\hat{\Theta}_1 = g_1(X_1), \hat{\Theta}_2 = g_2(X_1, X_2), \dots, \hat{\Theta}_n = g_n(X_1, \dots, X_n)$$

wird als **konsistent** in Bezug auf den Parameter ϑ bezeichnet, wenn die Wahrscheinlichkeit, mit der der Schätzfehler $|\hat{\Theta}_n - \vartheta|$ ein vorgegebenes kleines Niveau $\varepsilon > 0$ übersteigt (oder mit ihm identisch ist), mit zunehmendem Stichprobenumfang n gegen Null strebt, d.h.

$$\lim_{n \rightarrow \infty} P(|\hat{\Theta}_n - \vartheta| \geq \varepsilon) = 0. \quad (\text{III1: 37})$$

Anders formuliert können wir auch sagen, dass für einen konsistenten Schätzer durch die Wahl eines hinreichend großen Stichprobenumfangs die Wahrscheinlichkeit, dass der Schätzer $\hat{\Theta}_n$ ein vorgegebenes kleines Intervall $[\vartheta - \varepsilon; \vartheta + \varepsilon]$ um ϑ nicht trifft, beliebig klein gemacht werden kann.

Konsistenz liegt etwa vor, wenn die $\hat{\Theta}_n$ erwartungstreu oder zumindest asymptotisch erwartungstreu für ϑ sind und die Varianzen $\text{Var}(\hat{\Theta}_n)$ bei zunehmendem Stichprobenumfang gegen Null gehen.

¹⁰ In diesem Zusammenhang sei noch einmal darauf hingewiesen, dass eine breite Dichtefunktion eine hohe Varianz und eine schmale Dichtefunktion niedrige Varianz anzeigt.

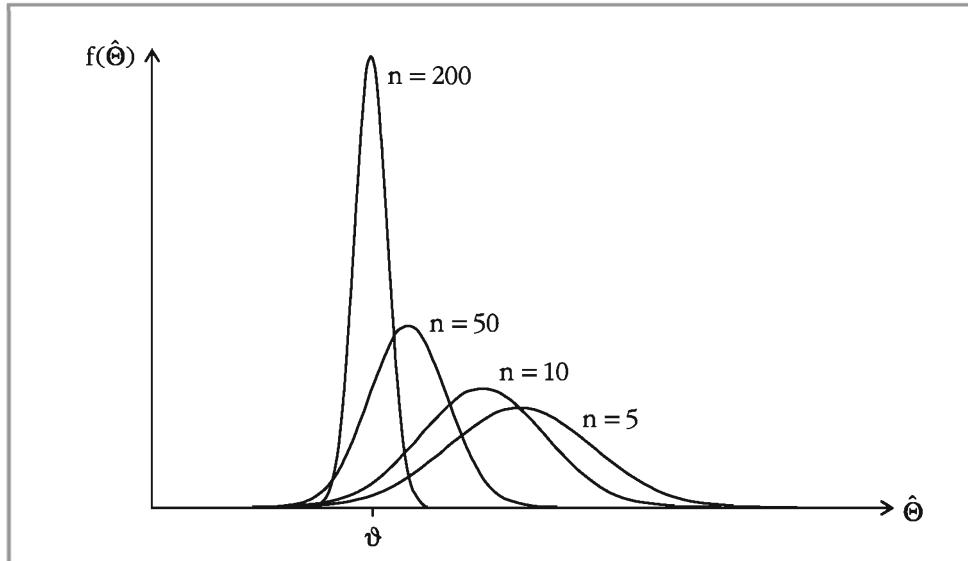


Abbildung III 4: Dichten einer konsistenten Schätzfunktion für verschiedene n

Abbildung III 4 zeigt einen nicht erwartungstreuen (jedoch asymptotisch erwartungstreuen) und konsistenten Schätzer. Wir erkennen, dass die vom Schätzer erzeugten Schätzwerte bei laufender Vergrößerung des Stichprobenumfangs n im Durchschnitt den zu schätzenden Parameter ϑ treffen. Außerdem ist zu erkennen, wie mit wachsendem n die Varianz immer stärker abnimmt.

Konsistent sind z.B. das Stichprobenmittel und der Stichprobenanteilswert, da sie bereits erwartungstreue Schätzer darstellen und wir an ihren Varianzen erkennen können (n im Nenner), dass diese mit zunehmendem Stichprobenumfang immer geringer werden. Die Stichprobenvarianz ist erwartungstreu und konsistent.¹¹ Die für die Schätzung der Varianz der Grundgesamtheit naheliegende mittlere quadratische Abweichung S^2 ist asymptotisch erwartungstreu und konsistent.

Die Konsistenz ist eine *Minimalanforderung* an einen Schätzer. Sie kann als Versicherung dafür angesehen werden, dass ein großer Aufwand bei der Datenbeschaffung belohnt wird. Für hinreichend große Stichprobenumfänge n dürfen wir erwarten, dass konsistente Schätzer passable Schätzwerte liefern. Wenn ein Schätzer nicht einmal konsistent ist, dann ist er meist nicht empfehlenswert.

1.3.5 Mittlerer quadratischer Fehler

Bei manchen Schätzproblemen ist ein erwartungstreuer Schätzer nicht immer der beste, da Erwartungstreue des Öfteren mit einer allzu großen Varianz erkauft wird. Um Varianzen zu verringern, ist man häufig sogar bereit kleine Verzerrungen in Kauf zu nehmen. Letztendlich sollte die erwartete Nähe des Schätzwertes zu seinem wahren Wert Entscheidungskriterium für die Wahl eines Schätzers sein.

¹¹ Hinweis: Es gilt $E(\hat{\sigma}^2) = \sigma^2$ und $\text{Var}(\hat{\sigma}^2) = \frac{1}{n} (\eta_4 - \frac{n-3}{n-1} \sigma^4)$ für $n > 1$, wobei η_4 abkürzend für das 4. zentrale Moment $E[(X-\mu)^4]$ von X steht.

Erwartungswert und Varianz eines Schätzers, also

$$E(\hat{\Theta}) \text{ und } \text{Var}(\hat{\Theta}) = E[(\hat{\Theta} - E(\hat{\Theta}))^2],$$

gehen in ihren sog. **mittleren quadratischen Fehler** (MQF) ein. Dieser kann als Maß für die Güte einer Schätzfunktion angesehen und durch Anwendung des Verschiebungssatzes auf die Zufallsvariable $\hat{\Theta}$ in folgende Form gebracht werden:

$$\underbrace{E[(\hat{\Theta} - \vartheta)^2]}_{\text{Mittlerer quadratischer Fehler}} = \underbrace{\text{Var}(\hat{\Theta})}_{\text{Varianz des Schätzers}} + \underbrace{[E(\hat{\Theta}) - \vartheta]^2}_{\text{Quadrat der Verzerrung}} \quad (\text{III1: 38})$$

Der MQF berücksichtigt also sowohl die Varianz als auch die *Verzerrung* (engl. bias). Wir erkennen an (III1: 38) und Abbildung III 5, dass der Übergang zu einem leicht verzerrten Schätzer $\hat{\Theta}_B$ vorteilhaft sein kann, wenn damit zugleich eine wirkliche Verkleinerung der Varianz erreicht wird.

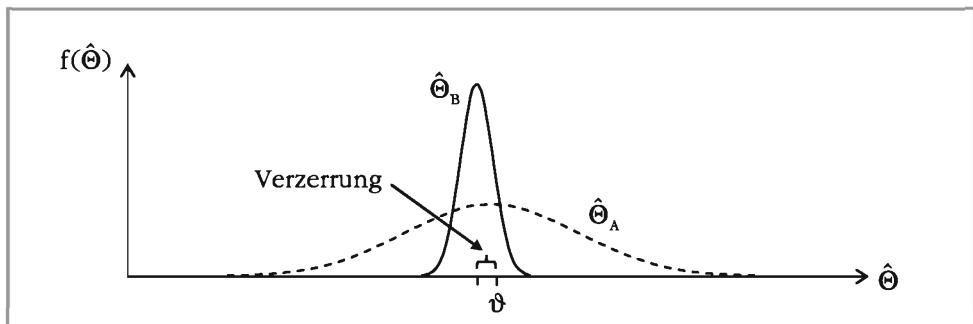


Abbildung III 5: Vorteilhaftigkeit leicht verzerrter Schätzer

Für einen erwartungstreuen Schätzer ist die Verzerrung Null und somit entspricht der MQF gerade der Varianz des Schätzers. Für das Stichprobenmittel gilt also beispielsweise $\text{MQF}(\bar{X}) = \sigma^2 / n$. Bei einem konsistenten Schätzer geht der MQF bei $n \rightarrow \infty$ gegen Null. Streben mit wachsendem Stichprobenumfang sowohl Varianz als auch die Verzerrung gegen Null, so weist der Schätzer auch die Eigenschaft der Konsistenz auf.

1.4 Konstruktion von Schätzern

Zur Konstruktion von Schätzern liefert die statistische Methodenlehre zahlreiche Verfahren, wie z.B. die Momenten-Methode, die Maximum-Likelihood-Methode, die Methode der Kleinsten Quadrate, die Bayes-Methode, die Minimax-Methode usw. Wir wollen im Folgenden exemplarisch die **Methode der kleinsten Quadrate** (engl. ordinary least squares, OLS) und die **Maximum-Likelihood-Methode** (ML) vorstellen.¹² Die Maximum-Likelihood-Methode ist ein zentrales Konzept der statistischen Schätztheorie. Die Methode der kleinsten Quadrate ist vor allem aufgrund ihrer Einfachheit interessant.

¹² OLS werden wir im Rahmen des Kapitels IV noch intensiver behandeln.

1.4.1 Methode der kleinsten Quadrate

Mittels der Methode der kleinsten Quadrate können in vielen Fällen brauchbare Schätzer gewonnen werden. Sind etwa x_1, \dots, x_n die Beobachtungswerte aus einer einfachen Stichprobe aus einer bestimmten Grundgesamtheit, so kann nach der Methode der kleinsten Quadrate als Schätzwert für den Erwartungswert μ dieser Grundgesamtheit derjenige Wert $\hat{\mu}$ verwendet werden, durch den die Summe der quadrierten Abstände des Schätzwertes $\hat{\mu}$ zu den Beobachtungswerten x_i minimal wird. Wir erhalten den Schätzwert also, indem wir den Term

$$\sum_{i=1}^n (x_i - \hat{\mu})^2$$

nach $\hat{\mu}$ ableiten, gleich Null setzen und schließlich nach $\hat{\mu}$ auflösen. Da die zweite Ableitung dieses Terms nach $\hat{\mu}$ positiv ist, erhalten wir das Minimum

$$\begin{aligned} \frac{d \sum_{i=1}^n (x_i - \hat{\mu})^2}{d \hat{\mu}} &= \sum_{i=1}^n 2(x_i - \hat{\mu})(-1) = -2 \sum_{i=1}^n (x_i - \hat{\mu}) = 0 \\ \sum_{i=1}^n x_i - \sum_{i=1}^n \hat{\mu} &= \sum_{i=1}^n x_i - n\hat{\mu} = 0 \\ \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}. \end{aligned}$$

Wir können also das Stichprobenmittel \bar{x} als Kleinst-Quadrate-Schätzfunktion für μ bezeichnen.

1.4.2 Maximum-Likelihood-Methode

Da in der induktiven Statistik die Verteilung der Grundgesamtheit und somit auch die gemeinsame Verteilung der StichprobenvARIABLEN X_i unbekannt sind, ist es von Vorteil alle als möglich erachteten Wahrscheinlichkeitsverteilungen zu betrachten. Ist lediglich ein Parameter unbekannt, charakterisiert

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n | \vartheta) = f_X(x_1 | \vartheta) \cdot \dots \cdot f_X(x_n | \vartheta) = \prod_{i=1}^n f_X(x_i | \vartheta) \quad (\text{III1: 39})$$

die gemeinsame Wahrscheinlichkeitsfunktion (bzw. -dichte) der n *unabhängigen* und *identisch verteilten* StichprobenvARIABLEN X_1, \dots, X_n unter der Bedingung, dass der unbekannte Parameter den Wert ϑ besitzt. (III1: 39) wird als sog. **Likelihood-Funktion** bezeichnet.¹³ Sie ist für jeden festen Wert von ϑ eine n -dimensionale Wahrscheinlichkeitsfunktion (bzw. -dichte) und für feste x_1, \dots, x_n einer konkreten Stichprobe eine Funktion von ϑ . Im letztgenannten Fall gibt die Funktion dann die

¹³ Unter dem Begriff Likelihood versteht man die Wahrscheinlichkeit, dass ein Stichprobenergebnis bei einem angenommenen Wert der (des) zu schätzenden Parameter(s) eintritt. Dieses Konzept unterscheidet sich dadurch von dem der Wahrscheinlichkeit, dass eine Wahrscheinlichkeit auf das Auftreten zukünftiger Ereignisse ausgerichtet ist, während ein Likelihood sich auf vergangene Ereignisse mit bekannten Ergebnissen bezieht.

Wahrscheinlichkeit (bzw. Dichte) dieser Stichprobe für alle möglichen Werte von ϑ an. Die Likelihood-Funktion fassen wir im Folgenden als Funktion von ϑ auf und schreiben für sie

$$L(\vartheta | x_1, \dots, x_n) = \prod_{i=1}^n f_x(x_i | \vartheta) \quad (\text{III1: 40})$$

oder kurz $L(\vartheta)$. Zur Schätzung des unbekannten Parameters ϑ der gemeinsamen Wahrscheinlichkeitsfunktion der Stichprobenvariablen fordert nun das **Maximum-Likelihood-Prinzip** bei festem Stichprobenergebnis x_1, \dots, x_n denjenigen Wert $\hat{\vartheta}$ zu wählen, der die Likelihood-Funktion $L(\vartheta)$ (bzw. die Wahrscheinlichkeit für das Eintreten des beobachteten Stichprobenresultats) maximiert. Der auf diese Art und Weise nach dem Prinzip "Nichts ist wahrscheinlicher als die Realität" konstruierte Schätzer $\hat{\Theta} = g(X_1, \dots, X_n)$ heißt **Maximum-Likelihood-Schätzer**.

Wenn sie existieren, sind Maximum-Likelihood-Schätzer i.d.R. *konsistent*, erwartungstreu oder *zumindest asymptotisch erwartungstreu*, effizient oder *zumindest asymptotisch effizient* und unterliegen dem Zentralen Grenzwertsatz, d.h. sie sind *asymptotisch normalverteilt*.

Beispiel III 1-9

Schätzung des Parameters p einer $B(n,p)$ -Verteilung

Im Anschluss an eine Studentenparty werden 10 Studenten von der Polizei kontrolliert. Dabei wird bei den Studenten Nawid S. und Christian J. festgestellt, dass sie mehr als den zur Fahrtauglichkeit erlaubten Promillewert Alkohol im Blut haben. Der Anteil der nicht fahrtauglichen Studenten in der Stichprobe liegt damit bei $2/10 = 0,2$. Betrachten wir diese Stichprobe als mit Zurücklegen gezogen und damit binomialverteilt, so ergibt sich die Wahrscheinlichkeit dafür, dass die Anzahl X der nicht fahrtauglichen Studenten in der Auswahl bei 2 liegt, in Abhängigkeit vom Anteilswert p der Grundgesamtheit als

$$P(X=2) = \binom{10}{2} p^2 (1-p)^8.$$

Genau dies ist die Likelihood-Funktion $L(p)$ für dieses Fallbeispiel. Man spricht hier gelegentlich auch von der *Likelihood-Funktion der konkreten Stichprobe*. Für ausgewählte Werte von p ergeben sich folgende Wahrscheinlichkeiten:

p	$L(p) = P(X=2 p)$	p	$L(p) = P(X=2 p)$
0,0	0,0000000	0,6	0,0106168
0,1	0,1937102	0,7	0,0014467
0,2	0,3019899	0,8	0,0000737
0,3	0,2334744	0,9	0,0000004
0,4	0,1209324	1,0	0,0000000
0,5	0,0439453		

Wir erkennen, dass die Likelihood-Funktion bei $p = 0,2$ ihr Maximum erreicht. Der Maximum-Likelihood-Schätzwert für den unbekannten Anteilswert der Grundgesamtheit ist somit gleich dem Stichprobenanteilswert $\hat{p} = 0,2$.

Wir wollen nun die Funktion $L(p)$ allgemein betrachten. Es gilt zunächst

$$L(p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Zur Bestimmung des Maximums dieser Funktion ist nach den Regeln der Differentialrechnung ihre erste Ableitung nach p zu bilden und gleich Null zu setzen. Da die Likelihood-Funktion $L(p)$ und ihr natürlicher Logarithmus $\ln L(p)$, die sog. Log-Likelihood-Funktion, jedoch an derselben Stelle ihr Maximum besitzen, maximiert man meist nicht $L(p)$ sondern $\ln L(p)$. Der Grund dafür ist darin zu finden, dass wir durch Logarithmieren das vorliegende Produkt in eine Summe überführen können, die sich leichter ableiten lässt. Wir erhalten in unserem Fall

$$\ln L(p) = \ln \binom{n}{x} + x \ln p + (n-x) \ln(1-p),$$

womit sich die erste Ableitung zu

$$\frac{d \ln L(p)}{d p} = \frac{x}{p} - \frac{(n-x)}{(1-p)}$$

ergibt. Durch Nullsetzen und Auflösen dieses Terms und unter Beachtung der Tatsache, dass die zweite Ableitung stets negativ ist, ergibt sich für den Maximum-Likelihood-Schätzwert des Stichprobenanteilswerts

$$\hat{p} = \frac{x}{(n-x)} = 0 \quad \rightarrow \quad x - x\hat{p} = n\hat{p} - x\hat{p} \quad \leftrightarrow \quad \hat{p} = \frac{x}{n} = \bar{x}.$$

Die Zufallsvariable $\hat{p} = \bar{x}$ stellt also den Maximum-Likelihood-Schätzer für p dar.

Beispiel III 1-10

Schätzung des Parameters λ einer $Po(\lambda)$ -Verteilung

Liegt eine konkrete Stichprobe x_1, \dots, x_n vom Umfang n aus einer Poisson-verteilten Grundgesamtheit $X \sim Po(\lambda)$ mit der Wahrscheinlichkeitsfunktion

$$f_{Po}(x) = \begin{cases} \frac{\lambda^x}{x!} \cdot e^{-\lambda} & \text{für } x = 0, 1, 2, \dots \\ 0 & \text{sonst} \end{cases}$$

vor, so erhalten wir den ML-Schätzwert für den unbekannten Parameter $\lambda > 0$ wie folgt:

Die Likelihood-Funktion $L(\lambda) \equiv L(\lambda | x_1, \dots, x_n)$ der Stichprobenrealisation ist

$$L(\lambda) = \prod_{i=1}^n f_X(x_i | \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}.$$

Durch Logarithmieren erhalten wird daraus die Log-Likelihood-Funktion

$$\begin{aligned} \ln L(\lambda) &= \sum_{i=1}^n \ln \left(\frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda} \right) = \sum_{i=1}^n (\ln(\lambda^{x_i}) - \ln(x_i!) + \ln(e^{-\lambda})) \\ &= \sum_{i=1}^n (x_i \cdot \ln \lambda - \ln(x_i!) - \lambda) = \ln \lambda \cdot \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!) - n\lambda. \end{aligned}$$

Deren ersten beiden Ableitungen lauten

$$\frac{d \ln L(\lambda)}{d \lambda} = \frac{1}{\lambda} \sum_{i=1}^n x_i - n \quad \text{und} \quad \frac{d^2 \ln L(\lambda)}{d \lambda^2} = -\frac{1}{\lambda^2} \sum_{i=1}^n x_i.$$

Nullsetzen der ersten Ableitung liefert

$$\frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0 \quad \rightarrow \quad \frac{1}{\lambda} \sum_{i=1}^n x_i = n \quad \leftrightarrow \quad \frac{1}{\lambda} \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} n = 1 \quad \leftrightarrow \quad \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Da die zweite Ableitung stets negativ ist, liegt ein Maximum von $\ln L(\lambda)$ vor. Wir erhalten also \bar{x} als ML-Schätzwert von λ . \bar{X} ist demnach der ML-Schätzer für den Parameter λ einer Poissonverteilung.

Beispiel III 1-11

Schätzung des Parameters λ einer $\text{Ex}(\lambda)$ -Verteilung

Liegt eine konkrete Stichprobe x_1, \dots, x_n vom Umfang n aus einer exponentialverteilten Grundgesamtheit $X \sim \text{Ex}(\lambda)$ mit der Dichte

$$f_{\text{Ex}}(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

vor, so können wir den ML-Schätzwert für den unbekannten Parameter $\lambda > 0$ folgendermaßen bestimmen:

Die Likelihood-Funktion $L(\lambda) \equiv L(\lambda | x_1, \dots, x_n)$ der vorliegenden Stichprobe lautet

$$L(\lambda) = \prod_{i=1}^n f_X(x_i | \lambda) = \prod_{i=1}^n \lambda \cdot e^{-\lambda \cdot x_i} = \lambda^n \cdot \prod_{i=1}^n e^{-\lambda \cdot x_i}.$$

Daraus erhalten wir die Log-Likelihood-Funktion

$$\ln L(\lambda) = \ln \left(\lambda^n \cdot \prod_{i=1}^n e^{-\lambda \cdot x_i} \right) = \ln \lambda^n + \ln \left(\prod_{i=1}^n e^{-\lambda \cdot x_i} \right) = n \cdot \ln \lambda + \sum_{i=1}^n \ln e^{-\lambda \cdot x_i} = n \cdot \ln \lambda - \lambda \cdot \sum_{i=1}^n x_i$$

und deren Ableitungen

$$\frac{d \ln L(\lambda)}{d \lambda} = n \cdot \frac{1}{\lambda} - \sum_{i=1}^n x_i \quad \text{und} \quad \frac{d^2 \ln L(\lambda)}{d \lambda^2} = -\frac{n}{\lambda^2}.$$

Die Nullstelle der ersten Ableitung ist

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

Da die zweite Ableitung stets negativ ist, liegt ein Maximum von $\ln L(\lambda)$ vor. Wir erhalten also den Kehrwert des realisierten Stichprobenmittels als ML-Schätzwert für λ . Der ML-Schätzer für λ ist damit $1/\bar{X}$. Dieser Schätzer ist asymptotisch erwartungstreu, konsistent und asymptotisch effizient.

Analog zu den vorhergehenden Beispielen kann gezeigt werden,¹⁴ dass für eine **Normalverteilung** mit dem Erwartungswert μ und der Varianz σ^2 als Maximum-Likelihood-Schätzer für μ ebenfalls das Stichprobenmittel \bar{X} resultiert. Als Schätzer für die Varianz ergibt sich nach der Maximum-Likelihood-Methode bei unbekanntem μ der verzerre, aber asymptotisch erwartungstreue, konsistente und asymptotisch effiziente Schätzer S^2 .

¹⁴ Wir überlassen dies in Aufgabe PS10 dem Leser.

2.

Intervallschätzung

Im vorhergehenden Abschnitt haben wir Stichprobenergebnisse verwendet, um unbekannte Parameter der Grundgesamtheit zu schätzen. Wir hatten uns dabei auf die Punktschätzung beschränkt, die einen einzigen Schätzwert liefert. Da die in einer Stichprobe realisierten Merkmalsausprägungen zufallsabhängig sind, wird aber ein gefundener Punktschätzwert nur in den seltensten Fällen genau mit dem gesuchten Parameter der Grundgesamtheit übereinstimmen. Um wenigstens Aussagen über die Genauigkeit der Schätzung machen zu können, bedienen wir uns der Intervallschätzung. Es handelt sich dabei um ein Verfahren, das ein Intervall liefert, in dem der unbekannte Parameter mit einer vorgegebenen Wahrscheinlichkeit liegt.

2.1 Grundlagen

Ziel einer Intervallschätzung ist es, einen unbekannten Parameter ϑ einer Grundgesamtheit so zu schätzen, dass als Schätzergebnis ein Intervall auftritt, das den wahren ϑ -Wert mit einer vorgegebenen Wahrscheinlichkeit $1 - \alpha$ bzw. mit einer Wahrscheinlichkeit von α nicht überdeckt. Die Untergrenze v_u und die Obergrenze v_o eines solchen Intervalls wird durch das Stichprobenergebnis festgelegt. Dies bedeutet, dass v_u und v_o Realisationen gewisser Stichprobenfunktionen $V_u = g(X_1, X_2, \dots, X_n)$ und $V_o = g(X_1, X_2, \dots, X_n)$ sind. Zwischen diesen muss die Beziehung

$$V_u \leq V_o \quad (\text{III2: 1})$$

gelten. Es muss also für alle möglichen Stichprobenrealisierungen $v_u \leq v_o$ sein. Zudem muss die Wahrscheinlichkeit, dass der unbekannte Parameter ϑ der Grundgesamtheit von V_u und V_o eingeschlossen wird, gleich $1 - \alpha$ sein.

$$P(V_u \leq \vartheta \leq V_o) = 1 - \alpha \quad (\text{III2: 2})$$

Erfüllen zwei Stichprobenfunktionen V_u und V_o diese Voraussetzungen (III2: 1) und (III2: 2), so wird das Zufallsintervall $[V_u, V_o]$ als **Konfidenzintervall** für ϑ zum Konfidenzniveau $1 - \alpha$ bezeichnet. Die Realisierung $[v_u, v_o]$ heißt Ergebnis der Intervallschätzung oder auch **Schätzintervall**.

Das **Konfidenzniveau** (oder die **Vertrauenswahrscheinlichkeit**) $1 - \alpha$ sagt grob gesprochen aus, bei welchem Anteil von Stichproben ein Schätzintervall entsteht, das den unbekannten Parameter ϑ enthält. Dagegen misst α den Anteil der Stichproben, die ein Intervall liefern, das ϑ nicht enthält. Wir bezeichnen α deshalb auch als **Irrtumswahrscheinlichkeit**. Die Wahl des Konfidenzniveaus – man wählt i.d.R. $1 - \alpha = 0,90, 0,95$ oder $0,99$ – liegt im Ermessen des Statistikers. Er kann dieses seinen Genauigkeitsbedürfnissen entsprechend festlegen. Eine Vergrößerung des Konfidenzniveaus (d.h. eine Verkleinerung der Irrtumswahrscheinlichkeit)

keit) hat nämlich eine Vergrößerung des Schätzintervalls und somit eine "gröbere" Schätzung zur Folge.

Ein Konfidenzintervall $[V_u, V_o]$ für ϑ wird als **symmetrisch** bezeichnet, wenn

$$P(\vartheta < V_u) = P(\vartheta > V_o) = \frac{\alpha}{2}. \quad (\text{III2: 3})$$

Das zweite Gleichheitszeichen gilt, da nach (III2: 2) das Komplementärereignis zu $V_u \leq \vartheta \leq V_o$ die Wahrscheinlichkeit α besitzt und sich dieses Komplementärereignis aus den beiden disjunkten Ereignissen $\vartheta < V_u$ und $\vartheta > V_o$ ergibt.

Als Intervallschätzung verstehen wir im Folgenden immer die Realisation eines symmetrischen Konfidenzintervalls. Wir behandeln konkret Intervallschätzungen für μ , p und σ^2 . Im Rahmen dieser gilt es Funktionen V_u und V_o zu finden, die (zumindest näherungsweise) die Beziehungen (III2: 1) und (III2: 3) und folglich auch (III2: 2) für $\vartheta = \mu$, $\vartheta = p$ bzw. $\vartheta = \sigma^2$ erfüllen.

2.2 Konfidenzintervalle für das arithmetische Mittel

Im Abschnitt III 1.2.2.1 bzw. in (III1: 11) sahen wir, dass die Zufallsvariable \bar{X} bei $N(\mu, \sigma^2)$ -verteilter Grundgesamtheit mit dem Erwartungswert $E(\bar{X}) = \mu$ und der Varianz $\text{Var}(\bar{X}) = \sigma_{\bar{X}}^2$ normalverteilt ist, wobei beim *Ziehen mit Zurücklegen* (bzw. beim Ziehen ohne Zurücklegen und erfüllter Bedingung $n \leq 0,1N$)

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

galt. Das standardisierte Stichprobenmittel $(\bar{X} - \mu) / \sigma_{\bar{X}}$ folgte somit einer $N(0, 1)$ -Verteilung. Aus der grafischen Darstellung der Verteilung des Stichprobenmittels (vgl. Abbildung III 6) können wir daher unter Verwendung des $\alpha/2$ -Fraktils und des $(1-\alpha/2)$ -Fraktils der Standardnormalverteilung zunächst

$$P\left(\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \quad \text{und} \quad P\left(\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} > z_{1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \quad (\text{III2: 4})$$

festhalten. Die Wahrscheinlichkeit, dass das standardisierte Stichprobenmittel nun in ein Intervall zwischen diesen beiden Fraktilen fällt, ergibt sich aufgrund der vorherrschenden Komplementärbeziehung zu

$$P\left(z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha. \quad (\text{III2: 5})$$

Einige weitere Umformungen in (III2: 5) erlauben schließlich die Darstellung

$$P(\mu + z_{\frac{\alpha}{2}} \sigma_{\bar{X}} \leq \bar{X} \leq \mu + z_{1-\frac{\alpha}{2}} \sigma_{\bar{X}}) = 1 - \alpha.$$

Da aufgrund der Symmetrie der Standardnormalverteilung $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$ gilt, wollen wir im Folgenden nur noch $z_{1-\frac{\alpha}{2}} = z$ bzw. $z_{\frac{\alpha}{2}} = -z$ schreiben. Wir erhalten damit

$$P(\mu - z \sigma_{\bar{X}} \leq \bar{X} \leq \mu + z \sigma_{\bar{X}}) = 1 - \alpha. \quad (\text{III2: 6})$$

Wir sprechen bei dem in (III2: 6) vorkommenden Intervall von einem sog. **Wahrscheinlichkeitsintervall** um den Mittelwert μ der Grundgesamtheit. In dieses fällt

der in einer *Stichprobe gefundene Mittelwert* \bar{X} bei bekannter Grundgesamtheit mit einer Wahrscheinlichkeit von $1 - \alpha$. Wie Abbildung III 6 zeigt, liegen rechts des $(1 - \alpha/2)$ -Fraktils der Standardnormalverteilung genau $(\alpha/2) \cdot 100\%$ der gesamten Fläche unter der Dichtefunktion. Da die Verteilung symmetrisch ist, liegen auch links des $(\alpha/2)$ -Fraktils $(\alpha/2) \cdot 100\%$. Das Wahrscheinlichkeitsintervall nimmt somit $(1 - \alpha) \cdot 100\%$ der Fläche ein.

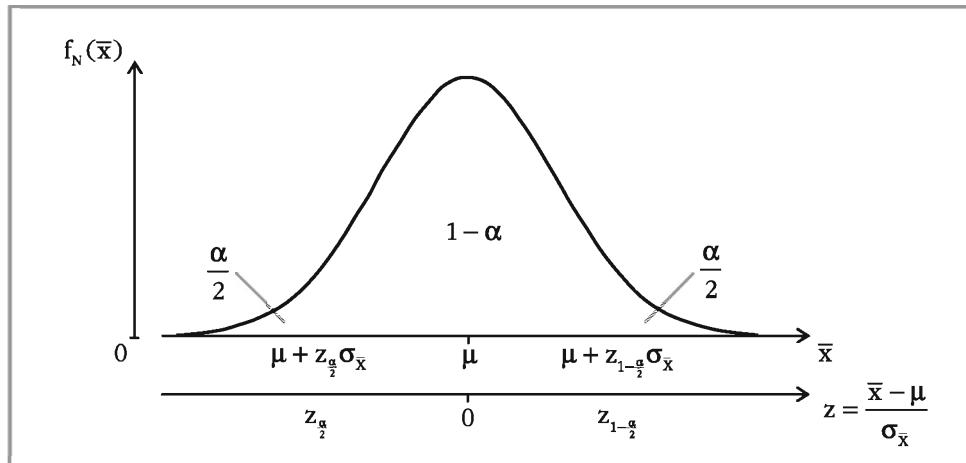


Abbildung III 6: Wahrscheinlichkeitsintervall

Beispiel III 2-1

Wahrscheinlichkeitsintervall

Für eine Irrtumswahrscheinlichkeit von 5 % ($\alpha = 0,05$) erhalten wir für eine mit Erwartungswert $\mu = 2$ und Varianz $\sigma^2 = 16$ normalverteilte Grundgesamtheit bei einem Stichprobenumfang von $n = 100$ und Ziehen mit Zurücklegen, das Wahrscheinlichkeitsintervall

$$\begin{aligned} 2 - z_{1-\frac{\alpha}{2}} \cdot \frac{4}{\sqrt{100}} &\leq \bar{X} \leq 2 + z_{1-\frac{\alpha}{2}} \cdot \frac{4}{\sqrt{100}} \\ 2 - 1,96 \cdot 0,4 &\leq \bar{X} \leq 2 + 1,96 \cdot 0,4 \\ 1,22 &\leq \bar{X} \leq 2,78 \end{aligned}$$

mit der Eigenschaft

$$P(1,22 \leq \bar{X} \leq 2,78) = 0,95.$$

Wird also aus der genannten Grundgesamtheit eine Stichprobe vom Umfang $n = 100$ gezogen, so liegt der resultierende Stichprobenmittelwert mit einer Wahrscheinlichkeit von 95 % zwischen 1,22 und 2,78.

Unsere bisherigen Ausführungen bezogen sich auf Wahrscheinlichkeitsintervalle für das Stichprobenmittel. Wir interessieren uns nun aber für ein Verfahren, dass ein Intervall liefert, in das der *wahre Mittelwert* μ und nicht das Stichprobenmittel \bar{X} mit einer Wahrscheinlichkeit von $1 - \alpha$ fällt. Um ein solches zu finden, müssen wir hinsichtlich der Verteilung der Grundgesamtheit und der Bekanntheit ihrer Varianz drei Fälle (III 2.2.1 bis III 2.2.3) unterscheiden. Für jeden der Fälle erhalten wir ein etwas anders definiertes Konfidenzintervall.

2.2.1 Normalverteilte Grundgesamtheit mit bekannter Varianz

Wissen wir, dass die *Grundgesamtheit normalverteilt* und außerdem die *Varianz σ^2 der Grundgesamtheit bekannt* ist, so können wir ein Konfidenzintervall für μ aus der Beziehung (III2: 6) ableiten. Wir nehmen dazu an dem darin enthaltenen Wahrscheinlichkeitsintervall folgende Umformungen vor:

$$\begin{aligned}\mu - z\sigma_{\bar{x}} &\leq \bar{X} \leq \mu + z\sigma_{\bar{x}} \\ -z\sigma_{\bar{x}} &\leq \bar{X} - \mu \leq z\sigma_{\bar{x}} \\ -z\sigma_{\bar{x}} - \bar{X} &\leq -\mu \leq z\sigma_{\bar{x}} - \bar{X} \\ z\sigma_{\bar{x}} + \bar{X} &\geq \mu \geq -z\sigma_{\bar{x}} + \bar{X} \\ \bar{X} - z\sigma_{\bar{x}} &\leq \mu \leq \bar{X} + z\sigma_{\bar{x}}\end{aligned}$$

Mit $V_u = \bar{X} - z\sigma_{\bar{x}}$ und $V_o = \bar{X} + z\sigma_{\bar{x}}$ erhalten wir so das **Konfidenzintervall**

$$[V_u, V_o] = [\bar{X} - z\sigma_{\bar{x}}, \bar{X} + z\sigma_{\bar{x}}], \quad (\text{III2: 7a})$$

für welches

$$P(\bar{X} - z\sigma_{\bar{x}} \leq \mu \leq \bar{X} + z\sigma_{\bar{x}}) = 1 - \alpha \quad (\text{III2: 7b})$$

gilt. Da \bar{X} eine Zufallsvariable ist, sind die Grenzen des Konfidenzintervalls zufallsabhängig. Bei mehreren Stichproben j ($j = 1, 2, \dots$) im Umfang von jeweils n werden sich unterschiedliche Stichprobenmittel \bar{x}_j und damit auch unterschiedliche *Schätzintervalle* für μ ergeben (vgl. Abbildung III 7).

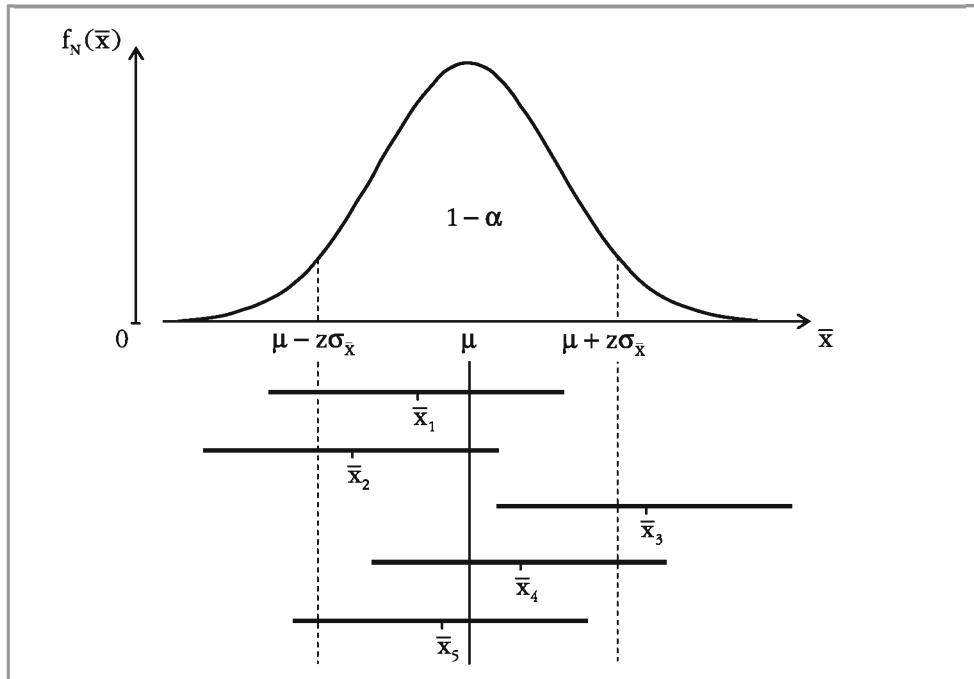


Abbildung III 7: Intervallschätzungen für den Mittelwert

Legen wir um den Mittelwert der Grundgesamtheit (unter der Annahme, dass dieser bekannt ist) ein Wahrscheinlichkeitsintervall, in das der Stichprobenmittelwert \bar{X} , mit einer Wahrscheinlichkeit von $1 - \alpha$ fällt (vgl. gestrichelte Linien in Abbildung III 7), so umschließt eine Intervallschätzung um \bar{x} , nur dann den gesuchten Parameter μ , wenn dieses \bar{x} , innerhalb des Wahrscheinlichkeitsintervalls um μ liegt. Die Wahrscheinlichkeit $1 - \alpha$ kann daher als Wahrscheinlichkeit dafür interpretiert werden, dass ein Konfidenzintervall um den Stichprobenmittelwert \bar{X} den gesuchten Mittelwert μ der Grundgesamtheit überdeckt. Betrachten wir eine durch eine konkrete Stichprobe bestimmte Intervallschätzung, so wird diese μ entweder einschließen oder nicht. Gerade daher liegt uns also mit einem Konfidenzintervall ein Verfahren vor, das bei einer großen Zahl von Stichproben in $(1 - \alpha) \cdot 100\%$ aller Fälle eine Intervallschätzung liefert, die den unbekannten Parameter μ umschließt.

Mit zunehmenden α verkürzt sich die Länge eines Konfidenzintervalls, sodass damit auch die Wahrscheinlichkeit, dass das Intervall den wahren Wert nicht überdeckt, steigt. Typische in der Praxis verwendete Irrtumswahrscheinlichkeiten und die dazugehörigen z-Werte sind

$$\begin{aligned}\alpha = 0,01 &\rightarrow z = 2,575 \\ \alpha = 0,05 &\rightarrow z = 1,96 \\ \alpha = 0,10 &\rightarrow z = 1,645.\end{aligned}$$

Beispiel III 2-2

Intervallschätzung für μ



Von $N = 9.000$ Haltern einer bestimmten Pkw-Marke werden $n = 50$ dahingegen befragt, wie lange sie ihren Neuwagen nutzen konnten, bis der erste Defekt auftrat. Das Ergebnis dieser Befragung war eine durchschnittliche Nutzungsdauer von $\bar{x} = 4,5$ Jahren. Die Standardabweichung der als normalverteilt angenommenen Grundgesamtheit ist bekannt und beträgt $\sigma = 0,5$ Jahre. Bestimmen wir nun ein 99 %-Schätzintervall ($\alpha = 0,01$) für die durchschnittliche Nutzungsdauer μ der Grundgesamtheit. Es empfiehlt sich dazu allgemein folgende Vorgehensweise:

- Bestimmung des $(1-\alpha/2)$ -Frakts der Standardnormalverteilung:

$$z = z_{\frac{1-\alpha}{2}} = z_{1-0,005} = z_{0,995} = 2,58$$

- Ermittlung des Stichprobenmittelwertes \bar{x} und der Standardabweichung $\sigma_{\bar{x}}$:

$$\bar{x} = 4,5$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0,5}{\sqrt{50}} = 0,07 \quad n \leq 0,1N \rightarrow \text{Der Endlichkeitskorrekturfaktor beim Ziehen ohne Zurücklegen kann vernachlässigt werden.}$$

- Bildung des Schätzintervalls:

$$\bar{x} - z\sigma_{\bar{x}} \leq \mu \leq \bar{x} + z\sigma_{\bar{x}} \quad \rightarrow \quad 4,5 - 2,58 \cdot 0,07 \leq \mu \leq 4,5 + 2,58 \cdot 0,07 \\ 4,32 \leq \mu \leq 4,68$$

Dieses Schätzintervall wird nun den unbekannten Mittelwert μ der Grundgesamtheit mit einer Wahrscheinlichkeit von 99 % einschließen.

2.2.2 Normalverteilte Grundgesamtheit mit unbekannter Varianz

Ausgehend von einer *normalverteilten Grundgesamtheit* mit bekannter Varianz, haben wir bereits mehrfach von der Tatsache Gebrauch gemacht, dass für eine solche das Stichprobenmittel \bar{X} mit dem Erwartungswert μ und der Varianz $\sigma_{\bar{X}}$ normalverteilt und damit das beim Ziehen mit Zurücklegen (bzw. beim Ziehen ohne Zurücklegen und $n \leq 0,1N$) resultierende standardisierte Stichprobenmittel

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0; 1)$$

einer Standardnormalverteilung folgt. Ist die *Varianz σ^2* einer solchen Grundgesamtheit jedoch *unbekannt*, was in praktischen Anwendungen der realistischere Fall ist, so muss diese geschätzt werden, um eine Schätzung für den Standardfehler des Stichprobenmittels $\sigma_{\bar{X}}$ zu erhalten. Es empfiehlt sich dabei, die Stichprobenvarianz $\hat{\sigma}^2$ heranzuziehen. Wir erhalten damit eine mit $n - 1$ Freiheitsgraden t-verteilte Zufallsvariable T (vgl. Abschnitt III 1.2.2.1) der Form

$$T = \frac{\bar{X} - \mu}{\hat{\sigma}_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t(n - 1).$$

Ist nun $t_{\frac{\alpha}{2}, n-1}$ das $\alpha/2$ -Fraktil und $t_{1-\frac{\alpha}{2}, n-1}$ das $(1 - \alpha/2)$ -Fraktil der $t(n - 1)$ -Verteilung, so können wir analog zur Einleitung von Abschnitt III 2.2 ein Intervall

$$t_{\frac{\alpha}{2}, n-1} \leq \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \leq t_{1-\frac{\alpha}{2}, n-1}$$

angeben, in das die Zufallsvariable T mit einer Wahrscheinlichkeit von $1 - \alpha$ fällt. Da wir aufgrund der Symmetrie der t-Verteilung vereinfacht $t = t_{1-\frac{\alpha}{2}, n-1} = -t_{\frac{\alpha}{2}, n-1}$ schreiben können, ergibt sich daraus durch Umstellung das **Konfidenzintervall**

$$[V_u, V_o] = \left[\bar{X} - t \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + t \cdot \frac{\hat{\sigma}}{\sqrt{n}} \right] \quad (\text{III2: 8a})$$

für welches

$$P\left(\bar{X} - t \cdot \frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{X} + t \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right) = 1 - \alpha \quad (\text{III2: 8b})$$

gilt.¹ Wichtig ist in diesem Zusammenhang besonders, dass durch Schätzung von σ^2 eine zusätzliche Fehlerquelle entsteht, die die Schätzungsgenauigkeit beeinträchtigt. Bei der nun verwendeten t-Verteilung sind deshalb die Fraktilen im Vergleich zur

¹ Bei der Aufstellung des Schätzintervalls für eine konkrete Stichprobe empfiehlt es sich auf eine Vereinfachung der Stichprobenvarianz zurückzugreifen, die aus dem Verschiebungssatz resultiert. Diese stellt sich wie folgt dar:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \bar{X}^2$$

Standardnormalverteilung absolut größer. Dies hat zur Folge, dass das Schätzintervall verbreitert wird.

Beispiel III 2-3

Intervallschätzung für μ



Eine Maschine füllt Zahnpasta in Tuben ab. Das Füllgewicht soll als normalverteilt betrachtet werden. Eine Stichprobe im Umfang von $n = 15$ aus der laufenden Produktion ergab ein durchschnittliches Füllgewicht von 35 g bei einer Standardabweichung $\sigma = 5$ g. Gesucht ist nun eine Intervallschätzung, die den tatsächlichen Mittelwert μ der Grundgesamtheit mit 95 % Wahrscheinlichkeit einschließt.

- Bestimmung des $(1-\alpha/2)$ -Fraktils der t-Verteilung mit $n - 1$ Freiheitsgraden:

$$t = t_{1-\frac{0,05}{2}; 14} = t_{0,975; 14} = 2,14$$

- Ermittlung des Stichprobenmittelwertes \bar{x} und der Standardabweichung $\hat{\sigma}_{\bar{x}}$:

$$\bar{x} = 35$$

$$\hat{\sigma}_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{15}} = 1,29 \quad n \leq 0,1N \rightarrow \text{Der Endlichkeitskorrekturfaktor beim Ziehen ohne Zurücklegen kann vernachlässigt werden.}$$

- Bildung des Schätzintervalls:

$$\bar{x} - t\hat{\sigma}_{\bar{x}} \leq \mu \leq \bar{x} + t\hat{\sigma}_{\bar{x}} \quad \rightarrow \quad 35 - 2,14 \cdot 1,29 \leq \mu \leq 35 + 2,14 \cdot 1,29 \\ 32,24 \leq \mu \leq 37,76$$

Wie wir wissen, kann die t-Verteilung bei Freiheitsgraden von $w > 30$ durch die Standardnormalverteilung approximiert werden. Dies bedeutet, dass bei hinreichend großem Stichprobenumfang $t \approx z$ gilt und somit durch Ersetzen von t durch z in (III2: 8a) ein *approximatives Konfidenzintervall* für μ bei normalverteilter Grundgesamtheit und unbekannter Varianz σ^2 resultiert.

2.2.3 Beliebig verteilte Grundgesamtheit

Ist die *Grundgesamtheit beliebig verteilt*, so besagt der *zentrale Grenzwertsatz*, dass wir bei ausreichend *großem Stichprobenumfang* ($n > 30$) die Verteilungen der Gauß- bzw. T-Statistik

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\sigma} \quad \text{bzw.} \quad T = \frac{\bar{X} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

in guter Näherung durch die Standardnormalverteilung approximieren können. Wir erhalten daraus analog zur Vorgehensweise in den vorhergehenden Abschnitten folgende **approximative Konfidenzintervalle**:

- bei *bekannter Varianz σ^2 der Grundgesamtheit*:

$$[V_u, V_o] = \left[\bar{X} - z \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z \cdot \frac{\sigma}{\sqrt{n}} \right] \quad (\text{III2: 9})$$

b) bei unbekannter Varianz σ^2 der Grundgesamtheit:

$$[V_u, V_o] = \left[\bar{X} - z \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + z \cdot \frac{\hat{\sigma}}{\sqrt{n}} \right] \quad (\text{III2: 10})$$

Die geforderten Zusammenhänge (III2: 2) und (III2: 3) sind jedoch im Falle einer beliebig verteilten Grundgesamtheit und Approximation durch die Standardnormalverteilung nur näherungsweise erfüllt.

Beispiel III 2-4

Intervallschätzung für μ



Die Anzahl der in der Telefonzentrale eines Unternehmens eingehenden Anrufe pro Tag wird als Poisson-verteilt angesehen. Zur Intervallschätzung des Erwartungswerts μ bzw. des unbekannten Parameters λ dieser Verteilung wurde eine Stichprobe im Umfang von 36 Tagen gezogen, die eine durchschnittliche Anruferzahl von 25 ergab. Die Varianz der Grundgesamtheit ist nicht bekannt. Gesucht ist nun ein 90 %-Schätzintervall für den unbekannten Parameter der Grundgesamtheit.

1. Test auf Anwendbarkeit des zentralen Grenzwertsatzes:

Die Bedingung $n > 30$ ist hier erfüllt.

2. Bestimmung des $(1 - \alpha/2)$ -Fraktils der Standardnormalverteilung:

$$z = z_{1-\frac{\alpha}{2}} = z_{1-0.05} = z_{0.95} = 1,64$$

3. Ermittlung des Stichprobenmittelwertes \bar{x} und der Standardabweichung $\hat{\sigma}_{\bar{x}}$:

$$\bar{x} = 25$$

Die Standardabweichung kann hier nicht aus Einzelwerten berechnet werden, da keine vorliegen. Auch wurde keine Stichprobenstandardabweichung genannt. Da aber für die Poissonverteilung $\mu = \sigma^2 = \lambda$ gilt, kann wie folgt vorgegangen werden:

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{5}{\sqrt{36}} = 0,83 \quad \text{mit} \quad \hat{\sigma} = \sqrt{\bar{x}} = \sqrt{25} = 5$$

4. Bildung des Schätzintervalls:

$$\bar{x} - z \hat{\sigma}_{\bar{x}} \leq \mu \leq \bar{x} + z \hat{\sigma}_{\bar{x}} \quad \rightarrow \quad 25 - 1,64 \cdot 0,83 \leq \mu \leq 25 + 1,64 \cdot 0,83$$

$$23,64 \leq \mu \leq 26,36$$

2.3 Konfidenzintervall für den Anteilswert

Wie wir im Abschnitt III 1.2.3 gesehen haben, ist der Anteilswert \hat{p} einer ausreichend großen Stichprobe aus einer *dichotomen Grundgesamtheit* (N Objekte, von denen M die Eigenschaft A besitzen $\rightarrow p = M/N$) bei erfüllter Bedingung $np(1-p) \geq 9$ mit dem Erwartungswert $E(\hat{p}) = p$ und der Varianz $\text{Var}(\hat{p}) = \sigma_{\hat{p}}^2$ approximativ normalverteilt. Beim Ziehen mit Zurücklegen (bzw. beim Ziehen ohne Zurücklegen mit erfüllter Bedingung $n \leq 0,1N$) gilt

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}.$$

Analog zu den vorhergehenden Abschnitten kann aus der standardisierten und approximativ standardnormalverteilten Zufallsvariablen

$$Z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$

ein **approximativer Konfidenzintervall** für den Anteilswert hergeleitet werden. Wir erhalten das Intervall

$$[V_u, V_o] = \left[\hat{p} - z \cdot \sqrt{\frac{p(1-p)}{n}}, \hat{p} + z \cdot \sqrt{\frac{p(1-p)}{n}} \right], \quad (\text{III2: 11})$$

in welches der unbekannte Anteilswert p mit einer Wahrscheinlichkeit von $1 - \alpha$ fällt. Da der Anteilswert p der Grundgesamtheit gerade unbekannt ist, muss als Schätzer für den Anteilswert der Grundgesamtheit der Stichprobenanteilswert \hat{p} verwendet werden. Auch die Überprüfung unserer Approximationsbedingung $np(1-p) \geq 9$ kann nicht direkt erfolgen. Auch hier muss die Schätzung aus der Stichprobe verwendet werden, sodass nun also $n\hat{p}(1-\hat{p}) \geq 9$ gelten muss.

Beim Konfidenzintervall für den unbekannten Anteilswert p handelt es sich um einen *Sonderfall* von (III2: 10). Der Anteilswert ist nämlich, wie wir bereits in den Abschnitten III 1.2.1 und III 1.2.3 gesehen haben, nichts anderes als der Mittelwert einer dichotomen Grundgesamtheit (zählt also zu den beliebig verteilten Grundgesamtheiten), deren Varianz unbekannt ist. Es wurde in (III2: 10) lediglich σ durch den Ausdruck $\sqrt{\hat{p}(1-\hat{p})}$ ersetzt.

Beispiel III 2-5

Intervallschätzung für p



In einem Wahlbezirk mit $N = 10.000$ Wahlberechtigten soll der Anteil p der Wähler einer Partei A geschätzt werden. In einer Stichprobe vom Umfang $n = 100$ befinden sich nun $x = 40$ Wähler, die für die Partei A stimmen würden. Bestimmen wir nun ein Intervall, in das der wahre Wähleranteil p für Partei A mit einer Wahrscheinlichkeit von 95 % fällt.

- Prüfung der Normalverteilungsbedingung $n\hat{p}(1-\hat{p}) \geq 9$:

$$100 \cdot 0,4 \cdot (1-0,4) = 24 > 9$$

- Bestimmung des $(1-\alpha/2)$ -Fraktils der Standardnormalverteilung:

$$z = z_{1-\frac{\alpha}{2}} = z_{1-0,025} = z_{0,975} = 1,96$$

- Ermittlung des Stichprobenanteilswertes \hat{p} und der Standardabweichung $\hat{\sigma}_{\hat{p}}$:

$$\hat{p} = \frac{40}{100} = 0,4$$

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0,4 \cdot 0,6}{100}} = 0,0490$$

$n \leq 0,1N \rightarrow$ Der Endlichkeitskorrektfaktor beim Ziehen ohne Zurücklegen kann vernachlässigt werden.

- Bildung des Schätzintervalls:

$$\hat{p} - z\hat{\sigma}_{\hat{p}} \leq p \leq \hat{p} + z\hat{\sigma}_{\hat{p}} \quad \rightarrow \quad 0,4 - 1,96 \cdot 0,0490 \leq p \leq 0,4 + 1,96 \cdot 0,0490 \\ 0,3040 \leq p \leq 0,4960$$

2.4 Konfidenzintervall für die Varianz

Unter III 1.2.4 hatten wir gesehen, dass bei $N(\mu, \sigma^2)$ -verteilter Grundgesamtheit die Zufallsvariable

$$B = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)\hat{\sigma}^2}{\sigma^2}$$

einer Chi-Quadrat-Verteilung mit $n - 1$ Freiheitsgraden folgt. Bezeichnen wir nun mit $\chi^2_{\frac{\alpha}{2}, n-1}$ bzw. $\chi^2_{1-\frac{\alpha}{2}, n-1}$ das $\alpha/2$ - bzw. $(1 - \alpha/2)$ -Fraktil einer Chi-Quadrat-Verteilung mit $n - 1$ Freiheitsgraden, so können wir analog zur bisherigen Vorgehensweise das Intervall

$$\chi^2_{\frac{\alpha}{2}, n-1} \leq \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2}, n-1}$$

finden. Aufgrund der fehlenden Symmetrie der Chi-Quadrat-Verteilung verwenden wir im Folgenden $c_1 = \chi^2_{\frac{\alpha}{2}, n-1}$ bzw. $c_2 = \chi^2_{1-\frac{\alpha}{2}, n-1}$. Weitere Umformungen liefern das symmetrische **Konfidenzintervall²** für die Varianz

$$[V_u, V_o] = \left[\frac{(n-1)\hat{\sigma}^2}{c_2}, \frac{(n-1)\hat{\sigma}^2}{c_1} \right], \quad (\text{III2: 12})$$

in welches die unbekannte Varianz σ^2 mit einer Wahrscheinlichkeit von $1 - \alpha$ fällt. Ein solches Konfidenzintervall ist jedoch nur bei tatsächlich normalverteilter Grundgesamtheit zu empfehlen. Bei nicht normalverteilter Grundgesamtheit erhalten wir selbst bei hohem Stichprobenumfang nicht generell eine akzeptable Intervallschätzung von σ^2 .

Beispiel III 2-6

Intervallschätzung für σ^2



Zur Bestimmung des Durchschnittsalters der Mitarbeiter in einem Großkonzern wurde eine Stichprobe im Umfang von $n = 20$ gezogen, die ein durchschnittliches Alter von $\bar{x} = 35$ Jahre bei einer Standardabweichung von $\hat{\sigma} = 3,5$ Jahren ergab. Das Lebensalter in der Grundgesamtheit sei normalverteilt. Die Varianz σ^2 dieser Grundgesamtheit ist jedoch nicht bekannt. Es soll daher eine Intervallschätzung für die Varianz σ^2 der Grundgesamtheit bei einer Irrtumswahrscheinlichkeit von 2 % durchgeführt werden.

1. Bestimmung des $(\alpha/2)$ - und des $(1 - \alpha/2)$ -Fraktils der $\chi^2(n-1)$ -Verteilung:

$$c_1 = \chi^2_{\frac{\alpha}{2}, n-1} = \chi^2_{0,01; 19} = 7,63 \quad c_2 = \chi^2_{1-\frac{\alpha}{2}, n-1} = \chi^2_{0,99; 19} = 36,19$$

2. Ermittlung der Stichprobenvarianz $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = 3,5^2 = 12,25$$

² Für die Bezeichnung "symmetrisches Konfidenzintervall" spielt es keine Rolle, ob die zugrundeliegende Verteilung symmetrisch (Normal- und t-Verteilung) oder nicht symmetrisch (Chi-Quadrat-Verteilung) ist. Wichtig ist hierfür nur, dass die Irrtumswahrscheinlichkeit auf beide Seiten des Konfidenzintervalls gleichmäßig ($\alpha/2$) aufgeteilt ist.

3. Bildung des Schätzintervalls:

$$\frac{(n-1)\hat{\sigma}^2}{c_2} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{c_1} \quad \rightarrow \quad \frac{19 \cdot 12,25}{36,19} \leq \sigma^2 \leq \frac{19 \cdot 12,25}{7,63}$$

$$6,43 \leq \sigma^2 \leq 30,50$$

2.5 Überblick über die behandelten Konfidenzintervalle

Abbildung III 8 bietet einen Überblick über die in diesem Kapitel behandelten Konfidenzintervalle. Es wird dabei vom Modell des Ziehens mit Zurücklegen bzw. des Modells des Ziehens ohne Zurücklegen mit erfüllter Bedingung $n \leq 0,1N$ ausgeganen (Vernachlässigung des Endlichkeitskorrekturfaktors $(N-n)(N-1)$). z ist dabei das $(1-\alpha/2)$ -Fraktil der Standardnormalverteilung, t das $(1-\alpha/2)$ -Fraktil einer t-Verteilung mit $n-1$ Freiheitsgraden sowie c_i das $\alpha/2$ -Fraktil und c_2 das $(1-\alpha/2)$ -Fraktil einer Chi-Quadrat-Verteilung mit $n-1$ Freiheitsgraden.

Konfidenzintervall für μ	
	Varianz σ^2
GG normalverteilt	bekannt unbekannt
GG beliebig verteilt, aber da $n > 30$ approxi- mativ normalverteilt	$\bar{X} - z \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z \frac{\sigma}{\sqrt{n}}$ $\bar{X} - t \frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{X} + t \frac{\hat{\sigma}}{\sqrt{n}}$
Konfidenzinverall für p	
Dichotome GG, aber da $n\hat{p}(1-\hat{p}) \geq 9$ approximativ normalverteilt	$\hat{p} - z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Konfidenzintervall für σ^2	
GG normalverteilt	$\frac{(n-1)\hat{\sigma}^2}{c_2} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{c_1}$

Abbildung III 8: Konfidenzintervalle im Überblick

2.6 Planung des Stichprobenumfangs

Bei der Planung von Stichprobenerhebungen stellt sich vorab die Frage, wie groß die Stichprobe zu wählen ist. Hierbei spielt natürlich die erforderliche Genauigkeit der Untersuchung die zentrale Rolle. Die Genauigkeit kann durch die tolerierte Länge L des Konfidenzintervalls festgelegt werden. Bei gegebenem Konfidenzniveau und gegebener Streuung der Grundgesamtheit hängt die Länge des Konfidenzintervalls nur noch vom Stichprobenumfang n ab. Durch die Wahl des Stichprobenumfangs kann somit über die Länge des Konfidenzintervalls die Untersuchungsgenauigkeit gesteuert werden.

2.6.1 Konfidenzintervall für das arithmetische Mittel

Die Länge eines Konfidenzintervalls für das arithmetische Mittel μ einer *normalverteilten* Grundgesamtheit mit *bekannter Varianz* ist festgelegt als

$$L = V_o - V_u = (\bar{X} + z\sigma_{\bar{X}}) - (\bar{X} - z\sigma_{\bar{X}}) = 2z\sigma_{\bar{X}} \quad (\text{III2: 13})$$

Der darin vorkommende Term $z\sigma_{\bar{X}}$ wird in der Literatur auch häufig als *absoluter Fehler ϵ* bezeichnet, sodass $L = 2\epsilon$ gilt. L hängt hier nicht vom Stichprobenergebnis ab, da σ bekannt ist. Bei gegebener Vertrauenswahrscheinlichkeit $1 - \alpha$ kann deshalb durch die Wahl eines geeigneten Stichprobenumfangs n erreicht werden, dass die Länge des Konfidenzintervalls höchstens gleich einem vorgegebenen Wert für L ist. Wir ermitteln den *mindestens* notwendigen Stichprobenumfang im Falle des *Ziehens mit Zurücklegen* wie folgt:

$$L = 2z \frac{\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{2\sigma z}{L} \right)^2 = \left(\frac{\sigma z}{\epsilon} \right)^2 \quad (\text{III2: 14})$$

Wir sehen daran, dass beispielsweise eine Halbierung der Länge des Konfidenzintervalls eine Vervierfachung des Stichprobenumfangs erfordert. Beim *Ziehen ohne Zurücklegen* (und $n > 0,1N$) ergibt sich (da anderes $\sigma_{\bar{X}}$) hingegen

$$n = \frac{z^2 \cdot N \cdot \sigma^2}{(L/2)^2(N-1) + z^2 \cdot \sigma^2} = \frac{z^2 \cdot N \cdot \sigma^2}{\epsilon^2(N-1) + z^2 \cdot \sigma^2}. \quad (\text{III2: 15})$$

Beispiel III 2-7

Notwendiger Stichprobenumfang

Im Beispiel III 2-2 hatten wir bei einem Stichprobenumfang von $n = 50$ eine Intervalllänge von $L = 4,68 - 4,32 = 0,36$ Jahren erhalten. Um ein Schätzintervall von höchstens 0,1 Jahren Länge zu erhalten, müsste ein Stichprobenumfang von mindestens

$$n = \left(\frac{2 \cdot 0,5 \cdot 2,58}{0,1} \right)^2 = 665,64$$

gewählt werden, d.h. es müssten mindestens 666 Personen befragt werden.

Im Falle *unbekannter Varianz* muss die Varianz in (III2: 14) bzw. (III2: 15) geschätzt werden (vgl. dazu Abschnitt III 2.6.2). Aufgrund dieser Schätzung ist die Berechnung eines exakten Mindeststichprobenumfangs n zur Festlegung einer Maxi-

mallänge L des Konfidenzintervalls nicht möglich. Der *erforderliche Stichprobenumfang* kann *nur approximativ* bestimmt werden.

2.6.2 Konfidenzintervall für den Anteils Wert

Die Länge des Konfidenzintervalls für den unbekannten Anteils Wert einer *dichotomen Grundgesamtheit* ergibt sich als

$$L = V_o - V_u = 2z\sigma_{\hat{p}}. \quad (\text{III2: 16})$$

Zur Berechnung der Varianz des Stichprobenanteils Wertes müsste man p kennen, was nun aber gerade nicht der Fall ist. Wäre **p bekannt**, so ließe sich zu einer vorgegebenen Länge L des Konfidenzintervalls der dazu notwendige Stichprobenumfang n ermitteln. Beim *Ziehen mit Zurücklegen* ergäbe sich

$$L = 2z\sqrt{\frac{p(1-p)}{n}} \rightarrow n = \frac{z^2 p(1-p)}{(L/2)^2} = \frac{z^2 p(1-p)}{\epsilon^2}, \quad (\text{III2: 17})$$

wohingegen wir beim *Ziehen ohne Zurücklegen* (und $n > 0,1N$) für die Länge (aufgrund des anderen $\sigma_{\hat{p}}$) den folgenden Wert erhalten:

$$n = \frac{z^2 N p(1-p)}{(L/2)^2(N-1) + z^2 p(1-p)} = \frac{z^2 N p(1-p)}{\epsilon^2(N-1) + z^2 p(1-p)} \quad (\text{III2: 18})$$

Bei **unbekanntem p** und damit unbekannter Varianz kann im Allgemeinen *kein* erforderlicher Stichprobenumfang n angegeben werden. In der Praxis haben sich aber verschiedene Möglichkeiten herausgebildet, trotzdem Werte für n zu erhalten. Sowohl beim Konfidenzintervall für das arithmetische Mittel als auch für den Anteils Wert wird häufig die unbekannte Varianz (bzw. p beim Konfidenzintervall für den Anteils Wert) zur Bestimmung von n durch eine Vorstichprobe kleinen Umfangs geschätzt oder aus früheren Erhebungen übernommen. Eine solche Vorgehensweise besitzt allerdings den Nachteil, dass der so gefundene notwendige Stichprobenumfang aufgrund der Schätzung der Varianz auch zu klein ausfallen kann. Stattdessen könnte man wie folgt vorgehen: Liegt eine Information vor, dass der Schätzwert für σ^2 mit Sicherheit kleiner oder gleich einer Konstanten k ist, so kann die Einhaltung einer maximalen Intervalllänge L dadurch garantiert werden, dass man mindestens

$$n = \left(\frac{2zk}{L}\right)^2 = \left(\frac{zk}{\epsilon}\right)^2 \quad (\text{III2: 19})$$

wählt. So ist beispielsweise die geschätzte Standardabweichung einer dichotomen Grundgesamtheit mit Sicherheit kleiner oder gleich 0,5, da $\sqrt{\hat{p}(1-\hat{p})}$ an der Stelle $\hat{p} = 0,5$ das Maximum 0,5 besitzt.

2.6.3 Konfidenzintervall für die Varianz

Die Länge des Konfidenzintervalls für die Varianz σ^2 ist wegen

$$L = V_o - V_u = (n-1)\hat{\sigma}^2 \cdot (1/c_1 - 1/c_2) \quad (\text{III2: 20})$$

selbst von der zu schätzenden Varianz abhängig. Die Bestimmung eines notwendigen Stichprobenumfangs n zur Erreichung einer festen Länge L ist also auch hier *nur approximativ* möglich.

3.

Testen von Hypothesen

In den vorhergehenden beiden Abschnitten (Punkt- und Intervallschätzung) hatten wir uns mit der Frage beschäftigt, wie man mit Hilfe von Zufallsstichproben unbekannte Parameter der Grundgesamtheit schätzen kann. Im Rahmen von Testverfahren soll nun die Frage beantwortet werden, wie man mit Hilfe von Zufallsstichproben testen (prüfen) kann, ob bestimmte Hypothesen (Annahmen, Behauptungen) über unbekannte Grundgesamtheiten richtig oder falsch sind. Dabei wird eine Hypothese als statistisch widerlegt angesehen (abgelehnt, verworfen), wenn das Stichprobenergebnis im deutlichen (signifikanten) Gegensatz dazu steht. Entsprechende Überprüfungsverfahren bezeichnet man daher als Hypothesentests oder kurz Tests. Wir unterscheiden allgemein Parametertests, die Hypothesen über unbekannte Parameter der Grundgesamtheit überprüfen, und Verteilungstests, die Hypothesen über die Verteilungsart einer Grundgesamtheit testen.

3.1 Allgemeines Testschema

Alle statistischen Tests sind nach einem *einheitlichen Schema* aufgebaut. Dieses setzt sich aus den folgenden vier Schritten zusammen:

1. Formulierung einer *Nullhypothese* und einer *Gegenhypothese* (Alternativhypothese) mit Festlegung eines Signifikanzniveaus α
2. Ableiten einer *Teststatistik* V und ihrer Verteilung
3. Aufstellung einer *Entscheidungsregel* (Definition des sog. Verwerfungsreiches) für die Ablehnung der Nullhypothese
4. *Entscheidung* und *Interpretation* auf Basis der berechneten Teststatistik

Die Bestandteile der einzelnen Testschritte dieses Standardschemas wollen wir im Folgenden näher betrachten:

1a. Die Hypothesen H_0 und H_A

Bei der Formulierung von Hypothesen sind allgemein zwei Typen, sog. Parameter- und Verteilungshypothesen, zu unterscheiden.

Verteilungshypothesen liegen vor, wenn eine Hypothese über die unbekannte Verteilungsform der Grundgesamtheit aufgestellt wird. So kann beispielsweise eine Behauptung "die Grundgesamtheit besitzt bezüglich des Untersuchungsmerkmals eine Normalverteilung" als Verteilungshypothese betrachtet werden. Mit Hypothesen dieser Art und den dazugehörigen Tests werden wir uns in Abschnitt III 3.4 näher befassen.

Von *Parameterhypothesen* sprechen wir, wenn sich die Hypothese auf einen unbekannten Parameter der Grundgesamtheit bezieht. An folgendem Beispiel wird zunächst erläutert, welche verschiedenen Formulierungen von Hypothesen möglich sind:

Nehmen wir an, der Vorsitzende einer großen Partei behauptet gegenüber einem Amtskollegen, dass der Anteil p seiner Wähler in der Bevölkerung bei p_0 liegt. Er stellt mit dieser Behauptung in der Terminologie der Testtheorie die sog. **Nullhypothese** (Ausgangshypothese, H_0) auf, die wir kurz wie folgt formulieren:

$$H_0: p = p_0$$

Wir sprechen in einem solchen Fall auch von einer *Punkthypothese*, da sie sich nur auf einen einzigen Wert, nämlich $p = p_0$, bezieht. Würde der Parteivorsitzende z.B. behaupten, der Wähleranteil sei mindestens p_0 , dann gilt $H_0: p \geq p_0$ und es läge eine sog. *Bereichshypothese* vor. Ebenso wäre $H_0: p \leq p_0$ eine Bereichshypothese.

Zu diesen Nullhypothesen kann der Amtskollege des Parteivorsitzenden nun jeweils eine **Gegenhypothese** (Alternativhypothese, H_A) formulieren. In der Regel ist die Gegenhypothese eine Bereichshypothese:

- a) $H_0: p = p_0 \quad -- \quad H_A: p \neq p_0$ (Gegenannahme des Amtskollegen: Wähleranteil entweder größer oder kleiner als p_0)
- b) $H_0: p \geq p_0 \quad -- \quad H_A: p < p_0$ (Gegenannahme des Amtskollegen: Wähleranteil ist kleiner als p_0)
- c) $H_0: p \leq p_0 \quad -- \quad H_A: p > p_0$ (Gegenannahme des Amtskollegen: Wähleranteil ist größer als p_0)

Als Gegenhypothese bei a) ist jedoch auch eine alternative Punkthypothese denkbar. So könnte der Amtskollege etwa behaupten, dass der Wähleranteil nicht bei p_0 sondern bei p_A liegt. Es ergäbe sich dann $H_A: p = p_A$. Einen solchen nicht sehr praxisrelevanten Fall behandeln wir exemplarisch in Beispiel III 3-3 und gehen sonst generell von den Fällen a) bis c) aus.

Liegen Hypothesen des Typs a) vor, sprechen wir von **zweiseitigen**, in den Fällen b) und c) von **einseitigen Tests**. Einige Autoren sprechen bei b) auch von *rechtsseitigen* und bei c) von *linksseitigen* Tests. Analog zu diesem Beispiel für den Anteilswert p sind auch die Hypothesen über den Mittelwert μ und die Varianz σ^2 der Grundgesamtheit möglich. Wie wir sehen werden, können auch über Differenzen und Quotienten von Parametern der Grundgesamtheit Hypothesen aufgestellt (und getestet) werden. Zur Überprüfung dieser Hypothesen wird eine Stichprobe gezogen, deren Ergebnis eine Wahrscheinlichkeitsaussage darüber ermöglichen soll, ob H_0 zutrifft oder nicht. Je nach Stichprobenergebnis werden wir H_0 entweder ablehnen (verwerfen) oder nicht ablehnen.

1b. Das Signifikanzniveau α

Die *Ablehnung* von H_0 anhand eines konkreten Stichprobenergebnisses ist entweder richtig oder falsch. Ist H_0 nicht korrekt, so haben wir mit der Ablehnung die richtige Entscheidung getroffen. Ist H_0 jedoch korrekt, war die getroffene Entscheidung falsch. Wir sprechen in einem solchen Fall von einem sog. **Fehler 1. Art** oder auch **α -Fehler** (vgl. Abbildung III 9), dem die Wahrscheinlichkeit α zugeordnet ist. Vor Durchführung des Tests muss der Anwender diese Irrtumswahrscheinlichkeit festlegen.

lichkeit (das sog. **Signifikanzniveau α**) festlegen. Er bestimmt also, mit welcher Wahrscheinlichkeit er die Fehlentscheidung zulassen will, dass H_0 zu Unrecht abgelehnt wird. Gebräuchliche Werte für α sind 0,1, 0,05 oder 0,01. Je kleiner α gewählt wird, desto unwahrscheinlicher wird die Fehlentscheidung. Damit wird jedoch auch die Chance geringer, bei falschem H_0 zu einer Ablehnung von H_0 zu kommen.

Auch die *Nichtablehnung von H_0* kann richtig oder falsch sein. Ist H_0 korrekt, so treffen wir die richtige Entscheidung. Bei nicht korrekter H_0 ist die Nichtablehnung ein Fehler. Diesen bezeichnen wir auch als **Fehler 2. Art** oder **β -Fehler**. Auch er tritt mit einer gewissen Wahrscheinlichkeit auf, welche nur indirekt vom Anwender beeinflusst werden kann. Wir werden dies z.B. im Abschnitt III 3.3.1.1 konkret sehen. Abbildung III 9 fasst diese Gedanken nochmals zusammen.

		Entscheidung	
Wahrer Zustand		Nichtablehnung von H_0	Ablehnung von H_0
H_0 korrekt		Richtige Entscheidung	α-Fehler (Fehler 1. Art)
H_0 nicht korrekt		β-Fehler (Fehler 2. Art)	Richtige Entscheidung

Abbildung III 9: Fehler 1. und 2. Art

Beispiel III 3-1

Fehler 1. und 2. Art

Eine Person wird eines Verbrechens angeklagt, wobei nicht bekannt ist, ob sie dieses auch wirklich begangen hat. Unter der Nullhypothese H_0 : "Der Angeklagte ist unschuldig." können folgende juristische Entscheidungssituationen auftreten:

Angeklagter ist ...	Entscheidung des Gerichts	
	Freispruch (Nichtablehnung von H_0)	Schuldspruch (Ablehnung von H_0)
... unschuldig	Richtige Entscheidung	Justizirrtum zugunsten des Angeklagten (α -Fehler)
... schuldig	Justizirrtum zugunsten des Angeklagten (β -Fehler)	Richtige Entscheidung

Ziel eines Statistikers ist es, einen Test zu konzipieren, bei dem die Wahrscheinlichkeit für den α -Fehler und auch die für den β -Fehler in vertretbaren Grenzen gehalten wird. Präziser ausgedrückt: Es soll für einen gegebenen α -Fehler der geringstmögliche β -Fehler (mächtigster Test) erzielt werden. Problematisch ist dabei jedoch: Je niedriger die Wahrscheinlichkeit für den α -Fehler ausfällt, desto höher ist ceteris paribus (c.p., unter Konstanthaltung aller anderen Faktoren) diejenige für den β -Fehler. Würden wir nämlich H_0 generell ablehnen (in Beispiel III 3-1: nie Freispruch), gilt

$$P(\alpha\text{-Fehler}) = P(\text{Ablehnung} \mid H_0 \text{ korrekt}) = 1$$

$$P(\beta\text{-Fehler}) = P(\text{Nichtablehnung} \mid H_0 \text{ nicht korrekt}) = 0.$$

Würden wir H_0 dagegen nie ablehnen (in Beispiel III 3-1: immer Freispruch), gilt

$$P(\alpha\text{-Fehler}) = P(\text{Ablehnung} \mid H_0 \text{ korrekt}) = 0$$

$$P(\beta\text{-Fehler}) = P(\text{Nichtablehnung} \mid H_0 \text{ nicht korrekt}) = 1.$$

Es ist daher in der Regel nicht möglich beide Fehlerwahrscheinlichkeiten simultan unter beliebig vorgegebene Schranken zu drücken.

2. Die Teststatistik V

Die *Teststatistik* oder *Testfunktion* V verarbeitet die vorliegende konkrete Stichprobe und liefert einen Zahlenwert, der einen Beitrag zur Entscheidung der Ablehnung oder Nichtablehnung von H_0 leistet. An eine solche Funktion V stellen wir folgende *Anforderungen*, die wir im Rahmen konkreter Tests noch ausführlicher betrachten:

- Die Wahrscheinlichkeitsverteilung von V muss davon abhängen, ob H_0 oder H_A richtig ist.
- Die Verteilung von V muss, sofern die Nullhypothese zutrifft, zumindest approximativ bestimmt werden können, da nur dann nachgeprüft werden kann, ob das gegebene Signifikanzniveau α eingehalten wird.

Je nach Testverfahren werden wir andere Funktionen für V erhalten. Wir können daher festhalten, dass V die entscheidende *Größe* ist, die die einzelnen Tests von einander unterscheidet.

3. Die Entscheidungsregel

Zur Entscheidung, ob die vorliegende Nullhypothese anhand einer konkreten Stichprobe bzw. des Wertes v (mit Daten der Stichprobe berechnete Realisation der Zufallsvariablen V), abgelehnt werden kann oder nicht, wird ein sog. **Verwerfungsbereich** (oder kritischer Bereich) definiert. Fällt v in den Verwerfungsbereich, so wird H_0 abgelehnt. Bei der Konstruktion dieses Bereichs wird aufgrund des Problems, dass die Wahrscheinlichkeiten für α -Fehler und β -Fehler nicht gleichzeitig unter eine vorgegebene Grenze gedrückt werden können, wie folgt vorgegangen:

1. Wir stellen sicher, dass die Wahrscheinlichkeit für den Fehler 1. Art maximal α ist. [In den Fällen b) und c) kann es vorkommen, dass diese kleiner als α ist. Vgl. dazu Abschnitt II 3.3.1.1.]
2. Der nach 1. noch verbleibende Spielraum wird dann dazu benutzt, die Wahrscheinlichkeit für den β -Fehler so gering wie möglich zu halten. Wir erreichen dies, indem wir den Verwerfungsbereich so legen, dass ein Testfunktionswert v, der in diesen fällt, stark gegen H_0 und für H_A spricht.

Diese recht allgemeine Beschreibung zur Konstruktion des Verwerfungsbereiches werden wir bei den einzelnen Tests detaillierter betrachten.

4. Entscheidung und Interpretation

Je nachdem, ob v in den Verwerfungsbereich fällt oder nicht, ergibt sich als Ergebnis eines Tests die Ablehnung bzw. Nichtablehnung der Nullhypothese.

Fällt v in den Verwerfungsbereich sagen wir, dass die Stichprobe in signifikantem Widerspruch zu H_0 steht. Die daraus folgende **Ablehnung von H_0** wird als **Annahme der gewählten Gegenhypothese H_A** interpretiert. Genau genommen stellen

wir mit der Ablehnung von H_0 die statistische Behauptung auf, dass H_A korrekt ist. Diese kann richtig oder falsch sein. Die Wahrscheinlichkeit dafür, dass der Test zu dieser Behauptung führt, obwohl sie falsch ist, liegt jedoch nur bei α .

Fällt v nicht in den Verwerfungsbereich, steht die Stichprobe nicht in signifikantem Widerspruch zu H_0 . H_0 kann demnach **nicht abgelehnt** werden. Dies bedeutet **nicht, dass die Nullhypothese bestätigt ist**, sondern nur, dass sie anhand der vorliegenden Stichprobe nicht abgelehnt werden kann.

v innerhalb des Verwerfungsbereichs	H_0 kann abgelehnt werden → H_A statistisch bestätigt
v außerhalb des Verwerfungsbereichs	H_0 kann nicht abgelehnt werden → H_0 nicht bestätigt

Abbildung III 10: Interpretation eines Testergebnisses

Es gilt also allgemein, dass Tests nur zu einer Widerlegung (Falsifizierung) einer Nullhypothese und nicht zu deren Bestätigung herangezogen werden können. In der **Praxis** wird daher die *Hypothese, die "bestätigt" werden soll, als Gegenhypothese* formuliert. Dementsprechend kann dann die Ablehnung von H_0 als statistische Bestätigung der H_A mit Signifikanzniveau α interpretiert werden.

3.2 Testklassifizierung

Neben der bereits unter III 3.1 erwähnten Unterscheidung von Parameter- und Verteilungstests ist eine weitere wichtige Klassifizierung von Tests vorzunehmen:

Ist die Grundlage des Tests *eine einfache Stichprobe aus einer* Grundgesamtheit, so sprechen wir von einem sog. **Einstichprobentest**. Zu dieser Klasse der Tests, gehören beispielsweise die Einstichprobentests für den Anteilswert p (vgl. III 3.3.1.1), das arithmetische Mittel μ (vgl. III 3.3.1.2) und die Varianz σ^2 (vgl. III 3.3.1.4) sowie der χ^2 -Anpassungstest (vgl. III 3.4.1).

Hypothesen beziehen sich oft auch auf den Vergleich von Parametern verschiedener Grundgesamtheiten. Wir wollen uns hier auf zwei Grundgesamtheiten beschränken und Behauptungen wie "Die mittlere Lebensdauer von Motorentyp A liegt mindestens bei der von Motorentyp B." oder "Die Anteile defekter Artikel in Lieferung A und B sind identisch." untersuchen. Sind die zum Test solcher Hypothesen gezogenen *Stichproben voneinander unabhängig*, sprechen wir von **Zweistichprobentests**. Zu diesen zählen etwa die Zweistichprobentests für die Differenz zweier arithmetischer Mittel (vgl. III 3.3.2.1), die Differenz zweier Anteilswerte (vgl. III 3.3.2.2), und den Quotienten zweier Varianzen (vgl. III 3.3.2.3).

Wird aus einer Grundgesamtheit eine Stichprobe im Umfang n gezogen, bei der von jedem entnommenen Objekt zu zwei relevanten Untersuchungsmerkmalen X und Y die beiden Merkmalsausprägungen registriert werden (sog. zweidimensionale Stichprobe), sprechen wir bei getrennter Betrachtung der Datenerhebung zu X und zu Y vom Vorliegen zweier *verbundener einfacher Stichproben* vom Umfang n. Ein Test, der Hypothesen bezüglich solcher Stichproben untersucht, wird auch

als **Zweistichprobentest bei verbundenen Stichproben** bezeichnet. Hypothesen können sich hierbei beispielsweise mit folgenden Fragestellungen befassen:

- Hängen die Studienleistungen von Studierenden einer Hochschule in den Fächern Mathematik und Statistik voneinander ab oder nicht?
- Ist der Ernteertrag bei Verwendung des Düngemittels A höher als der bei Verwendung von Düngemittel B? Um hier die erforderliche *Abhängigkeit* der Stichproben zu gewährleisten, wird z.B. ein Feld je zur Hälfte mit Dünger A und B behandelt. Würden verschiedene Felder gedüngt, läge ein normaler Zweistichprobentest (da Unabhängigkeit) vor.

Zu dieser Art von Tests zählen etwa die Differenzentests für zwei arithmetische Mittel (vgl. III 3.3.3.1) sowie der Korrelationstest (vgl. III 3.3.3.2) und der Kontingenztest (vgl. III 3.4.2).

3.3 Parametertests

3.3.1 Einstichprobentests

Bei Einstichprobentests wird die Entscheidung über die Ablehnung oder Nichtablehnung der Nullhypothese anhand einer *einzigsten* Stichprobe getroffen. Wir wollen im Folgenden den Einstichprobentest für den Anteilswert, der genau genommen nur ein Sonderfall des Einstichprobentests für das arithmetische Mittel bei dichotomer Grundgesamtheit ist, verwenden, um den konkreten Aufbau eines derartigen Tests im Detail zu beschreiben. Die darauf folgenden Testverfahren werden wir dann in kompakterer Form darstellen.

3.3.1.1 Einstichprobentest für den Anteilswert

Mit dem Einstichprobentest für den Anteilswert, der auch als *approximativer Gaußtest* bezeichnet wird, lassen sich Behauptungen hinsichtlich des unbekannten Anteilswerts p einer Grundgesamtheit untersuchen. So können z.B. Hypothesen wie "Der Anteil defekter Artikel in einer großen Lieferung liegt bei p_0 ." oder "Der Anteil der Raucher unter den Studenten einer Hochschule liegt bei mindestens p_0 ." getestet werden. Angelehnt am allgemeinen Schema aus III 3.1 gehen wir dabei konkret wie folgt vor:

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Wir können bei diesem Test allgemein folgende mögliche Paare von Hypothesen H_0 und H_A festhalten:

- a) $H_0: p = p_0 \quad -- \quad H_A: p \neq p_0$
 - b) $H_0: p \geq p_0 \quad -- \quad H_A: p < p_0$
 - c) $H_0: p \leq p_0 \quad -- \quad H_A: p > p_0$
- (III3: 1)

Bei der Formulierung der Nullhypotesen der einseitigen Tests b) und c) haben wir bewusst zwei mögliche Formulierungen angegeben. Es gilt nämlich, dass die Wahrscheinlichkeit für den Fehler 1. Art bei Test c) nur für $H_0: p = p_0$ genau beim gewählten Signifikanzniveau α liegt. Für jeden anderen Wert $p < p_0$ wäre die Irr-

tumswahrscheinlichkeit kleiner als das gewählte Niveau α . Das Signifikanzniveau α ist daher im Falle einer Formulierung $H_0: p \leq p_0$ als *maximale Irrtumswahrscheinlichkeit* aufzufassen. Dies gilt analog auch für den Fall b).

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Nach Wahl von H_0 und H_A sowie des Signifikanzniveaus α besteht der nächste Schritt darin, eine geeignete Teststatistik (zur Verarbeitung der vorliegenden Stichprobe) und ihre Verteilung, die sog. *Testverteilung*, herzuleiten. Unter der *Annahme*, dass die Nullhypothese zutrifft, kann die vorliegende Stichprobe als eine aus einer dichotomen Grundgesamtheit mit dem Anteilswert p_0 gezogene betrachtet werden. Der Stichprobenanteilswert \hat{p} ist dann nach den Erkenntnissen aus Abschnitt III 1.2.3 näherungsweise normalverteilt, wenn $np_0(1-p_0) \geq 9$ gilt. Für Erwartungswert, Varianz und Standardabweichung von \hat{p} gilt dann bei Ziehen mit Zurücklegen (bzw. Ziehen ohne Zurücklegen und $n \leq 0,1N$)

$$E(\hat{p}) = p_0,$$

$$\text{Var}(\hat{p}) = \frac{p_0(1-p_0)}{n} \quad \rightarrow \quad \sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}.$$

Als *Teststatistik* V verwenden wir nun nicht \hat{p} sondern zweckmäßigerweise den standardisierten Stichprobenanteilswert $(\hat{p} - p_0) / \sigma_{\hat{p}}$, sodass (bei Vernachlässigung einer Stetigkeitskorrektur) die Teststatistik

$$V = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (\text{III3: 2})$$

resultiert. V ist *bei richtiger Nullhypothese* aufgrund des Zentralen Grenzwertsatzes *approximativ standardnormalverteilt*. Wir sagen daher auch, dass die Standardnormalverteilung die *Testverteilung* ist.

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Wie wir wissen, ist der Stichprobenanteilswert \hat{p} ein geeigneter (erwartungstreuer und konsistenter) Schätzer für den unbekannten Anteilswert p einer Grundgesamtheit. Wir können also damit rechnen, dass die Realisierung von \hat{p} nicht allzu sehr von p abweicht. Deshalb werden wir uns, falls \hat{p} stark gegen die Nullhypothese H_0 und für die betrachtete Gegenhypothese H_A spricht, dazu entschließen, H_0 als statistisch widerlegt anzusehen. Dies bedeutet in den einzelnen Fällen aus (III3: 1):

H_0 wird gegenüber H_A abgelehnt

- a) $H_0: p = p_0, H_A: p \neq p_0 \quad | \hat{p} - p_0 | \text{ "sehr groß" ist.} \quad (\text{III3: 3})$
- b) $H_0: p \geq p_0, H_A: p < p_0 \quad \text{falls } \hat{p} \text{ "viel kleiner" als } p_0 \text{ ist.}$
- c) $H_0: p \leq p_0, H_A: p > p_0 \quad \hat{p} \text{ "viel größer" als } p_0 \text{ ist.}$

Diese Entscheidungsregel gilt es nun zu präzisieren, da noch nicht klar definiert ist, was unter "sehr groß", "viel kleiner" und "viel größer" konkret zu verstehen ist. Da die Fehlentscheidung, dass eine korrekte Nullhypothese abgelehnt wird, lediglich mit der als zulässig vorgegebenen kleinen Irrtumswahrscheinlichkeit α vorkommen darf, muss zudem dieses α in die Entscheidungsregel Eingang finden.

Ob eine Abweichung des Stichprobenwerts vom hypothetischen Wert als deutlich (signifikant, groß) angesehen werden kann, können wir unter Berücksichtigung der Streuung von \hat{p} beurteilen. Wir verwenden die Teststatistik (III3: 2), in die diese Streuung direkt einfließt. Da diese (unter der Annahme einer korrekten Nullhypothese) approximativ standardnormalverteilt ist, können wir unter Verwendung des $(1 - \alpha/2)$ -Fraktils $z_{1-\alpha/2}$ bzw. des $(1 - \alpha)$ -Fraktils $z_{1-\alpha}$ der Standardnormalverteilung und der Realisation v der Teststatistik die folgende **Entscheidungsregel**

- | | |
|--------------------------------------|---------------------------|
| H_0 wird gegenüber H_A abgelehnt | |
| a) $H_0: p = p_0, H_A: p \neq p_0$ | $ v > z_{1-\alpha/2}$ |
| b) $H_0: p \geq p_0, H_A: p < p_0$ | falls $v < -z_{1-\alpha}$ |
| c) $H_0: p \leq p_0, H_A: p > p_0$ | $v > z_{1-\alpha}$ |
- (III3: 4)

aufstellen. Die in (III3: 4) verwendeten Fraktile der Standardnormalverteilung bezeichnet man in der Testtheorie auch als **kritische Werte**. Sie trennen den **Ablehnungsbereich** (kritischen Bereich) für H_0 vom **Nichtablehnungsbereich** für H_0 . Abbildung III 11 verdeutlicht diese Bereiche für den Fall a). Fällt also v zwischen die beiden kritischen Werte, wird die Nullhypothese H_0 nicht abgelehnt, d.h. die Abweichung von \hat{p} von p_0 wird als nicht signifikant angesehen. Außerhalb dieses Bereiches ist H_0 zu verwerfen, d.h. die Abweichung von Stichprobe und Nullhypothese wird als signifikant bewertet. Ist die Nullhypothese korrekt, kommt es bei Verwendung der Entscheidungsregel (III3: 4) tatsächlich nur mit einer Wahrscheinlichkeit von *genau α* ($= \alpha/2 + \alpha/2$) zu einer fälschlichen Ablehnung von H_0 .

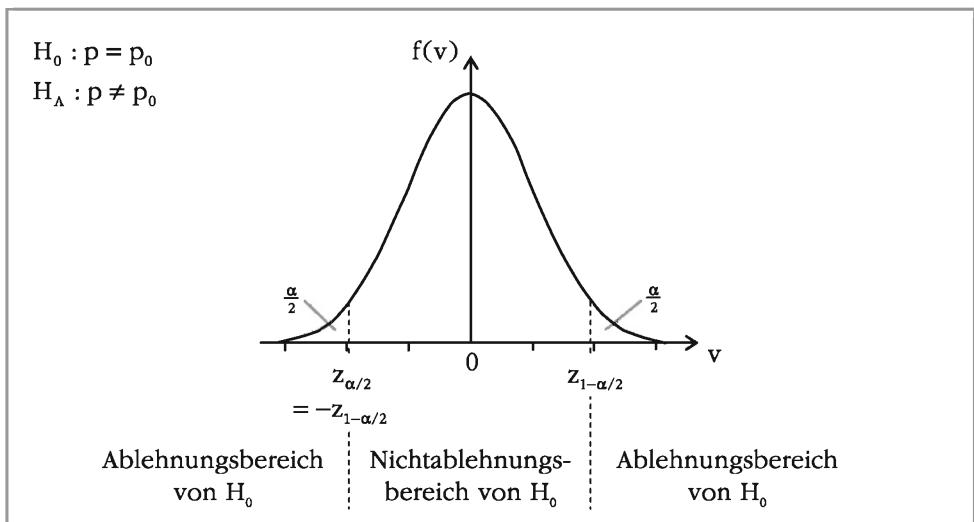


Abbildung III 11: Dichtefunktion von V bei korrekter Nullhypothese a)

Die entsprechenden Ablehnungs- und Nichtablehnungsbereiche für die Fälle b) und c) aus (III3: 4) sind in den beiden Abbildungen III 12 und III 13 dargestellt. Im Fall b) ist die Nullhypothese abzulehnen, wenn v unter $-z_{1-\alpha}$ liegt. Im Fall c) wird H_0 verworfen, wenn v über $z_{1-\alpha}$ liegt. Die Irrtumswahrscheinlichkeit beträgt in beiden Fällen *maximal α* . Bei diesen Entscheidungsregeln ist das Signifikanzniveau nur für $p = p_0$ genau α . Für die anderen möglichen Werte bei Richtigkeit der Null-

hypothesen (für den Fall b) $p > p_0$ und für den Fall c) $p < p_0$) wäre das Signifikanzniveau dann kleiner als α .

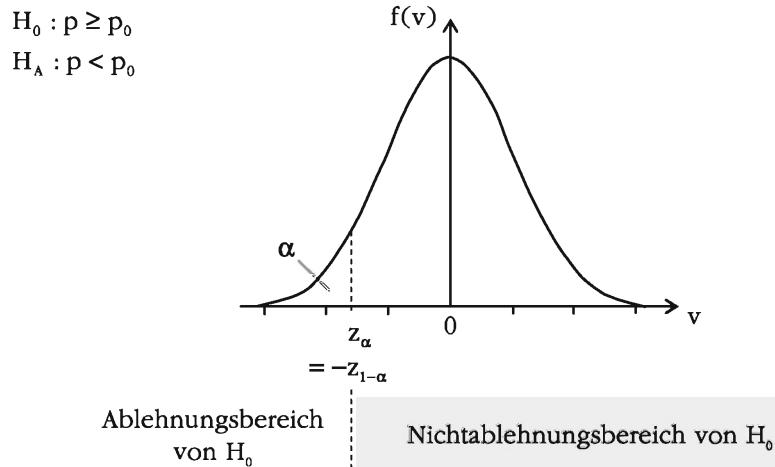


Abbildung III 12: Dichtefunktion von V bei korrekter Nullhypothese b)

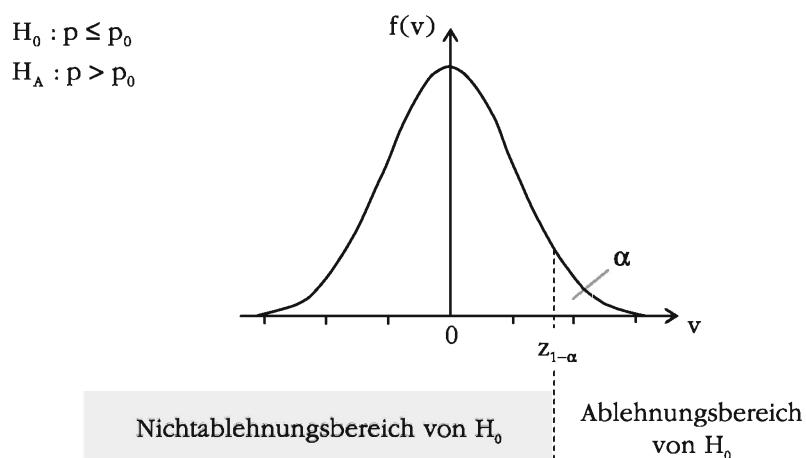


Abbildung III 13: Dichtefunktion von V bei korrekter Nullhypothese c)

(4) Entscheidung auf Basis der berechneten Teststatistik

Anhand des konkreten Wertes v und der Entscheidungsregel (III3: 4) kann letztlich eine Testentscheidung getroffen werden. Wird die Nullhypothese zum Signifikanzniveau α abgelehnt, sagen wir, dass der Test zu einem *signifikanten Ergebnis* gekommen ist. Kann die Nullhypothese nicht verworfen werden, so ist das Testergebnis nicht signifikant. Wie die Ablehnung oder Nichtablehnung zu werten sind, wurde bereits im Abschnitt III 3.1 unter Punkt 4 behandelt. Eine Ablehnung von H_0 bedeutet Bestätigung von H_A , die Nichtablehnung von H_0 jedoch keine Bestätigung von H_0 .

Alternativ könnte die Entscheidung über die Ablehnung oder Nichtablehnung von H_0 auch anhand des sog. *kritischen Anteilswerts* \hat{p}_{krit} erfolgen. Dieser ergibt sich in den Fällen b) und c) wie folgt:

Fall b): Ablehnung von H_0 , falls $v < -z_{1-\alpha}$

$$\frac{\hat{p}_{\text{krit}} - p_0}{\sigma_{\hat{p}}} = -z_{1-\alpha} \rightarrow \hat{p}_{\text{krit}} = p_0 - z_{1-\alpha} \sigma_{\hat{p}} \quad (\text{III3: 5})$$

Fall c): Ablehnung von H_0 , falls $v > z_{1-\alpha}$

$$\frac{\hat{p}_{\text{krit}} - p_0}{\sigma_{\hat{p}}} = z_{1-\alpha} \rightarrow \hat{p}_{\text{krit}} = p_0 + z_{1-\alpha} \sigma_{\hat{p}} \quad (\text{III3: 6})$$

Die Nullhypothese H_0 wird unter Verwendung dieser kritischen Anteilswerte verworfen, wenn im Fall b) die Bedingung $\hat{p} < \hat{p}_{\text{krit}}$ erfüllt ist. Im Fall c) muss entsprechend $\hat{p} > \hat{p}_{\text{krit}}$ gelten. Im Fall a) wird (III3: 5) mit \hat{p}_{krit}^u und (III3: 6) mit \hat{p}_{krit}^l bezeichnet, wobei an die Stelle von $z_{1-\alpha}$ jeweils das Fraktil $z_{1-\alpha/2}$ tritt. Eine Ablehnung der Nullhypothese H_0 erfolgt dann, wenn \hat{p} außerhalb des geschlossenen Intervalls $[\hat{p}_{\text{krit}}^l, \hat{p}_{\text{krit}}^u]$ liegt.

In Fall a) kann die Nullhypothese auch sehr einfach mit einem *Konfidenzintervall* geprüft werden. Wie wir aus (III3: 4) wissen, wird $H_0: p = p_0$ nicht abgelehnt, falls

$$|V| = \left| \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \right| \leq z_{1-\alpha/2}$$

erfüllt ist. Nehmen wir daran einige Umstellungen vor, erhalten wir:

$$-z_{1-\alpha/2} \leq \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \leq z_{1-\alpha/2}$$

$$\hat{p} - z_{1-\alpha/2} \sigma_{\hat{p}} \leq p_0 \leq \hat{p} + z_{1-\alpha/2} \sigma_{\hat{p}}$$

Dieser Ausdruck entspricht genau dem Konfidenzintervall für den Anteilswert p_0 . Wir können daher festhalten, dass H_0 abgelehnt werden kann, wenn der hypothetische Parameterwert p_0 außerhalb der vorliegenden Realisation dieses Konfidenzintervalls liegt. Diese alternativen Methoden der Testentscheidung sind jedoch nur zur Vollständigkeit und Veranschaulichung des Zusammenhangs zwischen Tests und Konfidenzintervallen aufgeführt. Im Folgenden beschränken wir uns auf die Anwendung der Entscheidungsregeln (III3: 4).

Beispiel III 3-2

Einstichprobentest für den Anteilswert



Ein Fabrikant behauptet, dass eine von ihm gelieferte Ladung eines Massenartikels einen Ausschussanteil p von höchstens 0,10 enthält. Der Empfänger ist aber der Meinung, dass der Anteil größer ist. Eine von ihm gezogene Stichprobe im Umfang von 100 Stück ergab 13 fehlerhafte Produkte. Kann aufgrund dieser Stichprobe die Behauptung des Lieferanten bei einem Signifikanzniveau von 0,05 verworfen werden?

(1) Hypothesen und Signifikanzniveau:

$$H_0: p \leq 0,10 \quad H_A: p > 0,10 \quad \alpha = 0,05$$

(2) Teststatistik und deren Verteilung:

Einstichprobentest für den Anteilswert:

$$V = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{a}{\sim} N(0;1), \text{ falls } H_0 \text{ richtig, da } np_0(1-p_0) \geq 9 \text{ (hier: } 100 \cdot 0,10 \cdot 0,90 = 9)$$

In konkreter Stichprobe:

$$v = \frac{\frac{13}{100} - 0,10}{\sqrt{\frac{0,10 \cdot 0,90}{100}}} = \frac{0,03}{0,03} = 1$$

(3) Entscheidungsregel:

Fall c) liegt vor. → H_0 ablehnen gegenüber H_A , wenn $v > z_{1-\alpha}$ gilt.

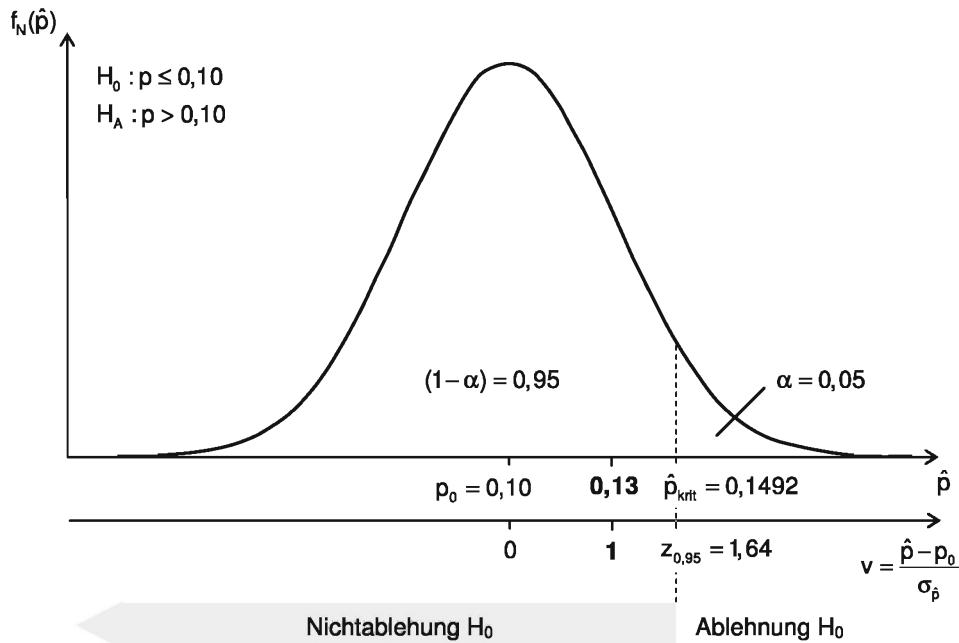
Dabei ist $z_{1-\alpha} = z_{1-0,05} = z_{0,95} = 1,64$.

(4) Entscheidung:

Es gilt hier ($v = 1$) $<$ ($z_{1-\alpha} = 1,64$).

Die Nullhypothese kann nicht abgelehnt werden. Es ist damit allerdings nicht verifiziert, dass der Ausschussanteil in der Lieferung tatsächlich höchstens 0,10 beträgt. Es ist lediglich die Aussage erlaubt, dass das Stichprobenergebnis mit der Behauptung des Lieferanten "nicht im Widerspruch" steht, eine beobachtete größere Abweichung zwischen \hat{p} und p_0 also möglicherweise dem Zufall zuzuschreiben ist.

Grafisch lässt sich dieser Test folgendermaßen skizzieren. Der kritische Anteilswert beträgt dabei $\hat{p}_{krit} = p_0 + z_{1-\alpha} \sigma_{\hat{p}} = 0,10 + 1,64 \cdot 0,03 = 0,1492$.



Wenden wir uns nun kurz dem Fall zu, dass sowohl H_0 als auch H_A Punkthypothesen sind. Dabei ist die konkrete Bestimmung der Wahrscheinlichkeit für den β -Fehler möglich. Dazu folgendes Beispiel III 3-3.

Beispiel III 3-3Wahrscheinlichkeit des β -Fehlers

Nehmen wir an, die Behauptung des Lieferanten aus Beispiel III 3-2 wäre $p = 0,1$ und die des Empfängers, dass der Ausschussanteil in der Lieferung bei genau $0,20$ liegt. Es wäre damit bei einem Signifikanzniveau von $\alpha = 0,05$ die Nullhypothese $H_0: p = 0,10$ gegenüber $H_A: p = 0,20$ zu testen.

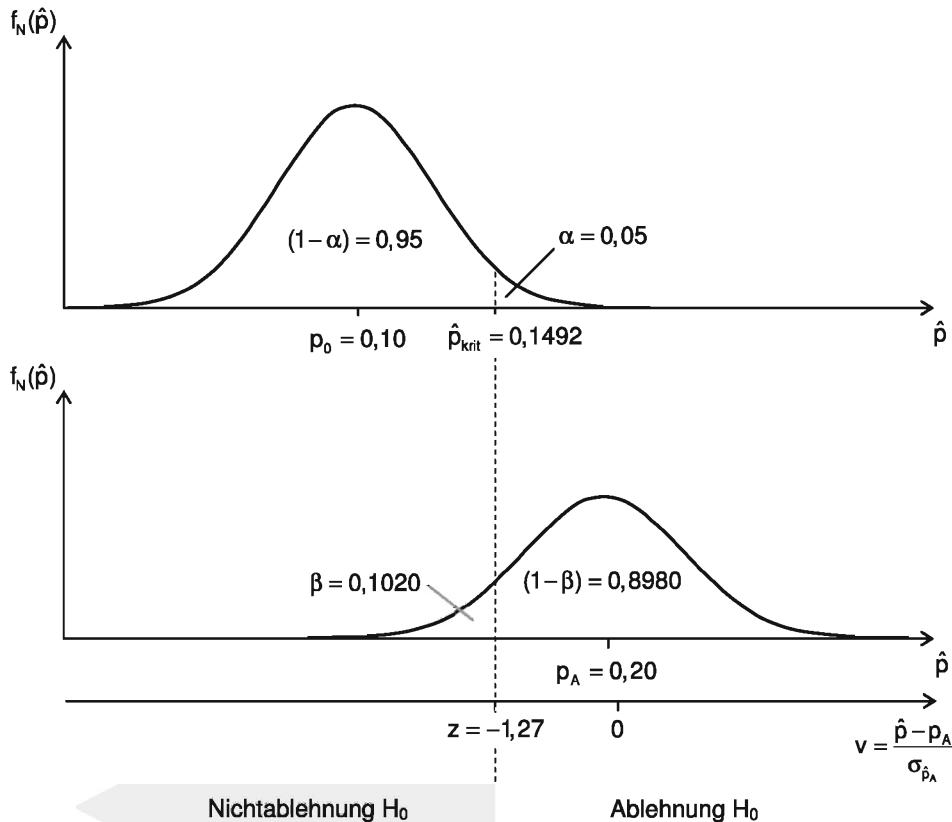
Sollte H_A korrekt sein, ist \hat{p} mit dem Erwartungswert $E(\hat{p}) = p_A = 0,20$ und der Varianz

$$\sigma_{\hat{p}_A}^2 = \frac{p_A(1-p_A)}{n} = \frac{0,20 \cdot 0,80}{100} = 0,0016 \quad \rightarrow \quad \sigma_{\hat{p}_A} = \sqrt{0,0016} = 0,04$$

normalverteilt. Die Wahrscheinlichkeit des β -Fehlers, also die Wahrscheinlichkeit dafür, dass die falsche H_0 nicht abgelehnt wird, lässt sich damit unter Verwendung des zum Signifikanzniveau α berechneten kritischen Anteilswertes \hat{p}_{krit} bestimmen. H_0 wird nicht abgelehnt, wenn die Bedingung $\hat{p} \leq \hat{p}_{krit}$ erfüllt ist. Es ergibt sich damit

$$\beta = P(\hat{p} \leq \hat{p}_{krit}) = \Phi\left(\frac{\hat{p}_{krit} - p_A}{\sigma_{\hat{p}_A}}\right) = \Phi\left(\frac{0,1492 - 0,20}{0,04}\right) = \Phi(-1,27) = 0,1020.$$

Die Teststatistik des Einstichprobentests für den Anteilswert besitzt bei Gültigkeit von H_0 eine andere Verteilung als bei Gültigkeit von H_A . Beide Verteilungen sowie die Bereiche von α - und β -Fehler sind im Folgenden skizziert:



Das Risiko eine korrekte Nullhypothese abzulehnen wird durch α , das Risiko eine falsche Nullhypothese nicht abzulehnen durch β ausgedrückt. Die Grafik zeigt sehr schön, dass bei Reduzierung des Signifikanzniveaus α ceteris paribus β zunimmt. Eine gleichzeitige Verminderung beider Fehlerrisiken wäre nur durch eine Erhöhung des Stichprobenumfangs n möglich, da dadurch die Varianzen der Schätzer sinken, die in die Berechnung der Teststatistiken eingehen.

3.3.1.2 Einstichprobentest für das arithmetische Mittel

Da der Einstichprobentest für den Anteilswert ein Sonderfall des Einstichprobentests für das arithmetische Mittel ist, ist die Vorgehensweise bei der Testdurchführung ähnlich. Wie bereits bei der Bestimmung des Konfidenzintervalls für das arithmetische Mittel (vgl. Abschnitt III 2.2), müssen wir jedoch wieder vier Fälle hinsichtlich der Verteilung der Grundgesamtheit und der "Bekanntheit" ihrer Varianz σ^2 unterscheiden. Je nach Fall, wird eine andere Teststatistik bzw. Testverteilung herangezogen. Auch die genaue Testbezeichnung ist jeweils anders.

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Es sind für diesen Test folgende Paare von Hypothesen H_0 und H_A denkbar:

- a) $H_0: \mu = \mu_0 \quad -\quad H_A: \mu \neq \mu_0$
- b) $H_0: \mu \geq \mu_0 \quad -\quad H_A: \mu < \mu_0$ (III3: 7)
- c) $H_0: \mu \leq \mu_0 \quad -\quad H_A: \mu > \mu_0$

Typische Behauptungen, die mit diesem Test untersucht werden können, sind also etwa "Eine Abfüllanlage hält die vorgeschriebene Sollabfüllmenge μ_0 ein." (a), "Das Durchschnittsgewicht von Mastschweinen konnte durch einen neuen Futterzusatz auf mindestens μ_0 gesteigert werden." (b) oder "Das Durchschnittseinkommen liegt in einem Landkreis bei höchstens μ_0 ." (c).

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Ist die **Grundgesamtheit normalverteilt** und deren **Varianz σ^2 bekannt**, ist die Zufallsvariable \bar{X} mit $E(\bar{X}) = \mu$ und $Var(\bar{X}) = \sigma^2 / \sqrt{n}$ (beim Ziehen mit Zurücklegen bzw. beim Ziehen ohne Zurücklegen und $n \leq 0,1N$) normalverteilt (vgl. Abschnitt III 1.2.2.1). Als Teststatistik V bietet sich daher die Gaußstatistik an. Diese ist unter der Bedingung, dass H_0 richtig ist, **standardnormalverteilt**.

$$V = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0; 1) \quad (\text{III3: 8})$$

Ist die **Grundgesamtheit normalverteilt**, ihre **Varianz σ^2 jedoch unbekannt**, verwenden wir als Schätzer für die Varianz der Grundgesamtheit die Stichprobenvarianz $\hat{\sigma}^2$. Als Teststatistik V bietet sich daher die t-Statistik an, welcher eine **t-Verteilung** mit $w = n - 1$ Freiheitsgraden zu Grunde liegt (vgl. Abschnitt III 1.2.2.1), sofern H_0 richtig ist.

$$V = \frac{\bar{X} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t(n - 1) \quad (\text{III3: 9})$$

Ist die **Grundgesamtheit beliebig verteilt**, der Stichprobenumfang n jedoch ausreichend groß, sind bei **bekannter Varianz σ^2**

$$V = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \stackrel{a}{\sim} N(0; 1) \quad (\text{III3: 10})$$

bzw. bei **unbekannter Varianz σ^2**

$$V = \frac{\bar{X} - \mu_0}{\hat{\sigma}_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}} \stackrel{a}{\sim} N(0; 1) \quad (\text{III3: 11})$$

geeignete Testfunktionen. Nach dem Zentralen Grenzwertsatz sind diese beiden Testfunktionen für $n > 30$ bei zutreffender Nullhypothese H_0 hinreichend genau **standardnormalverteilt**. Die beim Einstichprobentest für den Anteilswert verwendete Teststatistik V ist ein Sonderfall von (III3: 10).

Bei Fall (III3: 8) sprechen wir allgemein vom sog. *einfachen Gaußtest*. Der Fall (III3: 9) wird häufig als *t-Test* bezeichnet. Die Fälle (III3: 10) und (III3: 11) sind als *approximative Gaußtests* bekannt. Abbildung III 14 gibt einen zusammenfassenden Überblick.

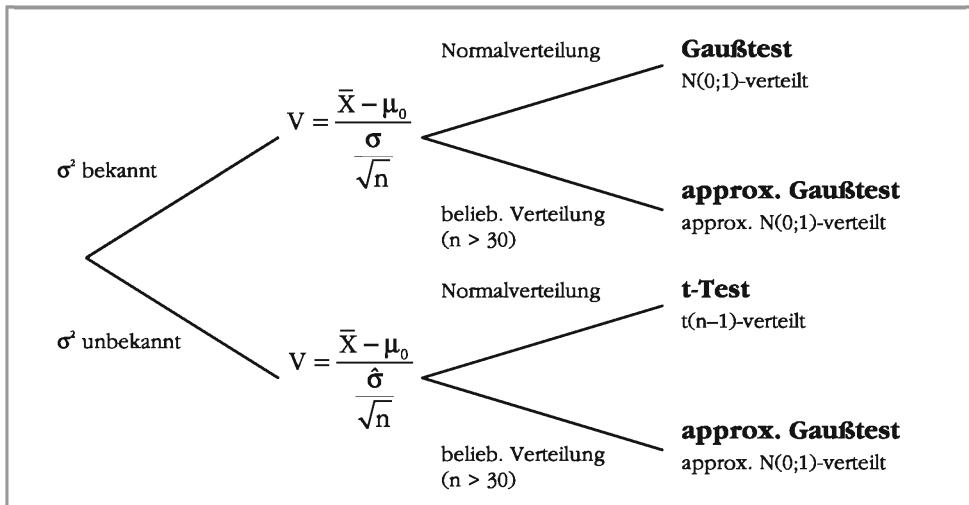


Abbildung III 14: Einstichprobentests für das arithmetische Mittel

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Für die Teststatistiken (III3: 8), (III3: 10) und (III3: 11) gilt unter Verwendung der konkreten Ausprägung v der Teststatistik analog zum Einstichprobentest für den Anteilswert

H_0 wird gegenüber H_A abgelehnt

- | | | |
|--------------------------------------------|---------------------------|--------------|
| a) $H_0: \mu = \mu_0, H_A: \mu \neq \mu_0$ | $ v > z_{1-\alpha/2}$ | . (III3: 12) |
| b) $H_0: \mu \geq \mu_0, H_A: \mu < \mu_0$ | falls $v < -z_{1-\alpha}$ | |
| c) $H_0: \mu \leq \mu_0, H_A: \mu > \mu_0$ | $v > z_{1-\alpha}$ | |

Im Falle von (III3: 9) können zur Entscheidung der Ablehnung oder Nichtablehnung der Nullhypothese nicht länger die Fraktile der Standardnormalverteilung verwendet werden. Stattdessen sind die entsprechenden Fraktile der ebenfalls symmetrischen t-Verteilung mit $n - 1$ Freiheitsgraden zu verwenden.

H ₀ wird gegenüber H _A abgelehnt		
a)	H ₀ : $\mu = \mu_0$, H _A : $\mu \neq \mu_0$	$ v > t_{1-\alpha/2, n-1}$
b)	H ₀ : $\mu \geq \mu_0$, H _A : $\mu < \mu_0$	falls $v < -t_{1-\alpha, n-1}$
c)	H ₀ : $\mu \leq \mu_0$, H _A : $\mu > \mu_0$	$v > t_{1-\alpha, n-1}$

(III3: 13)

Bei Freiheitsgraden von *mebr als 30*, kann die t-Verteilung bekanntlich durch die Standardnormalverteilung approximiert werden (vgl. Abschnitt II 3.3.2). Es können dann entsprechend deren Fraktile verwendet werden, sodass dann wiederum (III3: 12) gilt.

(4) Entscheidung auf Basis der berechneten Teststatistik

Die Entscheidungsfindung und Interpretation erfolgt wie bisher. Da bei allen Testverfahren gleich vorgegangen wird, wird der Punkt (4) des Testschemas bei der theoretischen Beschreibung der nachfolgenden Tests nicht mehr explizit aufgeführt.

Beispiel III 3-4

Einstichprobentest für den Mittelwert



Eine Stanzmaschine fertigt Metallteile, deren Fläche mit dem Mittelwert von $\mu = 0,30 \text{ cm}^2$ und einer Standardabweichung von $\sigma = 0,005 \text{ cm}^2$ normalverteilt ist. Aufgrund von kürzlich an der Maschine durchgeföhrten Wartungsarbeiten zweifelt der Maschinenführer nun an der Genauigkeit der Maschine. Eine Stichprobe im Umfang von $n = 50$ ergibt ein arithmetisches Mittel von $\bar{x} = 0,302 \text{ cm}^2$. Zu prüfen ist nun bei einem Signifikanzniveau von $\alpha = 0,05$ und unter der Annahme einer gleichgebliebenen Standardabweichung σ die Hypothese, dass die Maschine noch exakt arbeitet.

(1) Hypothesen und Signifikanzniveau:

$$H_0: \mu = 0,30 \quad H_A: \mu \neq 0,30 \quad \alpha = 0,05$$

(2) Teststatistik und deren Verteilung:

Einfacher Gaußtest:

$$V = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0;1), \text{ da Grundgesamtheit normalverteilt und Varianz } \sigma^2 \text{ bekannt.}$$

In konkreter Stichprobe:

$$V = \frac{0,302 - 0,30}{\frac{0,005}{\sqrt{50}}} = 2,83$$

(3) Entscheidungsregel:

Fall a) liegt vor. $\rightarrow H_0$ ablehnen gegenüber H_A , wenn $|V| > z_{1-\alpha/2}$ gilt.

Dabei ist $z_{1-\alpha/2} = z_{1-0,05/2} = z_{0,975} = 1,96$.

(4) Entscheidung:

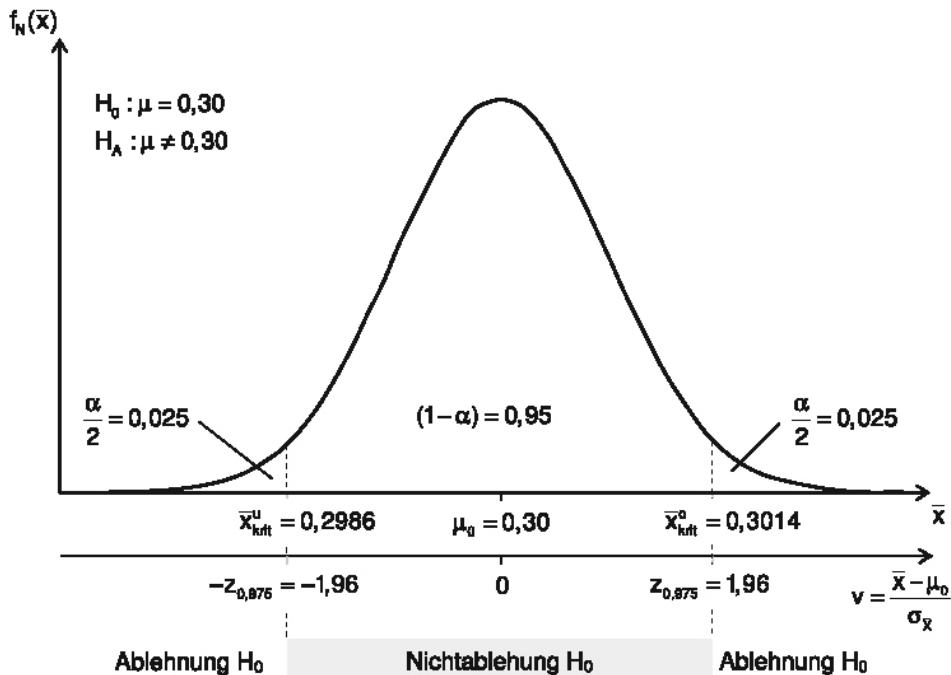
Es gilt hier ($|v| = 2,83 > (z_{1-\alpha/2} = 1,96)$.

Die Nullhypothese kann also abgelehnt werden. Es ist damit bei einem Signifikanzniveau von 5 % statistisch belegt, dass die Maschine nicht exakt (H_A) arbeitet.

Die nachfolgende Skizze veranschaulicht den durchgeföhrten Signifikanztest. Die beiden kritischen Mittelwerte \bar{x}_{krit}^u und \bar{x}_{krit}^o wurden darin wie folgt bestimmt:

$$\bar{x}_{\text{krit}}^u = \mu_0 - z_{1-\alpha/2} \sigma_{\bar{x}} = 0,30 - 1,96 \cdot \frac{0,005}{\sqrt{50}} = 0,2986$$

$$\bar{x}_{\text{krit}}^o = \mu_0 + z_{1-\alpha/2} \sigma_{\bar{x}} = 0,30 + 1,96 \cdot \frac{0,005}{\sqrt{50}} = 0,3014$$



Beispiel III 3-5

Einstichprobentest für den Mittelwert



Ein Bauer züchtet Mastschweine deren Gewicht mit einem Sollwert von $\mu = 110$ kg normalverteilt ist. Er setzt seit einiger Zeit einen Futterzusatz ein, der gewichtssteigernd wirken soll. Eine Stichprobe im Umfang von $n = 10$ liefert nun ein Durchschnittsgewicht von $\bar{x} = 117$ kg bei einer Standardabweichung von $\sigma = 7$ kg. Kann man aufgrund dieser Stichprobe auf einem Signifikanzniveau von $\alpha = 0,01$ schließen, dass sich das Durchschnittsgewicht in der Grundgesamtheit erhöht hat?

(1) Hypothesen und Signifikanzniveau:

$$H_0: \mu \leq 110$$

$$H_A: \mu > 110$$

$$\alpha = 0,01$$

(2) Teststatistik und deren Verteilung:

t-Test:

$$V = \frac{\bar{X} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t(n-1), \text{ da Grundgesamtheit normalverteilt, Varianz } \sigma^2 \text{ unbekannt und } (n-1) < 30.$$

In konkreter Stichprobe:

$$v = \frac{117 - 110}{\sqrt{10}} = 3,16$$

(3) Entscheidungsregel:

Fall a) liegt vor. $\rightarrow H_0$ ablehnen gegenüber H_A , wenn $v > t_{1-\alpha, n-1}$ gilt.Dabei ist $t_{1-\alpha, n-1} = t_{1-0,01; 10-1} = t_{0,99, 9} = 2,82$.

(4) Entscheidung:

Es gilt hier ($v = 3,16 > (t_{1-\alpha, n-1} = 2,82)$). H_0 kann abgelehnt werden. Es lässt sich also statistisch nachweisen, dass der Futterzusatz das Durchschnittsgewicht der Mastschweine erhöht hat.**3.3.1.3 Statistische Qualitätskontrolle**

In der Praxis kann der Einstichprobentest für das arithmetische Mittel dazu verwendet werden, kritische Verläufe in Produktionsprozessen zu identifizieren. Man nimmt dazu zunächst an, dass die gefertigten Werkstücke bzgl. eines Merkmals X (z.B. Füllmenge) mit Erwartungswert μ und Varianz σ^2 normalverteilt sind. Man entnimmt dann in regelmäßigen Zeitabständen Stichproben im Umfang von 5 bis 10 Stück und testet für jede dieser Stichproben die Hypothese, dass ein bestimmter Sollwert μ_0 eingehalten wird ($H_0: \mu = \mu_0$ gegen $H_A: \mu \neq \mu_0$). Diese Hypothese wird sowohl zum Signifikanzniveau $\alpha = 0,01$ als auch zu $\alpha = 0,05$ getestet.

Der Test erfolgt dabei implizit. Wie wir wissen, ist H_0 bei einem zweiseitigen Gaußtest bei Vorliegen einer konkreten Stichprobe nicht abzulehnen wenn

$$-z_{1-\alpha/2} \leq \left(v = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \right) \leq z_{1-\alpha/2}$$

gilt. Durch Umstellung erhalten wir daraus folgendes Intervall, welches für die Testentscheidung benutzt wird:

$$\mu_0 - z_{1-\alpha/2} \sigma_{\bar{x}} \leq \bar{x} \leq \mu_0 + z_{1-\alpha/2} \sigma_{\bar{x}}$$

Für $\alpha = 0,01$ ergibt sich das Intervall

$$\underbrace{\mu_0 - 2,576 \sigma_{\bar{x}}}_{\text{untere Kontrollgrenze}} \leq \bar{x} \leq \underbrace{\mu_0 + 2,576 \sigma_{\bar{x}}}_{\text{obere Kontrollgrenze}},$$

aus dem man die sog. **Kontrollgrenzen** gewinnt. Liegt der in der Stichprobe festgestellte Mittelwert \bar{x} außerhalb dieses Intervalls, ist die Nullhypothese H_0 bei einem Signifikanzniveau von $\alpha = 0,01$ abzulehnen. Ist dies der Fall, wird die Produktion sofort gestoppt und eine Maschinenüberprüfung vorgenommen.

Für $\alpha = 0,05$ erhält man das Intervall

$$\underbrace{\mu_0 - 1,96\sigma_{\bar{x}}}_{\text{untere Warngrenze}} \leq \bar{x} \leq \underbrace{\mu_0 + 1,96\sigma_{\bar{x}}}_{\text{obere Warngrenze}},$$

welches die sog. **Warngrenzen** liefert. Wird H_0 nicht zu $\alpha = 0,01$, sondern anhand dieses Intervalls zu einem Signifikanzniveau von $\alpha = 0,05$ abgelehnt, wird sofort eine weitere Stichprobe entnommen. Wird H_0 zu $\alpha = 0,05$ dann nicht mehr verworfen, lässt man die Produktion weiterlaufen, andernfalls stoppt man die Produktion.

Zur einfachen praktischen Arbeit mit dieser Methode werden die so definierten Warn- und Kontrollgrenzen einmal mit den Daten der Verteilung der Grundgesamtheit berechnet und in eine **Überwachungsgrafik** (vgl. Abbildung III 15) eingetragen. In diese werden auch die Mittelwerte der einzelnen Stichproben aufgenommen. Man sieht so relativ schnell, wann weitere Stichproben zu entnehmen sind, und wann die Produktion angehalten werden sollte. So ist in der untenstehenden Abbildung III 15 z.B. bei Stichprobe a eine weitere Stichprobe zu ziehen und nach Stichprobe b die Produktion anzuhalten.

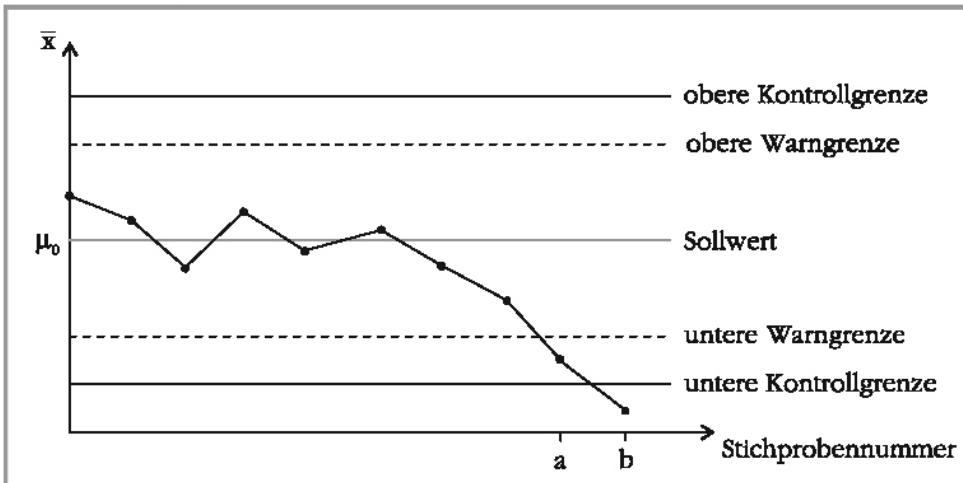


Abbildung III 15: Überwachungsgrafik zur statistischen Qualitätskontrolle

3.3.1.4 Einstichprobentest für die Varianz

Hypothesen hinsichtlich der unbekannten Varianz σ^2 einer **normalverteilten** Grundgesamtheit können mit Einstichprobentests für die Varianz untersucht werden. Aufgrund der Verteilung der dabei verwendeten Teststatistik bezeichnen wir diese auch als **Chi-Quadrat-Tests für die Varianz**. Man geht dabei wie folgt vor:

(1) Formulierung einer Nullhypothese und einer Gegenhypothese:

Die Null- und Gegenhypothesen haben die folgende gewohnte Form:

- a) $H_0: \sigma^2 = \sigma_0^2 \quad -- \quad H_A: \sigma^2 \neq \sigma_0^2$
- b) $H_0: \sigma^2 \geq \sigma_0^2 \quad -- \quad H_A: \sigma^2 < \sigma_0^2$ (III3: 14)
- c) $H_0: \sigma^2 \leq \sigma_0^2 \quad -- \quad H_A: \sigma^2 > \sigma_0^2$

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Wie wir unter III 1.2.4 gesehen haben, folgt die Größe $B = (n - 1)\hat{\sigma}^2 / \sigma^2$ bei *normalverteilter Grundgesamtheit* einer Chi-Quadrat-Verteilung mit $(n - 1)$ Freiheitsgraden. Für Tests, die sich auf die Varianz beziehen, kann diese Größe als Teststatistik verwendet werden. Sofern H_0 richtig ist, gilt also

$$V = \frac{(n - 1)\hat{\sigma}^2}{\sigma_0^2} = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n - 1). \quad (\text{III3: 15})$$

Die Stichprobenvarianz $\hat{\sigma}^2$ ist ein erwartungstreuer und konsistenter Schätzer für die unbekannte Varianz σ^2 der Grundgesamtheit. Es sprechen somit kleine Werte von V für $\sigma^2 < \sigma_0^2$ und große Werte von V für $\sigma^2 > \sigma_0^2$.

(III3: 15) wird nur bei unbekanntem μ der Grundgesamtheit verwendet. Ist μ bekannt, sollte folgende Größe verwendet werden, die bei korrekter Nullhypothese H_0 dann mit n Freiheitsgraden Chi-Quadrat-verteilt ist.

$$V = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n) \quad (\text{III3: 16})$$

Bei *nicht normalverteilter Grundgesamtheit* ist das Verfahren des Chi-Quadrat-Tests selbst bei hohem Stichprobenumfang n nicht generell anwendbar. Die Grundgesamtheit sollte zumindest approximativ normalverteilt sein und zusätzlich ein Stichprobenumfang von über 100 vorliegen.

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Da die Chi-Quadrat-Verteilung keine symmetrische Verteilung ist, sind bei diesem Test andere als die bisher behandelten Entscheidungsregeln für symmetrisch verteilte Teststatistiken zu definieren. Es gilt bei *unbekanntem μ* :

H_0 wird abgelehnt gegenüber H_A

- | | | |
|----------------------------------------------------------------|---------------------------------------------------------------------------|--|
| a) $H_0: \sigma^2 = \sigma_0^2, H_A: \sigma^2 \neq \sigma_0^2$ | $(0 \leq v < \chi_{\alpha/2, n-1}^2) \cup (v > \chi_{1-\alpha/2, n-1}^2)$ | |
| b) $H_0: \sigma^2 \geq \sigma_0^2, H_A: \sigma^2 < \sigma_0^2$ | falls $0 \leq v < \chi_{\alpha, n-1}^2$ | |
| c) $H_0: \sigma^2 \leq \sigma_0^2, H_A: \sigma^2 > \sigma_0^2$ | $v > \chi_{1-\alpha, n-1}^2$ | |
- (III3: 17)

Bei den χ^2 -Werten handelt es sich dabei jeweils um Fraktile einer Chi-Quadrat-Verteilung mit $(n - 1)$ Freiheitsgraden. Ist μ bekannt, sind die entsprechenden Fraktile einer Chi-Quadrat-Verteilung mit n Freiheitsgraden zu verwenden.

Beispiel III 3-6

Einstichprobentest für die Varianz



Die Varianz der normalverteilten Lebensdauer eines bestimmten Motorentyps lag in der Vergangenheit bei $\sigma^2 = 2,5$ Jahren². Nach Einführung eines neuen Produktionsverfahrens behauptet ein Mitarbeiter, dass sich dadurch die Varianz der Lebensdauer verringert hat. Lässt sich diese Hypothese anhand einer Stichprobe im Umfang von $n = 25$, die eine Varianz von $\hat{\sigma}^2 = 2,3$ Jahren² lieferte, bei einem Signifikanzniveau von $\alpha = 0,10$ statistisch nachweisen?

(1) Hypothesen und Signifikanzniveau:

$$H_0: \sigma^2 \geq 2,5 \quad H_A: \sigma^2 < 2,5 \quad \alpha = 0,10$$

(2) Teststatistik und deren Verteilung:

Chi-Quadrat-Test für die Varianz:

$$V = \frac{(n-1)\hat{\sigma}^2}{\sigma_0^2} \sim \chi^2(n-1), \text{ da Grundgesamtheit normalverteilt und } \mu \text{ unbekannt.}$$

In konkreter Stichprobe:

$$v = \frac{24 \cdot 2,3}{2,5} = 22,08$$

(3) Entscheidungsregel:

Fall b) liegt vor. → H_0 ablehnen gegenüber H_A , wenn $0 \leq v < \chi_{\alpha, n-1}^2$ gilt.

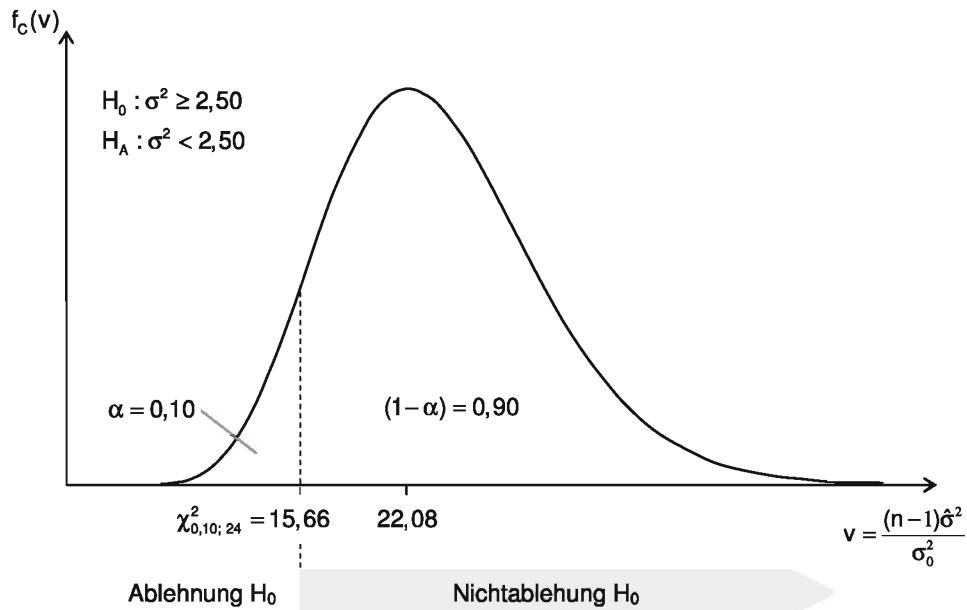
$$\text{Dabei ist } \chi_{\alpha, n-1}^2 = \chi_{0,10; 24}^2 = 15,66.$$

(4) Entscheidung:

Hier fällt $v = 22,08$ nicht in das Intervall $0 \leq v < 15,66$.

H_0 kann nicht abgelehnt werden. Es lässt sich also nicht statistisch nachweisen, dass das neue Produktionsverfahren die Varianz der Grundgesamtheit verringert hat.

Grafisch kann dieser Test bzw. die Dichte der nicht symmetrischen Chi-Quadrat-Verteilung folgendermaßen skizziert werden:



3.3.2 Zweistichprobentests

Bisher haben wir im Rahmen der Einstichprobentests ein Untersuchungsmerkmal aufgrund einer einfachen Stichprobe analysiert. Bei Zweistichprobentests werden nun aufgrund zweier *unabhängiger* einfacher Zufallsstichproben zwei Verteilungen bzw. deren Parameter verglichen. So vergleicht man in solchen Tests beispielsweise

das durchschnittliche Einkommen in verschiedenen Gruppen oder den Lernerfolg zweier Lehrmethoden. Ein solcher Test kann also angewendet werden, wenn

- zwei Grundgesamtheiten* auf ein gemeinsames Merkmal hin untersucht werden. (Bsp.: Einkommensstichproben aus einer Gruppe Frauen und einer Gruppe Männer)
- aus *einer Grundgesamtheit zwei unabhängige Stichproben* gezogen werden. (Bsp: Aus der Grundgesamtheit von Schülern wird eine einfache Stichprobe vom Umfang n rein zufällig auf zwei Gruppen mit unterschiedlichen Lehrmethoden aufgeteilt und die Leistungen jeder Gruppe untersucht.)

3.3.2.1 Vergleich zweier arithmetischer Mittel

Liegen aus zwei *unabhängigen* Stichproben der Umfänge n_1 und n_2 , die Stichprobenmittel \bar{x}_1 und \bar{x}_2 und damit auch deren Differenz $d = \bar{x}_1 - \bar{x}_2$ vor, können anhand dieser Stichprobenergebnisse Hypothesen hinsichtlich der Differenz $\mu_1 - \mu_2$ der arithmetischen Mittel zweier Grundgesamtheiten untersucht werden. Man bedient sich dazu des folgenden Testverfahrens:

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Die Hypothesen entsprechen bei diesem Test dem Schema (III3: 18). Mit einer Ablehnung der Nullhypothese im Fall a) lässt sich statistisch zeigen, dass die Differenz der arithmetischen Mittel in der Grundgesamtheit von Null verschieden ist ($\mu_1 - \mu_2 \neq 0$). Im Fall b) kann bei Ablehnung von H_0 eine negative Differenz ($\mu_1 - \mu_2 < 0$), im Fall c) eine positive Differenz ($\mu_1 - \mu_2 > 0$) nachgewiesen werden.

- $H_0: \mu_1 = \mu_2 \quad -- \quad H_A: \mu_1 \neq \mu_2$
 - $H_0: \mu_1 \geq \mu_2 \quad -- \quad H_A: \mu_1 < \mu_2$
 - $H_0: \mu_1 \leq \mu_2 \quad -- \quad H_A: \mu_1 > \mu_2$
- (III3: 18)

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Da es sich bei den Stichprobenmitteln \bar{X}_1 und \bar{X}_2 um erwartungstreue Schätzer für μ_1 und μ_2 handelt, ist die Differenz $D = \bar{X}_1 - \bar{X}_2$ zur Überprüfung der Hypothesen (III3: 18) geeignet. Ist \bar{x}_1 deutlich größer als \bar{x}_2 bzw. deutlich kleiner als \bar{x}_2 , ist zu vermuten, dass $\mu_1 > \mu_2$ bzw. $\mu_1 < \mu_2$ gilt. Werte für $d = \bar{x}_1 - \bar{x}_2$, die nahe Null liegen, sprechen für $\mu_1 = \mu_2$. Wieder müssen wir bei der Wahl der genauen Teststatistik die Verteilungsform der Grundgesamtheiten, sowie die Bekanntheit oder Unbekanntheit ihrer Varianzen σ_1^2 und σ_2^2 berücksichtigen.

Sind die beiden **Grundgesamtheiten** mit den Erwartungswerten μ_1 und μ_2 und den Varianzen σ_1^2 und σ_2^2 **normalverteilt** und sind die **Varianzen bekannt und ungleich** ($\sigma_1^2 \neq \sigma_2^2$), verwenden wir (beim Ziehen mit Zurücklegen bzw. Ziehen ohne Zurücklegen und $n \leq 0,1N$) die bei korrekter Nullhypothese standardnormalverteilte Teststatistik

$$V = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0; 1). \quad (\text{III3: 19})$$

Wir sprechen bei Verwendung von (III3: 19) und Vorliegen der genannten Voraussetzungen auch von einem sog. *Zweistichproben-Gaußtest*.¹ Sind die **ungleichen Varianzen** σ_1^2 und σ_2^2 **unbekannt**, können diese durch $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$ geschätzt werden. Dadurch ergibt sich die Teststatistik

$$V = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}}. \quad (\text{III3: 20})$$

Die exakte Verteilung dieser Statistik unter Gültigkeit von H_0 hängt allerdings von σ_1^2 und σ_2^2 ab und ist somit ebenso wie die beiden Varianzen unbekannt. Die Varianzen erweisen sich in diesem Zusammenhang als sog. störende Parameter, die die Herleitung eines exakten Tests für kleine Stichproben verhindern. Dieses Problem ist in der statistischen Literatur auch als *Behrens-Fisher-Problem* bekannt.²

Eine Näherungslösung der Problematik für kleine Stichproben ermöglicht der *Welch-Test*.³ Es lässt sich nämlich zeigen, dass die Verteilung von (III3: 20) unter H_0 durch eine t-Verteilung mit Freiheitsgraden in Höhe von

$$\frac{(1+Q)^2}{\frac{Q^2}{n_1-1} + \frac{1}{n_2-1}} \quad \text{mit} \quad Q = \frac{n_2 \cdot \hat{\sigma}_1^2}{n_1 \cdot \hat{\sigma}_2^2} \quad (\text{III3: 21})$$

hinreichend approximiert werden kann. Diese Verteilungsapproximation ermöglicht bei *kleinen Stichprobenumfängen* einen *approximativen Test*. Der Quotient zur Ermittlung der Freiheitsgrade ist i.d.R. nicht ganzzahlig. Wird der Welch-Test mit Hilfe der üblichen Quantil-Tabellen der t-Verteilung durchgeführt, muss der Quotient daher abgerundet werden.

Sind die betrachteten **Grundgesamtheiten beliebig verteilt**, wird bei $n_1 > 30$ und $n_2 > 30$ sowie **bekannten Varianzen** σ_1^2 und σ_2^2 die aufgrund des Zentralen Grenzwertsatzes *approximativ standardnormalverteilte* Teststatistik

$$V = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \stackrel{d}{\sim} N(0; 1) \quad (\text{III3: 22})$$

verwendet.⁴ Bei **unbekannten Varianzen** ersetzen wir in (III3: 22) σ_1^2 und σ_2^2 durch $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$. Auch hier sprechen wir von *approximativen Zweistichproben-Gaußtests*. Der Test für den Vergleich zweier Anteilswerte stellt einen bedeutenden

¹ Neben diesem Test, der von unterschiedlichen Varianzen der Grundgesamtheiten ausgeht, existiert in der Literatur der sog. *Zweistichproben-t-Test*. Dieser spielt jedoch aufgrund seiner restriktiven Voraussetzungen (normalverteilte Grundgesamtheiten, *gleiche Varianzen*) in der Praxis kaum eine Rolle und wird deshalb hier nicht behandelt. Interessierte finden eine Beschreibung des Tests in Bleymüller, J., Gehlert, G., Göllicher, H. (2004), Kapitel 17.

² Für Details hierzu vgl. z.B. Scheffé, H. (1970).

³ Vgl. dazu Welch, B.L. (1936, 1938).

⁴ Im Fall von $n_1 < 30$ und $n_2 < 30$ bietet sich die Verwendung des Rangsummentests von Wilcoxon an. Vgl. dazu z.B. Büning, H., Trenkler, G. (1998).

Sonderfall des letztgenannten approximativen Zweistichproben-Gaußtests dar und wird im Abschnitt III 3.3.2.2. separat behandelt.

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Da die Teststatistiken der dargestellten Zweistichprobentests für das arithmetische Mittel (abgesehen vom Welch-Test) einer Standardnormalverteilung folgen, gehen wir bei der Entscheidung über die Ablehnung oder Nichtablehnung von H_0 folgendermaßen vor:

H_0 wird gegenüber H_A abgelehnt	
a) $H_0: \mu_1 = \mu_2, H_A: \mu_1 \neq \mu_2$	$ v > z_{1-\alpha/2}$
b) $H_0: \mu_1 \geq \mu_2, H_A: \mu_1 < \mu_2$ falls	$v < -z_{1-\alpha}$
c) $H_0: \mu_1 \leq \mu_2, H_A: \mu_1 > \mu_2$	$v > z_{1-\alpha}$

Beispiel III 3-7

Approximativer Zweistichproben-Gaußtest

Zwei Gruppen von zufällig ausgewählten Studenten (Gruppe 1: $n_1 = 55$, Gruppe 2: $n_2 = 60$) wird eine Testklausur gestellt. Die durchschnittlichen Punkte und die dazu gehörigen Standardabweichungen betragen in Gruppe 1 $\bar{x}_1 = 50$ und $\hat{\sigma}_1 = 5$ und in Gruppe 2 $\bar{x}_2 = 54$ und $\hat{\sigma}_2 = 4$. Bei einem Signifikanzniveau von $\alpha = 0,05$ soll getestet werden, ob ein signifikanter Unterschied zwischen den Klausurergebnissen der beiden Studentengruppen besteht.

(1) Hypothesen und Signifikanzniveau:

$$H_0: \mu_1 = \mu_2 \quad H_A: \mu_1 \neq \mu_2 \quad \alpha = 0,05$$

(2) Teststatistik und deren Verteilung:

Approximativer Zweistichproben-Gaußtest

$$V = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}} \stackrel{a}{\sim} N(0;1), \text{ da Grundgesamtheit beliebig verteilt, } \sigma_1^2 \text{ u. } \sigma_2^2 \text{ unbekannt, } n_1 \text{ und } n_2 > 30.$$

In konkreter Stichprobe:

$$V = \frac{50 - 54}{\sqrt{\frac{5^2}{55} + \frac{4^2}{60}}} = \frac{-4}{0,85} = -4,71$$

(3) Entscheidungsregel:

Fall a) liegt vor. $\rightarrow H_0$ ablehnen gegenüber H_A , wenn $|v| > z_{1-\alpha/2}$ gilt.

Dabei ist $z_{1-\alpha/2} = z_{0,975} = 1,96$.

(4) Entscheidung:

Es gilt hier ($|v| = 4,71 > (z_{1-\alpha/2} = 1,96)$).

Die Nullhypothese kann abgelehnt werden. Es lässt sich damit statistisch nachweisen, dass ein signifikanter Unterschied zwischen den Leistungen beider Gruppen besteht.

3.3.2.2 Vergleich zweier Anteilswerte

Als Sonderfall von (III3: 23) wollen wir nun den Test für den Vergleich zweier Anteilswerte p_1 und p_2 näher betrachten. Auch er wird als *approximativer Zweistichproben-Gaußtest* bezeichnet, da die Grundgesamtheit nicht normal-, sondern binomialverteilt ist. Es empfiehlt sich bei diesem Test die folgende Vorgehensweise:

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Die Hypothesenpaare dieses Tests sind folgendermaßen aufgebaut:

- a) $H_0: p_1 = p_2 \quad -\quad H_A: p_1 \neq p_2$
- b) $H_0: p_1 \geq p_2 \quad -\quad H_A: p_1 < p_2$ (III3: 25)
- c) $H_0: p_1 \leq p_2 \quad -\quad H_A: p_1 > p_2$

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Die **Varianzen** σ_1^2 und σ_2^2 der Grundgesamtheiten sind hier **unbekannt**, da die Anteilswerte p_1 und p_2 der Grundgesamtheiten unbekannt sind. Die Teststatistik (III3: 26) kann als *approximativ standardnormalverteilt* angenommen werden, wenn $n_1 p_1 (1-p_1) \geq 9$ und $n_2 p_2 (1-p_2) \geq 9$ gilt.

$$V = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \stackrel{a}{\sim} N(0;1) \quad (\text{III3: 26})$$

Da die Werte p_1 und p_2 nicht bekannt sind, sind sie sowohl zur Prüfung der Approximationsbedingungen, als auch zur Berechnung von (III3: 26) mit den erwartungstreuen Schätzern \hat{p}_1 und \hat{p}_2 zu schätzen.

Wird die Nullhypothese a) überprüft, ist bei Gültigkeit von H_0 der Zusammenhang $p_1 = p_2 = p$ erfüllt und die Teststatistik (III3: 26) vereinfacht sich zu

$$V = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}} \stackrel{a}{\sim} N(0;1). \quad (\text{III3: 27})$$

Der unbekannte Parameter p wird dabei über den Schätzer (III3: 28) bestimmt:

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{\sum_{i=1}^{n_1} X_{1i} + \sum_{j=1}^{n_2} X_{2j}}{n_1 + n_2} \quad (\text{III3: 28})$$

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Da den Teststatistiken die Standardnormalverteilung zugrunde liegt, gilt

H_0 wird gegenüber H_A abgelehnt

- a) $H_0: p_1 = p_2, H_A: p_1 \neq p_2 \quad |v| > z_{1-\alpha/2}$ (III3: 29)
- b) $H_0: p_1 \geq p_2, H_A: p_1 < p_2 \quad \text{falls } v < -z_{1-\alpha}$
- c) $H_0: p_1 \leq p_2, H_A: p_1 > p_2 \quad v > z_{1-\alpha}$

Beispiel III 3-8**Approximativer Zweistichproben-Gaußtest**

In zwei Großstädten werden je $n_1 = n_2 = 500$ Personen hinsichtlich ihres Jahreseinkommens befragt. Es ergibt sich dabei, dass in Stadt A ein Anteil von $\hat{p}_1 = 0,05$ und in Stadt B ein Anteil von $\hat{p}_2 = 0,03$ der Befragten unter der Armutsgrenze lebt. Lässt sich anhand dieser Stichprobe bei einem Signifikanzniveau von $\alpha = 0,05$ widerlegen, dass der Anteil der armen Bevölkerung in beiden Städten gleich ist?

(1) Hypothesen und Signifikanzniveau:

$$H_0: p_1 = p_2 \quad H_A: p_1 \neq p_2 \quad \alpha = 0,05$$

(2) Teststatistik und deren Verteilung:

Approximativer Zweistichproben-Gaußtest:

$$V = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}} \stackrel{a}{\sim} N(0;1), \text{ da } \underbrace{n_1 \hat{p}_1 (1 - \hat{p}_1)}_{23,75} \geq 9 \text{ und } \underbrace{n_2 \hat{p}_2 (1 - \hat{p}_2)}_{14,55} \geq 9.$$

In konkreter Stichprobe:

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{500 \cdot 0,05 + 500 \cdot 0,03}{500 + 500} = 0,04$$

$$v = \frac{0,05 - 0,03}{\sqrt{0,04 \cdot 0,96}} \sqrt{\frac{500 + 500}{500 \cdot 500}} = \frac{0,02}{0,20 \cdot 0,06} = 1,67$$

(3) Entscheidungsregel:

Fall a) liegt vor. $\rightarrow H_0$ ablehnen gegenüber H_A , wenn $|v| > z_{1-\alpha/2}$ gilt.

Dabei ist $z_{1-\alpha/2} = z_{0,975} = 1,96$.

(4) Entscheidung:

Es gilt hier ($|v| = 1,67 < z_{1-\alpha/2} = 1,96$).

Die Nullhypothese kann nicht abgelehnt werden. Es lässt sich also anhand dieser Stichprobe nicht widerlegen, dass der Anteil der armen Bevölkerung in beiden Städten gleich ist.

3.3.2.3 Vergleich zweier Varianzen

Dieser Test, welcher aufgrund der Verteilung der Teststatistik auch als *Zweistichproben-F-Test* bezeichnet wird, dient der Überprüfung von Hypothesen bezüglich der unbekannten Varianzen σ_1^2 und σ_2^2 zweier Grundgesamtheiten. *Normalverteilte* Grundgesamtheiten sind dabei zwingende Voraussetzung für die Testanwendung.

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Hinsichtlich der unbekannten Varianzen werden folgende Fälle untersucht:

a) $H_0: \sigma_1^2 = \sigma_2^2 \quad -- \quad H_A: \sigma_1^2 \neq \sigma_2^2$

b) $H_0: \sigma_1^2 \geq \sigma_2^2 \quad -- \quad H_A: \sigma_1^2 < \sigma_2^2$

(III3: 30)

c) $H_0: \sigma_1^2 \leq \sigma_2^2 \quad -- \quad H_A: \sigma_1^2 > \sigma_2^2$

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Wie wir in Abschnitt III 1.2.5.3 gesehen haben, ist die Zufallsvariable

$$F = \frac{\hat{\sigma}_1^2 / \sigma_1^2}{\hat{\sigma}_2^2 / \sigma_2^2}$$

bei *normalverteilten* Grundgesamtheiten mit $u = (n_1 - 1)$ und $w = (n_2 - 1)$ Freiheitsgraden F-verteilt. Im Fall einer korrekten Nullhypothese ($\sigma_1^2 = \sigma_2^2$) ist somit auch die folgende Teststatistik V mit eben jenen Freiheitsgraden F-verteilt.

$$V = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F(n_1 - 1, n_2 - 1) \quad (\text{III3: 31})$$

V ist als Teststatistik für (III3: 30) geeignet, da ein hoher bzw. niedriger Wert der Statistik für $\sigma_1^2 > \sigma_2^2$ bzw. für $\sigma_1^2 < \sigma_2^2$ spricht.

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Da die F-Verteilung keine symmetrische Verteilung ist, sind für diesen Test die Entscheidungsregel wieder etwas zu modifizieren. Verwenden wir die Fraktile $f_{\alpha/2}$, $f_{1-\alpha/2}$, f_α und $f_{1-\alpha}$ der $F(n_1 - 1, n_2 - 1)$ -Verteilung, ergeben sich folgende Entscheidungsregeln:

H_0 wird abgelehnt gegenüber H_A

- a) $H_0: \sigma_1^2 = \sigma_2^2, H_A: \sigma_1^2 \neq \sigma_2^2 \quad (0 \leq v < f_{\alpha/2}) \cup (v > f_{1-\alpha/2})$ (III3: 32)
- b) $H_0: \sigma_1^2 \geq \sigma_2^2, H_A: \sigma_1^2 < \sigma_2^2 \quad \text{falls } 0 \leq v < f_\alpha$
- c) $H_0: \sigma_1^2 \leq \sigma_2^2, H_A: \sigma_1^2 > \sigma_2^2 \quad v > f_{1-\alpha}$

Für die Fraktile der F-Verteilung ist in diesem Zusammenhang zu beachten, dass

$$f_{\alpha/2, n_1 - 1, n_2 - 1} = \frac{1}{f_{1-\alpha/2, n_2 - 1, n_1 - 1}} \quad \text{und} \quad f_{\alpha, n_1 - 1, n_2 - 1} = \frac{1}{f_{1-\alpha, n_2 - 1, n_1 - 1}}.$$

Auf die *Voraussetzung der Normalverteilung* kann bei diesem Test auch bei großem Stichprobenumfang nicht verzichtet werden, da ansonsten das Testverfahren das vorgegebene Signifikanzniveau α nicht einhält.

Beispiel III 3-9

Zweistichproben-F-Test

Lässt sich bei einem vorgegebenen Signifikanzniveau von $\alpha = 0,05$ anhand zweier Stichproben (aus zwei normalverteilten Grundgesamtheiten), die bei Umfängen von $n_1 = 21$ und $n_2 = 11$ Standardabweichungen von $\hat{\sigma}_1 = 480$ Euro und $\hat{\sigma}_2 = 430$ Euro liefern, statistisch nachweisen, dass die Gehälter der Angestellten in einem Unternehmen Nr. 1 stärker streuen als in einem Unternehmen Nr. 2?

(1) Hypothesen und Signifikanzniveau:

$$H_0: \sigma_1^2 \leq \sigma_2^2 \quad H_A: \sigma_1^2 > \sigma_2^2 \quad \alpha = 0,05$$

(2) Teststatistik und deren Verteilung:

Zweistichproben-F-Test:

$$V = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F(n_1 - 1, n_2 - 1), \text{ da Grundgesamtheit normalverteilt.}$$

In konkreter Stichprobe:

$$v = \frac{480^2}{430^2} = 1,25$$

(3) Entscheidungsregel:

Fall c) liegt vor. → H_0 ablehnen gegenüber H_A , wenn $v > f_{1-\alpha, n_1-1, n_2-1}$ gilt.

Dabei ist $f_{1-\alpha, n_1-1, n_2-1} = f_{0,95; 20; 10} = 2,77$.

(4) Entscheidung:

Es gilt hier ($v = 1,25 < f_{1-\alpha, n_1-1, n_2-1} = 2,77$).

Die Nullhypothese kann nicht abgelehnt werden. Es lässt sich anhand dieser Stichprobe nicht nachweisen, dass die Streuung der Gehälter im Unternehmen A größer ist.

3.3.3 Parametertests bei verbundenen Stichproben

Bei den bisher behandelten Zweistichprobentests waren die Beobachtungen einer Stichprobe stets unabhängig von denen der anderen Stichprobe, d.h. es lagen **unabhängige (unverbundene) Stichproben** vor. Diese Unabhängigkeit wird meist dadurch gewährleistet, dass die Stichproben aus *unterschiedlichen Grundgesamtheiten* gezogen werden.

Abhängige (verbundene) Stichproben liegen vor, wenn die Beobachtungen der einen Stichprobe von denen der anderen abhängen. Konkret spricht man etwa von verbundenen Stichproben, wenn die Merkmalsausprägungen der *ersten* Stichprobe und die der *zweiten* Stichprobe jeweils an denselben Merkmalsträgern erhoben werden.

Verbundene Stichproben müssen aber nicht zwingend am selben Merkmalsträger erhoben werden. Sie können z.B. auch an Zwillings- oder Geschwisterpaaren oder an den Hälften von aufgeteilten Feldern durchgeführt werden. Gerade das letztgenannte Beispiel lässt einen klaren Vorteil verbundener Stichproben verdeutlichen. Bringt man etwa bei einer verbundenen Stichprobe Düngemittel, deren ertragsteigernde Wirkung untersucht werden soll, auf denselben Feldern (z.B. je zur Hälfte) auf, können keine zufälligen Unterschiede bei der Ackerqualität das Untersuchungsergebnis beeinträchtigen, wie es bei einer unverbundenen Stichprobe der Fall sein könnte. Die individuellen Unterschiede der Merkmalsträger werden also bei verbundenen Stichproben faktisch ausgeschaltet.⁵ Es ist allerdings nicht immer möglich, verbundene Stichproben zu erheben. Untersucht man etwa die Wirkung zweier Arzneimittel an einer kranken Person, macht es nach Heilung durch das erste Medikament keinen Sinn mehr, auch noch das zweite zu verabreichen.

⁵ Zur Effizienz verbundener und unverbundener Stichproben vgl. Pfanzagl, J. (1991).

Im Folgenden beschäftigen wir uns näher mit zwei typischen Vertretern dieser Testklasse, dem sog. Differenzentest (Zweistichprobentest für die Differenz zweier arithmetischer Mittel bei verbundenen Stichproben) und dem Korrelationstest.

3.3.3.1 Differenzentest

Beim Differenzentest handelt es sich um eine Anwendung des Einstichproben-t-Tests bzw. des approximativen Einstichproben-Gaußtests auf Differenzen von Stichprobenvariablen. Die Vorgehensweisen sind daher eng miteinander verwandt.

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Nehmen wir an, es liegt eine Grundgesamtheit vor, die bezüglich zweier Untersuchungsmerkmale X_1 und X_2 die Erwartungswerte μ_1 und μ_2 besitzt, so besteht zwischen diesen Erwartungswerten die Differenz $\delta = \mu_2 - \mu_1$. Die Hypothesen dieses Testverfahrens beziehen sich nun genau auf diese unbekannte Differenz der Grundgesamtheit. Wir interessieren uns im Speziellen dafür, ob mittels des Ergebnisses verbundener Stichproben festgestellt werden kann, ob diese ungleich, größer oder kleiner als Null ist. Ist beispielsweise eine positive (negative) Differenz statistisch nachweisbar, so bedeutet dies $\mu_2 > \mu_1$ ($\mu_2 < \mu_1$). Es sind somit folgende Hypothesenpaare denkbar:

- | | | | |
|-------------------------|----|----------------------|------------|
| a) $H_0: \delta = 0$ | -- | $H_A: \delta \neq 0$ | |
| b) $H_0: \delta \geq 0$ | -- | $H_A: \delta < 0$ | (III3: 33) |
| c) $H_0: \delta \leq 0$ | -- | $H_A: \delta > 0$ | |

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Wird eine verbundene Stichprobe im Umfang von n gezogen, erfassen wir für jeden Merkmalsträger ($i = 1, \dots, n$) die Ausprägungen x_{1i} und x_{2i} der Zufallsvariablen X_{1i} und X_{2i} . Wir wollen diese Zufallsvariablen im Folgenden als *normalverteilt* annehmen. Ihre Erwartungswerte sind $E(X_{1i}) = \mu_1$ und $E(X_{2i}) = \mu_2$. Damit stellen die Differenzen $D_i = X_{2i} - X_{1i}$ ebenfalls normalverteilte Zufallsvariablen dar. Sie messen für jedes untersuchte Objekt aus der Grundgesamtheit die Differenz der beiden Merkmalsausprägungen. Für ihre Erwartungswerte gilt

$$E(D_i) = E(X_{2i}) - E(X_{1i}) = \mu_2 - \mu_1 = \delta.$$

Für die Varianz schreiben wir

$$\text{Var}(D_i) = \sigma_D^2.$$

Das arithmetische Mittel der n Differenzen D_i ist definiert als

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i. \quad (\text{III3: 34})$$

Auch \bar{D} ist aufgrund der *Reproduktionseigenschaft* der Normalverteilung (beim Ziehen mit Zurücklegen bzw. Ziehen ohne Zurücklegen und $n \leq 0,1N$) mit folgenden Parametern *normalverteilt*:

$$E(\bar{D}) = \delta \quad (\text{III3: 35})$$

$$\text{Var}(\bar{D}) = \sigma_{\bar{D}}^2 = \frac{\sigma_D^2}{n} \quad (\text{III3: 36})$$

Da die Varianz der Differenzen σ_D^2 unbekannt ist, muss sie geschätzt werden. Es bietet sich dafür der Schätzer

$$\hat{\sigma}_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2 \quad (\text{III3: 37})$$

an. Wir erhalten damit eine geeignete Teststatistik, die aufgrund der geschätzten Varianz bei korrekter Nullhypothese und **normalverteilter Grundgesamtheit** einer ***t*-Verteilung** mit $(n - 1)$ Freiheitsgraden folgt:

$$V = \frac{\bar{D}}{\hat{\sigma}_D} \sim t(n-1) \quad (\text{III3: 38})$$

Sollte die **Grundgesamtheit beliebig verteilt** sein, kann bei erfüllter Bedingung $n > 30$ ebenfalls die Teststatistik (III3: 38) verwendet werden. Diese ist dann jedoch bei korrekter Nullhypothese *approximativ standardnormalverteilt*.⁶

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Je nach zugrunde liegender Verteilung der Teststatistik V sind folgende Entscheidungsregeln zu beachten:

H_0 wird gegenüber H_A abgelehnt	$N(0;1)$	$t(n-1)$
a) $H_0: \delta = 0, H_A: \delta \neq 0$	$ v > z_{1-\alpha/2}$	$ v > t_{1-\alpha/2, n-1}$
b) $H_0: \delta \geq 0, H_A: \delta < 0$	falls $v < -z_{1-\alpha}$	$v < -t_{1-\alpha, n-1}$
c) $H_0: \delta \leq 0, H_A: \delta > 0$	$v > z_{1-\alpha}$	$v > t_{1-\alpha, n-1}$

(III3: 39)

Beispiel III 3-10

Differenzentest

Um den Einfluss der Bereifung auf den Spritverbrauch von Pkws zu untersuchen, wurden 10 Fahrzeuge zunächst mit Reifensorte A bestückt und der Spritverbrauch auf 100 km gemessen. Auf dieselben Fahrzeuge wurden dann die Reifen der Sorte B aufgezogen und erneut der Spritverbrauch gemessen. Es ergab sich dabei (Liter je 100 km):

Pkw i	1	2	3	4	5	6	7	8	9	10
Verbrauch bei A	8,7	8,9	8,3	9,1	7,9	10,3	8,6	9,0	7,3	7,9
Verbrauch bei B	8,5	9,2	9,9	9,8	8,0	11,2	9,1	8,9	7,5	8,8

Lässt sich damit bei einem Signifikanzniveau von $\alpha = 0,05$ und angenommener Normalverteilung der Grundgesamtheit statistisch nachweisen, dass der Spritverbrauch bei Verwendung von Reifensorte B höher ist als bei Verwendung von Sorte A?

(1) Hypothesen und Signifikanzniveau:

$$H_0: \delta \leq 0 \quad H_A: \delta > 0 \quad \alpha = 0,05 \quad \text{wobei } \delta = \mu_B - \mu_A$$

(2) Teststatistik und deren Verteilung:

Einstichproben-t-Test als Differenzentest:

⁶ Im Fall von $n < 30$ bietet sich die Verwendung des Vorzeichen-Rangtests von Wilcoxon an. Vgl. dazu z.B. Büning, H., Trenkler, G. (1998).

$$V = \frac{\bar{D}}{\frac{\hat{\sigma}_D}{\sqrt{n}}} \sim t(n-1), \text{ da Grundgesamtheit normalverteilt und Varianz } \sigma_D^2 \text{ unbekannt.}$$

Um die konkrete Ausprägung v in der Stichprobe zu bestimmen, sind zunächst einige Berechnungen erforderlich:

i	x_{Ai}	x_{Bi}	$d_i = x_{Bi} - x_{Ai}$	d_i^2
1	8,7	8,5	-0,2	0,04
2	8,9	9,2	0,3	0,09
3	8,3	9,9	1,6	2,56
4	9,1	9,8	0,7	0,49
5	7,9	8,0	0,1	0,01
6	10,3	11,2	0,9	0,81
7	8,6	9,1	0,5	0,25
8	9,0	8,9	-0,1	0,01
9	7,3	7,5	0,2	0,04
10	7,9	8,8	0,9	0,81
			4,9	5,11

Es ergibt sich daraus

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = \frac{1}{10} \cdot 4,9 = 0,49 \text{ und}$$

$$\hat{\sigma}_D^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2 = \frac{n}{n-1} \left[\frac{1}{n} \sum_{i=1}^n d_i^2 - \bar{d}^2 \right] = \frac{10}{9} \left[\frac{5,11}{10} - 0,49^2 \right] = 0,30 \rightarrow \hat{\sigma}_D = 0,55.$$

Für die konkrete Ausprägung der Teststatistik gilt daher

$$v = \frac{0,49}{\frac{0,55}{\sqrt{10}}} = 2,82.$$

(3) Entscheidungsregel:

Fall c) liegt vor. $\rightarrow H_0$ ablehnen gegenüber H_A , wenn $v > t_{1-\alpha, n-1}$ gilt.

Dabei ist $t_{1-\alpha, n-1} = t_{0,95, 9} = 1,83$.

(4) Entscheidung:

Es gilt hier ($v = 2,82 > (t_{1-\alpha, n-1} = 1,83)$.

Die Nullhypothese kann abgelehnt werden. Damit kann es als statistisch gesichert angesehen werden, dass mit der Reifensorte B ein höherer Spritverbrauch entsteht.

Schreiben wir die Varianz der D_i in der Form $\text{Var}(D_i) = \sigma_D^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 - 2\sigma_{x_1 x_2}$, wird unmittelbar ein Vorteil verbundener Stichproben deutlich. Sind die Stichprobenvariablen nämlich positiv korreliert ($\sigma_{x_1 x_2} > 0$), gilt $\sigma_{x_1}^2 + \sigma_{x_2}^2 > \sigma_D^2$ und deshalb auch $\text{Var}(\bar{x}_1 + \bar{x}_2) > \text{Var}(\bar{D})$. Die geringere Streuung des Mittels der Differenzen gegenüber der Differenz der Stichprobenmittel erlaubt es also, Mittelwertunterschiede mit größerer Wahrscheinlichkeit aufzudecken.

3.3.3.2 Korrelationstest

Wie wir wissen, nimmt der Korrelationskoeffizient ρ bei unabhängigen Zufallsvariablen X_1 und X_2 einen Wert von Null an. Umgekehrt kann jedoch nicht generell ge-

sagt werden, dass ein Vorliegen von $\rho = 0$ Unabhängigkeit bedeutet. Ein solcher Rückschluss ist nur bei normalverteilten Zufallsvariablen möglich. Sind also zwei Zufallsvariablen X_1 und X_2 (gemeinsam) **normalverteilt**, kann aufgrund von n Realisationen dieser Zufallsvariablen über den Korrelationskoeffizienten ρ getestet werden, ob X_1 und X_2 *unabhängig* sind.

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Da es Sinn und Zweck dieses Testverfahrens ist, Aussagen über den unbekannten Korrelationskoeffizienten ρ der Grundgesamtheit zu machen, haben die Hypothesen folgenden Aufbau:

- a) $H_0: \rho = 0$ -- $H_A: \rho \neq 0$
- b) $H_0: \rho = 0$ (oder $\rho \geq 0$) -- $H_A: \rho < 0$ (III3: 40)
- c) $H_0: \rho = 0$ (oder $\rho \leq 0$) -- $H_A: \rho > 0$

In allen drei Fällen bedeutet eine Ablehnung von H_0 bei (gemeinsam) **normalverteilten Grundgesamtheiten**, dass wir auch den statistischen Nachweis der Abhängigkeit der betrachteten Variablen X_1 und X_2 erbringen können.

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Falls die beiden betrachteten Merkmale der Grundgesamtheit (gemeinsam) normalverteilt sind, eignet sich (III3: 41) als Teststatistik für dieses Testverfahren.⁷

$$V = \frac{R}{\sqrt{\frac{1-R^2}{n-2}}} \sim t(n-2) \quad (\text{III3: 41})$$

V folgt bei korrekter Nullhypothese einer t -Verteilung mit $(n - 2)$ Freiheitsgraden. R steht für den Schätzer des unbekannten Korrelationskoeffizienten ρ der Grundgesamtheit, dem der *Bravais-Pearson-Korrelationskoeffizient* (vgl. Abschnitt I 3.2.1) zu Grunde liegt.⁸

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Wird zum Signifikanzniveau α getestet, ist die Nullhypothese nach folgender Entscheidungsregel zu verwerfen:

H_0 wird gegenüber H_A abgelehnt		
a) $H_0: \rho = 0, H_A: \rho \neq 0$		$ v > t_{1-\alpha/2, n-2}$
b) $H_0: \rho \geq 0, H_A: \rho < 0$	falls	$v < -t_{1-\alpha, n-2}$
c) $H_0: \rho \leq 0, H_A: \rho > 0$		$v > t_{1-\alpha, n-2}$

⁷ Vgl. dazu z.B. Kanji, G.K. (2006).

⁸ Sollen Hypothesen des Typs $H_0: \rho = \rho_0$, $H_0: \rho \geq \rho_0$ und $H_0: \rho \leq \rho_0$ getestet werden, ist die folgende Teststatistik zu verwenden, die unter H_0 für $n > 30$ approximativ einer Standardnormalverteilung folgt:

$$V = 0,5 \cdot \left(\ln \frac{1+R}{1-R} - \ln \frac{1+\rho_0}{1-\rho_0} \right) \cdot \sqrt{n-3}$$

Beispiel III 3-11	Korrelationstest
--------------------------	------------------

Zwischen den normalverteilten Renditen zweier Wertpapiere A und B wurde für eine Zeitreihe von $n = 20$ Tagen ein Korrelationskoeffizient von $r_{AB} = 0,22$ festgestellt. Lässt sich damit bei einem Signifikanzniveau von $\alpha = 0,01$ statistisch nachweisen, dass die Renditen beider Aktien voneinander abhängen?

(1) Hypothesen und Signifikanzniveau:

$$H_0: \rho = 0 \quad H_A: \rho \neq 0 \quad \alpha = 0,01$$

(2) Teststatistik und deren Verteilung:

Korrelationstest:

$$V = \frac{R}{\sqrt{\frac{1-R^2}{n-2}}} \sim t(n-2)$$

In konkreter Stichprobe:

$$v = \frac{0,22}{\sqrt{\frac{1-0,22^2}{20-2}}} = 0,96$$

(3) Entscheidungsregel:

Fall a) liegt vor. $\rightarrow H_0$ ablehnen gegenüber H_A , wenn $|v| > t_{1-\alpha/2, n-2}$ gilt.

Dabei ist $t_{1-\alpha/2, n-2} = t_{0,995; 18} = 2,88$.

(4) Entscheidung:

Es gilt hier ($|v| = 0,96 < (t_{1-\alpha/2, n-2} = 2,88)$).

Die Nullhypothese kann nicht abgelehnt werden. Es lässt sich also anhand dieser Stichprobe nicht nachweisen, dass die Aktienrenditen voneinander abhängig sind.

Um eine Abhängigkeit zu bestätigen, wäre bei einer Stichprobe dieses Umfangs ($n = 20$) ein Korrelationskoeffizient r erforderlich, der die Bedingung

$$\left| \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \right| > t_{1-\alpha/2, n-2}$$

erfüllt. Es gilt also hier

$$\begin{aligned} \frac{r \cdot \sqrt{18}}{\sqrt{1-r^2}} &> 2,88 \leftrightarrow r^2 \frac{18}{1-r^2} > 2,88^2 \leftrightarrow r^2 > (1-r^2) \cdot 2,88^2 / 18 \leftrightarrow r^2 > (1-r^2) \cdot 0,46 \\ \leftrightarrow r^2 > 0,46 / (1+0,46) &\leftrightarrow r^2 > 0,32 \leftrightarrow |r| > 0,57. \end{aligned}$$

Erst wenn der Korrelationskoeffizient bei diesem Stichprobenumfang und vorgegebenem Signifikanzniveau von $\alpha = 0,01$ also nicht zwischen $-0,57$ und $+0,57$ liegt, kann dieser als signifikant angesehen und $H_0: \rho = 0$ verworfen werden.

Liegen ordinal skalierte Daten vor oder ist bei metrisch skalierten Daten keine Normalverteilung gegeben, kann der *Spearman-Rangkorrelationskoeffizient* R^s (vgl. Abschnitt I 3.2.2) verwendet werden. Die Zufallsvariable R^s ist dann für

$n > 30$ mit dem Erwartungswert $E(R^{sp}) = 0$ und der Varianz $\text{Var}(R^{sp}) = 1/(n - 1)$ hinlänglich normalverteilt.⁹ Als Testgröße eignet sich daher

$$V = R^{sp} \sqrt{n - 1} \stackrel{d}{\sim} N(0; 1). \quad (\text{III3: 43})$$

Die Testentscheidung ist in einem solchen Fall unter Verwendung der entsprechenden Fraktile der Standardnormalverteilung zu treffen.

3.3.4 Gütfunktionen von Parametertests

Wie wir bereits im Abschnitt III 3.1 gesehen haben, ist es nicht möglich die Wahrscheinlichkeiten für den Fehler 1. Art (Ablehnung von H_0 , obwohl H_0 richtig ist) und den Fehler 2. Art (Nicht-Ablehnung von H_0 , obwohl H_0 falsch ist) gleichzeitig unter ein vorgegebenes Niveau zu drücken, da mit sinkendem Fehler 1. Art ceteris paribus der Fehler 2. Art zunimmt. Aufgrund dieser Problematik haben wir bei der Definition des Verwerfungsbereiches die Anforderungen gestellt, dass

- a) beim Test die Wahrscheinlichkeit für den Fehler 1. Art maximal gleich einem vorgegebenem Signifikanzniveau α sein muss und
- b) unter Einhaltung dieser Bedingung die Wahrscheinlichkeit für den Fehler 2. Art so klein wie möglich gestaltet werden sollte.

Hauptaugenmerk lag in der bisherigen Betrachtung der Testverfahren beim Fehler 1. Art, den wir mit einem Wert α vorgegeben haben. Den Fehler 2. Art haben wir nur wenig berücksichtigt. Zur Beurteilung der Güte eines Tests ist aber zu prüfen, ob und inwieweit ein Test die *beiden* definierten Anforderungen erfüllt. Man bedient sich dazu in der Praxis der sog. Gütfunktion, der wir uns im Folgenden ausführlich widmen wollen:

Bezeichnen wir wie gewohnt einen beliebigen unbekannten Parameter der Grundgesamtheit mit ϑ , unsere Teststatistik mit V und den Verwerfungsbereich für H_0 mit B , so können wir eine Funktion

$$g(\vartheta) = P(H_0 \text{ ablehnen} \mid \vartheta) = P(V \in B \mid \vartheta) \quad (\text{III3: 44})$$

definieren, die (für ein vorgegebenes Signifikanzniveau α und festen Stichprobenumfang n) die Wahrscheinlichkeit dafür angibt, H_0 abzulehnen, wenn der unbekannte wahre Parameter der Grundgesamtheit gleich ϑ ist. Diese Funktion bezeichnen wir als **Macht (Mächtigkeit) oder Gütfunktion** eines Tests. Voraussetzung für die Bestimmung ihrer Funktionswerte ist, dass die Verteilung von V bekannt ist (z.B. nicht nur bei korrekter Nullhypothese). Die Bestimmung der Gütfunktion ist bei vielen Tests kompliziert. Häufig kann sie nur näherungsweise mit Hilfe von Simulationstechniken (z.B. Monte-Carlo-Verfahren) erfolgen.

Eine Gütfunktion kann z.B. einen Verlauf wie in Abbildung III 16 (graue Kurve) aufweisen. Es handelt sich dabei um die **Gütfunktion eines einfachen linksseitigen Gaußtests** ($H_0: \mu \leq \mu_0$), die wir nun näher betrachten. Ein linksseitiger Gaußtest erfüllt genau dann die obige Bedingung a), ist also ein *Signifikanztest zum Niveau α* , wenn der Maximalwert der Gütfunktion im Bereich $\Theta_0: \mu \leq \mu_0$ gleich α ist. Dies zeigt sich in Abbildung III 16 als $g(\mu_0) = \alpha$. Es handelt sich bei der dargestellten Gütfunktion also um diejenige eines Signifikanztests zum Niveau α .

⁹ Schlittgen, R. (2003) sieht R^{sp} bereits bei $n > 10$ als hinreichend genau normalverteilt an.

Ein wichtiger Indikator für die Erfüllung von Bedingung b) ist die Höhe der Funktionswerte $g(\mu)$ im Bereich $\Theta_A: \mu > \mu_0$. Die Wahrscheinlichkeit β dafür, H_0 nicht abzulehnen, wenn der Wert $\mu \in \Theta_A$ ist (also für den Fehler 2. Art), kann nämlich als

$$\beta = P(H_0 \text{ nicht ablehnen} | \mu) = 1 - g(\mu) \quad (\text{III3: 45})$$

bestimmt werden. Ein Fehler 2. Art ist also umso unwahrscheinlicher, je höher $g(\mu)$ im Bereich Θ_A ist. Eine steile Gütfunktion wäre also hier wünschenswert.

Die in Abbildung III 16 abgetragene (graue) Gütfunktion stellt die Gütfunktion eines **unverfälschten Tests** dar. Unverfälscht wird ein Gaußtest dann genannt, wenn es sich bei ihm um einen Signifikanztest zum Niveau α handelt, dessen Gütfunktion $g(\mu)$ für alle $\mu \in \Theta_A$ größer oder gleich α ist. Zusätzlich ist in der Abbildung die Gütfunktion eines **idealen Tests** enthalten. Ein Gaußtest wäre ideal, wenn die beiden Fehlerwahrscheinlichkeiten α und β sehr klein bzw. sogar beide Null sind. Dies würde bedeuten, dass die Nullhypothese nicht verworfen wird, solange der wahre Wert μ im Intervall Θ_0 liegt. Befindet sich μ im Bereich Θ_A , wird H_0 sicher verworfen. Kein Test besitzt jedoch eine derartige ideale Form.

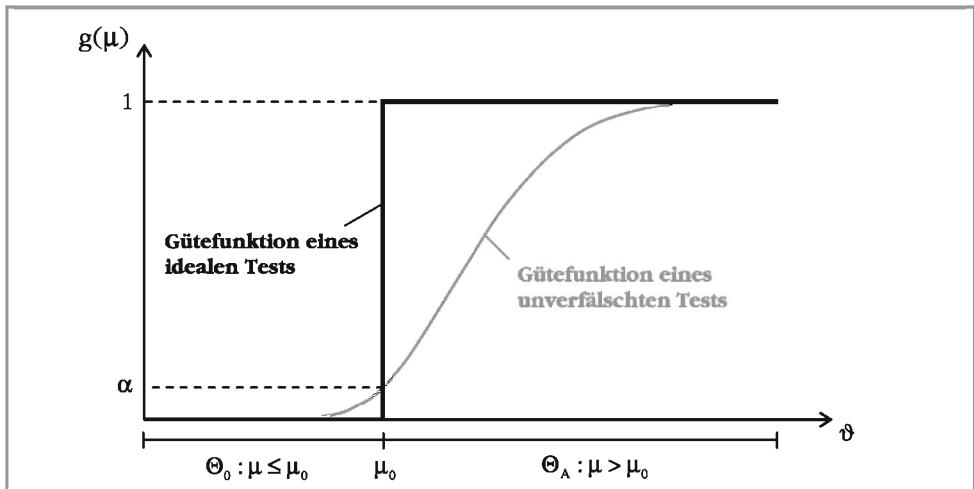


Abbildung III 16: Gütfunktion eines unverfälschten und eines idealen Tests

Die Differenz $1 - g(\mu) = \beta$ wird auch als *Operationscharakteristik* oder OC-Kurve bezeichnet. Sie gibt die Komplementärwahrscheinlichkeit zur Gütfunktion an. Für den einfachen linksseitigen Gaußtest hat sie im Vergleich zur Gütfunktion den in Abbildung III 17 dargestellten Verlauf. Beim Vergleich beider Funktionen wird schnell klar, dass auch die OC-Kurve zur Beurteilung der Güte eines Tests herangezogen werden könnte. Ein Signifikanztest zum Niveau α liegt vor, wenn der Minimalwert der OC-Kurve im Bereich Θ_0 gleich $1 - \alpha$ ist. Von einem unverfälschten Test können wir sprechen, wenn seine OC-Kurve für alle $\mu \in \Theta_A$ kleiner als $1 - \alpha$ ist. Nicht nur für die Gütfunktion erachten wir im Bereich Θ_A einen steilen Verlauf als wünschenswert, sondern auch für die OC-Kurve. Wir wollen bereits an dieser Stelle festhalten, dass der Verlauf von Gütfunktion und OC-Kurve allgemein umso steiler (und damit die Güte des Testverfahrens umso besser) ist, je größer der Stichprobenumfang n gewählt wird. In Beispiel III 3-12 werden wir dies anhand eines konkreten Zahlenbeispiels sehen.

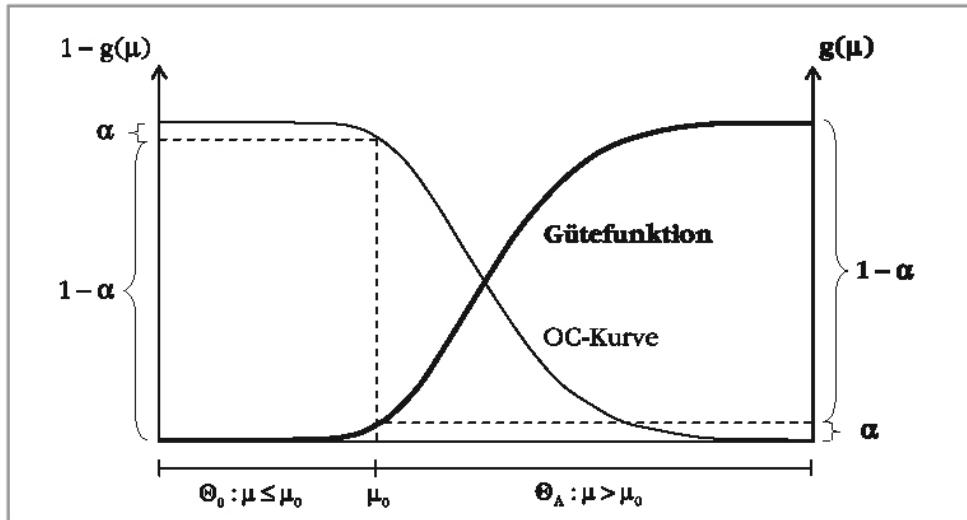


Abbildung III 17: Gütefunktion vs. OC-Kurve

Beispiel III 3-12**Berechnung einer Gütefunktion**

Im Folgenden werden wir die konkrete Gütefunktion für einen linksseitigen Gaußtest, d.h. einen Gaußtest mit $H_0: \mu \leq \mu_0$ gegenüber $H_A: \mu > \mu_0$, näher untersuchen. Aus der allgemeinen Definition der Gütefunktion (III3: 44) und den Testvorschriften des Gaußtests aus Abschnitt III 3.3.1.2 ergibt sich für die Gütefunktion dieses Tests allgemein als

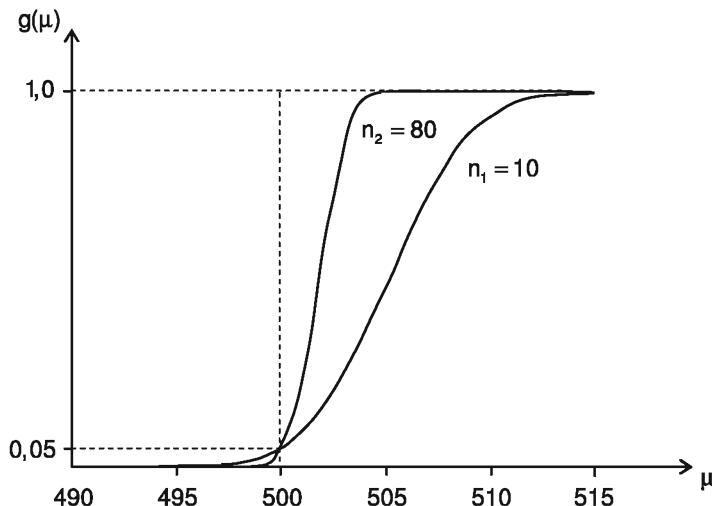
$$g(\mu) = P(\text{Ablehnung von } H_0 | \mu) = P(V > z_{1-\alpha} | \mu) = P\left(\frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} > z_{1-\alpha} \mid \mu\right).$$

Zur Berechnung konkreter Werte dieser Gütefunktion können wir ausnutzen, dass die Teststatistik V standardnormalverteilt ist. Dies führt zu folgender Rechnung:

$$\begin{aligned} g(\mu) &= P\left(\frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} > z_{1-\alpha} \mid \mu\right) = P\left(\frac{\bar{X} - \mu + \mu - \mu_0}{\sigma} \sqrt{n} > z_{1-\alpha} \mid \mu\right) \\ &= P\left(\frac{\bar{X} - \mu}{\sigma} \sqrt{n} > z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \sqrt{n} \mid \mu\right) = 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right) = \Phi\left(\frac{\mu - \mu_0}{\sigma} \sqrt{n} - z_{1-\alpha}\right) \end{aligned}$$

Für die numerischen Werte $\mu_0 = 500$, $\sigma = 10$, $\alpha = 0,05$ und die Stichprobenumfänge $n_1 = 10$ und $n_2 = 80$ ergeben sich damit die folgenden beiden Gütefunktionen $g_1(\mu)$ und $g_2(\mu)$ mit zugehörigen Werten:

μ	$g_1(\mu)$	$g_2(\mu)$	μ	$g_1(\mu)$	$g_2(\mu)$
499	0,0249	0,0056	503	0,2432	0,8505
500	0,0500	0,0500	505	0,4746	0,9977
501	0,0920	0,2265	510	0,9354	1,0000
502	0,1557	0,5573	515	0,9990	1,0000



Die genauere Betrachtung dieser konkreten Gütfunktion lässt folgende Schlüsse zu:

1. Grafik und Wertetabelle zeigen, dass für die Irrtumswahrscheinlichkeit α genau gilt:

$$g(\mu_0) = \Phi(-z_{1-\alpha}) = \alpha$$

2. Da es sich bei der Verteilungsfunktion der Standardnormalverteilung um eine monoton wachsende Funktion handelt, wächst auch die Gütfunktion monoton mit μ . Wir können generell sagen, dass die Gütfunktion des linksseitigen Gaußtests die gleiche Gestalt hat wie die Verteilungsfunktion einer Normalverteilung mit geeigneten Parametern μ und σ .
3. Je größer der Abstand des wahren Mittels μ vom hypothetischen Mittel μ_0 , d.h. $\mu - \mu_0$, ist, umso kleiner wird die Wahrscheinlichkeit $\beta = 1 - g(\mu)$ des Fehlers 2. Art. Für sehr große Abstände gilt schließlich $\beta \approx 0$. Eine Fehlentscheidung ist fast unmöglich.
4. Die Wahrscheinlichkeit des Fehlers 2. Art hängt vom Stichprobenumfang n ab. Für ein festes μ mit $\mu > \mu_0$ wird $g(\mu)$ mit wachsendem n größer, d.h. die Wahrscheinlichkeit des Fehlers 2. Art wird mit wachsendem n kleiner.
5. Es gilt a) $g(\mu) \leq \alpha$ für $\mu \leq \mu_0$ und b) $g(\mu) > \alpha$ für $\mu > \mu_0$. Dabei bedeutet a), dass die Wahrscheinlichkeit für den Fehler 1. Art höchstens α beträgt, d.h. das Signifikanzniveau α erweist sich als vorgebbare Obergrenze für $P(\text{Fehler 1. Art})$. Aus b) ergibt sich die Beziehung $P(\text{Fehler 2. Art}) < (1 - \alpha)$. Damit ist $P(\text{Fehler 2. Art})$ im Gegensatz zu $P(\text{Fehler 1. Art})$ nicht kontrolliert klein, sondern kann "unkontrolliert groß" werden.

Bei Anwendung des Gaußtests auf $H_0: \mu \leq \mu_0$ trifft man in der Praxis häufig auf die Frage, wie groß n sein muss, damit bei Zugrundelegung einer Irrtumswahrscheinlichkeit α eine Abweichung des tatsächlich vorliegenden Wertes μ_1 vom hypothetischen Wert μ_0 mit vorgebbar großer Wahrscheinlichkeit ξ bemerkt wird. Anders ausgedrückt, interessiert man sich dafür, wie groß n sein muss, damit die Gütfunktion an einer vorgegebenen Stelle μ_1 den vorgebbaren Wert ξ annimmt, d.h.

$$\Phi\left(\frac{\mu_1 - \mu_0}{\sigma} \sqrt{n} - z_{1-\alpha}\right) = \xi .$$

Zur Auflösung dieser Gleichung nach n ist zunächst in der Tabelle der Standardnormalverteilung diejenige Stelle z_ξ mit $\Phi(z_\xi) = \xi$ zu finden. Wir erhalten damit

$$\frac{\mu_1 - \mu_0}{\sigma} \sqrt{n} - z_{1-\alpha} = z_\xi,$$

woraus sich schließlich der gesuchte Stichprobenumfang n als

$$n = \frac{(z_{1-\alpha} + z_\xi)^2}{(\mu_1 - \mu_0)^2} \sigma^2$$

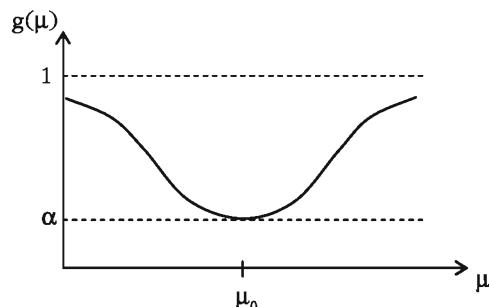
ergibt. Dieser Stichprobenumfang n besitzt eine interessante praxisrelevante Eigenschaft. Er ist nämlich der kleinste Stichprobenumfang mit dessen Hilfe jede positive Abweichung des tatsächlich vorliegenden Mittelwertes von μ_0 , die mindestens $d = \mu_1 - \mu_0 > 0$ beträgt, mindestens mit der Wahrscheinlichkeit ξ aufgedeckt wird.

Der Vollständigkeit halber zeigt Abbildung III 18 die Gestalt aller beim einfachen Gaußtest möglichen Gütfunktionen im Überblick. Dabei ist zu beachten, dass Θ_0 im Fall des zweiseitigen Tests nur aus einem Wert, nämlich μ_0 , besteht.

Zweiseitiger Test

$$H_0 : \mu = \mu_0 \quad H_A : \mu \neq \mu_0$$

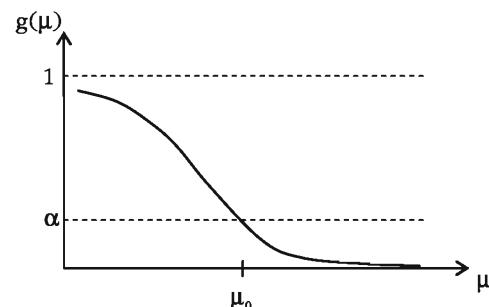
$$g(\mu) = \Phi\left(-z_{1-\alpha/2} + \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right) \\ + \Phi\left(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right)$$



Rechtsseitiger Test

$$H_0 : \mu \geq \mu_0 \quad H_A : \mu < \mu_0$$

$$g(\mu) = \Phi\left(-z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right)$$



Linksseitiger Test

$$H_0 : \mu \leq \mu_0 \quad H_A : \mu > \mu_0$$

$$g(\mu) = 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma} \sqrt{n}\right)$$

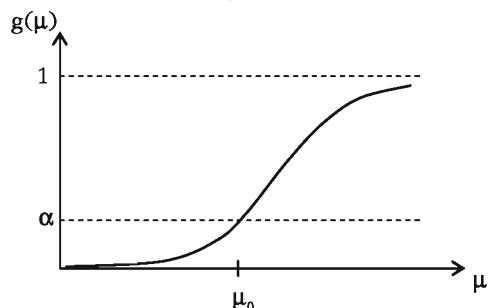


Abbildung III 18: Gütfunktionen beim einfachen Gaußtest

3.4 Verteilungstests

Die in den vorhergehenden Abschnitten behandelten Testverfahren dienten der Prüfung von Hypothesen, die die Parameter von Grundgesamtheiten betreffen. Wir haben sie daher auch als parametrische Testverfahren oder *Parametertests* bezeichnet. Im Folgenden wollen wir uns nun mit Tests beschäftigen, die Hypothesen überprüfen, die die unbekannte Verteilung von Grundgesamtheiten betreffen. Wir sprechen hier von nichtparametrischen Tests oder *Verteilungstests*.

3.4.1 Chi-Quadrat-Anpassungstest

Chi-Quadrat-Anpassungstests untersuchen, ob eine aus einer Zufallsstichprobe gewonnene Häufigkeitsverteilung mit einer für die unbekannte Verteilung der Grundgesamtheit gemachten Annahme vereinbar ist oder nicht. Es geht darum, eine in einer Stichprobe gewonnene Verteilung mit einer zu erwartenden Verteilung zu vergleichen. Da hier die Genauigkeit der Anpassung einer theoretischen Verteilung an eine empirische überprüft wird, bezeichnen wir den Test als Anpassungstest.¹⁰

Die Nullhypothese lautet hier immer, dass die Grundgesamtheit einer bestimmten Verteilung F_0 gehorcht. Je nachdem, ob es sich bei F_0 um eine diskrete Verteilung (mit wenigen Merkmalsausprägungen) oder eine stetige Verteilung (oder diskrete Verteilung mit vielen Merkmalsausprägungen) handelt, ist eine minimal andere Testvorgehensweise zu wählen.

3.4.1.1 Anpassungstest bei diskret verteilter Grundgesamtheit

Grundsätzlich gehen wir beim Chi-Quadrat-Anpassungstest davon aus, dass die Grundgesamtheit bezüglich eines Untersuchungsmerkmals X in k Klassen A_1, A_2, \dots, A_k eingeteilt werden kann und so jede Merkmalsausprägung a, der Grundgesamtheit zu genau einer der Klassen gehört. Wie groß diese Klassen sein sollten, werden wir im Folgenden noch detailliert behandeln. Im einfachsten Fall ist die Grundgesamtheit diskret verteilt und besitzt wenige Merkmalsausprägungen, sodass *jede Merkmalsausprägung a, als eigenständige Klasse A_i angesehen werden kann*. Beim Test empfiehlt sich dann folgende Vorgehensweise:

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Die Nullhypothese lautet stets, dass die unbekannte Verteilung F der Grundgesamtheit (bezüglich des Merkmals X) einer angenommenen theoretischen Verteilung F_0 entspricht, d.h.

$$H_0: F = F_0 \quad -- \quad H_A: F \neq F_0 . \quad (\text{III3: 46})$$

Bei einer diskret verteilten Grundgesamtheit könnte F_0 also beispielsweise eine Gleich-, Binomial- oder Poissonverteilung sein.

¹⁰ Neben dem Chi-Quadrat-Anpassungstest existiert der Kolmogorov-Smirnov-Anpassungstest (vgl. z.B. Massey, F. J. (1951), Miller, L. H. (1956), Stephens, M. A. (1970), Marsaglia, G., Tsang, W., Wang, J. (2003)). Er besitzt den Vorteil, dass er schon bei kleinen Stichprobenumfängen anwendbar ist.

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Zur Berechnung der Teststatistik wird zunächst die Anzahl h_j der Elemente der Stichprobe notiert, die in die einzelnen Klassen A_j (für $j = 1, 2, \dots, k$) fallen. Anschließend wird für jede der k Klassen eine *erwartete absolute Häufigkeit* h_j^e berechnet, die angibt, wie viele Elemente in die jeweilige Klasse fallen müssten, wenn die Grundgesamtheit wie in der Nullhypothese angenommen (also gemäß F_0) verteilt ist. Diese Häufigkeit, die nicht unbedingt ganzzahlig sein muss, ergibt sich als

$$h_j^e = np_j, \quad (\text{III3: 47})$$

wobei n der Stichprobenumfang ist und p_j die Wahrscheinlichkeit dafür angibt, dass ein Beobachtungswert x_i bei korrekter Nullhypothese in die Klasse A_j fällt. Es gilt also formal $p_j = P(X \in A_j | F_0)$.

Im Anschluss daran vergleichen wir die h_j^e mit den h_j in den einzelnen Klassen. Eine Zusammenfassung aller Differenzen zwischen h_j und h_j^e führt zur Teststatistik

$$V = \sum_{j=1}^k \frac{(h_j - h_j^e)^2}{h_j^e}, \quad (\text{III3: 48})$$

Es werden also die Differenzen zwischen den empirischen und theoretischen Häufigkeiten quadriert und durch die zugehörigen erwarteten Häufigkeiten dividiert. Die letztgenannte Operation dient der Gewichtung der Abweichungen. Eine Differenz von 10 ist nämlich z.B. bei einer erwarteten Häufigkeit von 20 als hoch einzustufen, während sie bei einer erwarteten Häufigkeit von 1.000 nur geringfügig ist. Bei der Berechnung der konkreten Realisation der Teststatistik wird folgendes Schema empfohlen:

Klasse	A_1	...	A_k	Σ
Häufigkeiten in Stichprobe	h_1	...	h_k	n
Wahrscheinlichkeiten unter H_0	p_1	...	p_k	1
Erwartete Häufigkeiten unter H_0	$h_1^e = np_1$...	$h_k^e = np_k$	n
$(h_j - h_j^e)^2$	$(h_1 - h_1^e)^2$...	$(h_k - h_k^e)^2$	-
$(h_j - h_j^e)^2 / h_j^e$	$(h_1 - h_1^e)^2 / h_1^e$...	$(h_k - h_k^e)^2 / h_k^e$	V

Abbildung III 19: Hilfstabelle für den Chi-Quadrat-Anpassungstest

Die hier verwendete Testgröße V zeigt deutlich die Grundidee des Chi-Quadrat-Anpassungstests: Es wird für jede Klasse A_j die Anzahl h_j der Beobachtungen, die tatsächlich in A_j fallen, mit der Anzahl h_j^e der Werte verglichen, die in A_j bei korrekter Nullhypothese zu erwarten sind. Ein hoher Testfunktionswert v ist daher ein Indiz dafür, dass die hypothetische Verteilungsfunktion F_0 nicht richtig ist.

Es lässt sich zeigen, dass V unter H_0 näherungsweise einer *Chi-Quadrat-Verteilung* mit $(k - 1)$ Freiheitsgraden folgt.¹¹ Voraussetzung für die Anwendung der Chi-Quadrat-Verteilung ist jedoch, dass die erwarteten absoluten Häufigkeiten h_j^e (oder beobachteten absoluten Häufigkeiten h_j) nicht zu klein sind. Als *Faustregel* dient

¹¹ Diese Anzahl von Freiheitsgraden ergibt sich, da bei k Klassen und vorgegebenem Stichprobenumfang n nur in $k - 1$ Klassen die Besetzungszahlen frei variieren können.

$$V \stackrel{a}{=} \chi^2(k-1) \quad \text{falls} \quad h_j^e \geq 5 \text{ für } j = 1, 2, \dots, k. \quad (\text{III3: 49})$$

Wird diese Bedingung nicht von jeder der k Klassen erfüllt, behilft man sich häufig damit, benachbarte Klassen zusammenzulegen. Dabei ist jedoch zu bedenken, dass sich im Allgemeinen durch Reduktion von k die Wahrscheinlichkeit für den Fehler 2. Art (d.h. eine falsche Nullhypothese nicht abzulehnen) erhöht. Gerade daher sollte man bei einem Testergebnis, welches nicht zur Ablehnung von H_0 führt, beachten, ob dies nicht evtl. auf eine zu grobe Klasseneinteilung zurückzuführen ist.

Bei der Berechnung der erwarteten Häufigkeiten h_j^e wird die hypothetische Verteilung F_0 benötigt. Die Verteilung der verwendeten Teststatistik ist jedoch von der hypothetischen Verteilung unabhängig. Man spricht daher beim Chi-Quadrat-Anpassungstest auch von einem *verteilungsfreien Test*.

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Zur Entscheidung der Ablehnung oder Nichtablehnung der Nullhypothese nutzen wir das $(1-\alpha)$ -Fraktil der Chi-Quadrat-Verteilung mit $(k-1)$ Freiheitsgraden. Es gilt

$$\begin{aligned} H_0 &\text{ wird abgelehnt gegenüber} \\ H_A: F &\neq F_0, \quad \text{falls} \quad v > \chi_{1-\alpha, k-1}^2. \end{aligned} \quad (\text{III3: 50})$$

Beispiel III 3-13

χ^2 -Anpassungstest (diskrete Verteilung)

Ein Würfel wird 120-mal geworfen. Es ergeben sich dabei die Augenzahlen von 1 bis 6 mit folgenden Häufigkeiten:

Augenzahl a_j	1	2	3	4	5	6
Häufigkeit h_j in der Stichprobe	24	18	19	17	22	20

Es soll nun anhand dieser Stichprobe bei einem Signifikanzniveau $\alpha = 0,01$ geprüft werden, ob es sich bei dem Würfel um einen fairen Würfel handelt, also die Augenzahlen gleichverteilt sind.

(1) Hypothesen und Signifikanzniveau:

$$H_0: \text{Die Augenzahlen sind gleichverteilt.} \quad \alpha = 0,01$$

$$H_A: \text{Die Augenzahlen sind nicht gleichverteilt.}$$

Alternativ lässt sich dieses Hypothesenpaar auch folgendermaßen formulieren, wobei die A_j gleich den einzelnen Augenzahlen a_j sind:

$$H_0: P(A_1) = 1/6, P(A_2) = 1/6, \dots, P(A_6) = 1/6 \text{ bzw. } P(1) = 1/6, P(2) = 1/6, \dots, P(6) = 1/6$$

$$H_A: P(A_1) \neq 1/6, P(A_2) \neq 1/6, \dots, P(A_6) \neq 1/6 \text{ bzw. } P(1) \neq 1/6, P(2) \neq 1/6, \dots, P(6) \neq 1/6$$

(2) Teststatistik und deren Verteilung:

Wie wir der Tabelle auf der Folgeseite entnehmen können, erhalten wir für die vorliegende Stichprobe eine konkrete Ausprägung der Teststatistik von $v = 1,70$. Wir erkennen außerdem, dass für alle j die Bedingung $h_j^e \geq 5$ erfüllt ist. Es gilt somit für die Verteilung der Teststatistik $V \stackrel{a}{=} \chi^2(k-1)$ bzw. konkret $V \stackrel{a}{=} \chi^2(5)$.

(3) Entscheidungsregel:

$$H_0 \text{ ablehnen gegenüber } H_A, \text{ wenn } v > \chi_{1-\alpha, k-1}^2 \text{ gilt.}$$

$$\text{Dabei ist } \chi_{1-\alpha, k-1}^2 = \chi_{0,99; 5}^2 = 15,09.$$

Augenzahl a_j bzw. Klasse j	1	2	3	4	5	6	Σ
h_j	24	18	19	17	22	20	120
p_j	1/6	1/6	1/6	1/6	1/6	1/6	1
$h^e_j = np_j$	20	20	20	20	20	20	120
$(h_j - h^e_j)^2$	16	4	1	9	4	0	-
$(h_j - h^e_j)^2 / h^e_j$	0,80	0,20	0,05	0,45	0,20	0,00	1,70

(4) Entscheidung:

Es gilt hier ($v = 1,70 < (\chi^2_{1-\alpha, k-1} = 15,09)$.

H_0 kann nicht abgelehnt werden. Die vorliegende Stichprobe liefert also keinen statistischen Nachweis dafür, dass der betrachtete Würfel manipuliert wurde.

Da die Chi-Quadrat-Verteilung zu den stetigen Verteilungen gehört, V aber eine diskrete Verteilung besitzt, wäre genau genommen eine *Stetigkeitskorrektur* (sog. Yates-Korrektur) erforderlich, sodass sich

$$v = \sum_{j=1}^k \frac{(|h_j - h_j^e| - 0,5)^2}{h_j^e}$$

ergibt. In der Praxis wird diese Stetigkeitskorrektur jedoch nur angewendet, wenn die *Anzahl der Freiheitsgrade* bei *Eins* liegt. Wir wollen sie daher im Folgenden vernachlässigen.

Müssen zur Ermittlung der erwarteten Häufigkeiten zunächst **g Parameter der hypothetischen Verteilung F_0 geschätzt** werden, vermindert sich die Anzahl der Freiheitsgrade der zu verwendenden Chi-Quadrat-Verteilung auf $k - g - 1$. Es muss $k - g - 1 > 0$ gelten, da die Anzahl der Freiheitsgrade einer Chi-Quadrat-Verteilung immer positiv sein muss.

Beispiel III 3-14 **χ^2 -Anpassungstest (diskrete Verteilung)**

Die Anzahl der an einem abgelegenen Grenzübergang in einem 10-Minuten-Intervall ankommenden Pkws, ist in folgender Tabelle wiedergegeben.

Klasse j	1	2	3	4	5	6	7
Anzahl der ankommenden Pkws a_j	0	1	2	3	4	5	≥ 6
Häufigkeit h_j in der Stichprobe	2	7	21	42	14	14	0

Es soll geprüft werden, ob die Anzahl der Pkw-Ankünfte als Poisson-verteilt betrachtet werden kann (Signifikanzniveau $\alpha = 0,05$).

(1) Hypothesen und Signifikanzniveau:

H_0 : Die Pkw-Ankünfte sind Poisson-verteilt. $\alpha = 0,05$

H_A : Die Pkw-Ankünfte sind nicht Poisson-verteilt.

(2) Teststatistik und deren Verteilung:

Zur Bestimmung der erwarteten Häufigkeiten h_j^e muss der Funktionalparameter λ der Poissonverteilung geschätzt werden. Da λ dem Erwartungswert der Poissonverteilung entspricht, können wir λ mit dem arithmetischen Mittel der Stichprobe schätzen. Es gilt

$$\hat{\lambda} = \bar{x} = \frac{\sum_{j=1}^7 a_j h_j}{\sum_{j=1}^7 h_j} = \frac{301}{100} = 3,01.$$

Mit Hilfe der Wahrscheinlichkeitsfunktion der Poissonverteilung (vgl. Abschnitt II 3.1.3)

$$f_{Po}(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

lassen sich unter Verwendung des geschätzten λ -Wertes die zur Ermittlung der h_j^e notwendigen Wahrscheinlichkeiten p_j berechnen. Wir erhalten die folgende Tabelle, in der sich der Wert 0,0849 als Komplementärwahrscheinlichkeit $1 - p_1 - \dots - p_6$ ergibt.

Klasse j a_j	1	2	3	4	5	6	7	Σ
h_j	2	7	21	42	14	14	0	100
$p_j = f_{Po}(a_j)$	0,0493	0,1484	0,2233	0,2240	0,1686	0,1015	0,0849	1
$h_j^e = np_j$	4,93	14,84	22,33	22,40	16,86	10,15	8,49	100

Wir stellen nun aber fest, dass die Klasse 1 die Bedingung $h_j^e \geq 5$ nicht erfüllt. Wir fassen sie daher mit der benachbarten Klasse 2 zusammen und erhalten so die folgende Tabelle, die uns v liefert.

Klasse j a_j	1 ≤ 1	2	3	4	5	6	Σ
h_j	9	21	42	14	14	0	100
$p_j = f_{Po}(a_j)$	0,1977	0,2233	0,2240	0,1686	0,1015	0,0849	1
$h_j^e = np_j$	19,77	22,33	22,40	16,86	10,15	8,49	100
$(h_j - h_j^e)^2$	115,99	1,77	384,16	8,18	14,82	72,08	-
$(h_j - h_j^e)^2 / h_j^e$	5,87	0,08	17,15	0,49	1,46	8,49	33,54

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > \chi^2_{1-\alpha, k-g-1}$ gilt.

Dabei ist $\chi^2_{1-\alpha, k-g-1} = \chi^2_{1-0,05; 6-1-1} = \chi^2_{0,95; 4} = 9,49$. Es gilt hier $g = 1$, da genau ein Parameter der hypothetischen Verteilung, d.h. konkret der Parameter λ der Poissonverteilung, geschätzt werden muss.

(4) Entscheidung:

Es gilt ($v = 33,54 > (\chi^2_{1-\alpha, k-g-1} = 9,49)$).

H_0 kann abgelehnt werden. Wir können die Anzahl der Pkw-Ankünfte also nicht als Poisson-verteilt betrachten.

3.4.1.2 Anpassungstest bei stetig verteilter Grundgesamtheit

Liegt ein stetiges Merkmal oder ein diskretes Merkmal mit sehr vielen Merkmalsausprägungen vor, ist eine Klassenbildung unerlässlich. Wir wollen dies am Beispiel des sog. *Normalitätstests* (Chi-Quadrat-Anpassungstest mit der Nullhypothese der Normalverteilung) veranschaulichen.

Beispiel III 3-15

χ^2 -Anpassungstest (stetige Verteilung)



Eine Stichprobe im Umfang von $n = 70$ lieferte für die Lebensdauer von Glühbirnen die folgende bereits in Klassen eingeteilten Werte (in Jahren) mit dazugehörigen Klassenhäufigkeiten. Der aus der Urliste berechnete Stichprobenmittelwert lag bei $\bar{x} = 3,5$ Jahren, die Stichprobenstandardabweichung bei $\hat{\sigma} = 0,7$ Jahren.

A _j	[0; 2,5]]2,5; 3,0]]3,0; 3,5]]3,5; 4,0]]4,0; 4,5]]5,0; ∞[
h _j	4	7	24	22	8	5

Wir wollen anhand dieser Stichprobenergebnisse auf einem Signifikanzniveau von 5 % testen, ob die Lebensdauer der Glühbirnen in der Grundgesamtheit normalverteilt ist.

(1) Hypothesen und Signifikanzniveau:

$$H_0: \text{Die Lebensdauer ist normalverteilt.} \quad \alpha = 0,05$$

$$H_A: \text{Die Lebensdauer ist nicht normalverteilt.}$$

(2) Teststatistik und deren Verteilung:

Zur Bestimmung der erwarteten Häufigkeiten h_j^* müssen die zwei Funktionalparameter μ und σ^2 der Normalverteilung geschätzt werden. Als erwartungstreue Schätzer eignen sich dafür der Stichprobenmittelwert und die Stichprobenvarianz. Unter Verwendung der Schätzwerte $\hat{\mu} = 3,5$ und $\hat{\sigma}^2 = 0,7^2$ bzw. der geschätzten hypothetischen Verteilungsfunktion $\hat{F}_0 = N(3,5; 0,7^2)$ erhalten wir daher aus folgender Tabelle:

a _j	2,5	3,0	3,5	4,0	4,5	∞	Σ
h _j	4	7	24	22	8	5	70
P(X ≤ a _j \hat{F}_0) = $\Phi(z_j)$	0,077	0,238	0,500	0,762	0,923	1,000	
p _j = $\Phi(z_j) - \Phi(z_{j-1})$	0,077	0,161	0,262	0,262	0,161	0,077	1
h _j *	5,359	11,267	18,373	18,373	11,267	5,359	70
(h _j - h _j *) ²	1,848	18,210	31,660	13,153	10,675	0,129	
(h _j - h _j *) ² / h _j *	0,345	1,616	1,723	0,716	0,947	0,024	5,37

Die darin vorkommenden Werte a_j entsprechen den Obergrenzen der Klassen aus der Aufgabenstellung. Jede Klassenobergrenze wird mittels der Schätzungen $\hat{\mu}$ und $\hat{\sigma}^2$ standardisiert. Aus den daraus entstehenden Werten z_j werden dann die dazu gehörigen Werte der Verteilungsfunktion der Standardnormalverteilung ermittelt. Aus zwei aufeinander folgenden Werten $\Phi(z_{j-1})$ und $\Phi(z_j)$ lässt sich schließlich durch Differenzbildung für jede Klasse j ein Wert p_j ermitteln. Alle weiteren Werte der Arbeitstabelle ergeben sich nach dem gewohnten Muster.

Es ergibt sich demnach die konkrete Ausprägung der Teststatistik als v = 5,37, wobei die Teststatistik V aufgrund der erforderlichen Schätzung zweier Verteilungsparameter

und der für alle j erfüllten Bedingung $h_j^e \geq 5$ näherungsweise einer Chi-Quadrat-Verteilung mit $k - g - 1$ (bzw. $6 - 2 - 1 = 3$) Freiheitsgraden folgt.¹² Wegen $h_j^e \geq 5$ für alle j ist keine weitere Klassenzusammenfassung notwendig.

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > \chi_{1-\alpha, k-g-1}^2$ gilt.

Dabei ist $\chi_{1-\alpha, k-g-1}^2 = \chi_{0,95; 3}^2 = 7,81$.

(4) Entscheidung:

Es gilt ($v = 5,37 < (\chi_{1-\alpha, k-g-1}^2 = 7,81)$).

H_0 kann nicht abgelehnt werden. Die vorliegende Stichprobe widerspricht also nicht der Normalverteilungshypothese. Auch bei Verwendung der Chi-Quadrat-Verteilung mit nur $(k - 1)$ Freiheitsgraden ergibt sich keine Ablehnung.

Bei der Anwendung des Chi-Quadrat-Anpassungstests in der **Praxis** ist zu beachten, dass er mitunter sehr *große β -Fehler* (geringe Macht) aufweist.¹³ Ein Anpassungstest ist deshalb i.d.R. kein geeignetes Verfahren zur Bestimmung des Typs einer Verteilung. In vielen Fällen können nämlich Nullhypthesen, wie etwa "X ist normalverteilt", "X ist lognormalverteilt" oder "X ist exponentialverteilt" allesamt nicht verworfen werden. Gerade deshalb wurden speziellere Testverfahren entwickelt. Dazu zählt beispielsweise der *Jarque-Bera-Test* auf Normalverteilung (vgl. Abschnitt IV 3.3).

3.4.2 Chi-Quadrat-Unabhängigkeitstest

Dieses auch als **Kontingenzttest** bezeichnete Testverfahren überprüft, ob zwei in der Grundgesamtheit beobachtete Merkmale X und Y voneinander unabhängig sind oder nicht. Seine Anwendung ist bereits bei nominal skalierten Daten möglich. Man geht bei ihm nach dem gleichen Prinzip vor, das auch dem Chi-Quadrat-Anpassungstest zugrunde liegt.

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Unter der Voraussetzung, dass eine zweidimensionale Stichprobe bzw. zwei verbundene einfache Stichproben aus einer größeren Grundgesamtheit mit den Stichprobenvariablen X_i ($i = 1, \dots, n$) zu X und Y_i ($i = 1, \dots, n$) zu Y gegeben sind, werden folgende Hypothesen getestet:

H_0 : Die beiden Merkmale X und Y sind unabhängig.

H_A : X und Y sind abhängig.

(III3: 51)

Anders als beim Korrelationstest (vgl. III 3.3.3.2) ist die (gemeinsame) Normalverteilung von X und Y keine Voraussetzung für diesen Test auf Unabhängigkeit.

¹² Werden die Schätzungen der unbekannten Verteilungsparameter aus der Urliste und nicht aus den klassierten Daten gewonnen, ist die Grenzverteilung der resultierenden Testgröße V genau genommen keine Chi-Quadrat-Verteilung mehr. Das zur Grenzverteilung von V gehörende $(1-\alpha)$ -Fraktile, liegt allerdings zwischen den entsprechenden Fraktilen der $\chi^2(k - g - 1)$ und der $\chi^2(k - 1)$ -Verteilung. Vgl. dazu Albrecht, P. (1980). Für hinreichend großes k sind diese drei Fraktile außerdem nahezu identisch.

¹³ Vgl. dazu z.B. Karlis, D., Xekalaki, E. (2000).

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Als Stichprobenergebnis erhalten wir Paare (x, y) , deren Häufigkeiten in einer Kontingenztabelle der in Abbildung III 20 dargestellten Form aufbereitet werden. Dabei werden zunächst die in der Stichprobe auftretenden Ausprägungen von X in $k \geq 2$ disjunkte, aneinander angrenzende Klassen A_1, \dots, A_k aufgeteilt. Gleiches gilt für die Ausprägungen von Y, die in $m \geq 2$ Klassen B_1, \dots, B_m aufgegliedert werden. Bei diskreten Merkmalen X und Y mit wenigen möglichen Ausprägungen kann es vorkommen, dass eine Klasse nur aus einer einzigen Merkmalsausprägung von X bzw. Y besteht. Die Anzahl derjenigen Paare (x, y) , die sowohl in Klasse A_i als auch B_j fallen, wird mit h_{ij} (gemeinsame Häufigkeit) bezeichnet. Interessieren wir uns dafür, wie viele Paare (x, y) in die Klasse A_i (ungeachtet von B) fallen, so ist dies aus der Randhäufigkeit $h(A_i)$ abzulesen. Dies gilt analog für B_j und $h(B_j)$.

Wir wollen zusätzlich zu den Stichprobenergebnissen auch gleich die Größen

$$h_{ij}^e = \frac{h(A_i) \cdot h(B_j)}{n} \quad (\text{III3: 52})$$

in die Kontingenztabelle aufnehmen. Es handelt sich dabei um die bei Gültigkeit der Nullhypothese in der Stichprobe *erwarteten absoluten Häufigkeiten*. Dies lässt sich wie folgt begründen: Da die relativen Häufigkeiten der Stichprobenelemente, die in die Klasse A_i bzw. B_j fallen, also

$$f(A_i) = \frac{h(A_i)}{n} \quad (i = 1, \dots, k) \quad \text{und} \quad f(B_j) = \frac{h(B_j)}{n} \quad (j = 1, \dots, m),$$

erwartungstreue *Schätzer* für die **unbekannten Randwahrscheinlichkeiten** in der Grundgesamtheit darstellen, muss bei Unabhängigkeit der beiden Merkmale X und Y nach dem Multiplikationssatz für unabhängige Ereignisse für alle i und j

$$f(A_i \cap B_j | H_0) = f(A_i) \cdot f(B_j) = \frac{h(A_i)}{n} \cdot \frac{h(B_j)}{n} \quad (i = 1, \dots, k; j = 1, \dots, m)$$

gelten. Wir erhalten daraus durch Multiplikation mit dem Stichprobenumfang n genau die unter (III3: 52) definierte Größe. Die erwarteten absoluten Häufigkeiten ergeben sich also allgemein als Produkt aus den jeweiligen Randhäufigkeiten dividiert durch den Stichprobenumfang n.

X	Y	B_1	B_2	...	B_m	Σ
A_1		h_{11} h_{11}^e	h_{12} h_{12}^e	...	h_{1m} h_{1m}^e	$h(A_1)$
A_2		h_{21} h_{21}^e	h_{22} h_{22}^e	...	h_{2m} h_{2m}^e	$h(A_2)$
\vdots		\vdots	\vdots	\ddots	\vdots	\vdots
A_k		h_{k1} h_{k1}^e	h_{k2} h_{k2}^e	...	h_{km} h_{km}^e	$h(A_k)$
Σ		$h(B_1)$	$h(B_2)$...	$h(B_m)$	n

Abbildung III 20: Kontingenztabelle für den Chi-Quadrat-Unabhängigkeitstest

Aus der Kontingenztabelle in Abbildung III 20 lässt sich schließlich, wie auch bereits beim Chi-Quadrat-Anpassungstest, die Teststatistik

$$V = \sum_{i=1}^k \sum_{j=1}^m \frac{(h_{ij} - h_{ij}^e)^2}{h_{ij}^e} \quad (\text{III3: 53})$$

aufstellen. Da bei Unabhängigkeit von X und Y zu erwarten ist, dass $h_{ij} = h_{ij}^e$ gilt, spricht ein kleiner Teststatistikwert v für Unabhängigkeit und ein hoher Wert v für Abhängigkeit. V folgt hier unter H_0 näherungsweise einer *Chi-Quadrat-Verteilung* mit $(k - 1)(m - 1)$ Freiheitsgraden. Diese Anzahl lässt sich wie folgt erklären: Die Testgröße V hat so viele Summanden, wie die Kontingenztabelle Zellen hat. Dies führt zunächst zu $k \cdot m - 1$ Freiheitsgraden, da bei Kenntnis von $k \cdot m - 1$ Feldinhaltten und bekanntem Stichprobenumfang der letzte Feldinhalt nicht frei variieren kann. Da unbekannte Parameter (Randwahrscheinlichkeiten) aus der Stichprobe geschätzt werden müssen, sind diese Freiheitsgrade um $k - 1$ Freiheitsgrade für die Randverteilung von X und $m - 1$ Freiheitsgrade für die Randverteilung von Y zu reduzieren. Daher gilt $km - 1 - (k - 1) - (m - 1) = km - k - m + 1 = (k - 1)(m - 1)$.

Von einer approximativen $\chi^2[(k - 1)(m - 1)]$ -Verteilung der Größe V kann aber wieder nur dann ausgegangen werden, wenn die *Faustregel*

$$h_{ij}^e \geq 5 \quad \text{für alle } i = 1, 2, \dots, k \text{ und } j = 1, 2, \dots, m \quad (\text{III3: 54})$$

erfüllt ist. Treten erwartete Häufigkeiten auf, die kleiner als 5 sind, sind wieder Spalten bzw. Zeilen zusammenzufassen.

Liegt eine Kontingenztabelle mit $k = m = 2$ vor, vereinfacht sich die Berechnung der Teststatistik. Es gilt dann einfach

$$V = \frac{n(h_{11}h_{22} - h_{12}h_{21})^2}{h(A_1) \cdot h(A_2) \cdot h(B_1) \cdot h(B_2)}. \quad (\text{III3: 55})$$

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Zur Entscheidung der Ablehnung oder Nichtablehnung der Nullhypothese bedienen wir uns des $(1 - \alpha)$ -Fraktils der Chi-Quadrat-Verteilung mit $(k - 1)(m - 1)$ Freiheitsgraden. Es gilt

H ₀ wird abgelehnt gegenüber	
H _A : X und Y sind abhängig, falls	$v > \chi^2_{1-\alpha, (k-1)(m-1)}$.

(III3: 56)

Beispiel III 3-16

Chi-Quadrat-Unabhängigkeitstest

Im Rahmen eines Einstellungstests werden 20 Ingenieure, 40 Betriebswirte und 20 Volkswirte einem Intelligenztest unterzogen, der ihre Eignung für das Unternehmen bestimmen soll. Es ergab sich dabei folgende Kontingenztabelle:

Abschluss Eignung	Ingenieur	Betriebswirt	Volkswirt	Σ
geeignet	5	25	10	40
nicht geeignet	15	15	10	40
Σ	20	40	20	80

Es soll anhand dieses Ergebnisses bei einem Signifikanzniveau $\alpha = 0,01$ getestet werden, ob die Eignung (Merkmal X) vom Studienabschluss (Merkmal Y) unabhängig ist oder nicht.

(1) Hypothesen und Signifikanzniveau:

$$H_0: \text{Die Merkmale } X \text{ und } Y \text{ sind unabhängig.} \quad \alpha = 0,01$$

$$H_A: \text{Die Merkmale } X \text{ und } Y \text{ sind abhängig.}$$

(2) Teststatistik und deren Verteilung:

Zur Berechnung der konkreten Ausprägung von V wollen wir zunächst die Kontingenztafel aus der Aufgabenstellung um die bei Gültigkeit der Nullhypothese erwarteten absoluten Häufigkeiten h_{ij}^e (z.B. $h_{11}^e = (20 \cdot 40) / 80 = 10$) erweitern.

X	Y	Ingenieur	Betriebswirt	Volkswirt	Σ
		5	25	10	
geeignet	10	20	10	40	40
	15	15	20	10	
Σ	20	40	20	20	80

Da alle h_{ij}^e größer als 5 sind, ist die Teststatistik V näherungsweise mit $(k - 1)(m - 1) = (2 - 1)(3 - 1) = 2$ Freiheitsgraden Chi-Quadrat-verteilt. Die konkrete Ausprägung v ergibt sich zu

$$\begin{aligned} v &= \sum_{i=1}^2 \sum_{j=1}^3 \frac{(h_{ij} - h_{ij}^e)^2}{h_{ij}^e} \\ &= \frac{(5-10)^2}{10} + \frac{(25-20)^2}{20} + \frac{(10-10)^2}{10} + \frac{(15-10)^2}{10} + \frac{(15-20)^2}{20} + \frac{(10-10)^2}{10} \\ &= 2,5 + 1,25 + 0 + 2,5 + 1,25 + 0 = 7,5. \end{aligned}$$

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > \chi^2_{1-\alpha, (k-1)(m-1)}$ gilt.

Dabei ist $\chi^2_{1-\alpha, (k-1)(m-1)} = \chi^2_{0,99; 2} = 9,21$.

(4) Entscheidung:

Es gilt hier ($v = 7,5 < \chi^2_{1-\alpha, (k-1)(m-1)} = 9,21$).

H_0 kann nicht abgelehnt werden. Wir können also nicht statistisch nachweisen, dass die Eignung vom Studienabschluss abhängig ist.

Bisher sind wir davon ausgegangen, dass die Randwahrscheinlichkeiten in der Grundgesamtheit unbekannt sind. Wir hatten sie daher mit den relativen Randhäufigkeiten in der Stichprobe geschätzt und konnten so die h_{ij}^e bestimmen. Sollten die **Randwahrscheinlichkeiten** $p_i = P(A_i)$ und $q_j = P(B_j)$ in der Grundgesamtheit **bekannt** sein, müssen zur Bestimmung der h_{ij}^e nicht länger die Randhäufigkeiten aus der Stichprobe verwendet werden. Wir rechnen in einem solchen Fall

$$h_{ij}^e = n \cdot p_i \cdot q_j. \quad (\text{III3: 57})$$

Bei bekannten Randwahrscheinlichkeiten der Grundgesamtheit ändert sich auch die Zahl der Freiheitsgrade der zu verwendenden Chi-Quadrat-Verteilung. Wir erhalten nun $(k \cdot m - 1)$ Freiheitsgrade. Der schätzungsbedingte Abzug, den wir im Fall unbekannter Randwahrscheinlichkeiten vorgenommen hatten, ist nun nicht mehr erforderlich.

Beispiel III 3-17

Chi-Quadrat-Unabhängigkeitstest

Eine Befragung unter 500 zufällig ausgewählten Auszubildenden, die Aufschluss über ihr Geschlecht (Merkmal Y) und ihre Ausbildungsrichtung (Merkmal X) geben soll, liefert die folgende Kontingenztabelle, in der an Stelle der Randhäufigkeiten der Stichprobe, die bekannten Randwahrscheinlichkeiten der Grundgesamtheit eingetragen sind. p_i bzw. q_j ist also dabei als die Wahrscheinlichkeit dafür zu interpretieren, dass ein zufällig ausgewählter Auszubildender beruflich bzw. geschlechtlich zur Gruppe i bzw. j gehört.

Ausbildungsrichtung X Geschlecht Y	männlich	weiblich	p_i
Handwerk	150	100	0,45
Industrie und Handel	100	50	0,30
öffentlicher Dienst	50	50	0,25
q_j	0,45	0,55	

Es soll auf einem Signifikanzniveau von $\alpha = 0,05$ geprüft werden, ob die Wahl der Ausbildungsrichtung (Merkmal X) vom Geschlecht (Merkmal Y) unabhängig ist.

(1) Hypothesen und Signifikanzniveau:

$$H_0: \text{Die Merkmale X und Y sind unabhängig.} \quad \alpha = 0,05$$

$$H_A: \text{Die Merkmale X und Y sind abhängig.}$$

(2) Teststatistik und deren Verteilung:

Zur Berechnung der konkreten Ausprägung von V werden zunächst die h_{ij}^e anhand der bekannten Randwahrscheinlichkeiten der Grundgesamtheit berechnet. Wir stellen diese zusammen mit den h_{ij} aus der Stichprobe in folgender Tabelle dar:

X	Y	männlich	weiblich
Handwerk		150 101,25	100 123,75
Industrie und Handel		100 67,50	50 82,50
öffentlicher Dienst		50 56,25	50 68,75

Die Teststatistik V ist näherungsweise mit $k \cdot m - 1 = 3 \cdot 2 - 1 = 5$ Freiheitsgraden Chi-Quadrat-verteilt. Die konkrete Ausprägung v lautet

$$v = \sum_{i=1}^3 \sum_{j=1}^2 \frac{(h_{ij} - h_{ij}^e)^2}{h_{ij}^e} = \frac{(150 - 101,25)^2}{101,25} + \frac{(100 - 123,75)^2}{123,75} + \frac{(50 - 67,50)^2}{67,50} + \frac{(50 - 82,50)^2}{82,50} + \frac{(50 - 56,25)^2}{56,25} + \frac{(50 - 68,75)^2}{68,75}$$

$$v = 23,47 + 4,56 + 15,65 + 12,80 + 0,69 + 5,11 = 62,28.$$

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > \chi^2_{1-\alpha, k-m-1}$ gilt.

Dabei ist $\chi^2_{1-\alpha, k-m-1} = \chi^2_{0,95; 5} = 11,07$.

(4) Entscheidung:

Es gilt hier ($v = 62,28 > (\chi^2_{1-\alpha, k-m-1} = 11,07)$).

H_0 kann abgelehnt werden. Man kann also annehmen, dass die Wahl der Ausbildungsrichtung vom Geschlecht des Auszubildenden abhängig ist.

Abschließend sei erwähnt, dass der Chi-Quadrat-Unabhängigkeitstest auch als Anpassungstest zu verstehen ist. Es wird nämlich die Anpassung an eine unabhängige Verteilung getestet. Dies erklärt auch die ähnliche Vorgehensweise und die Anzahl der zu verwendenden Freiheitsgrade der Testverteilung.

3.4.3 Chi-Quadrat-Homogenitätstest

Der Chi-Quadrat-Anpassungstest aus Abschnitt III 3.4.1 gehörte zur Gruppe der Einstichprobentests. Er untersucht, ob die in *einer Stichprobe* beobachtete Verteilung mit der für die unbekannte Verteilung der Grundgesamtheit gemachten Annahme in Widerspruch steht oder nicht. Der Chi-Quadrat-Homogenitätstest ist in die Gruppe der Zwei- und Mehrstichprobentests einzuordnen. Es wird nämlich hier untersucht, ob *mehrere unabhängige Stichproben* aus der gleichen Grundgesamtheit (oder aus verschiedenen Grundgesamtheiten mit gleicher Verteilung) stammen oder nicht. Wir werden daher eine ähnliche Vorgehensweise wie beim Chi-Quadrat-Unabhängigkeitstest feststellen.

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Die Nullhypothese beinhaltet bei diesem Test immer die Behauptung, dass die Verteilungen F_i ($i = 1, \dots, m$), aus denen die m vorliegenden Stichproben gewonnenen wurden, gleich sind.

$$\begin{aligned} H_0: F_1 &= F_2 = \dots = F_m \\ H_A: F_i &\neq F_j \text{ für mindestens ein Paar } (i,j) \end{aligned} \quad (\text{III3: 58})$$

Beim Vergleich der Verteilungen zweier Stichproben würden wir also $H_0: F_1 = F_2$ gegenüber $H_A: F_1 \neq F_2$ testen.

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Wie beim Chi-Quadrat-Unabhängigkeitstest empfiehlt sich auch hier die Verwendung einer Art von Kontingenztabelle. Diese kann wie in Abbildung III 21 gestaltet werden. Dort geben die h_{ij} an, wie oft in der j -ten Verteilung ein Wert aus der Klasse A_i vorkommt. Folglich sind die $h(SP_j)$ die Anzahl der Stichprobenbeobachtungswerte in der Verteilung j und $h(A_i)$ die Klassenhäufigkeiten insgesamt. Die Werte h_{ij}^e ergeben sich wie beim Chi-Quadrat-Unabhängigkeitstest. Auch die Teststatistik V ist identisch. Wieder ist V näherungsweise mit $(k - 1)(m - 1)$ Freiheitsgraden Chi-Quadrat-verteilt, wenn $h_{ij}^e \geq 5$ für alle i und j gilt.

X	Verteilung aus Stichprobe 1	Verteilung aus Stichprobe 2	...	Verteilung aus Stichprobe m	Σ
A_1	h_{11} h_{11}^e	h_{12} h_{12}^e	...	h_{1m} h_{1m}^e	$h(A_1)$
A_2	h_{21} h_{21}^e	h_{22} h_{22}^e	...	h_{2m} h_{2m}^e	$h(A_2)$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
A_k	h_{k1} h_{k1}^e	h_{k2} h_{k2}^e	...	h_{km} h_{km}^e	$h(A_k)$
Σ	$h(SP_1)$	$h(SP_2)$...	$h(SP_m)$	n

Abbildung III 21: Kontingenztabelle für den Chi-Quadrat-Homogenitätstest

Wir erkennen deutlich die Verwandtschaft des Chi-Quadrat-Homogenitätstests zum Chi-Quadrat-Unabhängigkeitstest. Die Nullhypothese beim Unabhängigkeitstest besagt nämlich, dass die bedingten Verteilungen alle gleich sind. Beim Homogenitätstest entsprechen die einzelnen Verteilungen F_1, F_2, \dots, F_m genau den bedingten Verteilungen des Chi-Quadrat-Unabhängigkeitstests.

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Über Ablehnung und Nicht-Ablehnung der Nullhypothese wird nach folgender Regel entschieden:

H_0 wird abgelehnt gegenüber

$$H_A : F_i \neq F_j \text{ für mindestens ein Paar } (i,j), \quad \text{falls} \quad v > \chi^2_{1-\alpha, (k-1)(m-1)} \quad (\text{III3: 59})$$

Beispiel III 3-18

Chi-Quadrat-Homogenitätstest

In einem Bundesland wurde bei 50 Kleinunternehmen, die seit mehr als zehn Jahren bestehen ("alte Unternehmen"), sowie bei 100 Kleinunternehmen, die seit weniger als zehn Jahren bestehen ("junge Unternehmen"), der Fremdkapitalanteil am Firmenkapital untersucht. Es ergab sich daraus folgende Kontingenztabelle:

Fremdkapitalanteil	alte Unternehmen	junge Unternehmen	Σ
0 - 25 %	20	15	35
25 - 50 %	15	15	30
50 - 75 %	10	35	45
75 - 100 %	5	35	40
Σ	50	100	150

Es soll bei einem Signifikanzniveau von $\alpha = 0,02$ untersucht werden, ob die beiden Unternehmensgruppen bezüglich ihrer Fremdfinanzierung homogen sind.

(1) Hypothesen und Signifikanzniveau:

H_0 : Der Verschuldungsgrad in beiden Gruppen stimmt überein. $\alpha = 0,02$

H_A : Der Verschuldungsgrad in beiden Gruppen stimmt nicht überein.

(2) Teststatistik und deren Verteilung:

Wir benötigen zunächst wieder die Werte h_{ij}^e , die wir nach der bereits bekannten Formel (III3: 53) errechnen können.

Fremdkapitalanteil	alte Unternehmen	junge Unternehmen
0 - 25 %	20 11,67	15 23,33
25 - 50 %	15 10	20 20
50 - 75 %	10 15	35 30
75 - 100 %	5 13,33	30 26,67

Die Teststatistik V ist näherungsweise mit $(k-1)(m-1) = (4-1)(2-1) = 3$ Freiheitsgraden Chi-Quadrat-verteilt. Die konkrete Ausprägung v ergibt sich wie folgt:

$$\begin{aligned}
 v &= \sum_{i=1}^4 \sum_{j=1}^2 \frac{(h_{ij} - h_{ij}^e)^2}{h_{ij}^e} = \frac{(20-11,67)^2}{11,67} + \frac{(15-23,33)^2}{23,33} + \frac{(15-10)^2}{10} + \frac{(20-20)^2}{20} \\
 &\quad + \frac{(10-15)^2}{15} + \frac{(35-30)^2}{30} + \frac{(5-13,33)^2}{13,33} + \frac{(30-26,67)^2}{26,67} \\
 &= 5,95 + 2,97 + 2,50 + 0,00 + 1,67 + 0,83 + 5,21 + 0,42 \\
 &= 19,55
 \end{aligned}$$

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > \chi^2_{1-\alpha, (k-1)(m-1)}$ gilt.

Dabei ist $\chi^2_{1-\alpha, (k-1)(m-1)} = \chi^2_{0,98; 3} = 9,84$.

(4) Entscheidung:

Es gilt hier ($v = 19,55 > \chi^2_{1-\alpha, (k-1)(m-1)} = 9,84$).

H_0 kann abgelehnt werden. Wir können also davon ausgehen, dass der Verschuldungsgrad in beiden Unternehmensgruppen nicht übereinstimmt.

3.5 Einfache Varianzanalyse

Unter einer Varianzanalyse (engl. analysis of variance, ANOVA) verstehen wir einen statistischen Test, der die Frage überprüft, ob die Differenz bzw. die Differenzen der arithmetischen Mittel von zwei oder mehr Grundgesamtheiten signifikant von Null verschieden ist bzw. sind oder nicht. Im Vergleich zum Zweistichprobentest für das arithmetische Mittel (vgl. Abschnitt III 3.3.2.1) können also mit diesem Test mehr als zwei arithmetische Mittel gleichzeitig auf *Homogenität* (Gleichheit) getestet werden. Wir könnten daher auch von einem "Mehrstichprobentest für die

Differenz arithmetisches Mittel" sprechen. Da in die Teststatistik dieses Tests die Stichprobenvarianzen eingehen, hat sich in der Literatur jedoch die Bezeichnung **Varianzanalyse** eingebürgert.

(1) Formulierung einer Nullhypothese und einer Gegenhypothese

Liegen $r \geq 2$ unabhängige einfache Stichproben vor, können die Stichprobenvariablen der j -ten Stichprobe, welche den Umfang n_j besitzt, mit X_{j1}, \dots, X_{jn_j} bezeichnet werden. Die arithmetischen Mittel der betrachteten Grundgesamtheiten wollen wir mit μ_j benennen.

Der sog. *einfachen Varianzanalyse*¹⁴ liegt die Modellannahme zu Grunde, dass die **Stichprobenvariablen** jeweils mit dem arithmetischen Mittel μ_i und der Varianz σ^2 **normalverteilt** sind. Die Varianzen sind also alle identisch. Wir sprechen dabei von **Varianzhomogenität** (auch häufig *Homoskedastizität* genannt).¹⁵ Die X_i sind also $N(\mu_i, \sigma^2)$ -verteilt.

Getestet wird nun unter diesen Annahmen, ob die Verteilungen, aus denen die r Stichproben stammen, alle denselben Erwartungswert besitzen oder nicht. Wir formulieren also das Hypothesenpaar

$$H_0: \mu_1 = \mu_2 = \dots = \mu_r \quad (III3: 60)$$

H_A : Mindestens zwei der μ_i sind verschieden.

(2) Ableiten einer Teststatistik V und ihrer Verteilung

Es gilt allgemein, dass sich die Gesamtstichprobenvarianz $\hat{\sigma}_{\text{Ges}}^2$ aller

$$n = \sum_{j=1}^r n_j$$

Beobachtungen als Summe der externen und internen Stichprobenvarianz darstellen lässt. Wir können also zunächst festhalten, dass

$$\hat{\sigma}_{Ges}^2 = \underbrace{\frac{1}{n-1} \sum_{j=1}^r n_j (\bar{X}_j - \bar{X}_{Ges})^2}_{\text{extern} \quad (\text{Variation zwischen den Gruppen- durchschnitten})} + \underbrace{\frac{1}{n-1} \sum_{j=1}^r \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2}_{\text{intern} \quad (\text{Variation innerhalb der Gruppen})}$$

gilt, wobei

$$\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji} \quad \text{und} \quad \bar{X}_{Ges} = \frac{1}{n} \sum_{j=1}^r n_j \bar{X}_j$$

das Stichprobenmittel der j-ten Stichprobe und das Gesamtstichprobenmittel aller n Beobachtungen sind. Diese **Streuungszerlegung** zeigt auf, ob sich die Varianz vor allem durch die Variation innerhalb der Gruppen oder durch Unterschiede zwischen den Gruppen ergibt. Definieren wir nun

$$Q_1 = \sum_{j=1}^r n_j (\bar{X}_j - \bar{X}_{Ges})^2 \quad (III3: 61)$$

¹⁴ Weitere Verfahren der Varianzanalyse sind z.B. in Hochstädter, D., Kaiser, U. (1988), Mickey, R.M., Dunn, O.J., Clark, V.A. (2004) oder Zöfel, P. (1992) zu finden.

¹⁵ Im Unterschied zu Varianzheterogenität (Heteroskedastizität).

sowie

$$Q_2 = \sum_{j=1}^r \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2 \quad (\text{III3: 62})$$

und verwenden diese Größen in der Teststatistik (III3: 63), genügt diese, sofern die Nullhypothese $\mu_1 = \mu_2 = \dots = \mu_r$ richtig ist, einer F-Verteilung mit $(r - 1)$ und $(n - r)$ Freiheitsgraden. Dies beruht darauf, dass sich zeigen lässt, dass Q_1 bzw. Q_2 nach Abschnitt III 3.3.1 mit $(r - 1)$ bzw. $(n - r)$ Freiheitsgraden Chi-Quadrat-verteilt sind und somit nach Abschnitt III 3.3.3 die Testgröße V F-verteilt ist.¹⁶

$$V = \frac{\frac{1}{r-1} Q_1}{\frac{1}{n-r} Q_2} \sim F(r-1, n-r) \quad (\text{III3: 63})$$

Da die \bar{X}_j erwartungstreue Schätzer für μ_j sind, ist bei korrekter Nullhypothese zu erwarten, dass in einer konkreten Stichprobe die Ausprägung q_1 von Q_1 und damit auch v gering ausfällt. Ein hoher Wert von q_1 führt umgekehrt zu einer hohen Ausprägung v und deutet damit darauf hin, dass die Gegenhypothese H_A zutrifft.

Die Berechnung von q_1 und q_2 in einer konkreten Stichprobe kann durch Anwendung des Verschiebungssatzes erheblich vereinfacht werden. Es gilt nämlich

$$q_1 = \sum_{j=1}^r n_j \bar{x}_j^2 - n \bar{x}_{Ges}^2$$

und

$$q_2 = \sum_{j=1}^r \sum_{i=1}^{n_j} x_{ji}^2 - \sum_{j=1}^r n_j \bar{x}_j^2.$$

(3) Aufstellung der Entscheidungsregel für die Ablehnung von H_0

Ablehnung oder Nichtablehnung wird bei diesem Testverfahren mittels des $(1 - \alpha)$ -Fraktils der $F(r - 1, n - r)$ -Verteilung entschieden. Es gilt

$$\begin{aligned} H_0 &\text{ wird abgelehnt gegenüber} \\ H_A: \text{Mindestens zwei der } \mu_j &\text{ sind verschieden, falls } v > f_{1-\alpha, r-1, n-r}. \end{aligned} \quad (\text{III3: 64})$$

Im Fall $r = 2$ ergibt sich bei der Varianzanalyse die gleiche Testentscheidung wie beim Zweistichproben-t-Test mit zweiseitiger Gegenhypothese, auf dessen Behandlung wir im Abschnitt III 3.3.2.1 verzichtet hatten.

Die Varianzanalyse führt auch dann noch zu brauchbaren Ergebnissen, wenn die Modellannahme einer normalverteilten Grundgesamtheit nicht exakt erfüllt ist. Beim F-Test handelt es sich nämlich um einen sog. **robusten Test**, der auf Abweichungen von der Normalverteilung nur schwach reagiert. Auch bei nur näherungsweise normalverteilter Grundgesamtheit ist also die Varianzanalyse anwendbar. Auf die Voraussetzung der Homoskedastizität kann dagegen nicht verzichtet werden.

¹⁶ Bei der Berechnung von Q_1 (Q_2) wird (werden) ein (r) Mittelwert(e) geschätzt. Deswegen reduziert sich die Anzahl der Freiheitsgrade auf $r - 1$ und $n - r$.

Beispiel III 3-19**Einfache Varianzanalyse**

In einem Unternehmen wurden jeweils 5 Bewerber der Hochschulen I, II, III und IV einem Eignungstest unterzogen, bei dem folgende Punktzahlen erreicht wurden:

Bewerber Hochschule	1	2	3	4	5
I	4	5	2	1	2
II	4	7	5	6	2
III	16	8	7	7	7
IV	8	9	8	7	11

Es soll anhand dieser Daten getestet werden, ob die Absolventen der einzelnen Hochschulen gleich qualifiziert sind ($\alpha = 0,01$). Als Qualifikationsmaß soll die mittlere Punktzahl dienen. Wir nehmen dabei an, dass die vorliegenden Ergebnisse Realisierungen normalverteilter unabhängiger Zufallsvariablen mit stets derselben Varianz sind.

(1) Hypothesen und Signifikanzniveau:

H_0 : Alle vier Hochschulen besitzen die gleiche mittlere Punktzahl. $\alpha = 0,01$

$$(\mu_1 = \mu_2 = \mu_3 = \mu_4)$$

H_A : Nicht alle Hochschulen besitzen die gleiche mittlere Punktzahl.

(Wenigstens zwei der μ_i sind ungleich.)

Alternativ:

H_0 : Sämtliche Studenten der drei Hochschulen bilden eine *einige Grundgesamtheit*, in der die Punktzahl $N(\mu, \sigma^2)$ -verteilt ist.

H_A : Es liegen *verschiedene Grundgesamtheiten* vor, in denen die Punktzahl $N(\mu_i, \sigma^2)$ -verteilt ist, wobei nicht alle μ_i gleich sind.

(2) Teststatistik und deren Verteilung:

Zur Berechnung der notwendigen Mittelwerte empfiehlt sich zunächst die Verwendung der folgenden Tabelle:

Bewerber (i) Hochschule (j)	1	2	3	4	5	Stichproben- summe	Stich- probenmittel
I	4	5	2	1	2	14	2,8 *
II	4	7	5	6	2	24	4,8
III	16	8	7	7	7	45	9,0
IV	8	9	8	7	11	43	8,6
Stichproben- gesamtsumme						126	
Stichproben- gesamtmittel							6,3 **

$$* 14 / 5 = 2,8 \quad ** 126 / (4 \cdot 5) = (2,8 + 4,8 + 9,0 + 8,6) / 4 = 6,3$$

Aus diesen Ergebnissen können wir folgende Werte bestimmen:

$$q_1 = \sum_{j=1}^4 n_j \bar{x}_j^2 - n \bar{x}_{\text{Ges}}^2 = (5 \cdot 2,8^2 + 5 \cdot 4,8^2 + 5 \cdot 9^2 + 5 \cdot 8,6^2) - (4 \cdot 5) \cdot 6,3^2 \\ = 929,2 - 793,8 = 135,4$$

$$q_2 = \sum_{j=1}^4 \sum_{l=1}^5 x_{jl}^2 - \sum_{j=1}^4 n_j \bar{x}_j^2 = (4^2 + 5^2 + 2^2 + \dots + 11^2) - 929,4 = 1026 - 929,2 = 96,8$$

Es ergibt sich so für die konkrete Ausprägung der Testgröße

$$v = \frac{(n-r)q_1}{(r-1)q_2} = \frac{(4 \cdot 5 - 4) \cdot 135,4}{(4 - 1) \cdot 96,8} = \frac{2.166,4}{290,4} = 7,46,$$

wobei V mit $r - 1 = 3$ und $n - r = 16$ Freiheitsgraden F-verteilt ist.

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > f_{1-\alpha, r-1, n-r}$ gilt.

Dabei ist $f_{1-\alpha, r-1, n-r} = f_{0,99; 3; 16} = 5,29$.

(4) Entscheidung:

Es gilt hier ($v = 7,46 > f_{1-\alpha, r-1, n-r} = 5,29$)

H_0 kann abgelehnt werden. Es kann also nicht angenommen werden, dass die Studenten der vier Hochschulen die gleiche Qualifikation aufweisen.

3.6 Überblick über die behandelten Testverfahren

Die folgende Tabelle bietet einen Überblick über die wichtigsten in diesem Kapitel besprochenen Testverfahren. Der Begriff Grundgesamtheit ist darin mit GG abgekürzt. Es wird generell davon ausgegangen, dass sehr große Grundgesamtheiten vorliegen, sodass Auswahlssätze von kleiner als 10 % resultieren.

Nullhypothese	Teststatistik V	Verteilung von V
$\mu = \mu_0$ (σ bekannt)	$\frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$	Standardnormalverteilung, wenn GG normalverteilt oder GG beliebig verteilt mit $n > 30$
$\mu = \mu_0$ (σ unbekannt)	$\frac{\bar{X} - \mu_0}{\hat{\sigma}} \sqrt{n}$	t-Verteilung mit $n - 1$ Freiheitsgraden, wenn GG normalverteilt Standardnormalverteilung, wenn GG beliebig verteilt und $n > 30$
$p = p_0$	$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	Standardnormalverteilung, wenn $np_0(1-p_0) \geq 9$
$\sigma^2 = \sigma_0^2$	$\frac{(n-1)\hat{\sigma}^2}{\sigma_0^2}$	Chi-Quadrat-Verteilung mit $n - 1$ Freiheitsgraden, wenn GG normalverteilt

$\mu_1 = \mu_2$ (σ_1, σ_2 bekannt und $\sigma_1 \neq \sigma_2$)	$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Standardnormalverteilung, wenn GG normalverteilt oder GG beliebig verteilt und $n_1 > 30$ und $n_2 > 30$
$\mu_1 = \mu_2$ (σ_1, σ_2 unbekannt und $\sigma_1 \neq \sigma_2$)	$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}}$	Standardnormalverteilung, wenn GG normal- oder beliebig verteilt und jeweils $n_1 > 30$ und $n_2 > 30$
$p_1 = p_2$	$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}}$ mit $\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$	Standardnormalverteilung, wenn $n_1 \hat{p}_1 (1 - \hat{p}_1) \geq 9$ und $n_2 \hat{p}_2 (1 - \hat{p}_2) \geq 9$
$\sigma_1^2 = \sigma_2^2$	$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$	F-Verteilung mit $n_1 - 1$ und $n_2 - 1$ Freiheitsgraden, wenn GG normalverteilt
$\delta = 0$	$\frac{\bar{D}}{\hat{\sigma}} \sqrt{n}$	t-Verteilung, mit $n - 1$ Freiheitsgraden, wenn GG normalverteilt und $n \leq 30$ Standardnormalverteilung, wenn $n > 30$
$\rho = 0$ (Bravais-Pearson)	$\frac{R \cdot \sqrt{n-2}}{\sqrt{1-R^2}}$	t-Verteilung mit $n - 2$ Freiheitsgraden, wenn GG normalverteilt
$\rho = 0$ (Spearman)	$R \sqrt{n-1}$	Standardnormalverteilung, wenn $n > 30$ und Merkmale mindestens ordinal skaliert
Stichprobe stammt aus einer GG mit bestimmter Verteilung.	$\sum_{j=1}^k \frac{(h_j - h_j^e)^2}{h_j^e}$	Chi-Quadrat-Verteilung mit $k - g - 1$ Freiheitsgraden, wenn $h_j^e \geq 5$ für alle j
Zwei Merkmale X und Y sind unabhängig voneinander.	$\sum_{i=1}^k \sum_{j=1}^m \frac{(h_{ij} - h_{ij}^e)^2}{h_{ij}^e}$	Chi-Quadrat-Verteilung mit $(k - 1)(m - 1)$ Freiheitsgraden, wenn $h_{ij}^e \geq 5$ für alle i und j Bei bekannten Randwahrscheinlichkeiten sind $km - 1$ Freiheitsgrade zu verwenden.
$\mu_1 = \mu_2 = \dots = \mu_r$	$\frac{(n - r)Q_1}{(r - 1)Q_2}$	F-Verteilung mit $r - 1$ und $n - r$ Freiheitsgraden, wenn GG normalverteilt und Homoskedastizität

4.

Aufgaben

Punktschätzung

Aufgabe PS1

Beweisen Sie rechnerisch den im Fall des Ziehens mit Zurücklegen geltenden Zusammenhang $E(\hat{\sigma}^2) = \sigma^2$!

Aufgabe PS2

Im Rahmen des Qualitätsmanagements wurde zwischen einem Unternehmen und einem Lieferanten die Vereinbarung getroffen, dass eine Lieferung (100.000 Stück) dann abgelehnt wird, wenn bei einer Stichprobe (ohne Zurücklegen) von 500 Stück mehr als 3 % der Artikel nicht die vorher vereinbarte Qualität aufweisen.

Wie groß ist die Wahrscheinlichkeit, dass eine bestimmte Lieferung abgelehnt wird, wenn sie einen Anteil von 5 % mangelhafter Artikel enthält?

Aufgabe PS3

Ein Wirtschaftsforschungsinstitut stellt bei Erhebungen zum Geschäftsklima fest, dass von 3.000 Unternehmen in Süddeutschland 1.200 einen Konjunkturaufschwung erwarten. Die Wahrscheinlichkeit dafür, dass in einer Stichprobe (ohne Zurücklegen) von 200 aus den 3.000 befragten Unternehmen der Anteil der optimistischen Unternehmer höchstens b beträgt, liege bei 80 %.

Bestimmen Sie b!

Aufgabe PS4

In einem Lager für Feingold gingen in einem Monat 350 Lieferungen ein. Nach den vorliegenden Materialeingangsscheinen ergibt sich ein durchschnittlicher Lieferwert von 1.550 Euro bei einer Standardabweichung von 150 Euro. Aus diesen Materialeingangsscheinen wird eine Stichprobe im Umfang von 50 (ohne Zurücklegen) entnommen.

Wie groß ist die Wahrscheinlichkeit, dass der durchschnittliche Lieferwert der Stichprobe zwischen 1.500 Euro und 1.600 Euro liegt?

Aufgabe PS5

In einer Schachtel befinden sich drei Diamanten mit den Gewichten 2, 4 und 6 Gramm. Es werden nun zwei Diamanten zufällig (mit Zurücklegen) gezogen.

- Bestimmen Sie Erwartungswert und Varianz der Grundgesamtheit und des Stichprobenmittelwertes!
- Geben Sie die Wahrscheinlichkeitsverteilung des Stichprobenmittels an!

Aufgabe PS6

Zwei Maschinen füllen unabhängig voneinander Konserven ab, wobei für das durchschnittliche Füllgewicht der Dosen von Maschine A und B gilt:

Maschine i	μ_i	σ_i
1	850 g	4 g
2	850 g	6 g

Es werden nun Stichproben im Umfang von $n_1 = 100$ und $n_2 = 120$ entnommen. Bestimmen Sie die Wahrscheinlichkeit dafür, dass sich die durchschnittlichen Füllgewichte in den Stichproben um höchstens 2 g unterscheiden!

Aufgabe PS7

Angenommen der Anteil der Koalitionswähler unter den Wahlberechtigten eines großen Wahlkreises betrage 52 %.

- Wie groß ist die Wahrscheinlichkeit, dass bei einer Zufallsstichprobe im Umfang von 500 Personen der Anteil der Koalitionswähler kleiner als 50 % ist?
- Wie groß müsste der Stichprobenumfang sein, wenn $P(\hat{p} < 0,5) = 0,0367$ gelten soll?

Aufgabe PS8

Eine Partei möchte in zwei Wahlkreisen A und B mit $N_A = 20.000$ und $N_B = 10.000$ Wahlberechtigten die Anteile p_A und p_B der ihr positiv gesinnten Wahlberechtigten und den unbekannten Gesamtanteil p schätzen. Dazu werden einfache Stichproben im Umfang von n_A und n_B gezogen, in der die Wahlberechtigten nach ihrer Einstellung zur Partei befragt werden (Befürwortung, keine Befürwortung). Ein Parteidirektor schlägt folgende Schätzfunktionen vor, in denen \hat{p}_A bzw. \hat{p}_B den zufallsabhängigen Anteil der Befürworter aus den jeweiligen Stichproben bezeichnen:

$$(1) \quad \hat{p} = \frac{1}{n_A + n_B} (n_A \hat{p}_A + n_B \hat{p}_B)$$

$$(2) \quad \hat{p} = \frac{1}{N_A + N_B} (N_A \hat{p}_A + N_B \hat{p}_B)$$

$$(3) \quad \hat{p} = \frac{1}{2} (\hat{p}_A + \hat{p}_B)$$

Beurteilen Sie die Güte dieser Schätzfunktionen hinsichtlich ihrer Erwartungstreue und bestimmen Sie für zwei Stichproben des Umfangs $n_A = 80$ (30 Befürworter) und $n_B = 20$ (10 Befürworter) mittels der erwartungstreuen Schätzfunktion die Schätzwerte für p_A , p_B und p !

Aufgabe PS9

Zeigen Sie, dass \bar{X} und S^2 die Maximum-Likelihood-Schätzer für die Parameter μ und σ^2 einer Normalverteilung sind!

Intervallschätzung

Aufgabe IS1

In einer Großbäckerei sticht eine Maschine aus einer durchlaufenden Teigmasse runde Plätzchen aus. Aufgrund langer Erfahrung sieht man die Durchmesser der dabei entstehenden Plätzchen als normalverteilt mit einem bestimmten Erwartungswert von μ mm und einer Standardabweichung von $\sigma = 0,5$ mm an. Aus der laufenden Produktion werden nun 10 Plätzchen zufällig entnommen und ihr Durchmesser festgehalten. Dabei ergeben sich folgende mm-Werte:

20,1	20,4	20,3	20,7	20,1
21,0	20,3	20,2	20,1	19,9

- Führen Sie eine Intervallschätzung für den Erwartungswert μ zu einer Irrtumswahrscheinlichkeit von 5 % durch!
- Nehmen Sie unter sonst unveränderten Bedingungen an, die Standardabweichung σ der Grundgesamtheit sei unbekannt! Wie verändert sich dadurch die Intervallschätzung aus a)?

Aufgabe IS2

Das Durchschnittsgewicht von 2.000 Konservendosen soll bei einem Konfidenzniveau von 95 % und einem absoluten Fehler von 1 g geschätzt werden.

Wie groß muss eine ohne Zurücklegen gezogene Stichprobe dazu mindestens sein, wenn die Varianz der Grundgesamtheit mit $\sigma^2 = 35 \text{ g}^2$ bekannt ist?

Aufgabe IS3

In einer Studie zum Spritverbrauch von Pkws wurden 60 Pkws eines bestimmten Modells mit 50 Liter Benzin ausgestattet und anschließend auf einer Rennstrecke bei 100 km/h gefahren. Mit dieser Füllung legten die Fahrzeuge im Durchschnitt $\bar{x} = 550$ km bei einer Standardabweichung von $\hat{\sigma} = 25$ km zurück.

- Berechnen Sie das Ergebnis einer Intervallschätzung (Vertrauenswahrscheinlichkeit 95 %) für die durchschnittliche Kilometerleistung μ und interpretieren Sie das geschätzte Intervall!
- Ist es möglich, einen exakten Mindeststichprobenumfang n anzugeben, der garantiert, dass bei gegebener Vertrauenswahrscheinlichkeit die Intervallschätzung für die durchschnittliche Kilometerleistung μ eine bestimmte Länge aufweist?

Aufgabe IS4

Ein Wirtschaftsforschungsinstitut fragt in einer Region 700 von 5.000 Unternehmen nach ihrem Monatsumsatz. Es ergab sich dabei ein durchschnittlicher Monatumsatz von $\bar{x} = 25.550 \text{ €}$ bei einer Standardabweichung von $\hat{\sigma} = 3.120 \text{ €}$.

- Bestimmen Sie ein Schätzintervall für den Mittelwert μ der Grundgesamtheit mit einer Vertrauenswahrscheinlichkeit von 99 %!
- In welchen Grenzen wird der Umsatz aller Unternehmen der Region mit 99 % Wahrscheinlichkeit liegen?

Aufgabe IS5

Von 50.000 Gästen der internationalen Autoausstellung in Frankfurt am Main wurden 200 zufällig hinsichtlich des PKW befragt, den sie halten. Unter den befragten Personen befanden sich 50, die die Marke XY besaßen.

Berechnen Sie das 95,45 % - Schätzintervall für den Anteil p der Personen der Grundgesamtheit, die die Marke XY besitzen!

Aufgabe IS6

In einem Versandhaus wurde für 40 Tage die Anzahl X der Bestellungen erfasst, die für einen bestimmten Artikel pro Tag eingehen. Es ergab sich

$$\begin{aligned} & 8; 3; 10; 9; 13; 15; 17; 8; 9; 10; 12; 15; 5; 7; 8; 20; 10; 2; 5; 18; \\ & 7; 5; 17; 1; 12; 14; 11; 1; 2; 19; 13; 10; 9; 6; 6; 21; 11; 2; 5; 22. \end{aligned}$$

Führen Sie eine Intervallschätzung zum Konfidenzniveau von 95 % für die unbekannte Wahrscheinlichkeit p durch, dass mindestens 10 Bestellungen pro Tag für den Artikel eingehen!

Aufgabe IS7

Eine Blitzumfrage unter 1.000 zufällig ausgewählten wahlberechtigten Bundesbürgern soll Aufschluss über den derzeitigen Wähleranteil einer Partei A geben. Die Befragung ergab, dass im Moment nur 80 Personen für diese Partei stimmen würden. Der Rest würde sich für andere Parteien entscheiden.

- Bestimmen Sie ein Schätzintervall zum Konfidenzniveau von 99 % für den Anteil p derjenigen Wahlberechtigten, auf die sich die Partei verlassen kann und interpretieren Sie dieses!
- Aufgrund des Befragungsergebnisses wird als sicher angesehen, dass sich der Stichprobenanteilswert \hat{p} bei Erhöhung des Stichprobenumfangs n höchstens noch auf den Wert 0,1 erhöht. Welcher Stichprobenumfang garantiert eine Länge des Konfidenzintervalls von 1 %?

Aufgabe IS8

Eine Untersuchung zu den Kontoführungsgebühren bei Kreditinstituten in der Bundesrepublik ergab für die jährlichen Kosten eines "Modellkontos" bei 20 Banken die folgenden Euro-Werte:

10	46	86	33	51	99	47	90	91	59
72	83	22	15	54	52	65	58	46	75

Es wird angenommen, dass die Grundgesamtheit normalverteilt ist.

- Schätzen Sie die Varianz σ^2 durch ein Konfidenzintervall zur Vertrauenswahrscheinlichkeit von 90 %!
- Bestimmen Sie für $\alpha = 0,05$ ein Schätzintervall für die mittleren Kosten μ !

Testen von Hypothesen

Aufgabe T1

Für eine bevorstehende Bundestagswahl wurde von einem Meinungsforschungsinstitut für eine Partei eine Umfrage unter 1.500 Wählern durchgeführt. Diese ergab, dass von diesen Wählern derzeit genau 105 für diese Partei stimmen würden.

- Testen Sie die Hypothese H_0 , dass der Stimmenanteil p dieser Partei unter den Wahlberechtigten höchstens 5 % beträgt! Das zu verwendende Signifikanzniveau sei $\alpha = 0,02$.
- Ab welchem Wähleranteil in der Stichprobe kann die unter a) verwendete Gegenhypothese als bestätigt angesehen werden?

Aufgabe T2

Zehn Tragseile eines bestimmten Herstellers wurden der Produktion zufällig entnommen und einem Belastungstest unterzogen. Die Traglast wurde dabei so lange erhöht, bis das Seil riss. Der Riss geschah bei folgenden kg-Gewichten x_i :

2.000, 2.030, 2.010, 2.040, 2.035, 1.995, 2.025, 2.005, 1.975, 2.050

- Testen Sie unter der Annahme, dass die Seile hinsichtlich des Reißgewichts X einer Normalverteilung genügen, die Hypothese H_0 , dass das mittlere Reißgewicht μ mindestens 2.000 kg beträgt! Gehen Sie dabei von einem Signifikanzniveau von $\alpha = 0,01$ aus! Interpretieren Sie Ihr Ergebnis, wenn der Test von einem Konkurrenzunternehmen dazu verwendet wird, zu beweisen, dass die Seile des Herstellers bereits bei Gewichten von unter 2.000 kg reißen!
- Überprüfen Sie ebenfalls bei $\alpha = 0,01$ und der Annahme einer normalverteilten Grundgesamtheit die Hypothese $H_0: \sigma^2 = 440$ gegen die Hypothese $H_A: \sigma^2 \neq 440$! σ^2 ist dabei die Varianz des Reißgewichts in der Grundgesamtheit aller produzierten Seile.

Aufgabe T3

Durch den Bau einer Umgehungsstraße soll das Verkehrsaufkommen an einer vielbefahrenen Kreuzung reduziert werden. Vor Bau der Umgehungsstraße (Situation 1) wurden bei 100 zufälligen Beobachtungen im Durchschnitt 25 Fahrzeuge/Minute bei einer Varianz von 7 an der Kreuzung registriert. Nach Bau der Umgehungsstraße (Situation 2) wird eine erneute Stichprobe gezogen, die Folgendes ergibt:

Anzahl passierender Fahrzeuge:	5	10	15	20	21	25	28	30	34
--------------------------------	---	----	----	----	----	----	----	----	----

Häufigkeit in der Stichprobe:	2	2	3	7	10	7	6	2	1
-------------------------------	---	---	---	---	----	---	---	---	---

Testen Sie $H_0: \mu_1 \leq \mu_2$ gegen $H_A: \mu_1 > \mu_2$ bei einem Signifikanzniveau von $\alpha = 0,01$ und interpretieren Sie Ihr Ergebnis im Sinne der geschilderten Situation!

Aufgabe T4

Zwei Gruppen von Studenten ($n_1 = 16$, $n_2 = 10$) absolvieren einen Test, wobei sich in Gruppe 1 eine mittlere Punktzahl von $\bar{x}_1 = 40$ bei einer Standardabweichung von $\hat{\sigma}_1 = 4,5$ Punkten ergibt. In Gruppe 2 stellt man $\bar{x}_2 = 43,5$ bei $\hat{\sigma}_2 = 2,2$ fest.

Prüfen Sie auf einem Signifikanzniveau von $\alpha = 0,02$, ob der beobachtete Unterschied der beiden Stichprobenstandardabweichungen signifikant ist! Welche Annahme bzgl. der Verteilung der Grundgesamtheiten müssen Sie treffen, damit der Test, den Sie für die Prüfung verwenden, angewendet werden kann?

Aufgabe T5

Kürzlich wurde in einer Fernsehshow untersucht, ob die Betrachtung einer erotischen Darbietung den Puls im gleichen Ausmaß ansteigen lässt, wie sportliche Betätigung. Dazu wurden 50 Männer für 10 Minuten auf ein Laufband geschickt und dauernd ihr Puls festgehalten. Danach durften dieselben Männer eine erotische Tanzeinlage betrachten. Auch hier wurde 10 Minuten lang der Puls registriert. Diese Stichprobe ergab, dass der Puls durch die sportliche Betätigung stärker angestiegen war. Der Unterschied lag im Durchschnitt bei 20 Schlägen pro Minute bei einer Standardabweichung von 8 Schlägen pro Minute.

Testen Sie bei einem Signifikanzniveau von $\alpha = 0,01$ die Nullhypothese, dass sich in der Grundgesamtheit die Mittelwerte der Pulsfrequenzen in beiden Situationen nicht unterscheiden! Erklären Sie, warum sich hier durch verbundene Stichproben im Vergleich zu unabhängigen Stichproben ein Effizienzgewinn erzielen lässt!

Aufgabe T6

In der Filiale einer Bank wurden zufällig 40 Kunden ausgewählt und gebeten hinsichtlich ihrer allgemeinen Zufriedenheit mit der Filiale und der Freundlichkeit der Angestellten Noten von 1 bis 4 zu vergeben. Es ergab sich aus den Befragungsergebnissen ein Rangkorrelationskoeffizient von 0,46.

Testen Sie auf einem Signifikanzniveau von $\alpha = 0,05$, ob diese Korrelation signifikant von Null verschieden ist!

Aufgabe T7

Eine Stichprobe im Umfang von $n = 100$ ergibt für zwei normalverteilte Zufallsvariablen X und Y einen Korrelationskoeffizienten von $r = 0,08$.

- Kann hieraus mit einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ die Behauptung $\rho > 0$ statistisch abgesichert werden?
- Wie groß müsste der Stichprobenumfang n mindestens sein, damit der Korrelationskoeffizient $r = 0,08$ mit $\alpha = 0,05$ signifikant größer als Null ist?

Aufgabe T8

Eine Münze, die Kopf oder Zahl zeigen kann, wird 200-mal geworfen. Es fällt dabei 80-mal Kopf und 120-mal Zahl.

Kann durch diese Stichprobe nachgewiesen werden, dass es sich um eine manipulierte Münze handelt (Signifikanzniveau $\alpha = 0,01$)? Geben Sie außerdem das Testverfahren an, das neben dem von Ihnen verwendeten ebenfalls zur Beantwortung dieser Fragestellung herangezogen werden könnte!

Aufgabe T9

Bei Verwendung eines bestimmten Werkstoffes ergaben sich in der Vergangenheit pro Tag bei einer bestimmten Fertigungsmaschine Störfälle, deren Anzahl als Zufallsvariable X' betrachtet werden kann. Ihre Wahrscheinlichkeitsverteilung konnte aufgrund langer Erfahrung wie folgt bestimmt werden:

j	0	1	2	3	4
$p_j = P(X' = j)$	0,15	0,20	0,30	0,20	0,15

Nach Wechsel des Werkstofflieferanten soll auf einem Signifikanzniveau von $\alpha = 0,05$ getestet werden, ob die Anzahl X der nun auftretenden Störfälle noch immer wie X' verteilt ist. Es wurden dazu in einem Beobachtungszeitraum von 100 Tagen folgende Häufigkeiten h_j festgestellt, mit denen j Störfälle auftraten:

j	0	1	2	3	4
h_j	11	28	33	18	10

Aufgabe T10

Eine exakte und eine kostengünstige, ungenauere Messmethode sollen miteinander verglichen werden. Es werden dazu 80 Messungen vorgenommen und jeweils die Unterschiede der Ergebnisse beider Methoden X (in mg) aufgezeichnet. Es wurde dabei eine Einteilung in Fehlerklassen vorgenommen, die folgende Häufigkeiten aufweisen:

A_j	$]-\infty; -1]$	$] -1; -0,5]$	$] -0,5; -0,1]$	$] -0,1; 0,1]$	$] 0,1; 0,5]$	$] 0,5; 1]$	$[1; \infty[$
h_j	8	10	17	12	16	8	9

Überprüfen Sie die Hypothese, dass der auftretende Messfehler einer Standardnormalverteilung genügt (Signifikanzniveau $\alpha = 0,05$)!

Aufgabe T11

An einer Mathematikklausur nahmen 200 männliche und 100 weibliche Studierende teil. Es wurden dabei folgende Ergebnisse erzielt:

Note	Männer	Frauen	Σ
1	15	5	20
2	35	15	50
3	50	20	70
4	60	10	70
5	40	50	90
Σ	200	100	300

Entsprechen die Leistungen der Frauen denjenigen der Männer ($\alpha = 0,05$)?



Aufgabe T12

Prüfen Sie bei einem Signifikanzniveau von $\alpha = 0,01$ die Behauptung, dass die folgenden zufällig erzeugten Zahlen gleichverteilt sind!

2	5	9	3	8	2
1	3	1	8	6	4
10	3	3	5	1	4
2	6	4	8	5	3
7	9	1	1	10	3
2	7	8	1	6	5
8	8	10	2	3	10
5	2	1	6	1	3
2	10	2	4	5	8
5	7	8	10	8	10

Aufgabe T13

Bei einem Unternehmen bewerben sich auf offene Stellen 60 Jugendliche mit und 40 Jugendliche ohne Ausbildung. 70 Bewerber werden nicht als geeignet empfunden, davon haben 40 eine Ausbildung.

Lässt sich mit diesem Ergebnis bei einem Signifikanzniveau von $\alpha = 0,05$ statistisch nachweisen, dass Berufsausbildung und Prüfungsergebnis voneinander abhängig sind?

Aufgabe T14

Einer einfachen Varianzanalyse lag eine Untersuchung zugrunde, die bei jeweils 5 Schülern aus 6 Klassen zur Feststellung des technischen Verständnisses die Zeit (in Minuten) erfassste, die zum Zusammenbau einer technischen Apparatur benötigt wurde. Das Testverfahren ergab, dass die Schüler der 6 Klassen nicht als gleich geschickt angesehen werden können.

Ändert sich das Testergebnis, wenn die Zeit nicht in Minuten, sondern in Stunden gemessen wird? Begründen Sie Ihre Antwort!

Aufgabe T15

Auf 5, 6 bzw. 7 gleich großen Feldern werden drei verschiedene Weizensorten W_1 , W_2 und W_3 angebaut, wobei sich folgende Ernteerträge (in Tonnen) ergaben:

Weizen W_1	12	13	10	9	11	
Weizen W_2	9	11	10	11	13	12
Weizen W_3	10	9	11	13	10	9

Testen Sie, ob die drei Sorten im Mittel den gleichen Ertrag abwerfen! Nehmen Sie dabei an, dass die Ergebnisse Realisierungen normalverteilter, unabhängiger Zufallsvariablen mit gleicher Varianz sind (Signifikanzniveau $\alpha = 0,05$)! Interpretieren Sie Ihr Testergebnis!

IV EINFÜHRUNG IN DIE ÖKONOMETRIE

Mit diesem letzten Kapitel wollen wir methodische Grundlagen legen, die vor allem für den empirisch arbeitenden Wirtschaftswissenschaftler und für Studenten von Interesse sind, die sich in ihren Diplom- bzw. Bachelor- oder Masterarbeiten mit empirischen Fragestellungen auseinandersetzen wollen.

Die Ökonometrie (engl. *econometrics* = *economic measurement*) bzw. die Regressionsanalyse als eines ihrer bedeutendsten Teilgebiete befasst sich mit der Untersuchung quantitativer Zusammenhänge zwischen Variablen. Sie liefert uns ein Instrumentarium, mit dem wir die Veränderungen einer Variablen durch Veränderung einer oder mehrerer anderer Variablen erklären können. Durch eine derartige Möglichkeit der Erklärung der Entwicklung einer Variablen ermöglicht uns die Ökonometrie auch die Prognose zukünftiger Veränderungen. Dies hat vor allem im Zusammenhang mit wirtschafts- und unternehmenspolitischen Entscheidungen große Bedeutung.

Die empirische Wirtschaftsforschung liefert zahlreiche Beispiele für die Schätzung ökonomischer Zusammenhänge mit Hilfe von Daten. So möchten Regierungen beispielsweise die kurz- und langfristigen Effekte von Steuerreformen auf wichtige Variablen wie Beschäftigung, Inflation und Bruttoinlandsprodukt abschätzen. Sie interessieren sich außerdem dafür, inwieweit eine Erhöhung der Tabaksteuer den Zigarettenkonsum reduziert. Dafür wird eine Nachfragefunktion für den Tabakkonsum geschätzt.

1.

Grundlagen

In diesem Abschnitt beschäftigen wir uns damit, was allgemein unter dem Begriff Regressionsanalyse zu verstehen ist, stellen ihre Ziele vor und den Bezug zu verwandten Termini wie Korrelation und Kausalität her. Wir bieten außerdem eine Einführung in eines der wichtigsten ökonometrischen Schätzverfahren, das Prinzip der kleinsten Quadrate. Dieses ermöglicht es uns, funktionale Zusammenhänge zwischen Variablen bei gegebenem Datenmaterial zu charakterisieren und ist daher im Rahmen dieses Abschnitts IV 1 als rein deskriptives Instrument zu verstehen. Seine induktiven Qualitäten werden wir im Rahmen der Abschnitte IV 2 und IV 3 behandeln.

1.1 Was ist Regressionsanalyse?

1.1.1 Ziele der Regressionsanalyse

Unter *Regressionsanalyse* verstehen wir allgemein die quantitative Analyse von Zusammenhängen zwischen Variablen. Sie versucht durch Quantifizierung theoretischer Modelle die Brücke zwischen abstrakter Modelltheorie und realer Welt zu schlagen. Da bei solchen Analysen unterschiedliche *Ziele* verfolgt werden können, ist eine Betrachtung dieser Ziele ein erster wichtiger Schritt zu ihrem Verständnis.

Die Regressionsanalyse dient erstens dazu, die *Realität zu beschreiben*. Eine Quantifizierung wirtschaftswissenschaftlicher Zusammenhänge ist möglich, da die Regressionsanalyse erlaubt, konkrete Beziehungen zu schätzen, für die ursprünglich nur abstrakte Parameter in Modellgleichungen eingesetzt wurden.

Beispiel IV 1-1	Nachfragefunktion
Die von Konsumenten nachgefragte Menge X wird häufig als lineare Funktion der Preise des betrachteten Gutes (P), des Substitutionsgutes (P^S) und des Komplementärgutes (P^K) sowie des verfügbaren Haushaltseinkommens (Y) abgebildet. Steigende Preise P und P^K verursachen bei normalen Gütern einen Rückgang, Zunahmen von P^S und Y eine Erhöhung der nachgefragten Menge.	$X = f(P, P^S, P^K, Y) = \beta_0 + \beta_1 P + \beta_2 P^S + \beta_3 P^K + \beta_4 Y$

$$X = 22,3 - 0,63P + 0,15P^S - 0,32P^K + 0,23Y$$

Die Regressionsanalyse erlaubt nun die obige allgemeine Gleichung unter Verwendung von Stichprobendaten konkret zu schätzen. So kann sich dabei z.B. ein Zusammenhang

$$X = 22,3 - 0,63P + 0,15P^S - 0,32P^K + 0,23Y$$

ergeben. Während wir mit der allgemeinen Nachfragefunktion nur aussagen konnten, dass bei einer Erhöhung des Preises des Substitutionsgutes (P^S) um eine Einheit die nachge-

fragte Menge um β_1 Einheiten steigt, können wir nun mit der geschätzten Nachfragefunktion argumentieren, dass eine Erhöhung des Preises P^S um eine Einheit (bei Konstanthaltung aller anderen Einflussfaktoren) konkret eine Erhöhung der nachgefragten Menge um 0,15 Einheiten auslöst. Wir bezeichnen den Wert 0,15 als geschätzten Regressionskoeffizienten oder -parameter.

Zweitens erweist sich die Regressionsanalyse beim *Testen von Hypothesen* über die Parameter der Grundgesamtheit als äußerst wertvoll. Bevor nämlich für ein Modell ausgesagt werden kann, dass es die Realität (Grundgesamtheit) gut beschreibt und es damit zur Prognose verwendet werden kann, gilt es die empirische Gültigkeit des Modells zu überprüfen.

Beispiel IV 1-2

Nachfragefunktion

Im obigen Beispiel IV 1-1 könnte z.B. die Hypothese getestet werden, ob es sich beim betrachteten Gut um ein normales Gut (Gut bei dem mit zunehmendem Einkommen Y auch die Nachfrage nach dem Gut steigt) handelt. Auf den ersten Blick würde der geschätzte Koeffizient 0,23 für das Haushaltseinkommen Y diese Hypothese unterstützen, da er positiv ist. Bevor eine solche Schlussfolgerung aber wirklich gerechtfertigt ist, muss mit einem entsprechenden Testverfahren geprüft werden, ob der geschätzte Koeffizient signifikant von Null verschieden ist. Erst dann kann von einem normalen Gut gesprochen werden.

Ist die empirische Gültigkeit des Modells durch entsprechende Tests untermauert, besteht das dritte Ziel der Regressionsanalyse in der *Abgabe von Prognosen*.¹ Aus historischem Datenmaterial geschätzte Modelle werden von Ökonomen nämlich häufig benutzt, um z.B. Prognosen für das künftige BIP-Wachstum oder die Inflationsrate abzugeben. Die Genauigkeit solcher Prognosen ist natürlich zum Großteil davon abhängig, ob die vergangene Entwicklung solcher Variablen wirklich Informationen über ihre zukünftige Entwicklung liefern kann.

Beispiel IV 1-3

Nachfragefunktion

Greifen wir noch einmal auf unsere geschätzte Nachfragefunktion aus den vorhergehenden Beispielen zurück und nehmen an, dass ein Unternehmer, der das betroffene Gut herstellt, eine Preiserhöhung beabsichtigt. Stünde ihm das geschätzte Regressionsmodell zur Verfügung und hätte sich dieses auch in einer Reihe von empirischen Tests bewährt, so könnte er damit die zu erwartenden Umsätze mit und ohne Preiserhöhung in den Folgeperioden prognostizieren. Dies stellt für ihn sicherlich eine wichtige Planungs- und Entscheidungsgrundlage für den Mitteleinsatz in zukünftigen Perioden dar. Ist z.B. $P = 1 \text{ €}$, $P^S = 2 \text{ €}$, $P^K = 1 \text{ €}$ und $Y = 1 \text{ Tsd. €}$, so ergibt sich eine prognostizierte Absatzmenge von $X = 21,88 \text{ Stück}$. Ist hingegen $P = 2 \text{ €}$, so resultieren $X = 21,25 \text{ Stück}$.

¹ Anders als die meisten Lehrbücher werden wir das Thema "Prognose" erst am Schluss behandeln (vgl. Abschnitt IV 6), da eine Prognose ohne genaue Untersuchung der Modellgüte und -spezifikation (vgl. insbesondere Abschnitt IV 4) oft fehlerbehaftet ist.

1.1.2 Grundgedanken und Abgrenzungen

Wie wir gesehen haben, bedienen sich Wirtschaftswissenschaftler der Regressionsanalyse, um quantitative Schätzungen der Zusammenhänge zwischen Variablen zu gewinnen, die vorher nur rein theoretisch formuliert wurden. Die **Regressionsanalyse** ist ein statistisches Verfahren, welches versucht die Veränderung einer sog. **erklärten**, abhängigen oder endogenen **Variablen** (Regressand) über die Veränderungen einer Reihe sog. **erklärender**, unabhängiger oder exogener **Variablen** (Regressoren) durch Quantifizierung *einer einzelnen Gleichung* zu erklären. Dieses Verfahren kommt sehr häufig zum Einsatz, da sich viele wirtschaftswissenschaftliche Zusammenhänge durch eine einzige Gleichung abbilden lassen (vgl. etwa Beispiel IV 1-1).

Eine Regression kann feststellen, ob eine quantitative Beziehung zwischen den erklärenden Variablen und der erklärten Variable besteht. Ein Regressionsergebnis kann aber, egal wie statistisch signifikant es auch sein mag, keine **Kausalität** beweisen. Es vermag nur Stärke und Richtung eines quantitativen Zusammenhangs und niemals Kausalität aufzuzeigen. Eine Kausalität einer erklärenden Variablen für die erklärte Variable kann man bereits vor Durchführung der Regressionsanalyse aus der wirtschaftswissenschaftlichen Theorie oder dem gesunden Menschenverstand heraus vermuten. Es sind nur solche erklärende Variablen in die Regressionsanalyse einzubeziehen, die auch wirklich für die erklärte Variable kausal sind. Das wohl bekannteste Beispiel lieferte in diesem Zusammenhang der englische Ökonom Stanley Jevons. Er untersuchte die Hypothese, dass Sonnenflecken (erklärende Variable) das Bruttoinlandsprodukt (erklärte Variable) beeinflussen. Dazu führte er eine Regressionsanalyse mit entsprechendem Datenmaterial durch und stellte dabei fest, dass ein signifikant positiver Zusammenhang vorliegt.² Da wir keine inhaltliche Begründung für eine kausale Beziehung zwischen den Variablen haben, könnten wir fälschlicherweise folgern, dass eine Veränderung der Sonnenfleckenzahl kausal für eine Veränderung des Bruttoinlandsproduktes ist.

Eng verwandt, jedoch konzeptionell von der Regressionsanalyse sehr verschieden, ist die **Korrelationsanalyse** (vgl. Abschnitt I 3.2). Ihr Primärziel ist die Messung der Stärke und Richtung eines *linearen* Zusammenhangs zwischen zwei Variablen (z.B. Mathematik- und Statistiknoten). Zudem werden die zwei betrachteten *Variablen symmetrisch* behandelt, d.h. es wird nicht zwischen erklärten und erklärenden Variablen unterschieden. Der Korrelationskoeffizient zwischen einer Variablen A und einer Variablen B ist demnach identisch mit dem zwischen B und A. In der Regressionsanalyse unterscheiden wir bewusst zwischen erklärter und erklärender Variable und es wird vielmehr versucht den Wert einer Variablen auf Basis gegebener Werte erklärender Variablen zu prognostizieren (z.B. die nachgefragte Menge in Abhängigkeit vom Einkommen).

² Vgl. Jevons, W. S. (1987). In einer vorgelagerten Studie untersuchte Jevons, W.S. (1985) außerdem den Einfluss von Sonnenflecken auf den Preis von Getreide. Beide Papiere bildeten den Ausgangspunkt der bekannten "sunspot theory of the business cycle".

1.2 Das Prinzip der kleinsten Quadrate

In Beispiel IV 1-1 haben wir gesehen, dass mittels Stichprobendatenmaterial in Modellen abstrakte Parameter durch konkrete Zahlen ersetzt werden können. Um derartige Schätzungen zu gewinnen, bedienen wir uns im Rahmen der Regressionsanalyse häufig dem *Prinzip der kleinsten Quadrate*, kurz *OLS* (engl. ordinary least squares). Dieses ist sowohl auf Modelle mit einer als auch auf Modelle mit mehreren erklärenden Variablen anwendbar und wird im Folgenden zunächst rein deskriptiv behandelt.

1.2.1 OLS bei Modellen mit einer erklärenden Variablen

Würde zwischen zwei Variablen Y und X *in einem vorliegenden Datensatz* bestehend aus n Wertepaaren (X_i, Y_i) ein exakter linearer Zusammenhang der Form

$$Y_i = \beta_0 + \beta_1 X_i \quad (\text{IV1: 1})$$

bestehen, so bedeutet dies (vgl. Abbildung IV 1), dass sämtliche Wertepaare (X_i, Y_i) auf einer Geraden liegen. Ihre Lage wird im Koordinatensystem durch die Parameter β_0 und β_1 eindeutig beschrieben. Die Konstante β_0 ist der Y -Achsen-Abschnitt der Geraden. Sie gibt den Wert an, den Y annimmt, wenn X gleich Null ist. β_1 ist die Geradensteigung und gibt Auskunft darüber, um wie viele Einheiten Y steigt, wenn X um eine Einheit steigt. Dies können wir uns mittels Abbildung IV 1 schnell klar machen. Hier führt ein Anstieg von X_1 auf X_2 (ΔX) zu einer Erhöhung des Wertes von Y von Y_1 auf Y_2 (ΔY). Es gilt

$$\frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\Delta Y}{\Delta X} = \beta_1, \quad (\text{IV1: 2})$$

sodass die Reaktion des Wertes Y auf Änderungen von X konstant und gleich dem Steigungskoeffizienten β_1 ist.

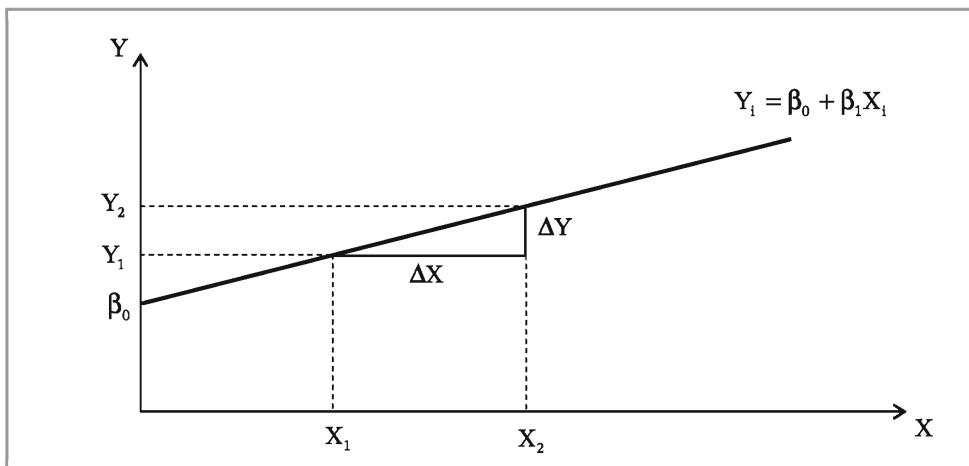


Abbildung IV 1: Linearer Zusammenhang zweier Variablen

Die Frage ist nun, wie wir in einer vorliegenden Stichprobe von n Wertepaaren (X_i, Y_i) die Parameter β_0 und β_1 bestimmen können. Da wir zunächst angenommen haben, dass alle Wertepaare exakt auf einer Geraden liegen, wäre es kein schweres Unterfangen, die Geradengleichung mittels zweier beliebiger Punkte aufzustellen. In der Praxis sind wir nun jedoch mit dem Problem konfrontiert, dass die Wertepaare in gegebenen Streudiagrammen häufig nur in etwa einen linearen Verlauf aufweisen (vgl. Abbildung IV 2). Es gilt dann ein Verfahren zu finden, mit dem eine Gerade $f(X_i) = \beta_0 + \beta_1 X_i$ optimal in dieses Streudiagramm eingepasst wird. Wir bezeichnen die Parameter β_0 und β_1 dieser optimalen Geraden als **Regressionsparameter** oder **-koeffizienten**. β_0 ist die Regressionskonstante, β_1 der Steigungsparameter. Y ist die erklärte, X die *einige* erklärende Variable. Wir sprechen daher bei der Schätzung von β_0 und β_1 auch von einer **einfachen Regression**:

Nehmen wir zunächst an, uns lägen zwei mögliche Schätzungen $\hat{\beta}_0$ und $\hat{\beta}_1$ vor. Wir wären damit in der Lage für unsere Stichprobe die Gerade in Abbildung IV 2 zu konstruieren. Es ist deutlich zu erkennen, dass unsere Stichprobenwertepaare (abgebildet durch die Punkte in der Abbildung) nicht auf der Geraden liegen, da zwischen ihnen kein exakt linearer Zusammenhang besteht. Vielmehr weichen die einzelnen Y_i von den geschätzten Werten \hat{Y}_i (sog. "gefittete Werte") ab, die sich aus den Schätzungen $\hat{\beta}_0$ und $\hat{\beta}_1$ sowie den X_i in der Form

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (\text{IV1: 3})$$

ergeben. Wir sprechen bei (IV1: 3) von der **geschätzten Regressionsgeraden oder -gleichung**. Die Abweichungen e_i der Y_i von dieser Geraden bzw. \hat{Y}_i ,

$$e_i = Y_i - \hat{Y}_i, \quad (\text{IV1: 4})$$

nennen wir **Residuen**.

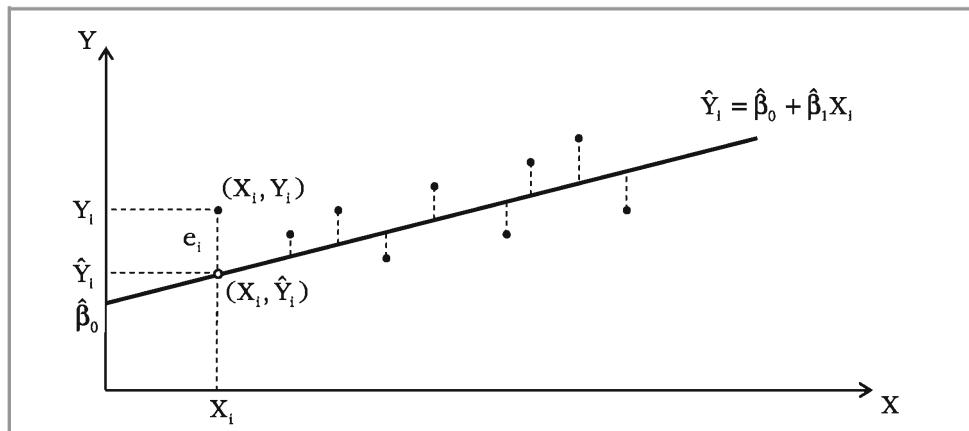


Abbildung IV 2: OLS-Schätzung

Ziel eines guten Schätzverfahrens sollte es nun sein, die Schätzungen $\hat{\beta}_0$ und $\hat{\beta}_1$ so zu bestimmen, dass diese Abweichungen e_i so gering wie möglich ausfallen. Genau hier setzt OLS an und bestimmt die Schätzwerte so, dass sie die Quadratesumme der Residuen minimieren. Dies ist ein vernünftiges Ziel für eine Schätztechnik.

Durch die Quadrierung der Residuen werden nämlich ihre Vorzeichen eliminiert, sodass sich positive und negative Residuen nicht gegenseitig aufheben.³

Das Ziel von OLS können wir also formal wie folgt formulieren:

$$\text{OLS minimiert } \sum_{i=1}^n e_i^2 \quad (\text{IV1: 5})$$

Da sich die Residuen gerade nach $e_i = Y_i - \hat{Y}_i$ ergeben und zugleich $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ gilt, minimiert OLS durch Wahl der entsprechenden Werte für $\hat{\beta}_0$ und $\hat{\beta}_1$ konkret den Term

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2, \quad (\text{IV1: 6})$$

in dem wir, wie auch im Folgenden, auf die Angabe des *Summationsindex* verzichten, da dieser *immer in den Grenzen i = 1 bis n* verläuft. Bilden wir die partiellen Ableitungen von (IV1: 6) nach den beiden Parameterschätzern und setzen diese gleich Null, erhalten wir die sog. *Normalgleichungen* zu

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_0} = \sum 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-1) = -2 \sum e_i = 0 \quad (\text{IV1: 7a})$$

und

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = \sum 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-X_i) = -2 \sum e_i X_i = 0. \quad (\text{IV1: 7b})$$

OLS-Schätzungen sind nun nichts anderes als die Werte von $\hat{\beta}_0$ und $\hat{\beta}_1$, die diese beiden Gleichungen erfüllen. Nach Division beider Seiten durch -2 und einigen Umformungen (u. a. $\sum X_i = n\bar{X}$) ergibt sich aus (IV1: 7a)

$$\sum Y_i - \sum \hat{\beta}_0 - \hat{\beta}_1 \sum X_i = 0 \rightarrow \bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} = 0 \quad (\text{IV1: 8a})$$

und aus (IV1: 7b)

$$\sum Y_i X_i - \hat{\beta}_0 \sum X_i - \hat{\beta}_1 \sum X_i^2 = 0 \rightarrow \frac{1}{n} \sum X_i Y_i - \hat{\beta}_0 \bar{X} - \hat{\beta}_1 \frac{1}{n} \sum X_i^2 = 0. \quad (\text{IV1: 8b})$$

Lösen wir dieses Gleichungssystem aus (IV1: 8a) und (IV1: 8b) nach $\hat{\beta}_0$ und $\hat{\beta}_1$ auf, erhalten wir schließlich⁴

³ Dies wäre z.B. der Fall, wenn wir die Summe der Residuen minimieren. Wir erhielten dann keine eindeutige Lösung. Eine mögliche "Regressionsgleichung" ist einfach $\hat{Y}_i = \bar{Y}$ und als Minimum der Residuensumme ergibt sich Null, da die Summe der Abweichungen einer Variablen von ihrem Mittelwert stets Null ist. Dies ist jedoch keine vernünftige Charakterisierung eines ggf. vorliegenden funktionalen Zusammenhangs.

⁴ Bei näherer Betrachtung von (IV1: 9) fällt auf, dass es sich dabei um nichts anderes als den Quotienten aus der geschätzten Kovarianz zwischen X und Y und der geschätzten Varianz von X handelt. Es ist somit für $\hat{\beta}_1$ auch die Schreibweise $\hat{\beta}_1 = \widehat{\text{Cov}}(X, Y) / \widehat{\text{Var}}(X)$ denkbar.

$$\hat{\beta}_1 = \frac{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n} \sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i}{\sum x_i^2} \quad \text{mit} \quad \begin{aligned} x_i &= X_i - \bar{X} \\ y_i &= Y_i - \bar{Y} \end{aligned} \quad (\text{IV1: 9})$$

und damit

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \quad (\text{IV1: 10})$$

Für die nach (IV1: 9) und (IV1: 10) bestimmten Werte nimmt also die Summe der quadrierten Abweichungen von der Regressionsgeraden ihr Minimum an. Betrachten wir (IV1: 9) nun etwas näher. Da die Varianz der Variable X im Nenner auftaucht, dürfen die *Ausprägungen von X* in einer Stichprobe *keinesfalls alle gleich* sein. Wäre dies der Fall, so wäre eine Schätzung von β_1 (und damit auch von β_0) unmöglich, da die Varianz dann den Wert Null annehmen würde. Außerdem verdient der Zähler besondere Aufmerksamkeit. Da \bar{Y} nicht vom Summationsindex abhängt und die Summe der Abweichungen einer Variable von ihrem Mittelwert immer Null ist (hier: $\sum x_i = \sum (X_i - \bar{X}) = 0$), gilt nämlich

$$\sum x_i y_i = \sum x_i (Y_i - \bar{Y}) = \sum x_i Y_i - \bar{Y} \sum x_i = \sum x_i Y_i. \quad (\text{IV1: 11})$$

Es ist also möglich (IV1: 9) alternativ als

$$\hat{\beta}_1 = \frac{\sum x_i Y_i}{\sum x_i^2} \quad \text{mit} \quad x_i = X_i - \bar{X} \quad (\text{IV1: 12})$$

darzustellen. Diese Form werden wir speziell in den Abschnitten IV 2.3 und 4.1 benötigen.

Neben dem eingangs angeführten mehr berechnungstechnischen Grund für eine Rechtfertigung von OLS als Schätztechnik wird OLS aus zwei weiteren Gründen häufig in der Praxis eingesetzt. Erstens handelt es sich bei OLS um das wohl einfachste aller ökonometrischen Schätzverfahren. So können OLS-Schätzungen (in einfachen Modellen mit bis zu zwei erklärenden Variablen) sogar noch "von Hand" ausgerechnet werden. Zweitens hat OLS in einem speziellen Modellkontext eine Reihe nützlicher *statistischer Eigenschaften* (vgl. Abschnitt IV 2.3) und die folgenden interessanten **mathematischen Eigenschaften**:

1. Die geschätzte Regressionsgerade verläuft durch die arithmetischen Mittel der erklärenden Variable X und der erklären Variable Y . Dies ergibt sich aufgrund von (IV1: 10). Dessen Umformung liefert nämlich

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}. \quad (\text{IV1: 13a})$$

Es gilt damit natürlich auch

$$\hat{Y} = \bar{Y}. \quad (\text{IV1: 13b})$$

2. Da die OLS-Schätzungen die Normalgleichung (IV1: 7a) erfüllen müssen, ist die Summe der Residuen und damit auch ihr Mittelwert \bar{e} stets gleich Null.

$$\sum e_i = 0 \rightarrow \bar{e} = \frac{1}{n} \sum e_i = 0 \quad (\text{IV1: 13c})$$

Aufgrund von (IV1: 13c) sind die Residuen auch als $e_i = e_i - \bar{e}$ interpretierbar, sodass die Summe der Residuenquadrate (nach Division durch n) auch die Varianz der Residuen verkörpert:

$$\text{Var}(e) = \frac{1}{n} \sum e_i^2 = \frac{1}{n} \sum (e_i - \bar{e})^2 \quad (\text{IV1: 13d})$$

3. Die Residuen sind nicht mit der erklärenden Variablen korreliert.

$$\sum e_i X_i = 0 \quad \rightarrow \quad \text{Cov}(e, X) = 0 \quad (\text{IV1: 13e})$$

Dies können wir wie folgt zeigen:

$$\text{Cov}(e, X) = \frac{1}{n} \sum e_i X_i = \frac{1}{n} \sum e_i (X_i - \bar{X}) = \frac{1}{n} \sum e_i X_i - \frac{1}{n} \bar{X} \sum e_i = 0 - 0 = 0$$

$\sum e_i X_i = 0$ folgt direkt aus Gleichung (IV1: 7b). Da die OLS-Schätzungen diese Normalgleichung erfüllen müssen, wird der Term $\sum e_i X_i$ im Rahmen der Ermittlung der Schätzungen automatisch Null.

4. Die Residuen sind nicht mit den geschätzten (sog. "gefitteten") Werten von Y korreliert.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad \rightarrow \quad \text{Cov}(e, \hat{Y}) = 0 \quad (\text{IV1: 13f})$$

Da \hat{Y} eine lineare Transformation von X ist und X nach (IV1: 13e) nicht mit e korreliert ist, muss auch \hat{Y} unkorreliert mit e sein.

Nach dieser theoretischen Einführung in die Vorgehensweise von OLS wollen wir uns nun einem konkreten Anwendungsbeispiel widmen. Davor ist jedoch noch eine für das Beispiel relevante Notation zu erklären: Wenn wir bei einer Regression mit Zeitreihendatenmaterial arbeiten, ersetzen wir das Subskript i gewöhnlich durch t mit $t = 1, 2, \dots, n$ (oder $t = 1, 2, \dots, T$), um den Zeitbezug einer Beobachtung darzustellen. Die verschiedenen Beobachtungen einer Stichprobe (z.B. Kurse) beziehen sich dann auf die *gleiche ökonomische Einheit* (z.B. Aktie), jedoch zu *unterschiedlichen Zeitpunkten*. Wir sprechen in einem solchen Fall von einer **Zeitreihenregression**. Werden im Datenmaterial hingegen von *verschiedenen ökonomischen Einheiten* Merkmalsausprägungen zum *selben Zeitpunkt* festgehalten (z.B. Marktkapitalisierung aller Aktien im DAX), sprechen wir von einer **Querschnittsregression**. Hier repräsentiert das Subskript i die i-te ökonomische Einheit.

Beispiel IV 1-4

OLS-Schätzung mit einer erklärenden Variablen



Wie einfach OLS auch ohne ökonometrische Softwarepakete anwendbar ist, wollen wir in diesem ersten Anwendungsbeispiel aufzeigen. Wir nutzen dazu eine fiktive Datentabelle, die für die Jahre 2000 bis 2010 den Teekonsum Y der Bürger eines Landes pro Person in Tassen je Tag und den realen Einzelhandelspreis für 100 g Tee in X Geldeinheiten beinhaltet. Wir haben die Tabelle bereits um die Berechnungen ergänzt, die wir für die Schätzung der Parameter der möglichen Beziehung

$$Y_t = \beta_0 + \beta_1 X_t$$

zwischen Teekonsum und realem Einzelhandelspreis benötigen. Darüber hinaus sind in den letzten beiden Spalten Angaben enthalten, die uns die Veranschaulichung der Gültigkeit einiger der oben aufgeführten mathematischen Eigenschaften von OLS erlauben.

Rohdaten			Berechnungen zur OLS-Schätzung				Ergebnisse	
Jahr	Konsum	Preis	$(Y_t - \bar{Y})$	$(X_t - \bar{X})$	$(X_t - \bar{X})^2$	$(X_t - \bar{X})(Y_t - \bar{Y})$	\hat{Y}_t	e_t
2000	3,15	0,94	0,48	-0,28	0,08	-0,13	2,81	0,34
2001	3,04	0,88	0,37	-0,34	0,11	-0,12	2,84	0,20
2002	2,81	0,86	0,14	-0,36	0,13	-0,05	2,85	-0,04
2003	2,77	0,88	0,10	-0,34	0,11	-0,03	2,84	-0,07
2004	2,73	0,93	0,06	-0,29	0,08	-0,02	2,82	-0,09
2005	2,65	0,94	-0,02	-0,28	0,08	0,01	2,81	-0,16
2006	2,57	1,33	-0,10	0,11	0,01	-0,01	2,62	-0,05
2007	2,36	2,12	-0,31	0,90	0,81	-0,28	2,23	0,13
2008	2,41	1,69	-0,26	0,47	0,22	-0,12	2,44	-0,03
2009	2,43	1,43	-0,24	0,21	0,05	-0,05	2,57	-0,14
2010	2,48	1,39	-0,19	0,17	0,03	-0,03	2,59	-0,11
Σ	29,40	13,39	0,00	0,00	1,72	-0,85	29,40	0,00
Mittel	2,67	1,22	0,00	0,00			2,67	0,00

Mit diesem Zahlenmaterial bzw. den Spaltensummen können wir die Regressionskoeffizienten der vermuteten funktionalen Beziehung durch Einsetzen in die Formeln (IV1: 9) und (IV1: 10) schätzen. Wir geben im Folgenden die exakten Schätzergebnisse und nicht die aus den gerundeten Tabellenwerten entstandenen an:

$$\hat{\beta}_1 = \frac{-0,85}{1,72} = -0,4956$$

$$\hat{\beta}_0 = 2,67 - (-0,4956 \cdot 1,22) = 3,2760$$

Für unsere geschätzte Regressionsgerade gilt also $\hat{Y}_t = 3,2760 - 0,4956X_t$. Besonders der geschätzte Koeffizient $\hat{\beta}_1$ ist hier für uns von Interesse. Er gibt nämlich an, um wie viele Einheiten sich der Konsum bei Erhöhung des Preises um eine Einheit veränderte. In diesem Fall war ein Rückgang des Konsums um 0,4956 Tassen pro Tag die Folge. Wir sprechen hier noch bewusst in Vergangenheitsform, da OLS zunächst nur rein deskriptiv das vorliegende Datenmaterial beschreibt.

Wir können die obige Tabelle außer zur Schätzung der Regressionsparameter auch dazu nutzen, einige bereits genannte mathematische Eigenschaften des Prinzips der kleinsten Quadrate zu veranschaulichen:

1. Durch Einsetzen von \bar{X} und \bar{Y} in $\hat{Y}_t = 3,2760 - 0,4956X_t$ können wir zeigen, dass diese beiden Mittelwerte auf der geschätzten Regressionsgeraden liegen.
2. Die Summe der Residuen liegt bei Null, woraus auch ein Mittelwert von Null folgt.
3. Die Summe der Y_t ist mit 29,40 gleich der Summe der \hat{Y}_t , sodass auch ihre Mittelwerte identisch sind.
4. Sowohl $\text{Cov}(e, X) = 0$ und $\text{Cov}(e, \hat{Y}) = 0$ sind erfüllt. Dies möge der Leser selbst nachrechnen oder sich dessen in der mitgelieferten Excel-Datei im Online-Service vergewissern.

1.2.2 OLS und Linearität

Dem Begriff der Linearität kommt im Rahmen des Prinzips der kleinsten Quadrate eine besondere Bedeutung zu. OLS ist nämlich nur in der Lage parameterlineare Beziehungen zwischen Variablen aus Stichprobendatenmaterial zu schätzen. Eine Variablenlinearität der Beziehung ist nicht zwingend erforderlich.

Variablenlinearität heißt, dass die Variablen einer Funktion Exponenten von Eins aufweisen und sie nicht mit anderen Variablen multipliziert werden. Eine Funktion

$$Y_i = \beta_0 + \beta_1 X_i$$

ist deshalb variablenlinear, wohingegen wir dies z.B. von der Funktion

$$Y_i = \beta_0 + \beta_1 X_i^2, \quad (\text{IV1: 14a})$$

die grafisch eine Parabel liefert, nicht sagen können. *Fehlende Variablenlinearität*, bedeutet nun aber nicht, dass die Parameter in (IV1: 14a) nicht trotzdem mit OLS geschätzt werden können. Es könnte hier nämlich z.B. eine neue Variable $Z_i = X_i^2$ definiert werden, die zu einer neuen Gleichung

$$Y_i = \beta_0 + \beta_1 Z_i \quad \text{mit} \quad Z_i = X_i^2 \quad (\text{IV1: 14b})$$

führt. Diese Gleichung ist variablenlinear und kann mittels OLS geschätzt werden. Da eine derartige Transformation für jede nicht variablenlineare Gleichung möglich ist, sind wir in der Lage die Parameter einer Vielzahl funktionaler Formen mit OLS zu schätzen. Nachfolgend findet sich davon eine kleine Auswahl, auf deren besondere Eigenschaften wir im Abschnitt IV 4.2 noch näher zu sprechen kommen.

Lineare Form:	$Y_i = \beta_0 + \beta_1 X_i$
Quadratische Form: ⁵	$Y_i = \beta_0 + \beta_1 X_i^2$
Logarithmische Formen:	$\ln Y_i = \beta_0 + \beta_1 \ln X_i$
	$Y_i = \beta_0 + \beta_1 \ln X_i$
	$\ln Y_i = \beta_0 + \beta_1 X_i$
Inverse Form:	$Y_i = \beta_0 + \beta_1 \frac{1}{X_i}$

Für die Auswahl der funktionalen Form, die einen vorliegenden Datensatz optimal beschreibt, gibt es eine Reihe von Kriterien, die wir im Abschnitt IV 1.2.4 noch näher betrachten werden. Im nachfolgenden Beispiel wollen wir zunächst nur die Residuenquadratesumme als Kriterium betrachten.

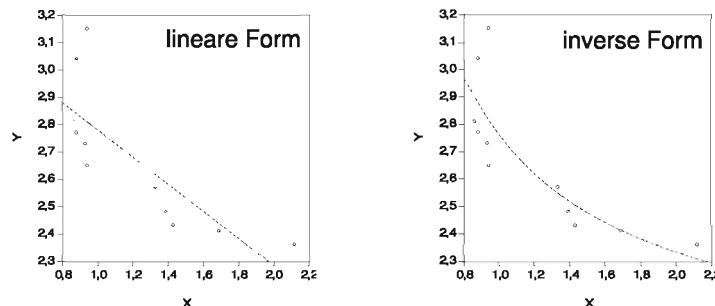
Beispiel IV 1-5

OLS-Schätzung, Linearität und Funktionswahl



Der aufmerksame Leser hat sich sicher gefragt, warum wir bei Beispiel IV 1-4 auf eine grafische Darstellung des Datenmaterials und der geschätzten Regressionsgerade verzichtet haben. Dies begründet sich dadurch, dass eine lineare Funktion für die Beschreibung des vorliegenden Zahlenmaterials nicht unbedingt optimal ist. Eine inverse Form würde hier z.B. bessere Dienste leisten, wie die nachfolgenden Abbildungen zeigen.

⁵ In der Praxis werden wir eine quadratische Polynomform als Modell mit zwei erklärenden Variablen, d.h. als $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$, spezifizieren.



Wie die Grafiken vermuten lassen, wird die Summe der quadrierten Abweichungen der Beobachtungen von den geschätzten Funktionen für die inverse Form $Y_t = \beta_0 + \beta_1 \cdot 1/X_t$ geringer sein. Prüfen wir dies durch Schätzung der Parameter der inversen Form nach. Wir substituieren dazu $Z_t = 1/X_t$ und schätzen $Y_t = \beta_0 + \beta_1 Z_t$:

Rohdaten		Berechnungen zur OLS-Schätzung				Ergebnisse			
Jahr	Konsum	Y_t	Z_t	$(Y_t - \bar{Y})$	$(Z_t - \bar{Z})$	$(Z_t - \bar{Z})^2$	$(Z_t - \bar{Z})(Y_t - \bar{Y})$	\hat{Y}_t	e_t
2000	3,15	1,06	0,48	0,17	0,03	0,08	0,08	2,81	0,34
2001	3,04	1,14	0,37	0,24	0,06	0,09	0,09	2,88	0,16
2002	2,81	1,16	0,14	0,27	0,07	0,04	0,04	2,90	-0,09
2003	2,77	1,14	0,10	0,24	0,06	0,02	0,02	2,88	-0,11
2004	2,73	1,08	0,06	0,18	0,03	0,01	0,01	2,82	-0,09
2005	2,65	1,06	-0,02	0,17	0,03	0,00	0,00	2,81	-0,16
2006	2,57	0,75	-0,10	-0,15	0,02	0,02	0,01	2,55	0,02
2007	2,36	0,47	-0,31	-0,43	0,18	0,13	0,13	2,31	0,05
2008	2,41	0,59	-0,26	-0,31	0,09	0,08	0,08	2,41	0,00
2009	2,43	0,70	-0,24	-0,20	0,04	0,05	0,05	2,50	-0,07
2010	2,48	0,72	-0,19	-0,18	0,03	0,03	0,03	2,52	-0,04
Σ	29,40	9,87	0,00	0,00	0,64	0,54	0,54	29,40	0,00
Mittel		2,67	0,90	0,00	0,00			2,67	0,00

Analog zu Beispiel IV 1-4 erhalten wir damit die Schätzung 0,8522 für β_1 und 1,9079 für β_0 , d.h. eine geschätzte Regressionsfunktion $\hat{Y}_t = 1,9079 + 0,8522 \cdot 1/X_t$. Bestimmen wir nun für die lineare Form aus Beispiel IV 1-4 die Residuenquadrate summe, so erhalten wir den Wert 0,2464. Für die inverse Form ergibt sich 0,2046. Die inverse Form beschreibt die Daten also besser. Wir sagen hier auch, dass die inverse Form eine bessere Anpassungs-güte an die Daten oder einen besseren "fit" aufweist.

Wir sehen an diesem Beispiel, dass wir unsere Regressionsfunktion optimal in unser Datenmaterial einpassen können, indem wir das Datenmaterial betrachten und damit eine geeignet erscheinende funktionale Form auswählen. Handelt es sich bei dem Datenmaterial um eine Stichprobe aus einer größeren Grundgesamtheit, ist eine solche Vorgehensweise problematisch, wenn wir uns für die Eigenschaften der Grundgesamtheit interessieren, da wir ja nur die Stichprobe optimal beschreiben. In solchen Fällen kann mit der geschätzten Regressionsgleichung evtl. nur schlecht prognostiziert werden. Im Abschnitt IV 2 und den weiteren Abschnitten des Kapitels IV werden wir sehen, dass es sehr wichtig ist, für Aussagen zur Grundgesamtheit (Rückschlüsse, Prognosen) bei der Auswahl der funktionalen Form auch inhaltliche Begründungen (z.B. Gesetz des abnehmenden Grenzertrages) zu berücksichtigen.

Eine variablenlineare oder nicht variablenlineare Funktion ist nur dann mit OLS schätzbar, wenn sie parameterlinear ist.⁶ Bei der *Parameterlinearität* liegt das Hauptaugenmerk auf den Exponenten der Koeffizienten einer Gleichung. Sind sie gleich Eins und werden keine Koeffizienten miteinander multipliziert, so kann die Gleichung unabhängig von den Exponenten der Variablen als parameterlinear bezeichnet werden. Während also z.B. $Y_i = \beta_0 + \beta_1 X_i^2$ nicht variablenlinear ist, ist die Funktion dennoch parameterlinear und ihre Parameter daher mittels OLS-schätzbar.

Problematisch ist nun aber *fehlende Parameterlinearität*. Ohne diese kann OLS nicht angewendet werden. Es gibt jedoch Fälle, in denen einfache Transformationen aus einer nicht parameterlinearen Funktion eine parameterlineare Funktion erzeugen können. Betrachten wir beispielsweise die nicht parameterlineare Funktion

$$Y_i = \beta_0 X_i^{\beta_1}. \quad (\text{IV1: 15a})$$

Logarithmieren wir diese Funktion, so erhalten wir eine parameterlineare Form

$$\ln Y_i = \ln \beta_0 + \beta_1 \ln X_i, \quad (\text{IV1: 15b})$$

die sich mit OLS schätzen lässt.

1.2.3 OLS bei Modellen mit mehreren erklärenden Variablen

In der Praxis werden funktionale Beziehungen mit nur einer erklärenden Variablen nur in den seltensten Fällen geschätzt. Vielmehr befassen wir uns mit Formen wie

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K X_{Ki}, \quad (\text{IV1: 16})$$

wobei X_{1i} die i-te Beobachtung der ersten erklärenden Variablen und X_{2i} die i-te Beobachtung der zweiten erklärenden Variablen und K die Anzahl der erklärenden Variablen darstellen. Y wird also hier von mehreren erklärenden Variablen X_k mit $k = 1, 2, \dots, K$ bestimmt. Wir sprechen auch von **multipler Regression**.

Die OLS-Schätzung der Parameter derartiger funktionaler Formen folgt den gleichen mathematischen Prinzipien wie bei solchen mit nur einer einzigen erklärenden Variablen. Auch hier werden die $\hat{\beta}$ s gesucht, die die Quadratesumme der Residuen, $\sum e_i^2$, minimieren. Nur stammen die Residuen diesmal eben aus einer Gleichung mit mehreren erklärenden Variablen, d.h.

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_K X_{Ki}.$$

Zudem ist auch hier wieder Parameterlinearität zwingende Voraussetzung für die Anwendung der OLS-Schätzung.

Im Fall von *zwei* erklärenden Variablen, also

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}, \quad (\text{IV1: 17})$$

gewinnen wir die OLS-Schätzungen der Parameter β_0 , β_1 und β_2 über⁷

$$\hat{\beta}_1 = \frac{(\sum y_i x_{1i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2}, \quad (\text{IV1: 18})$$

⁶ Dies war bei allen bisher vorgestellten Funktionstypen der Fall.

⁷ Für eine Herleitung dieser Schätzformeln vgl. z.B. Gujarati, D.N. (2003), Kapitel 7.

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{1i}^2) - (\sum y_i x_{1i})(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2}, \quad (\text{IV1: 19})$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2, \quad (\text{IV1: 20})$$

wobei die Abweichungen der Variablen von ihrem arithmetischen Mittel wieder in Kleinbuchstaben, also $y_i = Y_i - \bar{Y}$, $x_{1i} = X_{1i} - \bar{X}_1$ und $x_{2i} = X_{2i} - \bar{X}_2$, ausgedrückt werden und die Summation in den Grenzen $i = 1$ bis n erfolgt. Die mathematischen Eigenschaften der Schätzer ergeben sich analog zu denen in Modellen mit nur einer erklärenden Variablen. So ist z.B. die Summe der Residuen wieder Null und die erklärenden Variablen sind nicht mit den Residuen korreliert. Die OLS-Schätzer bei mehr als zwei erklärenden Variablen lassen sich nur noch mit Hilfe von Matrizen sinnvoll darstellen, wobei sich aber keine neuen Erkenntnisse ergeben. Zudem erledigt die moderne Statistiksoftware die Ermittlung der Schätzungen für uns, weshalb wir uns nicht weiter damit beschäftigen müssen.

Bei mehr als einer erklärenden Variablen sind die Regressionskoeffizienten β_k nun aber etwas anders zu interpretieren. β_1 ist nun der Einfluss einer Erhöhung von X_1 um eine Einheit auf Y unter Konstanthaltung aller anderen berücksichtigten erklärenden Variablen. Wir sagen auch, dass der Parameter β_1 den partiellen Einfluss von X_1 auf Y angibt. Analog sind auch die anderen Regressionskoeffizienten zu deuten. Die multiple Regression ermöglicht also auch in den Wirtschaftswissenschaften, was sonst nur in einem kontrollierten Laborexperiment möglich wäre: Die Isolation weiterer Einflussfaktoren auf Y .

Beispiel IV 1-6

OLS-Schätzung mit zwei erklärenden Variablen



Eine im April 2009 von uns durchgeführte Erhebung zum Preis von Luxus-Villen im Raum Frankfurt am Main lieferte 10 Objekte, deren Preise (Y), Zimmerzahl (X_1) und Wohnfläche in m^2 (X_2) in der nachfolgenden Tabelle enthalten sind. Wir interessieren uns nun dafür, welchen Einfluss Zimmerzahl und Wohnfläche auf den Preis von Luxusimmobilien in Frankfurt haben, und möchten dazu die funktionale Beziehung

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

im vorliegenden Datenmaterial schätzen. Man beachte hierbei, dass wir nun das Subskript i verwenden, da wir mit Querschnittsdatenmaterial arbeiten. Die Schätzung dieser einfachen Beziehung wollen wir ein letztes Mal "von Hand" aufzeigen. In der Praxis bedienen wir uns spezieller ökonometrischer Software wie z.B. LIMDEP, SHAZAM, MICRO TSP, MINITAB, SAS, SPSS, STATA, Microfit, BMD oder PcGive. Die Schätzungen für dieses Lehrbuch werden mit Hilfe von EViews 6.0 durchgeführt. In künftigen Beispielen werden wir uns daher darauf beschränken, EViews-Regressionsoutputs zu präsentieren. Derartige ökonometrische Software soll uns ja schließlich zeitraubende Rechenarbeit abnehmen.

Zur Schätzung der Koeffizienten bedienen wir uns der Summen und Mittelwerte der nachfolgenden Tabellen. Setzen wir diese in die Formeln (IV1: 18) bis (IV1: 20) ein, erhalten wir

$$\hat{\beta}_1 = 100.108,41 \quad \hat{\beta}_2 = 3.095,92 \quad \hat{\beta}_0 = -431.689,56.$$

Dieses Ergebnis bedeutet, dass eine Erhöhung der Anzahl der Zimmer um Eins bei unveränderter Wohnfläche X_2 durchschnittlich zu einem um 100.108,41 Euro höheren Preis führte. Eine Erhöhung der Wohnfläche um einen m^2 hatte bei unveränderter Zimmerzahl X_1 eine durchschnittliche Preissteigerung von 3.095,92 Euro zur Folge.

Villa i	Preis Y_i	Zimmer X_{1i}	m^2 -Zahl X_{2i}	Abweichungen vom Mittel		
				y_i	X_{1i}	X_{2i}
1	725.000	6	175	-722,800	-2,3	-163,7
2	1.170.000	8	332	-277,800	-0,3	-6,7
3	1.190.000	7	260	-257,800	-1,3	-78,7
4	1.250.000	13	410	-197,800	4,7	71,3
5	3.500.000	12	400	2.052.200	3,7	61,3
6	1.700.000	6	390	252.200	-2,3	51,3
7	1.399.000	11	378	-48.800	2,7	39,3
8	1.490.000	6	492	42.200	-2,3	153,3
9	1.300.000	7	360	-147.800	-1,3	21,3
10	754.000	7	190	-693.800	-1,3	-148,7
Mittelwert	1.447.800	8,3	338,7			
i	$y_i X_{1i}$	$y_i X_{2i}$	$X_{1i} X_{2i}$	X_{1i}^2	X_{2i}^2	
1	1.662.440	118.322.360	376,51	5,29	26.797,69	
2	83.340	1.861.260	2,01	0,09	44,89	
3	335.140	20.288.860	102,31	1,69	6.193,69	
4	-929.660	-14.103.140	335,11	22,09	5.083,69	
5	7.593.140	125.799.860	226,81	13,69	3.757,69	
6	-580.060	12.937.860	-117,99	5,29	2.631,69	
7	-131.760	-1.917.840	106,11	7,29	1.544,49	
8	-97.060	6.469.260	-352,59	5,29	23.500,89	
9	192.140	-3.148.140	-27,69	1,69	453,69	
10	901.940	103.168.060	193,31	1,69	22.111,69	
Summe	9.029.600	369.678.400	843,90	64,10	92.120,10	

1.2.4 Güte einer geschätzten Regressionsgleichung

Haben wir mittels OLS die Parameter einer unterstellten Beziehung zwischen Variablen geschätzt, gilt es die Güte der geschätzten Regressionsfunktion zu untersuchen. Es gibt hierzu eine Reihe von Maßzahlen. Alle diese Maße machen nur Aussagen über den "fit" der Gleichung, d.h. sie beurteilen lediglich, wie nahe die \hat{Y}_i bei den tatsächlichen Y_i liegen bzw. wie gut sich die Regressionsfunktion ins vorliegende Datenmaterial einpasst. Davon haben wir bereits im Beispiel IV 1-5 Gebrauch gemacht, in dem wir zwei verschiedene funktionale Formen anhand der Residuenquadrate summe verglichen haben. Je niedriger diese nämlich ist, desto besser passt sich die unterstellte Funktion in das vorliegende Datenmaterial ein. Im Folgenden stellen wir Maße, wie das Bestimmtheitsmaß, den Korrelationskoeffizienten und das angepasste Bestimmtheitsmaß vor, die sich direkt oder indirekt ebenfalls dieser Quadrate summe bedienen.

1.2.4.1 Das Bestimmtheitsmaß

Zur Bestimmung des Ausmaßes an Variabilität von Y , das durch die Regression bzw. die Regressoren erklärt werden kann, können wir die mit der Beobachtungszahl n multiplizierte Varianz von Y , die auch als Gesamtquadratsumme (TSS,

engl. total sum of squares) bezeichnet wird, in zwei Komponenten aufzuteilen.⁸ Dies sind die erklärte Quadratsumme (ESS, engl. explained sum of squares), die nichts anderes als die mit n multiplizierte Varianz von \hat{Y} darstellt,⁹ und die bereits behandelte Residuenquadratsumme (RSS, engl. residual sum of squares):

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{TSS}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{ESS}} + \underbrace{\sum_{i=1}^n e_i^2}_{\text{RSS}} \quad (\text{IV1: 21})$$

Abbildung IV 3 veranschaulicht diese Zerlegung. Sie zeigt uns, dass die Abweichung von Y von seinem Mittelwert \bar{Y} , durch zwei Komponenten bedingt ist: Die Abweichung von \hat{Y} von seinem Mittelwert \bar{Y} und das Residuum. ESS misst also den Teil der Variabilität von Y , der durch die geschätzte Regressionsgerade erklärt wird, wohingegen RSS den nicht erklärten Teil darstellt.

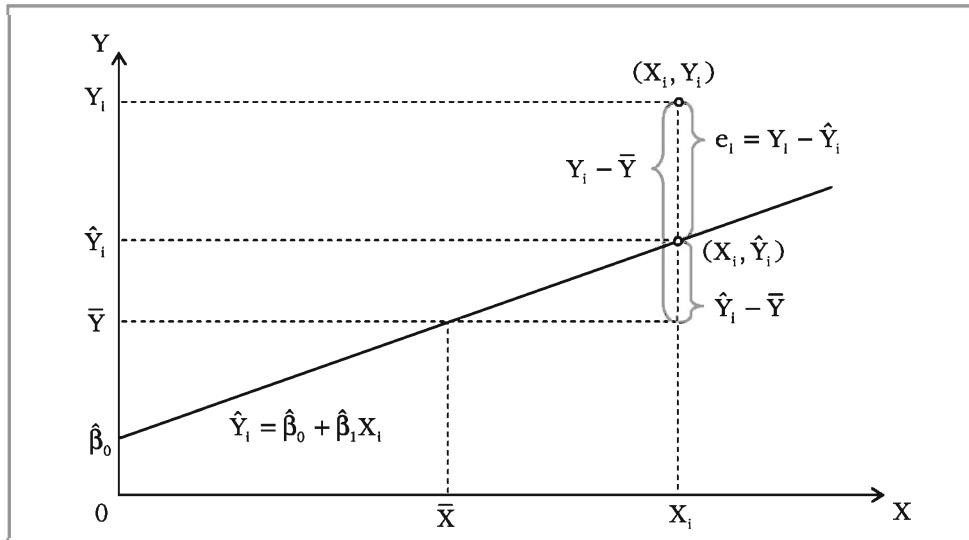


Abbildung IV 3: Zerlegung der Varianz von Y

Wir erkennen aus dieser Zerlegung, dass der "fit" einer Regressionsgeraden bzw. die Fähigkeit der Regressionsgeraden, die Veränderung von Y zu erklären, umso besser ist, je kleiner RSS relativ zu TSS ist. Da OLS RSS minimiert, erhalten wir durch dieses Verfahren (für eine gegebene funktionale Form in einem gegebenen Datensatz) stets das maximal mögliche ESS. Setzen wir ESS ins Verhältnis zu TSS, so erhalten wir das sog. **Bestimmtheitsmaß**

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (\text{IV1: 22a})$$

⁸ Für eine formale Herleitung dieser Zerlegung vgl. z.B. in Gujarati, D.N. (2003), Kapitel 3. Die Zerlegung ist nur möglich, wenn im Modell eine Konstante enthalten ist.

⁹ Es gilt schließlich (IV1: 13b), d.h. $\hat{Y} = \bar{Y}$.

welches uns den Anteil der durch die Regression erklären Variation von Y angibt. Da ESS hier im Zähler auftaucht, ist unmittelbar klar, dass OLS natürlich auch (für eine gegebene funktionale Form in einem gegebenen Datensatz) für das maximal mögliche R^2 sorgt. Wir erkennen außerdem, dass mit einem hohen R^2 auch ein besserer "fit" der geschätzten Regressionsgeraden verbunden ist.

TSS, RSS und ESS sind jeweils Quadratesummen, d.h. stets positiv, und es gilt zudem $ESS \leq TSS$, sodass R^2 im Intervall

$$0 \leq R^2 \leq 1 \quad (\text{IV1: 22b})$$

liegen muss. Ein R^2 -Wert von nahe Eins weist auf einen sehr guten "fit" hin, wohingegen bei einem Wert nahe Null die geschätzte Regressionsfunktion nicht besser in der Lage ist Y zu erklären als es das Stichprobenmittel \bar{Y} könnte. Abbildung IV 4 veranschaulicht diese Fälle.

In der linken Grafik besteht zwischen X und Y keinerlei Beziehung. Da die Regression keinerlei Erklärungsgehalt für die Veränderung von Y besitzen kann, ist ESS gleich Null, wodurch auch R^2 einen Wert von Null annimmt. Die geschätzte Regressionsgleichung könnte genauso gut $\hat{Y}_i = \bar{Y}$ lauten.

In der rechten Grafik kann die bestehende Beziehung zwischen X und Y gut durch die geschätzte Regressionsfunktion beschrieben werden. R^2 nimmt einen Wert von 0,95 an. Ein solcher Wert für das Bestimmtheitsmaß ist ein typisches Ergebnis einer Zeitreihenregression mit gutem "fit". Der Großteil der Variabilität von Y wird erklärt, wobei trotzdem ein zufälliger, unerklärter Teil verbleibt. Ein R^2 von genau Eins wird bei wirtschaftswissenschaftlichen Fragestellungen nie erreicht.

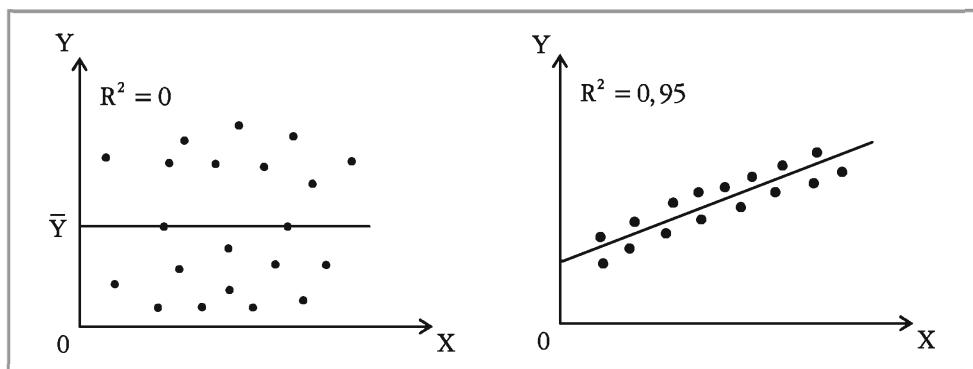


Abbildung IV 4: Bestimmtheitsmaß

Besonders bei *Zeitreihenregressionen* ist mit dem Bestimmtheitsmaß mit Vorsicht umzugehen. Hier ist ein hohes R^2 nämlich meist Resultat von gleichgerichteten signifikanten Trends der erklären und der erklärenden Variablen.¹⁰ Obwohl ein hohes R^2 festgestellt wird, kann es vorkommen, dass keine Kausalität von X für Y gegeben ist. Wir sollten nämlich immer beachten, dass ein hohes R^2 nicht impliziert,

¹⁰ R^2 wäre in solchen Fällen aussagekräftiger, wenn wir mit trendbereinigten Daten arbeiten. Das Thema Trends werden wir im Abschnitt IV 4.7.2 eingehend behandeln.

dass Änderungen von X zu Änderungen von Y führen, da es eine dritte Variable geben könnte, die X und Y gleichzeitig beeinflusst.

Bei *Querschnittsregressionen* erhalten wir häufig ein niedriges R^2 , da sich die Beobachtungen in ihren Werten stark unterscheiden. Nehmen wir beispielsweise an, wir wollen die Ausgaben für Jeans, die von Single-Haushalten im Monat März getätigt wurden, durch das Haushaltseinkommen, das Alter und Geschlecht der Singles, die Saison sowie zufällige Schwankungen erklären. Kauft ein Haushalt nun in einem Monat Jeans, so werden die Ausgaben für Jeans im Folgemonat i.d.R. bei Null liegen. Da zudem nicht alle Haushalte im selben Monat und im gleichen Wert Jeans kaufen, wird es z.B. selbst bei Haushalten mit gleichem Einkommen deutliche Unterschiede bei den Jeansausgaben des Monats März geben. Bei Querschnittsregressionen werden wir daher z.B. bereits ein R^2 von 0,5 als guten "fit" betrachten und unser Hauptaugenmerk auf die Identifikation von erklärenden Variablen legen, die einen deutlichen Einfluss auf die erklärte Variable haben. Die Beurteilung, ob ein R^2 -Wert als hoch oder niedrig einzustufen ist, ist daher studienspezifisch.

Es sei abschließend noch einmal klar hervorgehoben, dass OLS für eine *gegebene funktionale Form* in *gegebenem Zahlenmaterial* zum größtmöglichen R^2 führt. Durch die Wahl einer anderen funktionalen Form kann jedoch ggf. ein höheres R^2 erreicht werden. Wir wollen dies anhand eines einfachen Beispiels aufzeigen.

Beispiel IV 1-7

Bestimmtheitsmaß und Gütevergleiche



In Beispiel IV 1-5 hatten wir für eine lineare und eine inverse Form die Regressionsgüte anhand der Residuenquadratesummen miteinander verglichen und festgestellt, dass die inverse Form den besseren "fit" aufweist. Führen wir diesen Vergleich nun nochmals anhand der Bestimmtheitsmaße durch. Zu ihrer Bestimmung können wir direkt die bereits berechneten Residuenquadratesummen und zusätzlich TSS = 0,6682 heranziehen. Wir erhalten damit folgende Werte:

$$R_{\text{ln}}^2 = 1 - \frac{0,2464}{0,6682} = 0,6313 \quad R_{\text{inv}}^2 = 1 - \frac{0,2046}{0,6682} = 0,6938$$

Die inverse Form ist also in der Lage 69,38 % der Variation von Y zu erklären, während die lineare nur 63,13 % erklären kann.

Einen Vergleich wie in Beispiel IV 1-7 dürfen wir grundsätzlich nur durchführen, wenn beide Funktionen die *gleiche erklärte Variable* besitzen, d.h. auch Y in einer und ln Y in der anderen ist nicht zulässig. Außerdem sollte die *Zahl der erklärenden Variablen gleich* sein. Wollen wir Modelle mit gleicher erklärter Variable und unterschiedlicher Anzahl erklärender Variablen miteinander vergleichen, besitzt R^2 entscheidende Nachteile, die wir unter IV 1.2.4.3 näher betrachten und ein alternatives Güte- bzw. Vergleichsmaß vorstellen.

1.2.4.2 Einfacher Korrelationskoeffizient

Eng verwandt mit dem Bestimmtheitsmaß ist der *einfache Korrelationskoeffizient* r_{yx} , den wir aus der deskriptiven Statistik kennen. Er ist ein Maß für die Stärke und Richtung eines linearen Zusammenhangs zwischen zwei Variablen und ist im Falle einer Regressionsgleichung mit nur einer erklärenden Variablen als

$$r_{yx} = \frac{\sum (Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum (Y_i - \bar{Y})^2 \sum (X_i - \bar{X})^2}} \quad \text{mit } -1 \leq r_{yx} \leq +1$$

definiert. Je näher der Absolutbetrag von r_{yx} am Wert 1 liegt, desto stärker ist die Korrelation zwischen den betrachteten Variablen Y und X. Das Vorzeichen von r_{yx} gibt dabei die Richtung der Korrelation an.

Hat ein Modell *genau eine erklärende Variable*, entspricht das Bestimmtheitsmaß R^2 genau dem Quadrat des Korrelationskoeffizienten. Gerade daher sind die R^2 einer Regression $Y_i = \beta_0 + \beta_1 X_i$ und einer Regression $X_i = \alpha_0 + \alpha_1 Y_i$ auch immer identisch.

$$R^2 = (r_{yx})^2 \quad (\text{IV1: 23})$$

Im Falle eines Modells *mit z.B. zwei erklärenden Variablen* ergibt sich R^2 als

$$R^2 = (r_{yy})^2,$$

d.h. als Quadrat des Korrelationskoeffizienten zwischen Y und \hat{Y} .¹¹

1.2.4.3 Angepasstes Bestimmtheitsmaß

Ein ernstes Problem von R^2 liegt darin, dass das Einbeziehen weiterer erklärender Variablen in die Regression R^2 nie verringert. Vergleichen wir also zwei Gleichungen, die die *gleiche erklärte Variable* Y und die gleiche erklärende Variable X_1 besitzen, die zweite aber eine zusätzliche Variable X_2 einbezieht, so gilt immer

$$R^2_{\text{Gleichung 2}} \geq R^2_{\text{Gleichung 1}}.$$

Um dies zu verdeutlichen, betrachten wir noch einmal die formale Definition des Bestimmtheitsmaßes (IV1: 22a). Fügen wir eine weitere erklärende Variable in unsere Regressionsgleichung ein, so hat dies keinen Einfluss auf TSS, da die erklärte Variable unverändert bleibt. Da OLS sicherstellt, dass sich RSS durch eine zusätzliche Variable nicht erhöht, wird RSS entweder abnehmen oder gleich bleiben. Sollte sich nämlich RSS durch die zusätzliche Variable erhöhen, so wird OLS den geschätzten Koeffizienten der zusätzlichen Variable auf Null setzen, sodass das RSS vor Variablenhinzufügung beibehalten wird. Der Fall in dem der geschätzte Koeffizient der neuen Variable auf Null gesetzt wird, ist damit der einzige Fall bei dem R^2 durch Hinzufügung der neuen Variablen unverändert bleibt. In allen anderen Fällen wird R^2 mit einer neuen Variablen zunehmen.

Besonders Anfänger unter den Ökonometrikern neigen häufig dazu, das sog. "Game of Maximising R^2 " zu betreiben. Sie versuchen durch Hinzufügen weiterer Variablen in ihre Regressionsgleichungen ihr R^2 in die Höhe zu treiben und sie damit "besser" aussehen zu lassen. Daraus entstehen allerdings zwei Probleme. Erstens werden dadurch häufig theoretisch nicht taugliche erklärende Variablen aufgenommen¹² und zweitens wird dadurch die Schätzung weiterer Regressionskoeffizienten erforderlich, was zu einer Verringerung der **Freiheitsgrade** (Differenz zwischen der Anzahl vorliegender Beobachtungen n und der Anzahl der Regressionskoeffizienten $K + 1$ bzw. $DF = n - K - 1$; engl. degrees of freedom) führt. Lägen

¹¹ Beweise dieser Zusammenhänge finden sich z.B. in Gujarati, D.N. (2003), Kapitel 7.

¹² Die Folgen einer derartigen Vorgehensweise werden wir im Abschnitt IV 4.1.2 behandeln.

etwa für eine einfache Regression (1 + 1 Koeffizienten) genau $n = 15$ Beobachtungen vor, so würde durch eine zusätzliche Variable die Anzahl der Freiheitsgrade von 13 auf 12 sinken. *Je niedriger die Anzahl der Freiheitsgrade ist, desto geringer ist aber die Genauigkeit der Schätzungen.*

Zur Beurteilung, ob das Hinzufügen einer Variable die Fähigkeit einer geschätzten Regressionsgleichung zur Erklärung der Variablen Y verbessert, ist also R^2 nicht geeignet. Es empfiehlt sich vielmehr bei der Berechnung des Bestimmtheitsmaßes die Zahl der Freiheitsgrade zu berücksichtigen. Es entsteht somit das sog. **angepasste Bestimmtheitsmaß**

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n e_i^2 / (n - K - 1)}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n - 1)}. \quad (\text{IV1: 24a})$$

Betrachten wir den Bruch in (IV1: 24a) etwas näher: Der Nenner ist nichts anderes als der erwartungstreue Schätzer für die Varianz von Y. Seine Freiheitsgrade liegen bei $n - 1$, da durch die Schätzung des Erwartungswertes von Y durch \bar{Y} ein Freiheitsgrad verloren geht. Da zur Bestimmung der Residuen die Schätzung von $K + 1$ Regressionsparametern erforderlich ist, reduzieren sich die Freiheitsgrade im Zähler um $K + 1$, was insgesamt zu $n - K - 1$ führt. Ceteris paribus führt also ein höheres K zu einem niedrigeren \bar{R}^2 .

Wie sich zeigen lässt, kann das angepasste Bestimmtheitsmaß \bar{R}^2 wie in (IV1: 24b) mit dem Bestimmtheitsmaß R^2 in Verbindung gesetzt werden. Diese Darstellung zeigt, dass der größtmögliche Wert von \bar{R}^2 genau wie bei R^2 Eins ist. Der kleinstmögliche Wert ist jedoch nicht Null, da bei sehr geringem R^2 der Wert von \bar{R}^2 sogar negativ sein kann.

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - K - 1} \quad (\text{IV1: 24b})$$

Zusammenfassend misst das angepasste Bestimmtheitsmaß also unter Berücksichtigung der Anzahl der Feiheitsgrade den Teil der Varianz von Y, der durch die Schätzung erklärt wird. Es wird dazu verwendet, den "fit" von Gleichungen mit der gleichen erklären Variablen und unterschiedlicher Anzahl von erklärenden Variablen zu vergleichen. \bar{R}^2 wird bei Hinzufügen einer weiteren Variablen abnehmen oder zunehmen, je nachdem, ob der verbesserte "fit" durch die neue Variable den mit ihr verbundenen Verlust an Freiheitsgraden überwiegt oder nicht. Im "*Game of Maximising R^2* " ist es also sehr wahrscheinlich, dass durch eine neue Variable zwar R^2 steigt, \bar{R}^2 jedoch abnimmt, was uns zeigt, dass wir die Variable nie in die Gleichung hätten aufnehmen sollen.

Abschließend muss noch die Warnung ausgesprochen werden, dass anders als im deskriptiven *im induktiven Kontext* die Qualität des "fits" einer Schätzung nur ein Teilkriterium für die Gesamtqualität einer geschätzten Regressionsgleichung ist. Eine \bar{R}^2 -Maximierung sollte daher bei beabsichtigten Rückschlüssen auf die Grundgesamtheit oder Prognosen nie Ziel des Ökonometrikers sein. Was nützt beispielsweise ein hohes \bar{R}^2 , wenn ein Regressionskoeffizient ein für die zugrunde liegende

Modelltheorie absolut nicht plausibles Vorzeichen aufweist. Trotz gutem "fit" kann eine solche Gleichung nicht für Prognosen verwendet werden.

Beispiel IV 1-8**Grenzen des angepassten Bestimmtheitsmaßes**

Nehmen wir an, uns liegen zwei geschätzte Regressionsgleichungen für den Wasserkonsum (W) eines Bundeslandes in Abhängigkeit vom Wasserpreis je 5 Liter (P), seiner Bevölkerungszahl (B) und der Niederschlagsmenge (N) in Litern pro m^2 pro Jahr vor:

$$(1) \quad \hat{W} = 29,00 + 0,60B - 420,00N \quad \text{mit} \quad \bar{R}^2 = 0,855$$

$$(2) \quad \hat{W} = 24,50 + 47,00P + 0,45B - 360,00N \quad \text{mit} \quad \bar{R}^2 = 0,895$$

Bei Betrachtung dieser Gleichungen fällt sofort auf, dass durch das Hinzufügen der Variable P eine Erhöhung des angepassten Bestimmtheitsmaßes erreicht wurde. Gleichzeitig erkennen wir aber auch, dass der Koeffizient der neuen Variable P positiv ist, was bedeutet, dass bei Konstanthaltung aller anderen Variablen die Erhöhung des Wasserpreises unrealistisch zu einer Erhöhung des Wasserkonsums führt. Die geschätzte Gleichung (2) würde also bei Prognosen die Wassernachfrage tendenziell überschätzen, wenn das betrachtete Szenario von steigenden Preisen ausgeht (bzw. im umgekehrten Fall unterschätzen). Die Gleichung mit besserem "fit" hätte also eine schlechtere Prognosegüte.

2.

Das lineare Regressionsmodell und seine Annahmen

Bisher haben wir OLS nur als rein deskriptives Instrument zur Charakterisierung angenommener funktionaler Formen in Stichprobenzahlenmaterial herangezogen. In der Praxis wird OLS jedoch vor allem im Rahmen des sog. linearen Regressionsmodells als Schätzverfahren eingesetzt. Ziel ist es hier nicht mehr einen vorliegenden Datenbestand bestmöglich beschreiben, sondern anhand von diesem Rückschlüsse auf eine Grundgesamtheit ziehen zu können. Mit den Grundlagen dieses Modells werden wir uns unter IV 2.1 näher befassen. Sind die Annahmen des Abschnitts IV 2.2 für ein solches Modell erfüllt, so verfügen OLS-Schätzer über eine Reihe wertvoller Eigenschaften, die OLS zum "bestmöglichen" Schätzer im linearen Regressionsmodell machen. Das sog. Gauß-Markov-Theorem zeigt in einem solchen Fall, dass OLS jedem anderen linearen unverzerrten Schätzer "überlegen" ist.

2.1 Das lineare Regressionsmodell

2.1.1 Die Regressionsfunktion der Grundgesamtheit

Formal befasst sich die Regressionsanalyse mit der Untersuchung der Abhängigkeit einer Variablen (Y) von einer oder mehreren anderen Variablen (X_1, \dots, X_k) mit dem Ziel, den *Erwartungswert* der erklären Variablen *auf Basis* der bekannten Werte der erklärenden Variablen, also $E(Y|X_1, \dots, X_k)$ zu schätzen und/oder zu prognostizieren. Diese Kernaussage ist von so enormer Bedeutung, dass wir sie und alle weiteren Sachverhalte im Zusammenhang mit dem linearen Regressionsmodell zum besseren Verständnis anhand eines hypothetischen Beispiels erläutern werden.¹³

Wir gehen zunächst davon aus, dass wir die Grundgesamtheit kennen. Betrachten wir eine *Grundgesamtheit* aus $N = 60$ Familien, die nach ihrem wöchentlichen Einkommen (X) in 10 Gruppen (80 bis 260 Euro) eingeteilt und deren wöchentliche Konsumausgaben (Y , ebenfalls in Euro) in der Tabelle auf der Folgeseite gegeben sind. In den 10 Teilgesamtheiten, die durch die 10 X -Werte mit korrespondierenden Y -Werten entstehen, ist eine deutliche Variation der wöchentlichen Konsumausgaben zu erkennen. Trotz dieser Variabilität gewinnen wir aber den Eindruck, dass die wöchentlichen Konsumausgaben *im Durchschnitt* mit dem Einkommen zunehmen. Um dies zu verdeutlichen, haben wir sowohl in der Tabelle, als auch in der nachfolgenden Abbildung IV 5 die bedingten Erwartungswerte $E(Y|X)$ oder



¹³ Das hier verwendete Beispiel ist aus dem Lehrbuch von Gujarati, D.N. (2003), Kapitel 2, entnommen.

anders ausgedrückt die Mittelwerte der Teilgesamtheiten aufgenommen. Wir verwenden hier bewusst das Subskript i , da sich der Erwartungswert auf eine bestimmte Ausprägung der Variable X bezieht. Der erwartete Wert der Konsumausgaben einer Familie mit einem konkreten Einkommen X_i von 80,00 Euro ist demnach 65,00 Euro. Stellen wir also die Frage nach der besten (mittleren) Prognose der wöchentlichen Konsumausgaben einer Familie mit einem Einkommen von 80,00 Euro, so wäre die Antwort 65,00 Euro. Es ist wichtig, diese von den gegebenen X -Werten abhängenden Erwartungswerte vom unbedingten Erwartungswert $E(Y)$ abzugrenzen. Dieser repräsentiert nämlich die erwarteten Konsumausgaben aller 60 Familien und liegt hier bei 121,20 Euro.

$X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Y	55	65	79	80	102	110	120	135	137	150
\downarrow	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	-	88	-	113	125	140	-	160	189	185
	-	-	-	115	-	-	-	162	-	191
$E(Y X_i)$	65	77	89	101	113	125	137	149	161	173

Werfen wir nun einen genaueren Blick auf die grafische Darstellung dieses Zahlenmaterials. In dieser sind neben den einzelnen X - Y -Kombinationen auch die bedingten Erwartungswerte (hervorgehobene Punkte, die auf einer Geraden liegen) aufgenommen.

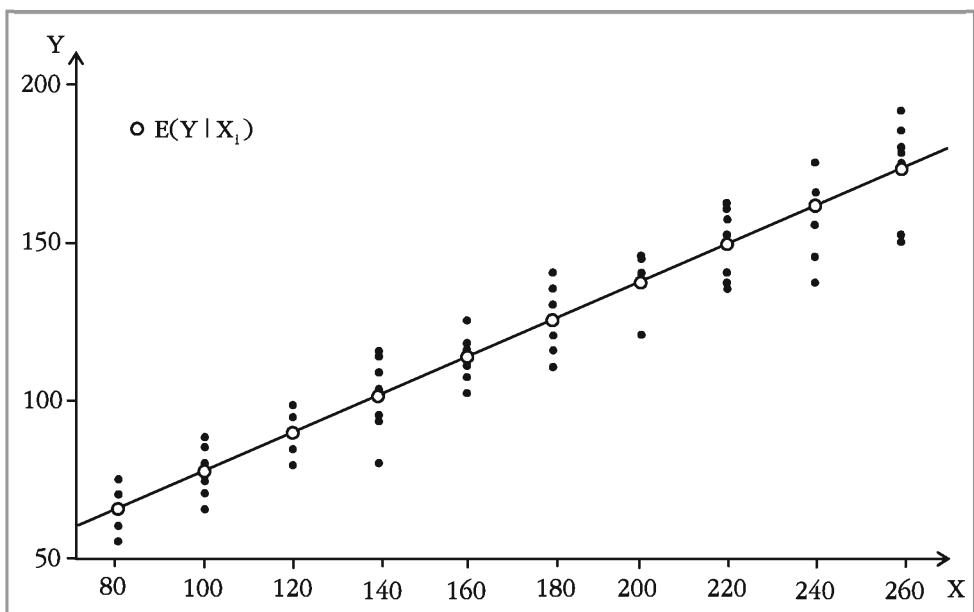


Abbildung IV 5: Regressionsfunktion der Grundgesamtheit I

Wir erkennen deutlich, dass sich die bedingten Erwartungswerte $E(Y|X_i)$ als eine Funktion von X_i ergeben, also

$$E(Y|X_i) = f(X_i) \quad (\text{IV2: 1})$$

gilt. Das Modell gilt also für jedes beliebige X_i aus der Grundgesamtheit. Da die Funktion (IV2: 1) die Grundgesamtheit charakterisiert, wird sie auch als **Regressionsfunktion der Grundgesamtheit** bezeichnet. Sie ist in diesem speziellen Fall eine Gerade durch die Punkte mit den Koordinaten $(X_i, E(Y|X_i))$ und lässt sich daher in der funktionalen Form

$$E(Y|X_i) = \beta_0 + \beta_1 X_i \quad (\text{IV2: 2})$$

mit β_0 und β_1 als den **Regressionskoeffizienten** oder **-parametern** darstellen. Die Regressionsfunktion der Grundgesamtheit gibt die *durchschnittliche* Reaktion von Y auf Veränderungen von X an. In unserem Beispiel sagt sie daher aus, dass bei steigendem Familieneinkommen *im Durchschnitt* auch die Konsumausgaben zunehmen. Können wir nun aber auch die generelle Aussage machen, dass mit zunehmendem Einkommen die Konsumausgaben jeder *einzelnen* Familie zunehmen? Betrachten wir die obige Tabelle, so wird deutlich, dass wir dies nicht können. Es gibt nämlich beispielsweise eine Familie, die bei einem wöchentlichen Einkommen von 100 Euro mit Konsumausgaben von 65 Euro weniger konsumiert als zwei Familien mit einem wöchentlichen Einkommen von 80 Euro. Es scheint also gewisse Faktoren zu geben, die dafür sorgen, dass bei einem gegebenen Einkommenslevel X_i die Konsumausgaben Y_i der individuellen Familien vom durchschnittlichen Konsum aller Familien mit diesem Einkommen abweicht. Diese Faktoren spiegeln sich im sog. **stochastischen Störterm** ϵ_i wieder. Er verkörpert die Abweichungen der individuellen Y_i von den zugehörigen bedingten Erwartungswerten $E(Y|X_i)$, also

$$\epsilon_i = Y_i - E(Y|X_i), \quad (\text{IV2: 3})$$

sodass sich Y_i durch eine Gleichung der Form

$$Y_i = E(Y|X_i) + \epsilon_i \quad (\text{IV2: 4})$$

beschreiben lässt. Für den Fall eines variablen- und parameterlinearen Modells mit nur einer erklärenden Variablen, wie es hier vorliegt, erhalten wir damit das sog. **einfache lineare Regressionsmodell**

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i. \quad (\text{IV2: 5})$$

Dieses besagt in unserem speziellen Beispiel, dass der Konsum einer individuellen Familie bei gegebenem Einkommenslevel durch zwei Komponenten beschrieben wird. Komponente 1 ist $E(Y|X_i)$, was nichts anderes als den durchschnittlichen Konsum aller Familien mit dem gleichen Einkommen darstellt. Wir bezeichnen diese Komponente auch als *systematischen* Teil des Regressionsmodells. Komponente 2 ist der *unsystematische* oder *zufällige* Teil ϵ_i des Regressionsmodells.

Eine weitere Betrachtung unserer Abbildung IV 5 zeigt, dass für jedes X_i eine Teilgesamtheit von Y -Werten existiert, die um ihren zugehörigen (bedingten) Erwartungswert streuen. Nehmen wir zur Vereinfachung an, dass diese Y -Werte in jeder Teilgesamtheit identisch und symmetrisch um ihren bedingten Erwartungswert verteilt sind, so finden wir in der Literatur für unser Beispiel auch häufig die folgende

Darstellung der Regressionsfunktion der Grundgesamtheit. Wir beschränken uns in Abbildung IV 6 auf die X-Werte 80, 140 und 220. Für alle anderen X-Werte gilt die Darstellung natürlich analog.

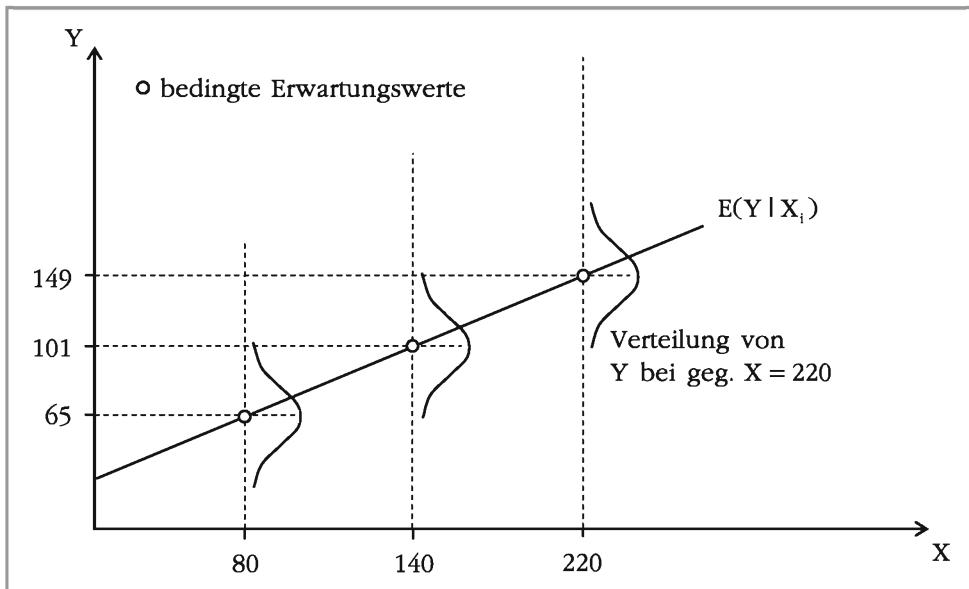


Abbildung IV 6: Regressionsfunktion der Grundgesamtheit II

Würden wir aus der hier dargestellten Grundgesamtheit in einer Stichprobe für jedes X_i ein zugehöriges Y_i ziehen, so wird klar, dass sich in einer neuen Stichprobe wohl andere Y_i ergeben. Wir haben in diesem Fall 10 verschiedene Ausprägungen von X vorliegen. Dies bedeutet, dass wir damit auch 10 Zufallsvariablen Y_i haben, deren Realisationen sich von Stichprobe zu Stichprobe ändern. Damit nehmen natürlich auch die stochastischen Störterme ϵ_i in jeder Stichprobe andere Werte an, da sie sich als Abweichung der Y_i von $E(Y|X_i)$ ergeben. Sie sind ebenfalls Zufallsvariablen. Veranschaulichen können wir dies sehr schön, indem wir die Teilgesamtheit von Y betrachten, die durch $X = 80$ bestimmt wird. Innerhalb dieser gelten nämlich aufgrund von (IV2: 5) für die einzelnen Y_i die folgenden Beziehungen:

$$Y_1 = 55 = \beta_0 + \beta_1 \cdot 80 + \epsilon_1$$

$$Y_2 = 60 = \beta_0 + \beta_1 \cdot 80 + \epsilon_2$$

$$Y_3 = 65 = \beta_0 + \beta_1 \cdot 80 + \epsilon_3$$

$$Y_4 = 70 = \beta_0 + \beta_1 \cdot 80 + \epsilon_4$$

$$Y_5 = 75 = \beta_0 + \beta_1 \cdot 80 + \epsilon_5$$

Wird also ein Wert für X vorgegeben (determiniert), so kann der Störterm einen zufälligen Wert annehmen. In diesem Beispiel gäbe es 5 mögliche Realisationen für den Störterm. Es ist nun von außerordentlicher Bedeutung für das Verständnis der noch folgenden Ausführungen (insbesondere unter IV 2.2), dass bei einem Stichprobenumfang von n (also n Werten von X, für die Y_i gezogen werden) genau n Zufallsvariablen ϵ_i vorliegen, die gewissen Verteilungen folgen und in der Stichpro-

be zufällige Realisationen annehmen. Wenn wir also von der Verteilung der ϵ_i sprechen, meinen wir dabei *nicht* die Verteilung, die die n Realisationen bildet, sondern die Verteilungen der einzelnen Zufallsvariablen ϵ_i für die Teilgesamtheiten von Y .

Da wir in der Praxis die Grundgesamtheit und damit die Parameter der Regressionsfunktion der Grundgesamtheit nicht kennen, ist der Term ϵ_i nicht nur eine Zufallsvariable, sondern kann zudem nicht beobachtet werden. Wir wissen jedoch, dass er sowohl positive, als auch negative Werte annehmen kann und er alle Veränderungen der erklärenen Variablen Y auffängt, die nicht durch die Variable X erklärt werden können. Dazu zählen z.B. jene Veränderungen, die auf vernachlässigte Variablen zurückzuführen sind. Selbst wenn diese ins Modell einbezogen werden, verbleiben jedoch immer noch Veränderungen von Y , die nicht durch die Variablen allein erkläbar sind. Allgemein lassen sich folgende Faktoren zusammenfassen, die sich im Störterm widerspiegeln:

- Es werden *erklärende Variablen vernachlässigt*, da z.B. das zugrunde liegende theoretische Modell nicht ausgereift ist oder bzgl. gewisser Variablen kein Datenmaterial zur Verfügung steht.
- Die Vermeidung von *Messfehlern* bei der Erfassung des Datenmaterials ist fast unmöglich.
- Die zu Grunde liegende theoretische Modellgleichung hat im Bezug auf die Variablen eine andere Form (z.B. quadratisch) als die für die Regression gewählte (z.B. linear).
- Bei der Modellierung von menschlichem Verhalten muss es immer einen Teil *rein zufälliger Veränderung* geben, den man nicht durch irgendwelche Variablen erklären kann.

Eine wichtige Eigenschaft des stochastischen Störterms ergibt sich aus der Annahme, dass die Regressionskurve der Grundgesamtheit durch die bedingten Erwartungswerte von Y verläuft. Dies impliziert, dass der bedingte Erwartungswert des stochastischen Störterms gleich Null ist:

$$E(\epsilon_i | X_i) = 0 \quad (\text{IV2: 6})$$

Berechnen wir nämlich den bedingten Erwartungswert von (IV2: 5), so erhalten wir zunächst¹⁴

$$E(Y_i | X_i) = \beta_0 + \beta_1 X_i + E(\epsilon_i | X_i). \quad (\text{IV2: 7})$$

Verläuft nun die Regressionsgerade durch die bedingten Erwartungswerte von Y , so gilt $E(Y_i | X_i) = \beta_0 + \beta_1 X_i$, was zu

$$\beta_0 + \beta_1 X_i = \beta_0 + \beta_1 X_i + E(\epsilon_i | X_i) \rightarrow E(\epsilon_i | X_i) = 0 \quad (\text{IV2: 8})$$

führt. Gilt also (IV2: 6), so sind (IV2: 2) und (IV2: 5) äquivalent. Die stochastische Spezifikation (IV2: 5) der Regressionsfunktion der Grundgesamtheit hat jedoch den Vorteil, dass neben den erklärenden Variablen im Modell auch der Einfluss anderer Faktoren erkennbar wird.

¹⁴ Es gilt generell $E(f(X) | X) = f(X)$. Da wir im Rahmen unseres Einführungsbeispiels die X_i als gegeben betrachten, gilt $E(\beta_1 X_i | X_i) = \beta_1 X_i$.

Die Gültigkeit von (IV2: 6) können wir auch in unserem hypothetischen Beispiel aufzeigen, da wir hier den Vorteil genießen, die Grundgesamtheit zu kennen, was natürlich in der Praxis nicht der Fall sein wird. Hier sind wir aber in der Lage die Parameter β_0 und β_1 durch Bestimmung einer Gleichung unter Verwendung zweier beliebiger Wertepaare $(X_i, E(Y|X_i))$ über das Gleichungssystem

$$(1) \quad 65 = \beta_0 + \beta_1 \cdot 80$$

$$(2) \quad 77 = \beta_0 + \beta_1 \cdot 100$$

zu ermitteln. Wir erhalten damit $\beta_0 = 17$ und $\beta_1 = 0,6$, was uns die folgenden möglichen Ausprägungen des Störterms bei z.B. $X = 80$ liefert:

$$\varepsilon_1 = Y_1 - (\beta_0 + \beta_1 \cdot 80) = 55 - (17 + 0,6 \cdot 80) = -10$$

$$\varepsilon_2 = Y_2 - (\beta_0 + \beta_1 \cdot 80) = 60 - (17 + 0,6 \cdot 80) = -5$$

$$\varepsilon_3 = Y_3 - (\beta_0 + \beta_1 \cdot 80) = 65 - (17 + 0,6 \cdot 80) = 0$$

$$\varepsilon_4 = Y_4 - (\beta_0 + \beta_1 \cdot 80) = 70 - (17 + 0,6 \cdot 80) = 5$$

$$\varepsilon_5 = Y_5 - (\beta_0 + \beta_1 \cdot 80) = 75 - (17 + 0,6 \cdot 80) = 10$$

Wie zu erkennen ist, kann der Störterm sowohl negative, als auch positive Realisationen annehmen, die sich jedoch innerhalb der Teilgesamtheit gegenseitig aufheben. Ihr Mittelwert liegt daher bei Null, sodass $E(\varepsilon_i | X=80) = 0$ gilt.

Zum Abschluss dieses Unterabschnittes wollen wir noch kurz darauf hinweisen, dass im Falle eines Regressionsmodells mit mehreren erklärenden Variablen, dem sog. **multiplen Regressionsmodell**, unsere Ausführungen analog gelten. Aus Gründen der Übersichtlichkeit verwenden wir hier für die Regressionsfunktion der Grundgesamtheit jedoch eine etwas andere Notation:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_K X_{iK} + \varepsilon_i \quad (\text{IV2: 9})$$

Hierbei ist X_{i1} die i-te Beobachtung der ersten erklärenden Variablen und X_{i2} die i-te Beobachtung der zweiten erklärenden Variablen und K die Anzahl der erklärenden Variablen. Y wird also hier von mehreren erklärenden Variablen X_k mit $k = 1, 2, \dots, K$ bestimmt.

2.1.2 Die Regressionsfunktion der Stichprobe

Da in der Regel nicht alle Daten der Grundgesamtheit bekannt sind, können die Parameter der Regressionsfunktion der Grundgesamtheit nicht exakt bestimmt werden. Nachdem die Gleichung des Regressionsmodells aufgestellt wurde (Wahl der funktionalen Form und der erklärenden Variablen) ist daher die Erhebung einer Stichprobe erforderlich, mit deren Hilfe die unbekannten Parameter schließlich geschätzt werden können. Die quantifizierte Version der Regressionsfunktion wird als **geschätzte Regressionsfunktion** (Regressionsfunktion der Stichprobe oder einfache geschätzte Regression) bezeichnet und hat im Fall eines linearen Modells mit nur einer erklärenden Variablen (analog bei mehreren Variablen) die Form

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i . \quad (\text{IV2: 10})$$

Gleichung (IV2: 10) stellt damit das Stichprobengegenstück zu (IV2: 2) dar. Sie ist uns bereits als (IV1: 3) aus unserem Abschnitt zu OLS bekannt. Liegt uns also

Stichprobenzahlenmaterial für X und Y vor, so können wir die unbekannten Parameter der Regressionsfunktion mittels OLS schätzen. $\hat{\beta}_0$ ist die OLS-Schätzung für β_0 und $\hat{\beta}_1$ die OLS-Schätzung für β_1 . Mittels dieser kann dann \hat{Y}_i , die Schätzung für $E(Y|X_i)$, bestimmt werden. Da wir die Parameter mit OLS schätzen, ist natürlich bei der Aufstellung des Modells bzw. der Wahl der funktionalen Form der Regressionsfunktion darauf zu achten, dass diese parameterlinear ist. Andernfalls können wir die Parameter nicht mit OLS schätzen.

Genau wie wir die Regressionsfunktion der Grundgesamtheit in zwei äquivalenten Formen (IV2: 2) und (IV2: 5) darstellen konnten, können wir dies auch für die geschätzte Regressionsfunktion (IV2: 10) tun:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \quad (\text{IV2: 11})$$

Der Term e_i ist darin unser bereits bekanntes Residuum. Es ist nach (IV1: 4) die Differenz zwischen Y_i und \hat{Y}_i , kann als Schätzung von ϵ_i betrachtet werden und wird aus den selben Gründen in die geschätzte Regressionsfunktion aufgenommen, aus denen ϵ_i in die Regressionsfunktion der Grundgesamtheit integriert wurde. Während der stochastische Störterm als $\epsilon_i = Y_i - E(Y_i|X_i)$ ein rein theoretisches Konzept und nicht empirisch beobachtbar war, können Residuen in einer Stichprobe als konkrete Zahlen bestimmt werden.

Zusammenfassend können wir sagen, dass es das primäre Ziel der einfachen linearen Regression (analog auch der multiplen Regression) ist, die Regressionsfunktion

$$Y_i = E(Y_i|X_i) + \epsilon_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

mit

$$Y_i = \hat{Y}_i + e_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$$

zu schätzen. Abbildung IV 7 zeigt die Regressionsfunktion der Grundgesamtheit und ihre Schätzung zur besseren Veranschaulichung graphisch. Darin sind die empirisch beobachteten Punkte (X_a, Y_a) aus einer Stichprobe, sowie die Graphen der wahren und der geschätzten Regressionsfunktion abgetragen. Es ist zu erkennen, dass die geschätzte Gerade nahe bei der wahren liegt, jedoch nicht mit ihr identisch ist. \hat{Y}_a , der geschätzte Wert von $E(Y|X_a)$ für die a-te Beobachtung von X, liegt auf der geschätzten Regressionsgeraden und unterscheidet sich von Y_a , dem tatsächlich beobachteten Wert von Y für die a-te Beobachtung. Die Differenz zwischen beiden ist das Residuum und wurde mit e_a bezeichnet. Obwohl der Störterm nicht beobachtet werden kann, haben wir diesen für die a-te Beobachtung von X unter Zuhilfenahme der (eigentlich auch nicht bekannten) wahren Regressionsgeraden als ϵ_a ebenfalls abgetragen.

Es ist deutlich zu erkennen, dass \hat{Y}_a den wahren Wert $E(Y|X_a)$ überschätzt. Dies gilt für jedes X_i rechts vom Punkt X_b . Links davon kommt es zu einer Unterschätzung. Derartige Über- und Unterschätzungen sind jedoch stichprobenspezifisch. Die Lage der geschätzten Regressionsfunktion ist von Stichprobe zu Stichprobe verschieden, da die geschätzten Regressionsparameter auf rein zufällig entstehendem Stichprobenzahlenmaterial beruhen. Obwohl wir die Lage der Regressionsfunktion der Grundgesamtheit nie kennen werden (außer bei bekannter Grundgesamtheit), werden wir uns in den folgenden Abschnitten damit befassen, wie wir die Regressionsfunktion so schätzen können, dass die Schätzung die Regressionsfunktion der Grundgesamtheit bestmöglich abbildet. Ein im Grunde faszinierendes Unterfangen!

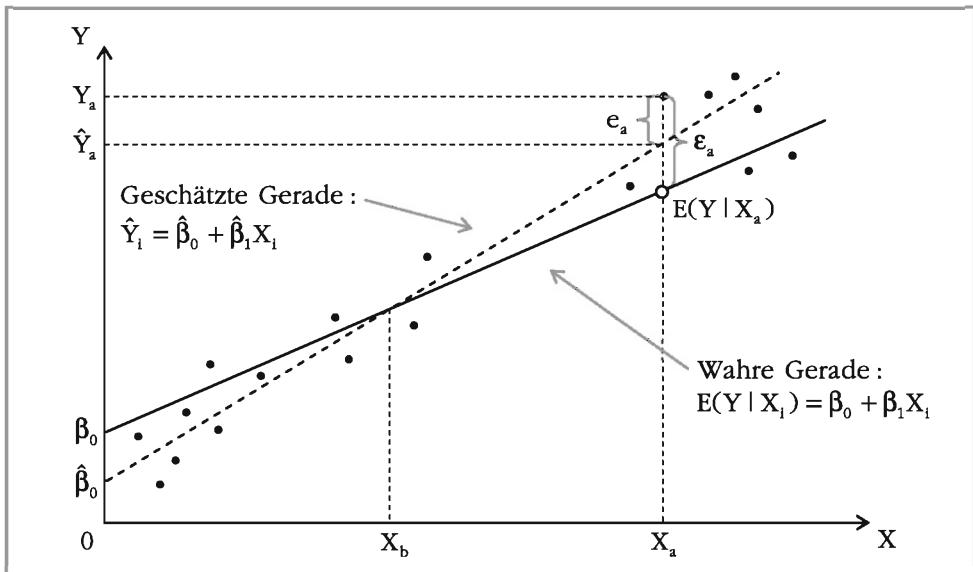


Abbildung IV 7: Wahre und geschätzte Regressionsgerade I

Zur besseren Veranschaulichung dieser Überlegungen greifen wir wieder auf unser Konsumbeispiel zurück. Da wir hier die Grundgesamtheit kennen (in der Praxis natürlich unrealistisch), konnten wir bereits zeigen, dass $E(Y|X_i) = 17 + 0,6X_i$ gilt. Ziehen wir nun aus der Grundgesamtheit eine zufällige Stichprobe, so könnten wir (bei vorgegebenen X_i) z.B. das in Abbildung IV zu findende Ergebnis erhalten. Verwenden wir diese Wertepaare zur OLS-Schätzung der Parameter β_0 und β_1 , so erhalten wir die Schätzung $\hat{Y}_i = 15,9818 + 0,5739X_i$, die sehr nahe an der Regressionsfunktion der Grundgesamtheit liegt.

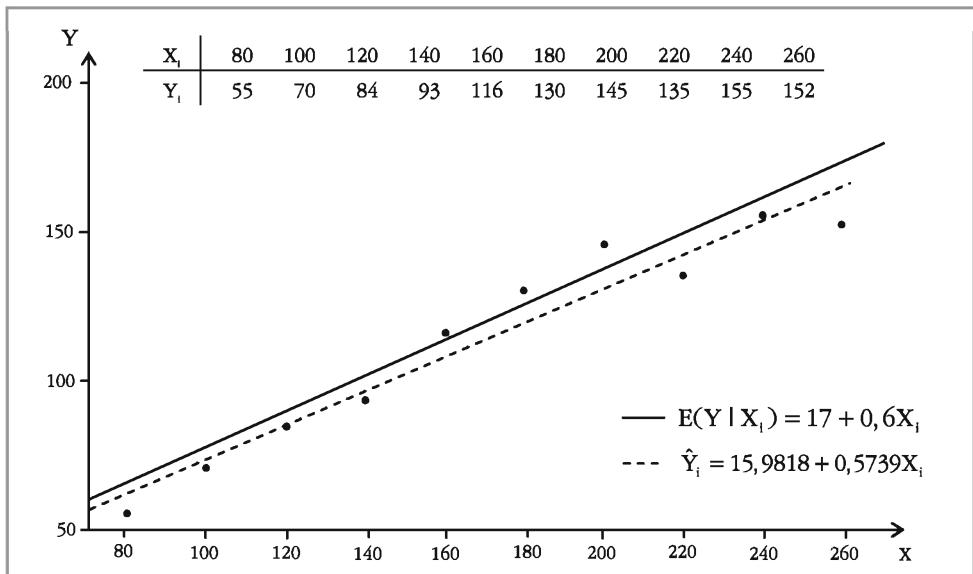


Abbildung IV 8: Wahre und geschätzte Regressionsgerade II

2.2 Klassische Annahmen

2.2.1 Annahmenkatalog

Das sog. **klassische lineare Regressionsmodell** (CLRM, engl. classical linear regression model) ist durch die nachfolgenden 6 Annahmen gekennzeichnet. Werden diese eingehalten, so liefert OLS nach dem sog. Gauß-Markov-Theorem (vgl. IV 2.3.2) die "bestmöglichen" Schätzer für die Koeffizienten des Modells. Aufgrund ihrer enormen Bedeutung, werden wir diese 6 Annahmen zum besseren Verständnis meist am einfachen Regressionsmodell veranschaulichen. Sie gelten natürlich analog auch für das multiple Modell.

Annahme 1: Das Regressionsmodell ist parameterlinear.

Wie bereits unter IV 1.2.2 besprochen wurde, erfordert die Anwendung der linearen Schätztechnik OLS die *Parameterlinearität der Regressionsfunktion der Grundgesamtheit*. Sie sollte also eine Form

$$f(Y_i) = \beta_0 + \beta_1 f(X_{1i}) + \beta_2 f(X_{2i}) + \dots + \beta_k f(X_{ki}) + \epsilon_i \quad (\text{IV2: 12})$$

aufweisen.¹⁵ Dies zeigt noch einmal, dass Variablenlinearität nicht zwingend erforderlich ist. Wurde eine nicht parameterlineare Form gewählt, so ist diese zu transformieren. So kann z.B. eine Exponentialfunktion

$$Y_i = e^{\beta_0} X_i^{\beta_1} e^{\epsilon_i}$$

mit e als Eulerscher Zahl durch logarithmische Transformation

$$\ln Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$$

in eine parameterlineare Form gebracht werden, die durch OLS schätzbar ist.

Annahme 2a: Der Erwartungswert des stochastischen Störterms bei gegebenen Werten der erklärenden Variablen ist gleich Null:

$$E(\epsilon_i | X_i) = 0 \quad (\text{IV2: 13})$$

Zum besseren Verständnis dieser zentralen Annahme des CRLM betrachten wir Abbildung IV 9. Diese zeigt für das einfache lineare Regressionsmodell für einige X_i die zugehörigen Y-Teilgesamtheiten. Jede Teilgesamtheit streut um den dazugehörigen bedingten Erwartungswert, wobei alle diese bedingten Erwartungswerte auf der Regressionsfunktion der Grundgesamtheit liegen. Innerhalb jeder Teilgesamtheit von Y liegen einige Werte von Y über, andere unter $E(Y|X_i)$. In Abbildung IV 9 sind zur Veranschaulichung zusätzlich die (als symmetrisch und gleich angenommenen) bedingten Verteilungen der ϵ_i skizziert. Sie haben ihre Mitte genau bei den bedingten Erwartungswerten $E(Y|X_i)$, bei denen der Störterm gerade Null ist.

¹⁵ Diese Annahme schließt also auch das Vorhandensein eines *additiven* Störterms ein.

Die Annahme 2a besagt nun einfach ausgedrückt, dass der stochastische Störterm innerhalb jeder Teilgesamtheit von Y einen Erwartungswert von Null haben soll. Positive und negative ϵ_i sollen sich für jedes gegebene X_i gegenseitig aufheben. Dies bedeutet nichts anderes, als dass alle Faktoren, die nicht explizit ins Modell aufgenommen wurden und sich daher in ϵ_i widerspiegeln, den bedingten Erwartungswert von ϵ_i nicht systematisch beeinflussen. Annahme 2a unterstellt damit, dass das Modell (IV2: 12) **korrekt spezifiziert** ist. Dies bedeutet, dass zum einen *keine bedeutenden erklärenden Variablen vernachlässigt* werden und zum anderen die *korrekte funktionale Form* für die Regressionsfunktion gewählt wurde.

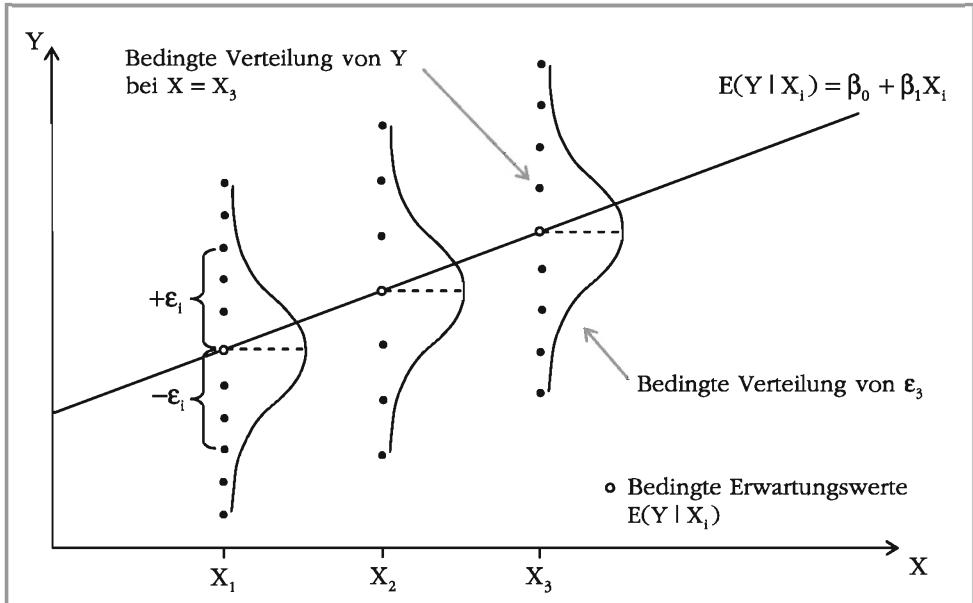


Abbildung IV 9: Bedingte Verteilung der ϵ_i

Falls wir nicht die korrekte funktionale Form gewählt (oder Variablen vernachlässigt) haben, gilt $E(\epsilon_i | X) \neq 0$. Betrachten wir dazu die Ausgangssituation

$$E(Y_i | X_i) = \beta_0 + \beta_1 X_i + E(\epsilon_i | X_i),$$

d.h. wir haben eine lineare funktionale Form zur Erklärung des bedingten Erwartungswertes von Y gewählt. Damit der Term $E(\epsilon_i | X)$ Null wird, muss sich die gewählte funktionale Form $\beta_0 + \beta_1 X_i$ mit der in der Grundgesamtheit geltenden Form decken, d.h. es muss hier $E(Y_i | X_i) = \beta_0 + \beta_1 X_i$ sein. Gilt nun in der Grundgesamtheit z.B. $E(Y_i | X_i) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2$, wählen wir jedoch eine lineare Form, so gilt

$$\beta_0 + \beta_1 X_i + \beta_2 X_i^2 = \beta_0 + \beta_1 X_i + E(\epsilon_i | X_i) \rightarrow E(\epsilon_i | X_i) = \beta_2 X_i^2 \neq 0.$$

Da es im Falle von $E(\epsilon_i | X) \neq 0$ zu einer verzerrten Schätzung der Regressionsparameter kommt, ist es von entscheidender Wichtigkeit die funktionale Form korrekt zu wählen. Welche funktionale Form jedoch konkret verwendet werden sollte, ist bei unbekannter Grundgesamtheit nicht gerade einfach zu sagen. Wichtige Anhaltspunkte kommen hier meist aus dem theoretischen Hintergrund, der zur Beschrei-

bung der Situation in der Grundgesamtheit herangezogen wird. *Keinesfalls* sollte die funktionale Form *allein anhand des "fits" in der Stichprobe* ausgewählt werden. Aus Annahme 2a folgt direkt, dass auch der *unbedingte Erwartungswert* des stochastischen Störterms einen Wert von Null annimmt:

$$E(\epsilon_i) = 0 \quad (\text{IV2: 14})$$

Die bedeutendste Folge von Annahme 2a ist die Unkorreliertheit des stochastischen Störterms und der erklärenden Variablen. Diese wollen wir im Folgenden als Annahme 2b aufnehmen, auch wenn sie mit 2a automatisch erfüllt ist. Wir wählen diese Vorgehensweise bewusst, da für den Fall, dass Annahme 2b verletzt wird, auch die Annahme 2a verletzt wird und es zur verzerrten und inkonsistenten Schätzung der Regressionskoeffizienten kommt. Zudem ist Annahme 2a nicht direkt mit statistischen Tests überprüfbar.

Annahme 2b: Der stochastische Störterm ist nicht mit der erklärenden Variable korreliert.

$$\text{Cov}(\epsilon_i, X_i) = 0 \quad (\text{IV2: 15})$$

Wenn wir ein Regressionsmodell aufstellen, nehmen wir an, dass die erklärende Variable und der stochastische Störterm einen *separaten Einfluss* auf die erklärtete Variable haben. Wären die erklärende Variable und der Störterm korreliert, so wäre es nicht möglich ihre individuellen Einflüsse auf Y zu untersuchen bzw. zu isolieren. OLS würde Veränderungen von Y, die auf den Störterm zurückzuführen sind, der erklärenden Variablen zuordnen. Läge z.B. eine positive Korrelation zwischen X und ϵ vor, so würde der geschätzte Koeffizient von X anders ausfallen als im Fall ohne Korrelation, da OLS die Variation von Y fälschlicherweise allein auf X und nicht auch auf ϵ zurückführt. Nur bei fehlender Korrelation besteht eine derartige Verzerrung nicht.

Ist $E(\epsilon_i | X_i) = 0$, so sind der stochastische Störterm und die erklärende Variable unkorreliert.¹⁶ Umgekehrt gilt dies jedoch *nicht*. Es gilt allgemein

$$\begin{aligned} E(\epsilon_i | X_i) = 0 &\rightarrow \text{Cov}(\epsilon_i, X_i) = 0 \\ \text{Cov}(\epsilon_i, X_i) \neq 0 &\rightarrow E(\epsilon_i | X_i) \neq 0 \end{aligned} \quad (\text{IV2: 16})$$

Aus der Annahme $\text{Cov}(\epsilon_i, X_i) = 0$ folgt *nicht* $E(\epsilon_i | X_i) = 0$, da die Kovarianz nur den linearen Zusammenhang zwischen zwei Variablen misst. Damit $E(\epsilon_i | X_i) = 0$ ist, muss auch die Kovarianz zwischen ϵ_i und jeder deterministischen Funktion von X_i , beispielsweise X_i^2 , Null sein. Es lässt sich aber im Falle $\text{Cov}(\epsilon_i, X_i) \neq 0$ zeigen, dass $E(\epsilon_i | X_i) \neq 0$ gilt. In der angewandten Ökonometrie wird in der Regel auf die Bedingung 2b abgestellt und diese anhand von Tests überprüft. Ist man nur an der Konsistenz der Schätzer und nicht zugleich an ihrer Erwartungstreue interessiert, so ist die Annahme 2a nicht erforderlich und es genügt die Annahme 2b.

Mit Annahme 2a ist also Annahme 2b automatisch erfüllt.¹⁷ Im Falle $E(\epsilon_i | X_i) = 0$ sagen wir, die Variable X ist **exogen**. Ist $\text{Cov}(\epsilon_i, X_i) \neq 0$, ist die Annahme der Exo-

¹⁶ Für einen Beweis hierfür vgl. z.B. Stock, J.H., Watson, M.W. (2003).

genität verletzt. X wird dann als **endogen** bezeichnet. Annahme 2b ist typischerweise bei vernächlässigten Variablen und falscher funktionaler Form verletzt und führt zur Verzerrung und Inkonsistenz der OLS-Schätzer.¹⁸

Annahme 3: Homoskedastizität: Bei gegebenen Werten der erklärenden Variable ist die Varianz von ϵ_i für alle Beobachtungen identisch.

$$\text{Var}(\epsilon_i | X_i) = \sigma^2 \quad (\text{IV2: 17a})$$

Der Begriff Homoskedastizität (homo = gleich, Skedastizität = Streuung) besagt zunächst nichts anderes als gleiche Varianz. Nehmen wir an, dass die Teilgesamtheiten von Y, die mit den verschiedenen X-Werten einhergehen, die gleiche Varianz besitzen, d.h. die Streuung der Y in den einzelnen Teilgesamtheiten identisch bleibt,¹⁹ so folgt daraus auch eine konstante Varianz der stochastischen Störterme. Grafisch lässt sich die Verteilung der ϵ_i in einem solchen Fall wie in Abbildung IV 10 darstellen. Darin beschreibt $f(\epsilon)$ die Wahrscheinlichkeitsdichte der Verteilungen des stochastischen Störterms. Wir erkennen, dass für jedes X_i die Verteilung des stochastischen Störterms aufgrund des gleichen Mittelwertes (Annahme 2a) von Null und der gleichen Varianz (Annahme 3) identisch ist.

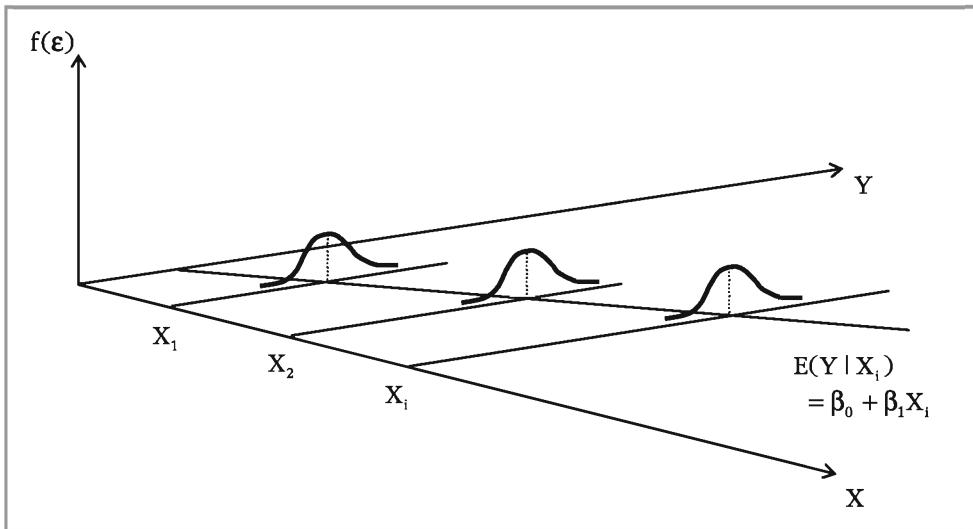


Abbildung IV 10: Homoskedastizität

¹⁷ Einen weiteren Fall, bei dem Annahme 2b automatisch erfüllt ist, werden wir im Abschnitt IV 2.2.2 im Zusammenhang mit deterministischen Regressoren kennenlernen.

¹⁸ Mit dem Problem der Verletzung von Annahme 2b werden wir uns unter IV 4.6 intensiv auseinandersetzen. Auch unter IV 4.1 und IV 4.2 werden wir ihm bereits begegnen.

¹⁹ Formal bedeutet dies $\text{Var}(Y | X_i) = \sigma^2$. Die bedingten Verteilungen von Y besitzen also die gleiche Streuung.

Analog zu den Folgen der Annahme 2a folgt aus Annahme 3 direkt, dass auch die *unbedingte Varianz* des stochastischen Störterms den Wert σ^2 besitzt. Es gilt also

$$\text{Var}(\varepsilon_i) = \sigma^2. \quad (\text{IV2: 17b})$$

Im Gegensatz zu Abbildung IV 10 zeigt Abbildung IV 11 eine Situation, die auch als **Heteroskedastizität** (hetero = ungleich) bezeichnet wird. Die Varianzen der Teilgesamtheiten von Y verändern sich nun mit X. Formal führt dies zu einer Abwandlung von (IV2: 17a) in (IV2: 18). Gerade daher ist auch das Subskript i in (IV2: 18) enthalten, welches andeutet, dass die Varianz nun nicht mehr konstant ist. Im Extremfall kann sich also für jedes X_i eine andere Varianz σ_i^2 ergeben.

$$\text{Var}(\varepsilon_i | X_i) = \sigma_i^2 \quad (\text{IV2: 18})$$

Abbildung IV 11 zeigt einen besonderen Fall von Heteroskedastizität, nämlich den, dass $\text{Var}(\varepsilon_1 | X_1) < \text{Var}(\varepsilon_2 | X_2) < \dots < \text{Var}(\varepsilon_i | X_i)$ gilt. Die Wahrscheinlichkeit in einer Stichprobe einen Y-Wert zu erhalten, der nahe an der Regressionsfunktion der Grundgesamtheit liegt ist somit in Teilgesamtheit $X = X_1$ höher als beispielsweise in $X = X_2$. Wir würden daher dazu neigen, Stichproben aus den Teilgesamtheiten zu ziehen, die weniger um ihren bedingten Erwartungswert streuen. Dies würde aber den Bereich möglicher X-Werte einschränken. Durch Annahme 3 sind alle Y-Werte innerhalb der verschiedenen Teilgesamtheiten gleichwertig, da ihre Streuung um die bedingten Erwartungswerte gleich ist.

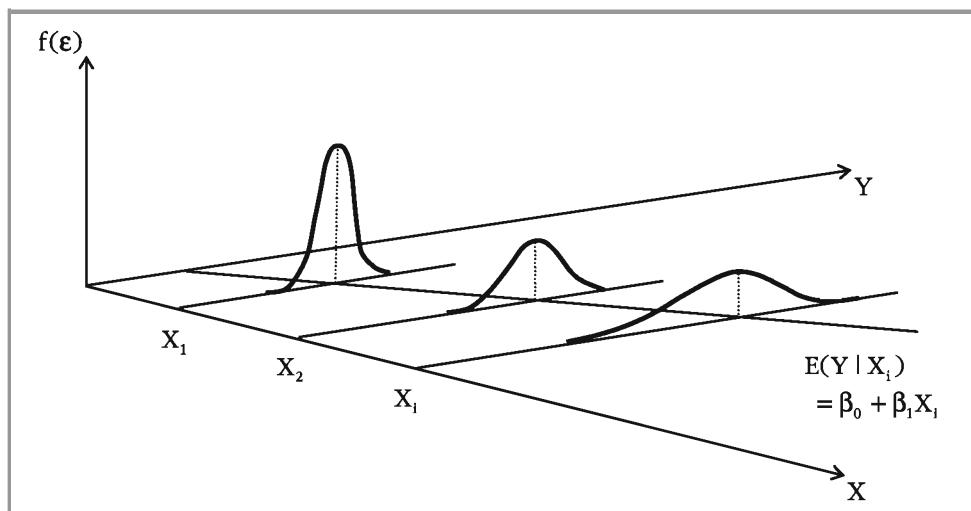


Abbildung IV 11: Heteroskedastizität

Zu einer Verletzung dieser Annahme 3 kommt es häufig in Querschnittsregressionen.²⁰ Nehmen wir z.B. an, wir untersuchen in einem Modell die persönlichen Ausgaben für Reisen in Abhängigkeit vom verfügbaren Einkommen. Da bei hohen Einkommen die Schwankungen der Ausgaben absolut betrachtet höchstwahrscheinlich größer sind als bei niedrigen Einkommen, ist auch mit zunehmender

²⁰ Mit dem Problem der Heteroskedastizität befassen wir uns im Abschnitt IV 4.4 im Detail.

Varianz des stochastischen Störterms zu rechnen. Wir haben es deshalb bei der Analyse von Querschnittsdaten sehr häufig mit dem Problem der Heteroskedastizität zu tun. Die OLS-Schätzer sind in einem solchen Fall immer noch erwartungstreu und konsistent, aber nicht mehr effizient.

Annahme 4: Keine Korrelation der Störterme: Bei gegebenen Werten der erklärenden Variablen ist die Kovarianz und damit die Korrelation zwischen zwei beliebigen ϵ_i und ϵ_j ($i \neq j$) gleich Null.

$$\text{Cov}(\epsilon_i, \epsilon_j | X_1, X_2) = 0 \quad (\text{IV2: 19})$$

Diese Annahme besagt, dass bei gegebenen Werten der erklärenden Variablen die Zufallsvariablen ϵ_i und ϵ_j , die durch zwei beliebige Y-Teilgesamtheiten bestimmt werden, nicht voneinander abhängen und nicht etwa Muster wie in den Streudiagrammen 1 und 2 in Abbildung IV 12 aufweisen. Wir haben in dieser Abbildung das Subskript t verwendet, da eine Verletzung von Annahme 4 häufig in Zeitreihenregressionen auftritt. Hier ist die Zufallsvariable ϵ_t häufig von der um eine Periode verzögerten Variablen ϵ_{t-1} abhängig. Wir sprechen dann auch von *Autokorrelation erster Ordnung*. Muster 1 verdeutlicht positive Autokorrelation. Auf positive Werte von ϵ in $t-1$ folgen wieder positive in t bzw. auf negative wieder negative. Bei negativer Autokorrelation in Muster 2 verhält es sich umgekehrt. Annahme 4 fordert nun die Abwesenheit derartiger Korrelationen. Muster 3 in Abbildung IV 12 zeigt, dass es hier keine systematische Beziehung zwischen den Störterms gibt, sodass keine Autokorrelation vorliegt.

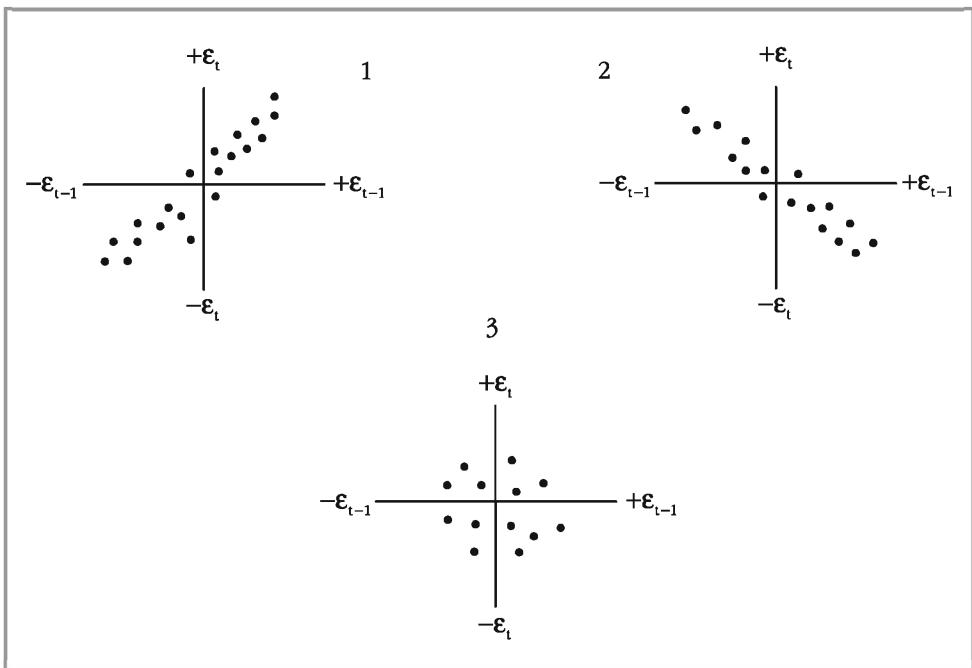


Abbildung IV 12: Korrelationsmuster des Störterms

Welchen Effekt hätte nun aber eine Verletzung dieser Annahme?²¹ Nehmen wir an, unsere Regressionsfunktion sei $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ und ε_i und ε_{i-1} positiv korreliert (vgl. Abbildung IV 12, Nr. 1). Y_i würde somit nicht nur von X_i abhängen, sondern auch von ε_{i-1} , da dieser in gewissem Ausmaß ε_i bestimmt. Durch Annahme 4 werden derartige Einflüsse ausgeschlossen, sodass Y_i nur durch den systematischen Einfluss von X_i bestimmt wird. Wie bereits erwähnt, ist diese Annahme besonders in Zeitreihenmodellen von Bedeutung. In diesem Zusammenhang besagt sie nämlich, dass eine Erhöhung des Störterms (z.B. durch einen zufälligen Schock) in einer Periode keinen Effekt auf den Störterm der Folgeperiode hat. In vielen Fällen kann eine solche Annahme jedoch unrealistisch sein, da die Auswirkungen eines zufälligen Schocks meist mehrere Perioden anhalten. Wir werden daher im weiteren Verlauf auch Verfahren betrachten, mit denen wir auch unter Berücksichtigung von Autokorrelation schätzen können. Eine OLS-Schätzung führt bei von Autokorrelation zu einem Effizienzverlust der OLS-Schätzer. In speziellen Modellen führt sie außerdem zu deren Verzerrung (vgl. Abschnitt IV 4.7.1.2).

Annahme 5: Die Anzahl n der Beobachtungen muss größer als die Zahl der zu schätzenden Parameter sein. Außerdem sind große Ausreißer bei den (X_i, Y_i) -Kombinationen unwahrscheinlich.

Wie wir es aus der Mathematik kennen, benötigen wir zur Aufstellung einer Gleichung mindestens 2 Punkte, die auf dieser Geraden liegen. Ähnlich verhält es sich bei der Regressionsanalyse. Zur Schätzung einer Regressionsfunktion mit zwei unbekannten Parametern benötigen wir beispielsweise mindestens 3 Beobachtungspaare (X_i, Y_i) . Andernfalls können wir keine Schätzung durchführen.

Der zweite Teil dieser Annahme bedeutet, dass Beobachtungen mit weit außerhalb des üblichen Datenbereichs liegenden Werten für X_i und/oder Y_i nicht vorkommen sollten. Sie können die Ergebnisse der OLS-Schätzung evtl. verfälschen (vgl. Abbildung IV 13). Derartige Ausreißer können ein Resultat von Datenerfassungsfehlern sein. So können sie etwa durch einfaches "Vertippen" oder durch die Verwendung unterschiedlicher Maßeinheiten für unterschiedliche Beobachtungen entstehen. Kann ein derartiger Fehler identifiziert werden, erfolgt eine einfache Korrektur.

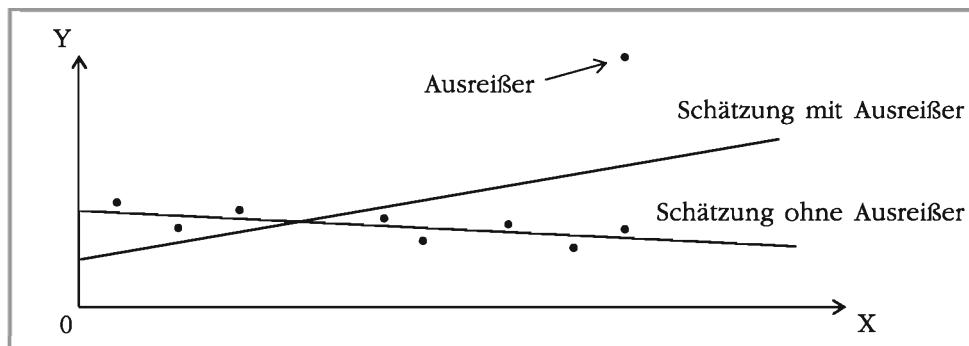


Abbildung IV 13: Einfluss von Ausreißern auf OLS-Schätzungen

²¹ Eine detaillierte Behandlung des Problems der Autokorrelation folgt unter IV 4.5 und speziell für sog. dynamische Modelle unter IV 4.7.1.2.

Bei der Untersuchung von Finanzmarktdaten (v.a. bei Aktienrenditen) kommt es häufig zu Ausreißern. Sie sind hier meist das Resultat ökonomischer Schocks. Wir könnten daher der Versuchung unterliegen die entsprechenden Beobachtungen aus dem Datensatz zu streichen, um die Schätzergebnisse zu "verbessern". Gerade dadurch gehen aber Informationen verloren, die die Grundgesamtheit beschreiben, was also nicht wirklich zu einer Verbesserung der Schätzungen führt. Vor Ausschluss sollten also die Ausreißer und vor allem deren Bedeutung für die Charakteristika der Grundgesamtheit genau untersucht werden.

Annahme 6: Keine erklärende Variable ist eine perfekte lineare Funktion einer oder mehrerer anderer erklärender Variablen (keine vollkommene Multikollinearität).

Diese Annahme 6 ist die einzige zusätzliche Annahme, die im multiplen Regressionsmodell getroffen wird. Sie ist für das einfache Regressionsmodell nicht relevant, da in diesem nur eine erklärende Variable vorhanden ist. Die Annahmen 1 bis 5 gelten für das multiple Regressionsmodell analog.²²

Vollkommene Kollinearität zwischen *zwei* erklärenden Variablen impliziert, dass sie in Wirklichkeit identisch sind, eine nur ein Vielfaches der anderen ist und/oder ein konstanter Term zu einer Variablen addiert wurde, um die andere zu erhalten. Formal können wir also sagen, dass es z.B. in einem Modell mit nur zwei erklärenden Variablen keine Beziehung $X_{1i} = \alpha_0 + \alpha_1 X_{2i}$ geben darf.²³ Existiert eine solche, kann OLS nicht zwischen den Variablen unterscheiden und eine Schätzung der Regressionsparameter ist unmöglich. Vielfach ist vollkommene Kollinearität das Resultat nicht beachteter Identitäten zwischen erklärenden Variablen. Der Ausschluss einer der vollkommen kollinearen Variablen aus der Regressionsfunktion ist daher oft (nicht generell) die einfachste Lösung dieses Problems.

Sind mehr als zwei erklärende Variablen betroffen und ergibt sich eine Variable als Linearkombination verschiedener anderer, so sprechen wir von **vollkommener Multikollinearität**. Sie kann beispielsweise auftreten, wenn sich eine erklärende Variable als Summe zweier anderer erklärender Variablen ergibt. Da es in der Literatur gebräuchlich ist, den Begriff Multikollinearität auch im Fall von Kollinearität zu verwenden, werden auch wir uns dieser Vorgehensweise anschließen.

Während es in der Praxis sehr unwahrscheinlich ist, auf vollkommene Multikollinearität zu treffen, kann *unvollkommene Multikollinearität* (hohe Korrelation zwischen den Variablen, aber keine exakte lineare Beziehung) zu Problemen bei der Schätzung führen. Da eine Unterscheidung der Variablen schwer fällt, ordnet OLS z.B. in einem Modell mit zwei erklärenden Variablen Veränderungen von Y, die eigentlich auf X_1 zurückzuführen sind, X_2 zu oder umgekehrt.²⁴

²² Unsere Annahmen 2a und 3 würden wir z.B. im Modell mit zwei Regressoren in der Form $E(\epsilon_i | X_{1i}, X_{2i}) = 0$ und $\text{Var}(\epsilon_i | X_{1i}, X_{2i}) = \sigma^2$ formulieren. Annahme 2b hätte die Form $\text{Cov}(\epsilon_i, X_{1i}) = \text{Cov}(\epsilon_i, X_{2i}) = 0$.

²³ Annahme 6 schließt nur perfekte lineare Beziehungen zwischen erklärenden Variablen aus. Polynomiale oder andere Verbindungen sind davon nicht betroffen.

²⁴ Mit derartigen Problemen werden wir uns unter IV 4.3 noch ausführlich beschäftigen.

2.2.2 Bedeutung deterministischer und stochastischer Regressoren

Bisher haben wir die Modellgleichungen und Annahmen des klassischen linearen Regressionsmodells für *stochastische Regressoren* präsentiert. Dies bedeutet, dass X eine Zufallsvariable darstellt und wir alle Aussagen bedingt auf eine spezielle Ausprägung dieser Zufallsvariablen formulieren müssen. In einer wiederholten Stichprobe können wir nämlich unterschiedliche Realisationen von X erhalten.

In vielen Einführungslehrbüchern zur Ökonometrie ist es Standard, die Annahmen des CLRM um die zusätzliche Annahme *deterministischer (nicht-stochastischer) Regressoren* zu erweitern. Dies bedeutet, dass davon ausgegangen wird, dass wir die Realisationen von X in wiederholten Stichproben fixieren können. Um dies zu veranschaulichen greifen wir auf unser Konsumbeispiel aus IV 2.1 zurück. In der dortigen Tabelle waren für 10 verschiedene X-Werte Teilgesamtheiten von Y aufgeführt. Werden nun zwei Stichproben aus dieser Grundgesamtheit gezogen, so könnten deren Ergebnisse folgendermaßen aussehen:²⁵

X	80	100	120	140	160	180	200	220	240	260
Y (Stichpr. 1)	55	65	79	80	102	110	120	135	137	150
Y (Stichpr. 2)	60	70	84	93	107	115	136	137	145	152

Wie zu erkennen ist, wurde für jede Stichprobe die gleiche Auswahl an X-Werten verwendet. Für jeden X-Wert ergab sich aber von Stichprobe zu Stichprobe ein anderer Y-Wert. Hier ist also lediglich Y als Zufallsvariable zu betrachten.

Ein Vorteil einer derartigen Zusatzannahme ist, dass wir unsere Modellannahmen unbedingt formulieren können, da sich die Bedingung (die X-Werte) schließlich nicht verändern kann. Aus den Annahmen 2a, 3 und 4 wird so für alle i (bzw. $i \neq j$)

$$E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{und} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0.$$

Die Annahme deterministischer Regressoren vereinfacht insbesondere die Herleitung zahlreicher statistischer Eigenschaften von OLS-Schätzern. Deswegen werden auch wir im weiteren Verlauf häufig davon Gebrauch machen (vgl. z.B. Abschnitt IV 2.3.1). Nachteil einer solchen Annahme ist zunächst, dass dadurch meist die zentrale Annahme vernachlässigt wird, dass der Störterm nicht mit der erklärenden Variable korreliert sein darf. Deterministische Regressoren erfüllen nämlich die Annahme 2b automatisch, da

$$\text{Cov}(\varepsilon_i, X_i) = E[(\varepsilon_i - E(\varepsilon_i))(X_i - E(X_i))] = [X_i - E(X_i)]E[\varepsilon_i] = 0.$$

Sind die X-Werte fest vorgegeben (determiniert), so gilt zudem, dass ε_i und X_i unabhängig voneinander sind und damit ist natürlich auch der bedingte Erwartungswert der ε_i für alle X_i konstant. Im Falle deterministischer Regressoren sind also sowohl Annahme 2b als auch 2a erfüllt.

In der empirischen Wirtschaftsforschung gibt es allerdings nur sehr selten deterministische Regressoren. X ist hier (anders als in kontrollierten Laborexperimenten) häufig stochastisch, also wie Y eine Zufallsvariable. Wollen wir beispielsweise die Rendite einer Aktie durch die eines Marktindex (z.B. CDAX) erklären, so können

²⁵ Zur besseren Zahlenvergleichbarkeit haben wir die ersten beiden Zeilen als Stichproben entnommen. Dies sind natürlich nur zwei von vielen möglichen Stichprobenausgängen.

wir in wiederholten Stichproben unmöglich die Rendite des Marktindex in der oben beschriebenen Art fixieren. Stattdessen werden wir von Stichprobe zu Stichprobe andere Werte für X erhalten.

Eine Konditionierung auf die X-Werte der Stichprobe bei statistischen Herleitungen bedeutet technisch gesprochen, dass wir die X-Werte als deterministisch in wiederholten Stichproben betrachten können. *Wenn wir also in einer Zufallsstichprobe annehmen können, dass $E(\epsilon_i | X_i) = 0$ gilt, können wir bei Herleitungen und Beweisen die X als deterministisch betrachten, obwohl sie stochastisch sind.*

Solange Annahme 2b erfüllt ist, stellt ein stochastisches X für die praktische Anwendung der Regressionsanalyse in den meisten Fällen kein Problem dar. Eine Verletzung von Annahme 2b führt bei stochastischem X jedoch zu verzerrten und inkonsistenten OLS-Schätzern.

2.2.3 Duplikation der Annahmen des CLRM durch OLS

Wie wir leicht erkennen, beziehen sich die Annahmen des CLRM auf die Regressionsfunktion der Grundgesamtheit und nicht die der Stichprobe. OLS besitzt nun die interessante Eigenschaft, dass es versucht, einige dieser Annahmen in der Stichprobe zu rekonstruieren, um die Regressionsfunktion der Grundgesamtheit so genau wie möglich schätzen zu können. So ist z.B. die mathematische OLS-Eigenschaft (IV1: 13c) bzw. $\sum \epsilon_i = 0 \rightarrow \bar{\epsilon} = 0$ auf Annahme 2a zurückzuführen. Genauso ist $\sum \epsilon_i X_i = 0$ ein Replikat von Annahme 2b. Es werden jedoch nicht alle Annahmen in der Stichprobe nachgebildet. Auch wenn $\text{Cov}(\epsilon_i, \epsilon_j | X_i, X_j) = 0$ ($i \neq j$) angenommen wird, trifft es *nicht* zu, dass auch $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ ($i \neq j$) gilt. Stattdessen werden wir noch feststellen, dass die Residuen häufig nicht nur autokorriert sondern auch noch heteroskedastisch sind.

Besonderes Augenmerk des klassischen Regressionsmodells liegt nun aber weniger auf den mathematischen, sondern vielmehr auf den *statistischen Eigenschaften* von OLS. Diese ergeben sich aus den Annahmen dieses Abschnitts und einer weiteren Annahme, die wir im folgenden Abschnitt IV 2.3 treffen.

2.3 Statistische Eigenschaften der OLS-Schätzer

Ziel der Regressionsanalyse ist es, nicht nur die geschätzte Regressionsfunktion aufzustellen, sondern daraus auch Rückschlüsse auf die Regressionsfunktion der Grundgesamtheit ziehen zu können. Wir interessieren uns also z.B. nicht nur für den Wert von $\hat{\beta}_1$, sondern wollen auch wissen, wie genau dieser den wahren Parameter β_1 schätzt. Um dies beurteilen zu können, müssen wir allerdings Kenntnis von der Verteilung der OLS-Schätzer haben, deren Werte sich von Stichprobe zu Stichprobe ändern und damit Zufallsvariablen sind.

2.3.1 Verteilung der OLS-Schätzer

Untersuchen wir zunächst, welchen Erwartungswert und welche Varianz unsere OLS-Schätzer für ein einfaches Regressionsmodell aufweisen. Wir setzen dazu

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

in unsere alternative OLS-Schätzformel (IV1: 12) ein und erhalten damit

$$\hat{\beta}_1 = \frac{\sum x_i(\beta_0 + \beta_1 X_i + \epsilon_i)}{\sum x_i^2} = \sum k_i(\beta_0 + \beta_1 X_i + \epsilon_i) \quad \text{mit} \quad k_i = \frac{x_i}{\sum x_i^2}.$$

Für diese k_i können wir festhalten, dass $\sum k_i = 0$ und $\sum k_i x_i = \sum k_i X_i = 1$ gilt. Unter Berücksichtigung dieser zwei Eigenschaften sind folgende Umformungen möglich:

$$\hat{\beta}_1 = \beta_0 \sum k_i + \beta_1 \sum k_i X_i + \sum k_i \epsilon_i = \beta_1 + \sum k_i \epsilon_i$$

Nach Resubstitution von k_i erhalten wir daraus die wichtige Darstellung (IV2: 20), auf die wir besonders im Abschnitt IV 4 noch häufig zurückgreifen werden.

$$\hat{\beta}_1 = \beta_1 + \frac{\sum x_i \epsilon_i}{\sum x_i^2} \quad (\text{IV2: 20})$$

Falls X nun deterministisch ist, liefert die Erwartungswertbildung von (IV2: 20) aufgrund von $E(\sum x_i \epsilon_i) = \sum E(x_i \epsilon_i) = \sum x_i E(\epsilon_i) = 0$ das Ergebnis

$$E(\hat{\beta}_1) = \beta_1 + E\left(\frac{\sum x_i \epsilon_i}{\sum x_i^2}\right) = \beta_1 + \frac{\sum x_i E(\epsilon_i)}{\sum x_i^2} = \beta_1, \quad (\text{IV2: 21})$$

was uns zeigt, dass $\hat{\beta}_1$ ein **erwartungstreuer Schätzer** für β_1 ist. Entstammen die X_i einer Zufallsstichprobe und gilt $E(\epsilon_i | X_i) = 0$, so können sie bei der Herleitung der Eigenschaften der OLS-Schätzer auch als feste Größen betrachtet werden. Es gilt auch in diesem Fall (X stochastisch), dass OLS-Schätzer erwartungstreue Schätzer sind.²⁶ Analoges gilt auch für den Schätzer der Regressionskonstanten, d.h.

$$E(\hat{\beta}_0) = \beta_0. \quad (\text{IV2: 22})$$

Neben ihrer Erwartungstreue sind die OLS-Schätzer außerdem **konsistent**, d.h. mit zunehmendem Stichprobenumfang nähert sich der Schätzer immer mehr dem wahren Parameter der Grundgesamtheit an.

Da OLS-Schätzungen von der zugrunde liegenden Stichprobe abhängig sind, können sie sich von Stichprobe zu Stichprobe verändern. Gerade daher ist ein Maß für die Präzision der OLS-Schätzer von besonderem Interesse. Unter Präzision verstehen wird dabei, wie stark die Schätzwerte um den wahren Wert der Grundgesamtheit streuen. Es empfiehlt sich daher die Varianz bzw. die Standardabweichung (auch häufig als *Standardfehler des Schätzers* bezeichnet) der Schätzer als Präzisionsmaß heranzuziehen. Für das einfache lineare Regressionsmodell lässt sich zeigen, dass für die Varianz des Schätzers des Steigungskoeffizienten

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2} \quad (\text{IV2: 23})$$

und für die des Schätzers der Regressionskonstanten

$$\text{Var}(\hat{\beta}_0) = \frac{\sum x_i^2}{n \sum x_i^2} \sigma^2 \quad (\text{IV2: 24})$$

²⁶ Für Details hierzu vgl. z.B. Wooldridge, J.M. (2006), Kapitel 2.

gilt.²⁷ Dabei ist $x_i = X_i - \bar{X}$ und σ^2 die konstante Varianz der stochastischen Störterme ϵ_i der Grundgesamtheit. Abgesehen von σ^2 liegen alle Daten für die Berechnung dieser Standardfehler mit einer Stichprobe vor. Das unbekannte σ^2 können wir mittels des erwartungstreuen Schätzers

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n - 2} \quad (\text{IV2: 25})$$

schätzen. Wir sprechen bei der Wurzel von (IV2: 25) vom sog. **Standardfehler der Regression**,²⁸ der auch häufig als Maß für den "fit" der geschätzten Regressionsfunktion eingesetzt wird. Wie wir in (IV1: 13d) gesehen haben, ist der Zähler nichts anderes als die Summe der quadrierten Abweichungen der Residuen von ihrem Mittelwert. Dividiert durch n würden wir also die Varianz der Residuen als Schätzer für die Varianz der stochastischen Störterme verwenden. Da dieser aber nicht erwartungstreu wäre, greifen wir zu einer Modifikation. Der hier verwendete Nenner $n - 2$ korrigiert eine leichte Abwärtsverzerrung, die durch die Schätzung von zwei Regressionskoeffizienten entsteht. Dies wird auch als Freiheitsgradkorrektur bezeichnet, da durch die Schätzung von β_0 und β_1 (also zwei Koeffizienten) zwei Freiheitsgrade beim Stichprobendatenmaterial verloren gehen. Bei großem Stichprobenumfang n macht es jedoch keinen Unterschied, ob durch n oder $n - 2$ dividiert wird.

Da auch die Standardfehler der Schätzer stichprobenabhängig sind, sind diese genau wie die Schätzer selbst Zufallsvariablen. Hinzu kommt, dass wir zu ihrer Bestimmung eine Schätzung von σ^2 benötigen. Die Standardfehler, die wir berechnen können, sind also genau genommen nur *Schätzungen der echten Standardfehler*.

Beispiel IV 2-1

Standardfehler von OLS-Schätzern



Berechnen wir zur Veranschaulichung die (Schätzung der) Standardfehler unserer Schätzer aus unserem Kaffeekonsumbeispiel IV 1-4. Dazu benötigen wir zunächst den Standardfehler der Regression, der sich als Quadratwurzel aus (IV2: 25) ergibt:

$$\hat{\sigma} = \sqrt{\frac{0,25^2 + 0,16^2 + \dots + (-0,11)^2}{11 - 2}} = 0,1287$$

Mit diesem können wir schließlich die Standardfehler der OLS-Schätzer als Wurzel von (IV2: 23) und (IV2: 24) schätzen:

$$\hat{\sigma}_{\beta_0} = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \hat{\sigma} = \sqrt{\frac{2,57^2 + 2,50 + \dots + 2,02^2}{11 \cdot 1,2741}} \cdot 0,1287 = 0,1216$$

$$\hat{\sigma}_{\beta_1} = \sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}} = \sqrt{\frac{0,1287}{1,2741}} = 0,1140$$

²⁷ Hinter der Herleitung der Varianzen der Schätzer stecken u.a. die Annahmen 3 und 4 des CLRM. Mit dieser werden wir uns unter IV 4.4.2 im Detail befassen.

²⁸ Eine Herleitung des Schätzers (IV2: 25) und der Beweis seiner Erwartungstreue sind z.B. in Gujarati, D.N. (2003), Kapitel 3, zu finden.

Im Regressionmodell mit *zwei* erklärenden Variablen ergeben sich die Standardfehler der Regressionskoeffizienten als positive Wurzeln der Varianzen

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1i}^2 (1 - r_{12}^2)}, \quad (\text{IV2: 26})$$

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{12}^2)} \quad (\text{IV2: 27})$$

und

$$\text{Var}(\hat{\beta}_0) = \left(\frac{1}{n} \frac{\bar{X}_1^2 \sum x_{2i}^2 + \bar{X}_2^2 \sum x_{1i}^2 - 2\bar{X}_1 \bar{X}_2 \sum x_{1i} x_{2i}}{\sum x_{1i}^2 \sum x_{2i}^2 - (\sum x_{1i} x_{2i})^2} \right) \cdot \sigma^2, \quad (\text{IV2: 28})$$

wobei σ^2 wieder die konstante Varianz der ϵ_i und r_{12} der empirische Korrelationskoeffizient zwischen X_1 und X_2 ist.²⁹ Als Schätzer für σ^2 verwenden wir hier

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n - 3}. \quad (\text{IV2: 29})$$

Im Vergleich zu (IV2: 25) fällt auf, dass die Anzahl der Freiheitsgrade nun bei $n - 3$ liegt, da wir zur Bestimmung der e_i zunächst β_0 , β_1 und β_2 schätzen müssen, wodurch drei Freiheitsgrade verloren gehen. Analog ergibt sich die Anzahl der Freiheitsgrade bei Modellen mit mehr als zwei erklärenden Variablen.

Nach unserem Ausflug ins multiple Regressionsmodell kehren wir nun aber zurück zum einfachen Regressionsmodell. In diesem wird in der Regel angenommen, dass die ϵ_i innerhalb einer Teilgesamtheit von Y einer **Normalverteilung** folgen. Wir hatten daher in den Abbildungen IV 9 und IV 10 die Verteilungen bereits als symmetrisch skizziert. Konkret sagen wir, dass die $\epsilon_i | X_i$ unabhängig und identisch mit Erwartungswert 0 und Varianz σ^2 normalverteilt sind.³⁰ Da die Verteilung von $\hat{\beta}_1$ und aufgrund von (IV1: 10) auch die von $\hat{\beta}_0$, wie wir in (IV2: 20) sehen, von der Verteilung von ϵ_i abhängen, folgen auch die OLS-Schätzer einer Normalverteilung. So gilt für den Schätzer des Steigungsparameters

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum x_i^2}\right) \quad (\text{IV2: 30})$$

und für den Schätzer der Konstanten

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2\right). \quad (\text{IV2: 31})$$

²⁹ Leser, die sich für eine Herleitung dieser Formeln (IV2: 26) bis (IV2: 28) interessieren, verweisen wir auf Gujarati, D.N. (2003), Kapitel 7.

³⁰ Durch die Annahme der Normalverteilung der $\epsilon_i | X_i$ sind auch die $Y_i | X_i$ jeweils mit $E(Y_i | X_i) = \beta_0 + \beta_1 X_i$ und $\text{Var}(Y_i | X_i) = \sigma^2$ unabhängig normalverteilt.

Die Normalverteilungsannahme ist jedoch in der Praxis nicht immer erfüllt. In einem solchen Fall können Transformationen der Variablen eine Verteilung liefern, die eher der Normalverteilung entspricht. So folgt z.B. der logarithmierte Kurs einer Aktie eher einer Normalverteilung als der Kurs selbst. Dies ist aber eine empirische Frage und führt nicht immer zum gewünschten Resultat. Bei *großen Stichprobenumfängen*³¹ ist die Verletzung der Normalverteilungsannahme aufgrund des Zentralen Grenzwertsatzes (vgl. Abschnitt II 3.4.1) kein schwerwiegendes Problem. Es gilt nämlich dann, dass die OLS-Schätzer *asymptotisch normalverteilt* sind.³²

2.3.2 Gauß-Markov-Theorem

Wir haben bereits erwähnt, dass OLS-Schätzer unter gewissen Voraussetzungen ideale Eigenschaften aufweisen und somit "bestmögliche" Schätzer sind. Das sog. **Gauß-Markov-Theorem** besagt in diesem Zusammenhang, dass bei erfüllten Annahmen des klassischen Regressionsmodells die OLS-Schätzer in der Klasse der unverzerrten linearen Schätzer minimale Varianz besitzen. Wir sagen kurz, sie sind **BLUE** (engl. best linear unbiased estimators) und erfüllen damit die folgenden Bedingungen:

1. Der Schätzer ist **linear**, d.h. er ist eine lineare Funktion der Zufallsvariablen ϵ oder Y .
2. Er ist **unverzerrt** und **konsistent**. Dies bedeutet, dass die OLS-Schätzungen der Koeffizienten um den wahren Wert streuen und diesen im Durchschnitt treffen. Mit zunehmendem Stichprobenumfang nähert sich der Schätzer außerdem dem wahren Parameter an.
3. Er weist minimale Varianz in der Klasse aller linearen unverzerrten Schätzer auf. In einem solchen Fall wird er auch als **effizienter Schätzer** bezeichnet.

Zum besseren Verständnis dieses Theorems betrachten wir Abbildung IV 14. Sie zeigt die Stichprobenverteilung des OLS-Schätzers $\hat{\beta}^{OLS}$ und die eines alternativen Schätzers $\hat{\beta}^{Alt.}$ für ein beliebiges β .

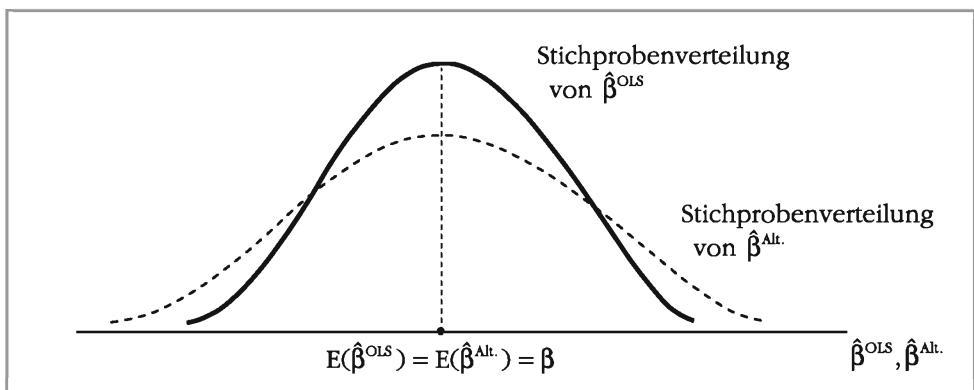


Abbildung IV 14: Veranschaulichung der BLUE-Eigenschaft von OLS

³¹ Häufig reicht bereits ein Stichprobenumfang von 30 Beobachtungen aus.

³² Vgl. z.B. Greene, W.H. (2008), Kapitel 2 oder Wooldridge, J.M. (2006), Kapitel 4.

Die Stichprobenverteilungen (hier als symmetrisch angenommen) zeigen die Verteilung der Werte, die die Schätzer in wiederholten Stichproben annehmen. Wir erkennen, dass beide Schätzer als unverzerrte Schätzer von β bezeichnet werden können, da $E(\hat{\beta}^{OLS}) = \beta$ und $E(\hat{\beta}^{Alt.}) = \beta$ gilt. Nehmen wir nun an, dass beide Schätzer lineare Funktionen von Y sind, so würden wir uns für den Schätzer entscheiden, dessen Varianz geringer ist, da dann die Stichprobenwerte des Schätzers näher bei β liegen. Kurz gesagt würden wir uns für den Schätzer entscheiden, der BLUE ist. Gauß und Markov haben gezeigt, dass eine solche Entscheidung unter den Annahmen 1 bis 5 immer zu Gunsten des OLS-Schätzers gefällt wird. Mit dem dahinter steckenden mathematischen Beweis wollen wir uns hier aber nicht näher befassen.³³

Besonders interessant an diesem Theorem ist, dass keinerlei Annahmen über die genaue Verteilung von ϵ_i getroffen werden. Allein die Erfüllung der Annahmen des klassischen Regressionsmodells reicht aus, damit das Theorem Gültigkeit besitzt. Wird jedoch zusätzlich die Annahme der Normalverteilung getroffen, so besitzen die OLS-Schätzer in der gesamten Klasse (linearer *und* nichtlinearer) unverzerrter Schätzer minimale Varianz. Sie sind also dann sogar **BUE** (engl. best unbiased estimators).

³³ Vgl. dazu z.B. Hill, R.C., Griffiths, W.E., Judge, G.G. (2001), Kapitel 4.

3.

Testen von Hypothesen und Konfidenzintervalle

Nachdem wir mit der Annahme der Normalverteilung der Störterme die Stichprobenverteilung der OLS-Schätzer festgelegt haben, sind wir nun in der Lage, Hypothesen über die unbekannten Parameter der Grundgesamtheit zu testen. Die folgenden Testverfahren werden uns helfen, anhand von Stichproben bzw. der geschätzten Regressionsfunktion Aussagen über die Grundgesamtheit zu treffen. Wir können durch sie prüfen, ob unser OLS-Schätzergebnis rein zufällig von einem bestimmten Wert abweicht und ob unsere Modelltheorien durch die vorliegende Stichprobe verworfen werden können oder nicht.

3.1 Testen einzelner Regressionsparameter – t-Test

Beim t-Test handelt es sich um den Test, den Ökonometriker wohl am häufigsten zum Testen von Hypothesen bezüglich einzelner Regressionsparameter heranziehen. Hypothesen zu mehr als einem Koeffizienten werden mit dem F-Test (vgl. Abschnitt IV 3.2) überprüft. Der t-Test wird so häufig verwendet, da er für den Fall, dass der *stochastische Störterm normalverteilt* ist und die *Varianz dieser Verteilung geschätzt* werden muss, der geeignete Test ist. Da auch bei nicht normalverteiltem Störterm in *großen Stichproben* aufgrund des Zentralen Grenzwertsatzes von *approximativer Normalverteilung* ausgegangen werden kann, hat sich dieser Test als Standard etabliert.

Bereits im Hinblick auf Abschnitt IV 4, in dem wir die Verletzungen der Annahmen des CLRM und ihre Folgen behandeln, sei an dieser Stelle erwähnt, dass das Testen von Hypothesen zunächst nur sinnvoll ist, wenn sichergestellt wurde, dass im geschätzten Regressionsmodell alle wichtigen Variablen enthalten sind und eine geeignete funktionale Form gewählt wurde. Außerdem ist anzunehmen, dass der Datensatz, der für die Schätzung verwendet werden soll, von ausreichendem Umfang und nicht andersweitig mängelbehaftet ist. Erst in derart spezifizierten Modellen sind Hypothesentests aussagekräftig.

3.1.1 Hypothesen, t-Statistik und Entscheidungsregel

In der Regressionsanalyse werden im Rahmen des t-Tests typischerweise folgende Arten von Hypothesen bzgl. der unbekannten Regressionsparameter der Grundgesamtheit untersucht:

- | | | | | |
|----|--------------------------------|-----------|--------------------------------|----------|
| a) | $H_0 : \beta_k = \beta_k^0$ | gegenüber | $H_A : \beta_k \neq \beta_k^0$ | |
| b) | $H_0 : \beta_k \geq \beta_k^0$ | " | $H_A : \beta_k < \beta_k^0$ | (IV3: 1) |
| c) | $H_0 : \beta_k \leq \beta_k^0$ | " | $H_A : \beta_k > \beta_k^0$ | |

Dabei ist $\hat{\beta}_k$ einer der K Steigungsparameter des Regressionsmodells und β_k^0 ein zugehöriger hypothetischer Wert. Aus Gründen, auf die wir im Abschnitt IV 4.2.1 zu sprechen kommen, empfiehlt es sich nicht, Hypothesentests für die Konstante des Regressionsmodells durchzuführen. Die folgenden Tests beziehen sich also stets auf die Parameter $\hat{\beta}_k$ mit $k = 1, 2, \dots, K$.

Zur Überprüfung von Nullhypotesen des Typs (IV3: 1) bedienen wir uns in der Praxis des t-Tests. Für ein typisches Regressionsmodell wird dazu zunächst für den interessierenden Parameter ein t-Wert bzw. der Wert der t-Statistik

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\hat{\sigma}_{\hat{\beta}_k}} \quad (k = 1, 2, \dots, K) \quad (\text{IV3: 2a})$$

berechnet, wobei $\hat{\beta}_k$ der geschätzte Regressionskoeffizient der k-ten Variable, β_k^0 der hypothetische Wert für β_k aus der Nullhypothese und $\hat{\sigma}_{\hat{\beta}_k}$ der geschätzte Standardfehler des k-ten OLS-Schätzers ist. Bei näherer Betrachtung dieser Statistik fällt auf, dass es sich bei t_k um nichts anderes als den mit dem hypothetischen Erwartungswert β_k^0 und der geschätzten Standardabweichung $\hat{\sigma}_{\hat{\beta}_k}$ standardisierten OLS-Schätzer $\hat{\beta}_k$ handelt und t_k damit ein Analogon zur Teststatistik des unter III 3.3.1.2 behandelten Gaußtests mit unbekannter Varianz (bzw. t-Tests) darstellt. Wäre die Standardabweichung des OLS-Schätzers bekannt und würde diese zur Standardisierung verwendet, so würde t_k (aufgrund der Normalverteilung der Störterme) bei korrekter Nullhypothese einer Standardnormalverteilung folgen. Da wir in der Praxis aber den Standardfehler des OLS-Schätzers nicht kennen und ihn schätzen müssen, folgt t_k einer t-Verteilung mit $n - K - 1$ Freiheitsgraden, da K Steigungsparameter und die Regressionskonstante geschätzt werden.³⁴

Da die meisten Hypothesentests prüfen, ob ein bestimmter Regressionskoeffizient von Null verschieden ist, ist β_k^0 gewöhnlich Null. Die Formel für die t-Statistik, die auch in den meisten Ökonometrie-Softwarepaketen implementiert ist, ist dann

$$t_k = \frac{\hat{\beta}_k}{\hat{\sigma}_{\hat{\beta}_k}} \quad (k = 1, 2, \dots, K) . \quad (\text{IV3: 2b})$$

Auch (IV3: 2b) folgt einer t-Verteilung mit $n - K - 1$ Freiheitsgraden. Im Fall $\beta_k^0 = 0$ hat der Wert der t-Statistik aber stets das gleiche Vorzeichen wie der geschätzte Regressionskoeffizient, da der zugehörige geschätzte Standardfehler immer positiv ist. Je größer der absolute Wert der t-Statistik ist, desto größer ist die Wahrscheinlichkeit, dass der betroffene Koeffizient signifikant von Null verschieden ist.

Beispiel IV 3-1

Geflügelkonsum in den USA



Studenmund, A.H. (2006) erklärt den Pro-Kopf-Geflügelkonsum in Pfund Y der US-Bürger durch das multiple Regressionsmodell

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$$

mit den Variablen X_1 : Geflügelpreis pro Pfund, X_2 : Rindfleischpreis pro Pfund und X_3 : verfügbares Einkommen in 100 Dollar. Bereits vor Schätzung des Modells werden wir vermuten, dass die Koeffizienten β_2 und β_3 positiv und der Koeffizient β_1 negativ sind. Eine

³⁴ Für den Beweis hierfür verweisen wir auf Hogg, R.V., McKean, J.W., Craig, A.T. (2004).

Schätzung mit Studenmunds Zahlenmaterial in EViews (Eingabe der Gleichung in der Form $y c x1 x2 x3$ im Menü Quick / Estimate Equation; c steht dabei für die Konstante im Regressionsmodell), scheint diese Vermutung auf den ersten Blick zu bestätigen, da die geschätzten Koeffizienten das gewünschte Vorzeichen aufweisen. Ein Blick auf die Koeffizienten allein reicht jedoch nicht aus. Erst entsprechende t-Tests können unserer Vermutung Nachdruck verleihen.

Dependent Variable: Y		Sample: 1960 1999		
Method: Least Squares		Included observations: 40		
	Coefficient	Std. Error	t-Statistic	Prob.
C	27.59394	1.584458	17.41539	0.0000
X1	-0.607160	0.157120	-3.864300	0.0004
X2	0.092188	0.039883	2.311452	0.0266
X3	0.244860	0.011095	22.06862	0.0000
R-squared	0.990391	Mean dependent var		50.56725
Adjusted R-squared	0.989590	S.D. dependent var		19.53879
S.E. of regression	1.993549	Akaike info criterion		4.312350
Sum squared resid	143.0726	Schwarz criterion		4.481238
Log likelihood	-82.24700	F-statistic		1236.776
Durbin-Watson stat	0.897776	Prob(F-statistic)		0.000000

Da wir im Laufe dieses Kapitels immer wieder auf EViews-Ergebnisse zurückgreifen, betrachten wir die obigen etwas näher. Jeder **EViews-Output** gibt in der "Kopfzeile" die erklärte Variable (Y), die Schätzmethode (Least Squares), den Zeitindex oder die laufende Nummer der ersten und letzten Beobachtung der Stichprobe (1960, 1999) und den Stichprobenumfang (40) an. Den "Mittelteil" des Outputs bilden die geschätzten Parameter (Coefficient), die zugehörigen geschätzten Standardfehler (Std. Error), bereits berechnete t-Werte für $\beta_k^0 = 0$ (t-Statistic) und die sog. p-Werte (Prob.), deren Bedeutung wir unter IV 3.1.2 noch näher behandeln. Der "Endteil" beinhaltet eine Reihe von Maßen, von denen wir einige bereits behandelt haben. Zu diesen zählt das Bestimmtheitsmaß (R-squared), das angepasste Bestimmtheitsmaß (Adjusted R-squared), der Standardfehler der Regression (S.E. of regression), die Residuenquadratesumme RSS (Sum squared resid), der Mittelwert der erklärten Variablen (Mean dependent var) und deren Standardabweichung (S.D. dependent var). Die verbleibenden Maße werden wir im Verlauf dieses Kapitels noch näher behandeln.

Ein **Regressionsoutput** in der ebenfalls sehr verbreiteten Software **Stata** hat hingegen die auf der Folgeseite zu sehende Gestalt. Hier befinden sich rechts oben die Anzahl der Beobachtungen (Number of obs), das Bestimmtheitsmaß (R-squared), das angepasste Bestimmtheitsmaß (Adj R-squared), der Standardfehler der Regression (Root MSE) sowie die F-Statistik und ihr p-Wert (vgl. dazu Abschnitt IV 3.2.2). Links oben findet sich die bereits unter IV 1.2.4.1 im Zusammenhang mit dem Bestimmtheitsmaß behandelte Varianzzerlegung. In der ersten Spalte (SS) sind darin TSS (Total), ESS (Model) und RSS (Residual) enthalten. Wir erkennen daher schnell, dass hier $R^2 = ESS / TSS = 14745,73 / 14888,80 = 0,9904$ gilt. Die weiteren Spalten zu Freiheitsgraden (df) und mittleren Summen (MS) werden wir dabei insbesondere unter IV 3.2.2 benötigen. Der untere Teil des Outputs enthält die klassischen Regressionsergebnisse, d.h. die Parameterschätzungen (Coef.), die dazu gehörigen geschätzten Standardfehler der Schätzer (Std. Err.), t-Werte für $\beta_k^0 = 0$ (t) und die entsprechenden p-Werte ($P > |t|$). Darüber hinaus sind außerdem 95 %-Konfidenzintervalle für die jeweiligen Parameter enthalten (vgl. dazu Abschnitt IV 3.1.4).

Source	SS	df	MS	Number of obs	=	40
Model	14745,73	3	4915,24	F(3, 36)	=	1236,78
Residual	143,07	36	3,97	Prob > F	=	0
Total	14888,80	39	381,76	R-squared	=	0,9904
				Adj R-squared	=	0,9896
				Root MSE	=	1,9935
yt	Coef.	Std. Err.	t	P > t	[95 % Conf. Interval]	
x1t	-0,6072	0,1571	-3,86	0	-0,9258	-0,2885
x2t	0,0922	0,0399	2,31	0,027	0,0113	0,1731
x3t	0,2449	0,0111	22,07	0	0,2224	0,2674
_cons	27,5939	1,5845	17,42	0	24,3805	30,8074

In der Literatur werden Regressionsergebnisse in der Regel nicht in derartig ausführlichen Formen dargestellt. Häufig finden wir die folgende kompakte Form:

$$\hat{Y}_t = 27,5939 - 0,6072X_{1t} + 0,0922X_{2t} + 0,2449X_{3t}, \quad n = 40 \text{ (36)}$$

$$(0,1571) \quad (0,0399) \quad (0,0111) \quad R^2 = 0,9896$$

$$t = -3,86 \quad 2,31 \quad 22,07$$

Die unter den geschätzten Regressionskoeffizienten stehenden Klammerwerte sind dabei die dazugehörigen geschätzten Standardfehler. Die aufgeführten t-Werte sind die Werte der t-Statistik für den betreffenden Koeffizienten. Der Wert, der in Klammern bei n angegeben ist, ist die Anzahl der Freiheitsgrade ($40 - 3 - 1 = 36$). Auf eine derartige Darstellung werden wir z.T. zum Vergleich verschiedener Regressionsergebnisse zurückgreifen.

Wir sehen hier, dass die Werte der t-Statistiken automatisch von EViews geliefert werden. Zum besseren Verständnis ihrer Entstehung wollen wir aber dennoch exemplarisch den Wert der t-Statistik für den zweiten Regressionskoeffizienten berechnen:

$$t_2 = \frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{-0,0922}{0,0399} = 2,31$$

Dabei ist zu beachten, dass es bei dieser manuellen Berechnung durch von Rundungsunterschieden zu geringfügigen Abweichungen ggü. den t-Werten aus EViews kommen kann.

Um zu entscheiden, ob eine Nullhypothese bzgl. eines Regressionskoeffizienten auf Basis eines berechneten Wertes der t-Statistik abgelehnt werden kann oder nicht, bedienen wir uns in Verbindung mit (IV3: 1) der allgemeinen **Entscheidungsregel**

Ablehnung von H_0 , wenn

- a) $|t_k| > t_{1-\alpha/2, n-K-1}$
- b) $t_k < -t_{1-\alpha, n-K-1}$
- c) $t_k > t_{1-\alpha, n-K-1}$

(IV3: 3)

Die darin enthaltenen Testschranken sind, wie wir es bereits aus der induktiven Statistik kennen, Fraktile der t-Verteilung mit $n - K - 1$ Freiheitsgraden. Diese können unter Verwendung des gewünschten Signifikanzniveaus α in der tabellierten t-Verteilung nachgeschlagen werden.

Es ist wichtig, dass die Entscheidungsregel (IV3: 3) auf der Normalverteilungsannahme von ϵ , beruht. Wird diese Annahme verletzt, ist nicht mehr sichergestellt, dass t_k einer t-Verteilung folgt. Einen formalen Test für die Erfüllung der Normalverteilungsannahme werden wir unter IV 3.3 präsentieren. In vielen Fällen wird die t-Verteilung aber auch bei Verletzung dieser Annahme als Approximation der wahren Verteilung von t_k verwendet, falls der Stichprobenumfang sehr groß ist.

Beispiel IV 3-2

Einseitiger t-Test

Im Beispiel IV 3-1 hatten wir vermutet, dass die Geflügelnachfrage wohl negativ mit dem Geflügelpreis in Verbindung steht. Allein ein negatives Vorzeichen des entsprechenden geschätzten Regressionskoeffizienten reicht nun aber nicht aus, um diese Vermutung zu bestätigen. Wir müssen wissen, ob er signifikant negativ ist. Gerade daher empfiehlt es sich, $H_0: \beta_1 \geq 0$ gegenüber $H_A: \beta_1 < 0$ zu testen. Wir erinnern uns, dass wir die Behauptung, die wir bestätigen wollen, immer als Gegenhypothese formulieren, da mit Falsifizierung von H_0 die Gegenhypothese bestätigt wird.

Entscheiden wir uns für ein Signifikanzniveau von 5 % und berücksichtigen die hier vorliegenden Freiheitsgrade von $40 - 3 - 1 = 36$, so erhalten wir durch Nachschlagen von $t_{0,95,36}$ den kritischen Wert 1,69. Wir kennen aus unserer EViews-Schätzung außerdem bereits den Wert der t-Statistik $t_1 = -3,86$.

Da nun $t_1 < -t_{1-\alpha/2,n-K-1}$ bzw. $-3,86 < -1,69$ erfüllt ist, kann H_0 nach (IV3: 3) abgelehnt werden. Es kann also mittels dieser Stichprobe gezeigt werden, dass (bei einem Signifikanzniveau von 5 %) der geschätzte Wert von β_1 signifikant negativ ist, also in der Grundgesamtheit ein negativer Zusammenhang zwischen Geflügelkonsum und Geflügelpreis besteht.

Beispiel IV 3-3

Zweiseitiger t-Test

Zweiseitige t-Tests werden in der Praxis häufig dann eingesetzt, wenn es in der dem Regressionsmodell zugrundeliegenden Modelltheorie keine genaue Aussage zum vermuteten Vorzeichen einzelner Regressionsparameter oder konkurrierende Hypothesen diesbezüglich gibt. In einem solchen Fall wird lediglich getestet, ob der betreffende Regressionskoeffizient *signifikant von Null verschieden* ist.

Nehmen wir an, dass in unserem Beispiel IV 3-1 Uneinigkeit darüber besteht, ob es sich bei Geflügel um ein normales Gut (steigende Nachfrage bei steigendem Einkommen) oder um ein sog. inferiores Gut (sinkende Nachfrage bei steigendem Einkommen) handelt. In einem solchen Fall würden wir $H_0: \beta_3 = 0$ gegenüber $H_A: \beta_3 \neq 0$ testen.

Da im vorliegenden Fall bei einem 5 %-Signifikanzniveau die Bedingung $|t_3| > t_{1-\alpha/2,n-K-1}$ bzw. nach Nachschlagen von $t_{0,975,36}$ genau $|22,07| > 2,03$ erfüllt ist, kann H_0 abgelehnt werden. Die Schätzung von β_3 ist demnach signifikant von Null verschieden.

3.1.2 Der p-Wert

Alternativ zur Entscheidungsregel (IV3: 3) kann die Testentscheidung beim t-Test auch mit dem sog. p-Wert durchgeführt werden, der in vielen Ökonometrieprogrammen Standardbestandteil des Regressionsoutputs ist. Der dort ausgegebene p-Wert dient zum Testen *zweiseitiger Hypothesen* mit $\beta_k^0 = 0$, d.h. $H_0: \beta_k = 0$ gegenüber $H_A: \beta_k \neq 0$. Er nimmt Werte zwischen 0 und 1 an und ist die Wahrscheinlichkeit, bei korrekter Nullhypothese für einen bestimmten Koeffizienten β_k einen t-

Wert zu erhalten, der größer als $|t_k|$ ist. Er liefert damit die geringste Irrtumswahrscheinlichkeit, bei der wir die Nullhypothese ablehnen können (p -Wert = Irrtumswahrscheinlichkeit, die wir hinnehmen müssten, wenn wir H_0 ablehnen). Ein kleiner p -Wert lässt Zweifel über die Nullhypothese aufkommen. Ist der p -Wert kleiner als die akzeptierte Irrtumswahrscheinlichkeit, so kann H_0 abgelehnt werden.

Die Bestimmung des p -Wertes überlassen wir generell EViews, da wir so nicht in der Tabelle der t-Verteilung genau die Irrtumswahrscheinlichkeit finden müssen, die mit dem Regressionsergebnis bzw. t_k einhergeht. Da EViews aber nur p -Werte für zweiseitige Tests ausgibt, müssen wir im Fall eines einseitigen t-Tests mit $\beta_k^0 = 0$ den von EViews gelieferten p -Wert durch Zwei dividieren. Als einfache Testentscheidungsregel mittels des p -Wertes halten wir damit

Ablehnung von H_0 , wenn

(IV3: 4)

$p_k < \alpha$ und Vorzeichen von $\hat{\beta}_k$ gemäß H_A

fest, wobei die Vorzeichenbedingung beim zweiseitigen Test irrelevant ist. p_k ist dabei der p -Wert des k -ten Koeffizienten und α das gewählte Signifikanzniveau.

Beispiel IV 3-4

Testentscheidung mittels p -Wert



Nehmen wir an wir wollen in unserem Beispiel IV 3-1 statistisch aufzugehen, dass der Regressionskoeffizient der erklärenden Variable "Rindfleischpreis" positiv ist, also $H_0: \beta_2 \leq 0$ gegenüber $H_A: \beta_2 > 0$ testen. EViews liefert für diesen Koeffizienten einen p -Wert von 0,0266 und unser Signifikanzniveau α sei 5 %. Da ($p = 0,0266/2 < 0,05$) gilt und $\hat{\beta}_2$ das Vorzeichen von H_A (positiv) aufweist, können wir unsere Nullhypothese ablehnen.

3.1.3 Beschränkungen des t-Tests

Nach Behandlung der allgemeinen Vorgehensweise beim t-Test gilt es, sich mit Beispielen zu befassen, die zeigen, wie der t-Test *nicht* eingesetzt werden sollte:

1. Statistische Signifikanz vs. theoretische Validität

Dass statistische Signifikanz nicht mit theoretischer Validität einhergeht, zeigt ein einfaches Regressionsmodell,³⁵ welches zur Erklärung des britischen Konsumenntenpreisindex Y geschätzt wurde:

$$\begin{aligned} \hat{Y}_t &= 10,9 - 3,20X_t + 0,39X_t^2 & n &= 21 (17) \\ &(0,23) \quad (0,02) & R^2 &= 0,982 \\ t &= -13,9 \quad 19,5 \end{aligned}$$

Auf den ersten Blick können wir sagen, dass X und X^2 statistisch signifikante Variablen bei der Erklärung von Y sind. Aufgrund der hohen t -Werte können wir nämlich die Nullhypotesen der entsprechenden zweiseitigen t-Tests ablehnen. Die Regressionsparameter der Grundgesamtheit können also als von Null verschieden betrachtet werden. Wenn wir nun aber erfahren, dass es sich bei X um die kumulierte Niederschlagsmenge in Großbritannien handelt, so können wir natürlich trotz

³⁵ Vgl. Hendry, D. F. (1980) und Studenmund, A. H. (2006), Kapitel 5.

vorliegender statistischer Signifikanz nicht auf die Gültigkeit dieses theoretischen Modells schließen.³⁶ Die vorliegende Signifikanz ergibt sich dadurch, dass die beiden Variablen jeweils einen Trend aufweisen.³⁷

2. t-Test und "Wichtigkeit" einer Variable

Statistische Signifikanz gibt Aufschluss darüber, dass ein bestimmtes Stichproben-schätzergebnis nicht reiner Zufall ist. Sie erlaubt keine Aussagen darüber, welche Variable den größten Teil der Variation von Y erklärt. Eine Schlussfolgerung, dass die Variable mit dem signifikantesten Koeffizienten (höchster t-Wert) gleichzeitig die "wichtigste" (größter Effekt auf Y) ist, wäre also falsch.

3. t-Test bei hohen Stichprobenumfängen

Mit zunehmendem Stichprobenumfang n liegt der unverzerrte Schätzwert eines Regressionsparameters immer näher am wahren Parameter der Grundgesamtheit. Dies führt über einen sinkenden (geschätzten) Standardfehler des OLS-Schätzers zu steigenden t-Werten. Dies lässt sich anhand von (IV3: 2a) wie folgt aufzeigen:

$$\frac{\hat{\beta}_k - \beta_k^0}{\hat{\sigma}_{\hat{\beta}_k}} = t_k \uparrow$$

Dies zeigt, dass sehr hohe Stichprobenumfänge hohe t-Werte haben und wir damit viele Nullhypotesen ablehnen können. Bei hohen Stichprobenumfängen sind nämlich aufgrund des kleineren Standardfehlers bereits kleine Abweichungen signifikant. Lautet beispielsweise die Nullhypothese einer bestimmten Theorie, dass eine Variable X keinen Einfluss auf die erklärte Variable Y hat, so werden bei sehr großen Stichproben noch so kleine Abweichungen des geschätzten Koeffizienten von Null häufig als signifikant ausgewiesen. Für die inhaltliche Erklärung von Y spielt dieser Regressor X aufgrund des kleinen Parameters aber kaum eine Rolle. Außerdem stellt sich die Frage, ob eine sehr kleine Abweichung des Parameters von Null überhaupt gegen die ökonomische Theorie spricht. Deshalb können wir die Bedeutung einer Variablen für die Erklärung von Y nicht alleine mit der Signifikanz beurteilen.

3.1.4 Konfidenzintervalle für Regressionsparameter

Wie wir wissen, liegt uns mit einem Konfidenzintervall ein Verfahren vor, das bei wiederholter Anwendung einen Bereich überdeckt, in dem der wahre Parameter der Grundgesamtheit mit einer vorgegebenen Wahrscheinlichkeit (Vertrauenswahrscheinlichkeit) $1 - \alpha$ liegt. Für einen Regressionskoeffizienten kann ein Konfidenzintervall wie folgt gebildet werden:³⁸

$$\hat{\beta}_k - t_{1-\alpha/2, n-K-1} \cdot \hat{\sigma}_{\hat{\beta}_k} \leq \beta_k \leq \hat{\beta}_k + t_{1-\alpha/2, n-K-1} \cdot \hat{\sigma}_{\hat{\beta}_k} \quad (\text{IV3: 5})$$

³⁶ Ökonometrisches Arbeiten ohne jegliche theoretische Fundierung ist in manchen empirischen Analysen ein Problem. Es wird hier nämlich nicht mehr theoretisch begründet, sondern nur noch geprüft, ob die Daten etwas aussagen können.

³⁷ Das Thema Trends werden wir im Abschnitt IV 4.7.2 eingehend behandeln.

³⁸ Analog zu den Herleitungen der Konfidenzintervalle im Kapitel III erfolgt jene von Intervall (IV3: 5) durch Umformungen der zweiseitigen Entscheidungsregel aus (IV3: 3).

Der darin enthaltene kritische t-Wert entspricht dem, der auch beim zweiseitigen t-Test verwendet wird. Konfidenzintervalle dieser Art werden in der Praxis meist dafür verwendet um eine Bandbreite möglicher Werte für β_k nennen und nicht nur eine Punktschätzung vorweisen zu können. Damit ist die Chance größer, mit der Schätzung den tatsächlichen Parameterwert der Grundgesamtheit zu treffen. Des Weiteren kann ein solches Konfidenzintervall, wie wir in Kapitel III gesehen haben, auch zum Testen zweiseitiger Hypothesen verwendet werden. Fällt der hypothetische Parameterwert β_k^0 in das Konfidenzintervall, so kann $H_0: \beta_k = \beta_k^0$ nämlich nicht gegenüber $H_1: \beta_k \neq \beta_k^0$ abgelehnt werden. Außerhalb des Konfidenzintervalls kommt es zur Ablehnung von H_0 .

Beispiel IV 3-5

Cobb-Douglas-Funktion



Die sog. Cobb-Douglas Produktionsfunktion ist in ihrer stochastischen Form als

$$Y_t = \beta_0 X_{1t}^{\beta_1} X_{2t}^{\beta_2} e^{\epsilon_t}$$

definiert, wobei Y den Output oder das Produktionsniveau einer Volkswirtschaft, X_1 den dafür notwendigen Arbeitsinput und X_2 den Kapitalinput symbolisiert. e ist hier die Eulersche Zahl (Basis des natürlichen Logarithmus). Besondere Eigenschaft einer solchen Funktion ist es, dass ihre Koeffizienten β_1 und β_2 als partielle Output-Elastizitäten interpretiert werden können. Das bedeutet, sie geben die prozentuale Veränderung des Outputs bei Erhöhung des zugehörigen Inputfaktors (bei Konstanz des anderen) um 1 % an.

Es ist deutlich erkennbar, dass die Beziehung zwischen Output und den beiden Inputs nicht parameterlinear ist. Durch logarithmische Transformation erhalten wir aber die Form

$$\ln Y_t = \ln \beta_0 + \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \epsilon_t,$$

welche parameterlinear und damit mit OLS schätzbar ist. Wir sprechen hier auch von einer sog. funktionalen Doppel-Log-Form, mit der wir uns im Abschnitt IV 4.2.2 noch genauer befassen werden. Die erklärte Variable ist nun $\ln Y$ und die beiden erklärenden Variablen sind $\ln X_1$ und $\ln X_2$. In β_0 ist die Konstante des Regressionsmodells.

Um dieses transformierte Modell für Deutschland zu schätzen, verwenden wir Datenmaterial des Statistischen Bundesamtes von 1970 bis 2008. Konkret belegen wir Y mit einem realen BIP-Index, X_1 mit der Zahl der Erwerbstätigen und X_2 mit dem realen Bruttoanlagevermögen. Die Schätzergebnisse sind auf der Folgeseite abgebildet. Den benötigten *natürlichen Logarithmus* der Variablen bilden wir in EViews durch den Befehl *LOG(...)* bei der Gleichungseingabe. Wiedervereinigungsbedingt (Veränderung der Struktur der Grundgesamtheit) haben wir uns hier entschieden, den Datenbestand in zwei Teile aufzuspalten und zwei Schätzungen durchzuführen.

Dependent Variable: LOG(Y)		Sample: 1970 1990		
Method: Least Squares		Included observations: 21		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.782437	1.197967	-6.496371	0.0000
LOG(X1)	0.681214	0.132706	5.133256	0.0001
LOG(X2)	0.595082	0.026142	22.76379	0.0000
R-squared	0.988977	Mean dependent var		4.295205
Adjusted R-squared	0.987752	S.D. dependent var		0.142581
S.E. of regression	0.015780	Akaike info criterion		-5.328623
Sum squared resid	0.004482	Schwarz criterion		-5.179406
Log likelihood	58.95054	F-statistic		807.4487
Durbin-Watson stat	0.711273	Prob(F-statistic)		0.000000

Dependent Variable: LOG(Y)		Sample: 1991 2008		
Method: Least Squares		Included observations: 18		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.268270	1.089139	-7.591566	0.0000
LOG(X1)	0.721091	0.117261	6.149436	0.0000
LOG(X2)	0.585267	0.022650	25.83952	0.0000
R-squared	0.993260	Mean dependent var		4.731507
Adjusted R-squared	0.992361	S.D. dependent var		0.079766
S.E. of regression	0.006972	Akaike info criterion		-6.942900
Sum squared resid	0.000729	Schwarz criterion		-6.794505
Log likelihood	65.48610	F-statistic		1105.193
Durbin-Watson stat	1.541683	Prob(F-statistic)		0.000000

Interessieren wir uns nun für ein 95 %-Konfidenzintervall für den Parameter des logarithmierten Kapitalinputs (im Modell von 1991 bis 2008), so hätte dieses die folgende Gestalt:

$$\hat{\beta}_2 - t_{0,975; 15} \cdot \hat{\sigma}_{\hat{\beta}_2} \leq \beta_2 \leq \hat{\beta}_2 + t_{0,975; 15} \cdot \hat{\sigma}_{\hat{\beta}_2}$$

$$0,5853 - 2,13 \cdot 0,0227 \leq \beta_2 \leq 0,5853 + 2,13 \cdot 0,0227$$

$$0,5369 \leq \beta_2 \leq 0,6337$$

Demnach liegt β_2 also mit einer Wahrscheinlichkeit von 95 % zwischen 0,5369 und 0,6337. Eine Nullhypothese $H_0: \beta_2 = 0,75$ könnte damit gegenüber $H_A: \beta_2 \neq 0,75$ abgelehnt werden, da der hypothetische Wert 0,75 außerhalb des Konfidenzintervalls liegt.

3.2 Simultanes Testen mehrerer Parameter – F-Test

Während sich der t-Test mit einer Hypothese bzgl. eines einzelnen Regressionsparameters beschäftigt, dient der F-Test zum Testen mehrerer Hypothesen bzgl. einer Gruppe von Parametern.

3.2.1 Hypothesen, F-Statistik und Entscheidungsregel

Die Vorgehensweise des F-Tests ist recht einfach. Zunächst zwingt er der Regressionsgleichung lineare Restriktionen in Form einer Nullhypothese auf. Die daraus resultierende "restiktive" Regressionsgleichung ist diejenige, die gelten würde, wenn die Nullhypothese korrekt ist. Im nächsten Schritt wird diese "restiktive" Gleichung mit OLS geschätzt und der "fit" dieser Gleichung mit dem "fit" der Regressionsgleichung ohne Restriktionen verglichen. Sind sie nicht signifikant verschieden, werden die Restriktionen der Nullhypothese nicht abgelehnt, andernfalls schon. Der Vergleich der "fits" erfolgt über die F-Statistik

$$F = \frac{(RSS_M - RSS) / M}{RSS / (n - K - 1)}, \quad (\text{IV3: 6})$$

wobei RSS_M die Residuenquadratesumme der "restiktiven" Gleichung, RSS die Residuenquadratesumme der Gleichung ohne Restriktionen, M die Anzahl der Restriktionen und $n - K - 1$ die Anzahl der Freiheitsgrade der Regressionsgleichung ohne Restriktionen darstellen.

RSS_M ist stets größer oder gleich RSS, da die Restriktion der Koeffizienten an Stelle einer Bestimmung der Koeffizienten durch OLS nie RSS verringern kann. Im Extremfall sind RSS_M und RSS gleich, was zu $F = 0$ führt. H_0 wird in einem solchen Fall nicht abgelehnt, da die Stichprobe andeutet, dass die Restriktionen korrekt zu sein scheinen. Mit zunehmendem F wird es hingegen immer weniger wahrscheinlich, dass H_0 korrekt ist. Überschreitet F den kritischen F-Wert, so kommt es zur Ablehnung von H_0 . Da die F-Statistik (IV3: 6) unter H_0 einer F-Verteilung folgt und die Freiheitsgrade ihres Zählers bei M, die des Nenners bei $n - K - 1$ liegen, ist der kritische F-Wert nichts anderes als das $(1 - \alpha)$ -Fraktil der $F(M, n - K - 1)$ -Verteilung.

$$\text{Ablehnung von } H_0, \text{ wenn } F > F_{1-\alpha, (M, n-K-1)} \quad (\text{IV3: 7})$$

Zur besseren Veranschaulichung des F-Tests werden wir im folgenden Abschnitt IV 3.2.2 einen Spezialfall betrachten und für diesen die Testprozedur im Detail (inkl. Restriktionen und "restriktiver" Modellgleichung) näher analysieren.

3.2.2 F-Test für die Gesamtsignifikanz

Der F-Test erlaubt es zu testen, ob die Steigungskoeffizienten $\beta_1, \beta_2, \dots, \beta_K$ in einer Regressionsgleichung mit K erklärenden Variablen alle gleichzeitig Null sind. Dazu werden die K Restriktionen $\beta_1 = 0, \beta_2 = 0, \dots, \beta_K = 0$ in einer Nullhypothese über die Signifikanz des Gesamtmodells zusammengefasst:

$$\begin{aligned} H_0: \beta_1 &= \beta_2 = \dots = \beta_K = 0 \\ H_A: \text{mindestens eines der } \beta_k &\text{ ist ungleich Null} \end{aligned} \quad (\text{IV3: 8})$$

Verwenden wir den F-Test zum Test der Gesamtsignifikanz, so vereinfacht sich die zu verwendende F-Statistik (IV3: 6) in diesem Spezialfall zu

$$F = \frac{\text{ESS} / K}{\text{RSS} / (n - K - 1)}, \quad (\text{IV3: 9})$$

wobei ESS die erklärte Quadratesumme ist, die wir im Zusammenhang mit dem Bestimmtheitsmaß kennengelernt haben. Diese neue Form ergibt sich, da $RSS_M = TSS$ und somit $RSS_M - RSS = TSS - RSS = ESS$ gilt.

Bei dieser Art von F-Test hat die "restriktive" Regressionsgleichung die Form

$$Y_i = \beta_0 + \varepsilon_i, \quad (\text{IV3: 10})$$

was zum Ausdruck bringt, dass $\hat{Y}_i = \bar{Y}$. Der F-Test für die Gesamtsignifikanz testet also die Nullhypothese, dass der "fit" der Regressionsgleichung nicht signifikant besser ist als der, den wir durch die Verwendung des arithmetischen Mittels allein erzielen können. Während R^2 also keinen formalen Test für den "Gesamtfit" eines Modells liefern konnte, wird ein solcher durch den F-Test zur Verfügung gestellt. Formen wir (IV3: 9) wie folgt um, so können wir den Zusammenhang

$$\begin{aligned} F &= \frac{\text{ESS} / K}{\text{RSS} / (n - K - 1)} = \frac{n - K - 1}{K} \frac{\text{ESS}}{\text{RSS}} = \frac{n - K - 1}{K} \frac{\text{ESS}}{TSS - ESS} \\ &= \frac{n - K - 1}{K} \frac{\text{ESS} / TSS}{1 - (\text{ESS} / TSS)} = \frac{n - K - 1}{K} \frac{R^2}{1 - R^2} \\ &= \frac{R^2 / K}{(1 - R^2) / (n - K - 1)} \end{aligned} \quad (\text{IV3: 11})$$

zwischen F und R^2 feststellen. Dieser zeigt, dass wenn R^2 den Wert Null annimmt, auch F Null wird. Mit steigendem R^2 steigt auch F. Ist R^2 gleich Eins, geht F gegen unendlich. Ein F-Test, welcher Aussagen über die Gesamtsignifikanz der geschätzten Regressionsgleichung ermöglicht, ist also gleichzeitig ein Signifikanztest für R^2 . Anders ausgedrückt entspricht ein Test der Nullhypothese (IV3: 8) einem Test der Nullhypothese, dass R^2 gleich Null ist.

Beispiel IV 3-6

F-Test für Gesamtsignifikanz



Untersuchen wir zur Veranschaulichung dieses Tests die Gesamtsignifikanz unseres Ge-flügelkonsummodells aus Beispiel IV 3-1. Mit drei erklärenden Variablen formulieren wir also das folgende Hypothesenpaar:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_A: \text{mindestens eines der } \beta_k \text{ ist ungleich Null}$$

Wir haben nun zwei Möglichkeiten um an den Wert für F zu gelangen. Die umständlichere wäre es, ihn "von Hand" zu berechnen. Wir wissen nämlich aus der Modellgleichung, dass K = 3 gilt, und können im Regressionsoutput von EViews n = 40 und $R^2 = 0,9904$ ablesen. Eingesetzt in (IV3: 11) liefern diese Werte den gesuchten F-Wert 1.236,78. Alternativ können wir F auch direkt aus dem EViews-Output ablesen.

Den kritischen F-Wert 2,87 erhalten wir bei einem angenommenen Signifikanzniveau von 5 % durch Nachschlagen des 95 %-Fraktils der F(3; 36)-Verteilung. Da nun hier F deutlich größer als dieser kritische Wert ist, können wir H_0 ablehnen. Das Modell hat also einen signifikanten "Gesamtfit".

Auch für den F-Test können wir die Testentscheidung mittels eines p-Werts durchführen. EViews liefert nämlich auch für den F-Test einen p-Wert (hier: 0,0000). Ist dieser kleiner als das Signifikanzniveau wird H_0 abgelehnt. Dies ist hier der Fall.

Im Rahmen des F-Tests wird in der Literatur häufig die sog. ANOVA (engl. analysis of variance) vorgestellt. Es handelt sich dabei um nichts anderes als eine tabellarische Darstellung der Dekomposition der Varianz von Y, aus der sich die F-Statistik ableiten lässt. Diese ist in der nachfolgenden Tabelle dargestellt. Der Quotient aus beiden durchschnittlichen Quadratesummen liefert (IV3: 9).

Variation von Y	Quadratesummen	Freiheitsgrade	Durchschnitt
erklärt	ESS	K	ESS/K
unerklärt	RSS	n - K - 1	RSS/(n - K - 1)
Summe	TSS	n - 1	

Im Zusammenhang mit dem F-Test auf Gesamtsignifikanz taucht außerdem häufig die Frage auf, warum ein F-Test nicht durch individuelle t-Tests aller Stellungsparameter ersetztbar ist. Nehmen wir an, wir wollen ein Regressionsmodell mit zwei erklärenden Variablen auf Gesamtsignifikanz prüfen, also $H_0: \beta_1 = \beta_2 = 0$ gegenüber $H_A: \text{"mindestens eines der } \beta_k \text{ ist ungleich Null"}$ testen. Führen wir an Stelle eines F-Tests zwei t-Tests mit $H_0: \beta_1 = 0$ gegenüber $H_A: \beta_1 \neq 0$ und $H_0: \beta_2 = 0$ gegenüber $H_A: \beta_2 \neq 0$ so ergibt sich daraus folgende Problematik:³⁹

³⁹ Vgl. Stock, J.H., Watson, M.W. (2003), Kapitel 7.

Sind t_1 und t_2 die t-Statistiken für die beiden Parameter und gehen wir bei einem Signifikanzniveau von 5 % und großen Stichprobenumfang n von einem kritischen Wert von 1,96 aus,⁴⁰ so würden wir die "kombinierte Nullhypothese" $H_0: \beta_1 = 0$ und $\beta_2 = 0$ " ablehnen, wenn entweder $|t_1|$ oder $|t_2|$ den Wert 1,96 überschreitet.

Gehen wir nun davon aus, dass die Zufallsvariablen t_1 und t_2 unkorreliert und damit aufgrund ihrer Normalverteilung auch unabhängig sind. Die "kombinierte Nullhypothese" wird nur dann nicht abgelehnt, wenn sowohl $|t_1|$ als auch $|t_2|$ kleiner oder gleich 1,96 ist. Die Wahrscheinlichkeit, die "kombinierte Nullhypothese" nicht abzulehnen, kann damit aufgrund der Unabhängigkeit von t_1 und t_2 als

$$P(|t_1| \leq 1,96 \wedge |t_2| \leq 1,96) = P(|t_1| \leq 1,96) \cdot P(|t_2| \leq 1,96) = 0,95^2 = 90,25\%$$

bestimmt werden. Die Wahrscheinlichkeit einer korrekten "kombinierten Nullhypothese" abzulehnen ergibt sich daraus als $1 - 0,95^2 = 9,75\%$. Dies entspricht klar nicht dem gewünschten Signifikanzniveau. Eine solche Testprozedur würde die "kombinierte Nullhypothese" also zu oft ablehnen. Dies liegt daran, dass selbst wenn die erste Nullhypothese mittels t_1 nicht abgelehnt werden kann, immer noch eine Ablehnung der zweiten Nullhypothese über t_2 zu einer Ablehnung der "kombinierten Nullhypothese" führen kann.

Sind die beiden t-Statistiken korreliert, verschlimmert sich das beschriebene Problem noch mehr. Je höher die Korrelation zwischen ihnen ist, desto höher wird die Wahrscheinlichkeit ausfallen, eine korrekte Nullhypothese abzulehnen. Das vorgegebene Signifikanzniveau α wird also noch stärker überschritten. Gerade daher empfiehlt es sich nicht die Gesamt signifikanz eines Modells mittels einzelner t-Tests zu untersuchen. Ein F-Test ist hier die einzige vernünftige Vorgehensweise.

3.2.3 Weitere Anwendungen des F-Tests und der Chow-Test

In der Empirie kann der F-Test für eine Vielzahl anderer Zwecke als dem unter IV 3.2.2 beschriebenen eingesetzt werden. Wir können ihn z.B. dazu verwenden **Cobb-Douglas Funktionen** (CD-Funktionen) auf *konstante Skalenerträge* (eine bestimmte prozentuale Erhöhung aller Inputs führt zu genau der gleichen prozentualen Erhöhung des Outputs) zu testen. Formal liegen diese vor, wenn sich die Koeffizienten der CD-Funktion zu Eins summieren.

Beispiel IV 3-7

Test auf konstante Skalenerträge



Wir testen, ob unsere Cobb-Douglas-Produktionsfunktion (nach der Wiedervereinigung) aus Beispiel IV 3-5 konstante Skalenerträge aufweist:

$$\begin{aligned} H_0: \beta_1 + \beta_2 = 1 &\rightarrow \beta_2 = 1 - \beta_1 \\ H_A: \beta_1 + \beta_2 &\neq 1 \end{aligned}$$

Die hinter der F-Statistik steckenden Regressionsgleichungen sind in diesem Fall

$$\ln Y_t = \ln \beta_0 + \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \varepsilon_t$$

und die "restriktive" Gleichung

⁴⁰ Wir verwenden hier ein Normalverteilungsfraktil, da die t-Verteilung bei großen Stichprobenumfängen bekanntlich durch die Normalverteilung approximiert werden kann.

$$\begin{aligned}\ln Y_t &= \ln \beta_0 + \beta_1 \ln X_{1t} + (1 - \beta_1) \ln X_{2t} + \varepsilon_t \\ \ln Y_t &= \ln \beta_0 + \beta_1 (\ln X_{1t} - \ln X_{2t}) + \ln X_{2t} + \varepsilon_t \\ (\ln Y_t - \ln X_{2t}) &= \ln \beta_0 + \beta_1 (\ln X_{1t} - \ln X_{2t}) + \varepsilon_t.\end{aligned}$$

Ein Blick auf den Regressionsoutput von EViews unter Beispiel IV 3-5 verrät uns, dass RSS der Regressionsgleichung ohne Restriktionen bei 0,0007 liegt. Unser RSS_M liegt bei 0,0012, wie uns der nachfolgende Regressionsoutput zeigt.

Dependent Variable: LNY_LNX2		Sample: 1991 2008		
Method: Least Squares		Included observations: 18		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.972579	0.028708	-173.2107	0.0000
LNX1_LNX2	0.370004	0.021093	17.54178	0.0000
R-squared	0.950574	Mean dependent var		-4.470232
Adjusted R-squared	0.947485	S.D. dependent var		0.037384
S.E. of regression	0.008567	Akaike info criterion		-6.577337
Sum squared resid	0.001174	Schwarz criterion		-6.478406
Log likelihood	61.19603	F-statistic		307.7142
Durbin-Watson stat	1.204284	Prob(F-statistic)		0.000000

Wir erhalten damit für unsere Teststatistik nach (IV3: 9)

$$F = \frac{(0,0012 - 0,0007) / 1}{0,0007 / (18 - 2 - 1)} = 10,71.$$

Bei einem Signifikanzniveau von 5 % finden wir in der F(1;15)-Verteilung den kritischen Wert 4,54. Da $F > 4,54$ ist, können wir H_0 ablehnen. Es liegen somit keine konstanten Skalenerträge vor. Eine Erhöhung des Arbeits- und Kapitalinputs um z.B. jeweils 1 % führt also nicht zu einer Erhöhung des Outputs um ebenfalls 1 %.

Aus didaktischen Gründen haben wir uns in diesem Beispiel dazu entschlossen, das "restiktive" Modell direkt zu schätzen und die F-Statistik von Hand zu berechnen. In EViews können wir den F-Wert jedoch auch auf schnelle und einfache Art und Weise erhalten. Wir müssen dazu im Schätzergebnisfenster des originären Modells lediglich View / Coefficient Tests / Wald - Coefficient Restrictions aufrufen, die Restriktion $c(2) + c(3) = 1$ (Hinweis: In EViews besitzt die Modellkonstante den Index $k = 1$, sodass durch diesen Befehl der erste und zweite Steigungsparameter angesprochen werden.) eingeben und erhalten direkt den F- und den dazugehörigen p-Wert. Man beachte dabei, dass es hier aufgrund von Rundungsgenauigkeiten zu einem leicht anderen Wert als oben berechnet kommt.

Als weitere Anwendungsmöglichkeit des F-Tests wollen wir noch kurz auf den sog. **Chow-Test** eingehen.⁴¹ Dieser testet, ob sich bei Schätzung einer Modellgleichung für zwei unterschiedliche Datensätze (z.B. Jahresdaten von 1970-1981 vs. Jahresdaten von 1982-1995) signifikant unterschiedliche Parameterschätzungen ergeben. Ein derartiger Test kann daher als Test für die Stabilität eines unterstellten Modells interpretiert werden, d.h. Hinweise auf Strukturbrüche in der Grundgesamtheit geben.⁴² Die Vorgehensweise des Chow-Tests ist die folgende:

⁴¹ Vgl. Chow, C. (1960).

⁴² Andersartige Strukturbruchtests sind beispielsweise der CUSUM- und CUSUMQ-Test nach Brown, R., Durbin, J., Evans, J. (1975).

1. Durchführung identisch spezifizierter Regressions für die beiden Datensätze und Aufzeichnung von RSS_1 und RSS_2 .
2. Zusammenfassung der Datensätze zu einem Gesamtdatensatz, Durchführung der gleichen Regression für den neuen Datensatz und Aufzeichnung von RSS_G .⁴³
3. Berechnung der F-Statistik

$$F = \frac{(RSS_G - RSS_1 - RSS_2)/(K+1)}{(RSS_1 + RSS_2)/(n_1 + n_2 - 2K - 2)}, \quad (\text{IV3: 12})$$

bei der K die Anzahl der erklärenden Variablen und n_1 und n_2 die Anzahl der Beobachtungen in Datensatz 1 bzw. 2 sind.

4. Ablehnung der Nullhypothese gleicher Parameterschätzungen, wenn die Teststatistik (IV3: 12) größer oder gleich dem kritischen Wert aus der $F(K+1, n_1 + n_2 - 2K - 2)$ -Verteilung ist.

Beispiel IV 3-8

Strukturbruch Wiedervereinigung



In Beispiel 3-5 hatten wir zur Schätzung der Cobb-Douglas Produktionsfunktion für Deutschland das uns zur Verfügung stehende Datenmaterial in zwei Teilperioden (vor und nach der Wiedervereinigung 1990) zerlegt. Wir wollen nun mittels des Chow-Tests prüfen, ob sich unsere Schätzung für den Zeitraum 1970-1990 signifikant von der für den Zeitraum 1991-2008 unterscheidet. Wäre dies nämlich nicht der Fall, könnten wir beide Teilperioden zu einer zusammenlegen.

Zur Durchführung des Chow-Tests benötigen wir zunächst die Schätzergebnisse, die sich bei Verwendung des Gesamtzeitraums 1970-2008 ergeben:

Dependent Variable: LOG(Y)		Sample: 1970 2008		
Method: Least Squares		Included observations: 39		
C	-2.728550	0.284330	-9.598426	0.0000
LOG(X1)	0.139004	0.044310	3.137107	0.0034
LOG(X2)	0.651441	0.022428	29.04592	0.0000
R-squared	0.995238	Mean dependent var	4.496575	
Adjusted R-squared	0.994973	S.D. dependent var	0.249197	
S.E. of regression	0.017668	Akaike info criterion	-5.160305	
Sum squared resid	0.011238	Schwarz criterion	-5.032339	
Log likelihood	103.6259	F-statistic	3761.719	
Durbin-Watson stat	0.935418	Prob(F-statistic)	0.000000	

Mit diesen Ergebnissen bzw. der darin enthaltenen Größe $RSS_G = 0,0112$ können wir dann zusammen mit $RSS_1 = 0,0045$ und $RSS_2 = 0,0007$ aus Beispiel IV 3-5 die Teststatistik

$$F = \frac{(0,0112 - 0,0045 - 0,0007) / (2+1)}{(0,0045 + 0,0007) / (21+18-2\cdot2-2)} = 12,69$$

⁴³ Im Rahmen dieses Testverfahrens ist das Modell, welches mittels des Gesamtdatensatzes geschätzt wird, als "restriktives" Modell zu betrachten.

berechnen. Da dieser größer als der kritische Wert 2,89 aus der F(3;33)-Verteilung bei einem Signifikanzniveau von 5 % ist, können wir H_0 ablehnen. Dies bedeutet, dass wir (trotz auf den ersten Blick recht ähnlicher Schätzungen) signifikant unterschiedliche Ergebnisse in beiden Zeiträumen erhalten und die Teildatensätze nicht zusammenlegen sollten. Bei Schätzung über den gesamten Zeitraum ist der geschätzte Koeffizient von X_1 deutlich geringer.

Wie bereits im Beispiel IV 3-7 bietet EViews auch für dieses Testverfahren eine schnellere Durchführungsmöglichkeit. Dazu müssen wir lediglich im Schätzergebnisfenster des Modells für den Gesamtzeitraum von 1970 bis 2008 den Befehl View / Stability Tests / Chow Breakpoint Test anwählen, den Bruchpunkt 1991 eingeben und schon erhalten wir den (unter Beachtung von Rundungsgenauigkeiten) den obigen F-Wert und zusätzlich einen p-Wert für unsere Testentscheidung.

3.3 Test der Normalverteilungsannahme

Die unter IV 3.1 und IV 3.2 beschriebenen Testverfahren besitzen nur Gültigkeit, wenn die stochastischen Störerme wirklich normalverteilt sind oder aus großen Stichproben stammen. Nur dann folgt nämlich die t-Statistik einer t-Verteilung und die F-Statistik einer F-Verteilung. In der Literatur wird daher häufig die Durchführung von Tests auf Normalverteilung empfohlen. Gebräuchlichster Test auf Normalverteilung der ϵ_i ist dabei der **Jarque-Bera-Test**.⁴⁴ Bei diesem wird in einem ersten Schritt das dritte und vierte Moment (Schiefe und Kurtosis) der Verteilung der ϵ_i mittels der Residuen e_i nach

$$\hat{\mu}_3 = \frac{1}{n} \sum_{i=1}^n e_i^3 \quad \hat{\mu}_4 = \frac{1}{n} \sum_{i=1}^n e_i^4 \quad (\text{IV3: 13})$$

geschätzt. Darin ist $\hat{\sigma}$ der Standardfehler der Regression. Weichen die Schätzwerte für Schiefe und Kurtosis deutlich von 0 bzw. 3 ab, die bei Normalverteilung gelten müssten, ist dies ein Indiz dafür, dass die Störgrößen nicht normalverteilt sind. Aufgrund dieser Überlegungen kombinierten die Entwickler des Jarque-Bera-Tests die Zufallsvariablen (IV3: 13) zu einer einzigen Zufallsvariable, der sog. *Jarque-Bera-Statistik*

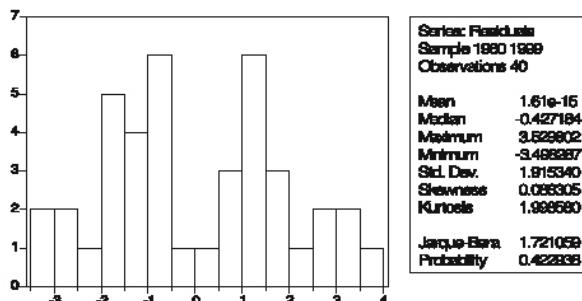
$$JB = n \left(\frac{\hat{\mu}_3^2}{6} + \frac{(\hat{\mu}_4 - 3)^2}{24} \right). \quad (\text{IV3: 14})$$

JB fällt umso größer aus, je stärker Schiefe und Kurtosis von 0 bzw. 3 abweichen. Durch die Quadrierung ist es nicht von Bedeutung, ob die Abweichungen positiv oder negativ sind. Bei Gültigkeit der Nullhypothese (Normalverteilung der ϵ_i) folgt die Jarque-Bera-Statistik *bei großem Stichprobenumfang* einer Chi-Quadrat-Verteilung mit zwei Freiheitsgraden. Dies gilt unabhängig von der Zahl der Freiheitsgrade des Regressionsmodells. Nach Nachschlagen des kritischen Wertes für das gewünschte Signifikanzniveau kann die Nullhypothese der Normalverteilung der ϵ_i abgelehnt werden, wenn dieser kleiner als der JB-Wert ist. Alternativ können wir ablehnen, wenn der zugehörige p-Wert kleiner als das Signifikanzniveau α ist.

⁴⁴ Vgl. Jarque, C. M., Bera, A. K. (1987) und Judge, G. G., et al. (1985).

Beispiel IV 3-9**Jarque-Bera-Test**

Testen wir für unser Geflügelkonsumbeispiel IV 3-1, ob wir von einer erfüllten Normalverteilungsannahme ausgehen können. In EViews ist dies relativ leicht, da dieser im Fenster der geschätzten Gleichung nach einem Mausklick unter View / Residual Tests / Histogram - Normality Test automatisch durchgeführt wird. Wir erhalten das folgende Ergebnis:



Da der p-Wert der Jarque-Bera-Statistik bei 0,4229 und damit deutlich über dem Signifikanzniveau liegt, das wir mit 5 % vorgeben, können wir hier die Nullhypothese der Normalverteilung nicht ablehnen. Dies ist verwunderlich, da die grafische Darstellung der vorliegenden bimodalen Verteilung recht stark von der einer Normalverteilung verschieden ist. Da sich der Jarque-Bera-Test nun aber nur an der Schiefe und Kurtosis orientiert, die bei der vorliegenden Verteilung nicht weit von den Werten einer Normalverteilung (0 und 3) entfernt liegen, kommt es zu keiner Ablehnung der Normalverteilungshypothese. Ergebnisse des Jarque-Bera-Tests sind daher stets kritisch zu betrachten und ggf. durch grafische Hilfsmittel wie beispielsweise Histogramme oder Quantil-Quantil-Plots⁴⁵ zu unterstützen.

Problematisch beim Jarque-Bera-Test ist neben einer im Beispiel IV 3-9 gesehenen alleinigen Fixierung auf Schiefe und Kurtosis, dass er erst bei großen Stichprobenumfängen zu verlässlichen Ergebnissen führt. Die Mächtigkeit des Jarque-Bera-Tests ist daher in kleinen Stichproben sehr gering, weshalb in der Praxis häufig auf einen *Normalverteilungstest verzichtet* wird. Darüber hinaus haben wir es in wissenschaftlichen Studien häufig mit relativ großen Stichprobenumfängen zu tun, sodass wir aufgrund des Zentralen Grenzwertsatzes bei vielen Parameterschätzungen von einer asymptotischen Normalverteilung ausgehen können.

⁴⁵ Bei diesem werden die geordneten Residuen e_1, \dots, e_n als i/n -Quantile ($i = 1, \dots, n$) der Residuenverteilung aufgefasst und den $(i - 0,5)/n$ -Quantilen einer Normalverteilung gegenübergestellt, deren Parameter aus den Residuen geschätzt werden. Der Wert 0,5 ist dabei als Stetigkeitskorrektur zu sehen, die der verbesserten Approximation der diskreten Residuenverteilung durch die stetige Normalverteilung dient. Tragen wir die n Quantilpaare in einem Koordinatensystem ab, erhalten wir den Q-Q-Plot. Erlaubt die Normalverteilung eine hinreichend gute Approximation der Residuenverteilung, dann liegen die Quantilpaare nahe an oder auf der Winkelhalbierenden des Koordinatensystems. Deutliche systematische Abweichungen der Punkte von der Geraden deuten auf eine Unverträglichkeit der Normalverteilungshypothese für die Residuen hin.

4.

Verletzungen der Annahmen des klassischen Regressionsmodells

Bisher sind wir bei den von uns durchgeführten Schätzungen und Tests davon ausgegangen, dass die dem CLRM zugrunde liegenden Annahmen alle erfüllt waren. Welche Folgen eine Verletzung dieser Annahmen nun aber für unsere Regressionsanalyse hat, ist Thema dieses Abschnitts. Wir werden uns hier im Speziellen mit Modellspezifikation (4.1, 4.2), Multikollinearität (4.3), Heteroskedastizität (4.4), Korrelation der Störterme (4.5) und Korrelation zwischen Störterm und erklärenden Variablen (4.6) beschäftigen. Wir schließen den Abschnitt mit einem Überblick über Besonderheiten ab, die bei Zeitreihendaten zu beachten sind (4.7).

4.1 Modellspezifikation I: Variablenwahl

Die Spezifikation eines Regressionsmodells umfasst allgemein zwei Teilbereiche: Die Wahl der korrekten erklärenden Variablen und der korrekten funktionalen Form. Dieser Abschnitt IV 4.1 befasst sich zunächst mit der Variablenwahl, wohingegen der funktionalen Form ein eigenständiger Abschnitt IV 4.2 gewidmet wird.

Allgemein sollten alle Variablen in die Regressionsgleichung aufgenommen werden, die aufgrund der zugrunde liegenden *Theorie* für das Modell essentiell sind. Sollte sich bei der Schätzung herausstellen, dass einige davon keine signifikanten Einflüsse auf die erklärte Variable haben, gilt es die Modelltheorie zu überarbeiten. Ist die zugrunde liegende Theorie mehrdeutig bzw. wenig ausdrücklich hinsichtlich der Variablenwahl ist dies problematisch. Wird eine relevante Variable vernachlässigt, führt dies zu Verletzungen der Annahmen 2a sowie 2b und verzerrt die verbleibenden Parameterschätzer sowie die Schätzer der Standardfehler. Wird eine irrelevante Variable integriert, führt dies zu keinen Verzerrungen der Parameterschätzer jedoch zu höheren Standardfehlern.

4.1.1 Vernachlässigte Variablen

Die Vernachlässigung einer *relevanten* erklärenden Variablen führt oft dazu, dass die verbleibenden Parameter *verzerrt* geschätzt werden. Betrachten wir dazu ein korrekt spezifiziertes Regressionsmodell der Form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \quad (\text{IV4: 1a})$$

in dem wir die Variable X_2 vernachlässigen, da wir z.B. kein Datenmaterial für sie zur Verfügung haben oder sie nicht kennen. Wir können die Vernachlässigung von X_2 auch als Auferlegung einer falschen Restriktion $\beta_2 = 0$ für (IV4: 1a) auffassen.

Die neue Regressionsgleichung hat dann die Form

$$Y_i = \beta_0^* + \beta_1^* X_{1i} + \varepsilon_i^* \quad \text{mit} \quad \varepsilon_i^* = \varepsilon_i + \beta_2 X_{2i}, \quad (\text{IV4: 1b})$$

da der stochastische Störterm die Effekte jeglicher vernachlässigter Variablen auf Y beinhaltet. Es ist bereits in (IV4: 1b) deutlich zu sehen, dass der Erwartungswert des neuen stochastischen Störterms nun von Null abweichen muss. Außerdem ist hier $E(\epsilon_i^* | X_{1i}) \neq 0$ und damit Annahme 2a verletzt.

Die Verzerrung des Schätzers $\hat{\beta}_1^*$ für den verbleibenden Koeffizienten β_1 entsteht nun dadurch, dass dieser zum Teil die Effekte der vernachlässigten Variablen auf Y aufnimmt. Dies lässt sich wie folgt erklären: Die erklärenden Variablen X_1 und X_2 sind in der Realität häufig korreliert. Bei der Schätzung einer Cobb-Douglas Produktionsfunktion ist etwa davon auszugehen, dass die Variablen Arbeit und Kapital korreliert sind. Eine Zunahme des Kapitals erfordert nämlich meist auch eine Zunahme von Arbeit, da z.B. neue Maschinen auch bedient werden müssen. Wird X_2 vernachlässigt, so wird diese Variable ein Teil des "neuen" stochastischen Störterms ϵ^* , sodass ϵ^* und X_1 nun nicht unabhängig sind. Gerade daher kommt es bei Änderungen von X_2 sowohl zu Änderungen von X_1 als auch von ϵ^* , was dazu führt, dass der stochastische Störterm und die erklärende Variable korreliert sind. Dies ist eine klare Verletzung der Annahme 2b des CLRM, die automatisch eine Verletzung von Annahme 2a mit sich bringt.

Sind Annahmen des CLRM nicht erfüllt, verliert das Gauss-Markov-Theorem seine Gültigkeit. Die OLS-Schätzer sind nicht länger BLUE, sodass ggf. eine andere Schätzmethode zu wählen ist. Die Verletzung der Annahme 2b führt allgemein dazu, dass $E(\hat{\beta}_1^*) \neq \beta_1$ gilt. Durch Vernachlässigung von X_2 wird OLS nämlich Veränderungen von Y, die eigentlich auf Änderungen von X_2 zurückzuführen sind, X_1 zuordnen. Formal können wir diese Verzerrung aufzeigen, indem wir zunächst in unsere Schätzformel (IV1: 12)

$$\hat{\beta}_1^* = \frac{\sum x_{1i} y_i}{\sum x_{1i}^2}$$

den Ausdruck (IV4: 1a) einsetzen und Umformungen analog zur Herleitung von (IV2: 20) vornehmen. Wir erhalten also aufgrund der Definition und der Eigenschaften von k_i aus Abschnitt IV 2.3.1 zunächst folgendes Ergebnis:

$$\begin{aligned}\hat{\beta}_1^* &= \frac{\sum x_{1i} (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i)}{\sum x_{1i}^2} = \sum k_i (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i) \\ &= \beta_0 \sum k_i + \beta_1 \sum k_i x_{1i} + \beta_2 \sum k_i x_{2i} + \sum k_i \epsilon_i \\ &= \beta_1 + \beta_2 \sum k_i x_{2i} + \sum k_i \epsilon_i\end{aligned}$$

Die Resubstitution von k_i ergibt schließlich¹

$$\hat{\beta}_1^* = \beta_1 + \beta_2 \frac{\sum x_{1i} x_{2i}}{\sum x_{1i}^2} + \frac{\sum x_{1i} \epsilon_i}{\sum x_{1i}^2}.$$

Nehmen wir wieder an, dass x_1 und x_2 im Gegensatz zu ϵ keine Zufallsvariablen sind und der Erwartungswert des stochastischen Störterms bei Null liegt, führt die Erwartungsbildung aufgrund von $E(\sum x_{1i} \epsilon_i) = \sum E(x_{1i} \epsilon_i) = \sum x_{1i} E(\epsilon_i) = 0$ zu folgender wichtigen Darstellung:

¹ Es ist dabei zu beachten, dass $\sum x_{1i} x_{2i} = \sum x_{1i} (X_{2i} - \bar{X}_2) = \sum x_{1i} X_{2i} - \bar{X}_2 \sum x_{1i} = \sum x_{1i} X_{2i}$ gilt.

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\sum x_{1i}x_{2i}}{\sum x_{1i}^2} = \beta_1 + \beta_2 \underbrace{\frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}}_{\text{Verzerrung}} \quad (\text{IV4: 2})$$

Dies verdeutlicht, dass der Schätzer für β_1 durch Vernachlässigung von X_2 verzerrt wird. Die Verzerrung nimmt mit zunehmendem Stichprobenumfang nicht ab, so dass es zusätzlich zur *Inkonsistenz* des Schätzers kommt. (IV4: 2) zeigt zudem, dass es trotz Vernachlässigung einer Variablen nicht unbedingt zu einer Verzerrung der übrigen Parameterschätzer kommen muss. Ist die vernachlässigte Variable nämlich nicht mit der verbleibenden Variable korreliert, d.h. $\text{Cov}(X_1, X_2) = 0$, so gibt es keine Verzerrung. Ein hoher Wert von $\text{Var}(X_1)$ kann eine bestehende Verzerrung verringern, aber nicht eliminieren, da $\text{Var}(X_1)$ endlich ist.

Formel (IV4: 2) ist besonders wertvoll, da sie uns die Richtung der Verzerrung in Abhängigkeit von der Korrelation zwischen X_1 und X_2 und dem Vorzeichen des wahren Parameters β_2 angeben kann. Ist z.B. der erwartete Koeffizient der vernachlässigten Variable und auch die Korrelation zwischen vernachlässigter und verbleibender Variable positiv, so wird der verbleibende OLS-Schätzer nach oben verzerrt.

Beispiel IV 4-1

Variablenvernachlässigung in Cobb-Douglas-Funktion



Betrachten wir unsere Cobb-Douglas Funktion aus Beispiel IV 3-5. Hier haben wir die erklärenden Variablen X_1 : Arbeitsinput und X_2 : Kapitalinput zur Erklärung des Outputs definiert. Dies lieferte für den Zeitraum 1991 bis 2008 die folgenden Regressionsergebnisse:

$$\begin{aligned} \ln \hat{Y}_t &= -8,2683 + 0,7211 \ln X_{1t} + 0,5853 \ln X_{2t} & n &= 18 (15) \\ &(0,1173) && R^2 = 0,9933 \\ t &= 6,15 & 25,84 \end{aligned}$$

Vernachlässigen wir nun die Variable X_2 im Modell, so würde die Schätzung des damit entstehenden Modells folgende Ergebnisse liefern:

$$\begin{aligned} \ln \hat{Y}_t &= -27,0824 + 3,0128 \ln X_{1t} & n &= 18 (15) \\ &(0,5011) && R^2 = 0,6740 \\ t &= 6,01 \end{aligned}$$

Abgesehen davon, dass sich dadurch der "Modelfit" verschlechtert (kleineres R^2), können wir feststellen, dass sich der geschätzte Koeffizient der verbleibenden Variable deutlich erhöht. Die Richtung dieser Veränderung hätten wir auch mittels (IV4: 2) schätzen können, da danach gilt: Verzerrung = (+) · (+) = (+). X_1 und X_2 sind nämlich positiv korreliert und die Theorie besagt, dass mit zunehmendem Kapitalinput auch der Output steigt. Es ist dabei aber unbedingt zu beachten, dass (IV4: 2) Aussagen über die erwartete Verzerrung und nicht über die beobachtete Veränderung der geschätzten Koeffizienten macht. Die Schätzung nach (IV4: 2) muss also nicht zwangsläufig die tatsächliche Änderung der geschätzten Koeffizienten widerspiegeln. Sie ist jedoch als Anhaltspunkt für die Tendenz der Entwicklung verwendbar.

Mit den Auswirkungen einer Vernachlässigung einer relevanten Variablen auf den Schätzer der Konstante β_0 wollen wir uns an dieser Stelle und auch bei der Behandlung der Verletzung der anderen Annahmen des CLRM nicht näher beschäftigen. Dies beruht einerseits auf der Tatsache, dass wir den Problemen, die mit der Schätzung von β_0 verbunden sind, einen eigenen Abschnitt (vgl. IV 4.2.1) widmen

und andererseits der Schätzer der Regressionskonstanten in praktischen Fragestellungen nahezu nie von Bedeutung ist.

Durch vernachlässigte Variablen kommt es außerdem dazu, dass die Standardfehler der OLS-Schätzer verzerrt geschätzt werden, was insgesamt zu wertlosen Hypothesentests führt, da diese verzerrten Werte z.B. in die Berechnung der t-Statistik eingehen. Wird X_2 vernachlässigt und sind X_1 und X_2 korreliert, so sind z.B. die Varianzen von $\hat{\beta}_1^*$ und $\hat{\beta}_1$ nicht gleich. Im Fall (IV4: 1a) bzw. (IV4: 1b) gilt konkret (vgl. (IV2: 23) und (IV2: 26))

$$\left(\text{Var}(\hat{\beta}_1^*) = \frac{\sigma^2}{\sum x_{1i}^2} \right) < \left(\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1i}^2 (1 - r_{12}^2)} \right).$$

Trotz Modellfehlspezifikation hat der Schätzer $\hat{\beta}_1^*$ eine geringere Varianz. Bei vernachlässigten relevanten Variablen ist zudem der Schätzer für σ^2 verzerrt, da dieser unter den Annahmen 2, 3 und 4 hergeleitet wird und wir bereits wissen, dass eine dieser Annahmen klar verletzt ist.²

Theoretisch ist es recht einfach den beschriebenen Verzerrungen der OLS-Schätzer entgegenzuwirken: Einfach die vernachlässigte Variable in die Regressionsgleichung aufnehmen! Dies ist jedoch einfacher gesagt als getan. Das Problem liegt nämlich darin, dass eine *Verzerrung* durch eine vernachlässigte Variable *schwer zu erkennen* ist. Dies gilt besonders dann, wenn wir der Überzeugung sind, dass das verwendete Modell korrekt spezifiziert ist. Wir werden uns meist erst dann Gedanken machen, wenn das Vorzeichen eines geschätzten Regressionsparameters signifikant anders ist, als wir es in der Modelltheorie erwarten. In einem solchen Fall ist es nämlich sehr wahrscheinlich, dass eine Verzerrung aufgrund vernachlässigter Variablen vorliegt. Nach derartigem Feststellen einer Verzerrung aufgrund einer vernachlässigten Variablen, stellt die *Frage, welche Variable* schließlich in die Gleichung *eingefügt* werden soll, ein weiteres Problem dar. Oft konzipieren wir unsere Modelle ja gerade so, dass sie alle relevanten Variablen berücksichtigen. Unerfahrene Ökonometriker neigen in einer solchen Situation dazu, ihre Modelle mit verschiedenen, theoretisch nicht immer unbedingt optimalen Zusatzvariablen zu testen und entscheiden sich letztendlich für diejenigen, die die Verzerrung beseitigen (korrekte Vorzeichen, wünschenswerte t-Werte). Die so geschätzten Gleichungen haben dann jedoch meist nur für den jeweiligen Datensatz Gültigkeit, für den sie geschätzt wurden. In anderen Stichproben erweisen sie sich als untauglich, da sie schließlich nicht die Charakteristika der Grundgesamtheit beschreiben. An Stelle eines derartigen "Rumprobierens" sollten wir besser Variablen suchen, die mit der Theorie im Einklang sind und daher auch sinnigerweise ins Modell aufgenommen werden sollten. Wichtige Hinweise geben dabei *anerkannte ökonomische Modelle* und auch Variablen, die sich in anderen *empirischen Untersuchungen* als wichtige Einflussfaktoren auf unsere zu erklärende Variable herauskristallisiert haben.

4.1.2 Überflüssige Variablen

Neben vernachlässigten Variablen können auch überflüssige Variablen das Regressionsergebnis beeinträchtigen. Während vernachlässigte Variablen bei $\text{Cov}(X_1, X_2) \neq 0$

² Vgl. dazu Rao, P., Miller, R. L. (1971) oder von Auer, L. (2007).

zur Verzerrung der Parameter- und Varianzschätzer führen, erhöhen überflüssige Variablen lediglich die Varianz der Parameterschätzer und führen nicht zu ihrer Verzerrung. Betrachten wir dazu das korrekt spezifizierte Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i, \quad (\text{IV4: 3a})$$

in welches jedoch eine überflüssige Variable eingefügt wurde, die zum Modell

$$Y_i = \beta_0^* + \beta_1^* X_{1i} + \beta_2^* X_{2i} + \varepsilon_i^* \quad \text{mit} \quad \varepsilon_i^* = \varepsilon_i - \beta_2^* X_{2i} \quad (\text{IV4: 3b})$$

führt. Wir haben unter IV 4.1.1 gesehen, dass die Auferlegung einer falschen Restriktion zu verzerrten Schätzern führt. Wir können das auch so interpretieren, dass unsere Schätzung von einer *falschen Information* beeinträchtigt wird. Durch die Aufnahme einer überflüssigen Variablen wird von uns hingegen eine *korrekte Information vernachlässigt*, und zwar, dass in der Grundgesamtheit $\beta_2^* = 0$ gilt. Dadurch ist (IV4: 3b) aber nicht falsch und hat daher auch keine Verzerrung (und keine Inkonsistenz) des Schätzers für β_1^* zur Folge.³

$$E(\hat{\beta}_1^*) = \beta_1 \quad (\text{IV4: 4})$$

Das Hinzufügen überflüssiger Variablen führt nicht zur Verzerrung der Schätzer der Steigungskoeffizienten des korrekten Modells. Dies lässt sich auch für den Schätzer der Konstanten im Regressionsmodell zeigen. Außerdem kann gezeigt werden, dass $E(\hat{\beta}_0^*) = 0$ gilt.⁴

Überflüssige Variablen führen bei korrelierten erklärenden Variablen dazu, dass die Varianz des Schätzers von β_1 erhöht wird, also (vgl. (IV2: 23) und (IV2: 26))

$$\left(\text{Var}(\hat{\beta}_1^*) = \frac{\sigma^2}{\sum x_{1i}^2 (1 - r_{12}^2)} \right) > \left(\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1i}^2} \right)$$

gilt.⁵ Dies führt wiederum dazu, dass die absolute Höhe der t-Werte abnimmt, was die Ablehnung von Hypothesen bzgl. Koeffizienten von Null erschwert. Wir können also sagen, dass die Schätzer des fehlspezifizierten Modells nicht effizient sind, da es andere Schätzer (im korrekten Modell) gibt, die eine geringere Varianz aufweisen. Zudem ist zu erwähnen, dass überflüssige Variablen auch oft zu einem sinkendem \bar{R}^2 (nicht jedoch R^2) führen, was das nachfolgende Beispiel zeigt.

Beispiel IV 4-2

Überflüssige Variable im Geflügelkonsummodell

In unserem Geflügelkonsumbeispiel IV 3-1 ergab sich die folgende Schätzung:

$$\begin{aligned} \hat{Y}_i &= 27,5939 - 0,6072 X_{1i} + 0,0922 X_{2i} + 0,2449 X_{3i}, & n &= 40 \quad (36) \\ (0,1571) &\quad (0,0399) & (0,0111) & \bar{R}^2 = 0,9896 \\ t = & -3,86 & 2,31 & 22,07 \end{aligned}$$

Fügen wir nun eine überflüssige Variable X_4 hinzu, die aus erzeugten Zufallszahlen besteht, so erhalten wir folgendes Ergebnis:

³ Für einen formalen Beweis vgl. Aufgabe I6.

⁴ Die Beweise und auch Hinweise zum Beweis der Konsistenz der Schätzer finden sich z.B. in Pindyck, R.S., Rubinfeld, D.L. (1998), Kapitel 7.

⁵ Der Schätzer von σ^2 bleibt unverzerrt, da keine der Annahmen 2, 3 und 4 verletzt wird.

$$\hat{Y}_t = 27,0824 - 0,5915X_{1t} + 0,0930X_{2t} + 0,2451X_{3t} + 0,5771X_{4t} \quad n = 40 (36)$$

(0,1624)	(0,0404)	(0,0112)	(1,2376)	$\bar{R}^2 = 0,9893$
t = -3,64	2,30	21,82	0,47	

Es fällt zunächst auf, dass \bar{R}^2 gefallen ist, was anzeigt, dass sich der "fit" des Modells verschlechtert hat. Außerdem ist festzustellen, dass sich keiner der geschätzten Regressionskoeffizienten signifikant verändert hat. Die Standardfehler der OLS-Schätzer haben sich jedoch alle erhöht, was zu durchgehend niedrigeren t-Werten geführt hat. Besonders auffällig ist der äußerst niedrige t-Wert der neuen Variablen. Dieser deutet an, dass der geschätzte Koeffizient der neuen Variablen nicht signifikant von Null verschieden ist. Sie sollte also nicht in die Gleichung aufgenommen werden.

4.1.3 Modellspezifikationskriterien und Spezifikationstests

Wie wir im letzten Beispiel gesehen haben, scheint es eine Reihe von Kriterien bzw. zu beantwortende Fragen zu geben, anhand derer wir beurteilen können, ob eine Variable in ein Regressionsmodell aufgenommen werden sollte oder nicht. Diese Spezifikationskriterien können wir zu den folgenden zusammenfassen:⁶

1. **Theorie:** Rechtfertigt die zugrunde liegende Modelltheorie die Aufnahme der Variable in die Regressionsgleichung?
2. **t-Test:** Ist der geschätzte Koeffizient der Variablen signifikant mit dem erwarteten Vorzeichen?
3. **\bar{R}^2 :** Verbessert sich der "Gesamtfit" der Gleichung, wenn die Variable hinzugefügt wird?
4. **Verzerrung:** Kommt es bei Hinzufügung der neuen Variablen zu signifikanten Veränderungen der bestehenden geschätzten Regressionskoeffizienten?

Können diese Fragen bejaht werden, gehört die betroffene Variable in die Regressionsgleichung. Da die Spezifikationskriterien 1 bis 4 jedoch schnell an ihre Grenzen geraten, wenn es z.B. darum geht, sich zwischen zwei Spezifikationen zu entscheiden, die beide alle 4 Punkte erfüllen, muss häufig auf mehr formale Spezifikationskriterien zurückgegriffen werden. In der Praxis wird hier häufig der sog. Ramsey RESET (engl. regression specification error test), das Akaike-Informationskriterium und das Schwarz-Kriterium herangezogen.

Der **Ramsey RESET**⁷ dient der Aufdeckung genereller Formen von funktionalen Fehlspezifikationen. Ist das Originalmodell (IV4: 5a) korrekt spezifiziert, dann sollten nichtlineare Funktionen der erklärenden Variablen nicht signifikant sein, wenn sie der Gleichung (IV4: 5a) hinzugefügt werden. \hat{Y} stellt eine Linearkombination der erklärenden Variablen dar. Potenzen von \hat{Y} agieren als Proxies für jede mögliche (unbekannte) vernachlässigte Variable (oder falsche funktionale Form). Kann durch einen F-Test gezeigt werden, dass eine derart erweiterte Gleichung einen besseren "Gesamtfit" aufweist als die originäre Gleichung, so ist dies ein Hinweis dafür, dass ein Spezifikationsfehler vorliegt. Liegt kein Spezifikationsfehler vor, sollten die geschätzten Koeffizienten dieser zusätzlichen Terme nicht signifikant von Null verschieden sein. Die generelle Vorgehensweise des Ramsey RESET können

⁶ Vgl. Studenmund, A.H. (2007), Kapitel 6.

⁷ Vgl. Ramsey, J.B. (1969).

wir im Falle eines Regressionsmodells mit zwei erklärenden Variablen folgendermaßen beschreiben: Zunächst wird das vorliegende Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \quad (\text{IV4: 5a})$$

geschätzt. Die daraus resultierenden Werte \hat{Y}_i werden dann zur Schätzung des erweiterten Modells

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + \beta_5 \hat{Y}_i^4 + \varepsilon_i \quad (\text{IV4: 5b})$$

verwendet. Anschließend wird der "Gesamtfit" beider Modelle durch Testen der Nullhypothese $H_0: \beta_3 = \beta_4 = \beta_5 = 0$ gegenüber $H_A: \text{"mindestens eines der } \beta_k \text{ mit } k = 3,4,5 \text{ ist ungleich Null"}$ mittels des F-Tests verglichen. Die zu verwendende Teststatistik und Testentscheidungsregel sind (IV3: 6) und (IV3: 7). Dabei bezieht sich $n - K - 1$ wieder auf die originäre Gleichung, während M die Anzahl der Zusatzterme der erweiterten Gleichung darstellt.

Beispiel IV 4-3

Ramsey RESET im Geflügelkonsummodell



Nehmen wir an, wir vernachlässigen in unserem Geflügelnachfragebeispiel IV 3-1 die Variable X_2 : Rindfleischpreis. Untersuchen wir nun mit Hilfe des Ramsey RESET, ob wir diese Fehlspezifikation aufdecken können. Die Schätzung des Modells ohne X_2 ergibt zunächst:

$$\begin{aligned} \hat{Y}_i &= 27,5415 - 0,4162 X_{1i} + 0,2682 X_{3i} & n = 40 (37) \\ &\quad (0,1413) \quad (0,0049) \quad R^2 = 0,9884 \\ t &= -2,95 \quad 55,00 \end{aligned}$$

Wir könnten nun daraus die zusätzlichen Terme \hat{Y}_i^2 , \hat{Y}_i^3 und \hat{Y}_i^4 berechnen, das erweiterte Modell schätzen und die RSS-Werte aus beiden Modellen zur Bestimmung der F-Statistik verwenden. Viel einfacher ist die Testdurchführung in EViews. Im Menü der geschätzten Gleichung ist dazu lediglich unter View / Stability Tests / Ramsey RESET Test die Anzahl der hinzuzufügenden "Proxy-Variablen" anzugeben. Dies liefert:

Ramsey RESET Test:

F-statistic	12.15741	Prob. F(3,34)	0.0000
Log likelihood ratio	29.15433	Prob. Chi-Square(3)	0.0000

Test Equation:

Dependent Variable: Y		Sample: 1960 1999		
Method: Least Squares		Included observations: 40		
	Coefficient	Std. Error	t-Statistic	Prob.
C	243.8141	36.75197	6.634041	0.0000
X1	-6.295676	0.984546	-6.394498	0.0000
X3	4.204614	0.662031	6.351085	0.0000
FITTED^2	-0.412208	0.070537	-5.843842	0.0000
FITTED^3	0.004891	0.000854	5.725530	0.0000
FITTED^4	-2.08E-05	3.71E-06	-5.605200	0.0000
R-squared	0.994676	Mean dependent var		50.56725
Adjusted R-squared	0.993893	S.D. dependent var		19.53879
S.E. of regression	1.526926	Akaike Info criterion		3.821871
Sum squared resid	79.27109	Schwarz criterion		4.075203
Log likelihood	-70.43742	F-statistic		1270.385
Durbin-Watson stat	1.217899	Prob(F-statistic)		0.0000000

Mit einem kritischen F-Wert von 2,88, der sich bei einem 5 %-Signifikanzniveau aus der F(3;34)-Verteilung ergibt, kann aufgrund des F-Werts des Ramsey RESET von 12,1574 die Nullhypothese $H_0: \beta_3 = \beta_4 = \beta_5 = 0$ abgelehnt werden. Auch anhand des p-Werts der F-Statistik des RESET-Tests von 0,0000 können wir H_0 bei jedem Signifikanzniveau ablehnen. Dies erlaubt uns die Aussage, dass es in der Modellgleichung einen Spezifikationsfehler zu geben scheint. Dieses Ergebnis ist nicht verwunderlich, da wir ja anfangs bewusst eine wichtige Variable ausgeschlossen haben. Wir haben hier \hat{Y}_t^2 , \hat{Y}_t^3 und \hat{Y}_t^4 gewählt. Der Test hätte die Fehlspezifikation aber auch bei z.B. nur \hat{Y}_t^2 und \hat{Y}_t^3 aufgedeckt.

Wichtig im Zusammenhang mit dem Ramsey RESET ist, dass dieser nur Aussagen dazu machen kann, ob ein Spezifikationsfehler vorliegt. Wo dieser genau liegt (vernachlässigte Variable, falsche funktionale Form, ...) wird nicht bestimmt. Dieser Test empfiehlt sich daher nicht zur konkreten Fehlersuche, sondern zur Sicherstellung, dass ein wohlspezifiziertes Modell auch als solches bezeichnet werden kann.

Eine andere Kategorie von Spezifikationskriterien befasst sich mit der Modifikation der Residuenquadratesumme RSS, um ein Maß für den "fit" einer Gleichung zu erhalten. Typischer Vertreter dieser Kategorie ist das bereits behandelte \bar{R}^2 . Neben diesem wurden auch das sog. **Akaike-Informationskriterium** (AIC) und das **Schwarz-Kriterium** (SC) entwickelt.⁸ Wie

$$\text{AIC} = \ln\left(\frac{\text{RSS}}{n}\right) + \frac{2 \cdot (K+1)}{n} \quad (\text{IV4: 6})$$

$$\text{SC} = \ln\left(\frac{\text{RSS}}{n}\right) + \frac{\ln(n) \cdot (K+1)}{n} \quad (\text{IV4: 7})$$

zeigen, wird bei diesen Maßen RSS hinsichtlich Stichprobenumfang n und Zahl der erklärenden Variablen K angepasst. Sie helfen also zu beurteilen, ob der verbesserte "fit", der durch Hinzufügen einer Variablen erreicht werden kann, die abnehmenden Freiheitsgrade und die zunehmende Modellkomplexität wert ist.

Die Anwendung beider Maße ist gleich. Es werden die alternativen Spezifikationen geschätzt und für jede der Wert von AIC bzw. SC berechnet. Je geringer AIC bzw. SC dabei ausfallen, desto besser ist die Spezifikation. Im Vergleich zu \bar{R}^2 kann gesagt werden, dass AIC und SC das Hinzufügen einer Variable stärker "betrachten". Bei einer Spezifikation, die AIC bzw. SC minimiert ist es daher weniger wahrscheinlich, dass sie eine irrelevante Variable enthält, als bei einer, die \bar{R}^2 maximiert.

Beispiel IV 4-4

AIC und SC im Geflügelkonsummodell

Betrachten wir die zwei möglichen Spezifikationen unseres Geflügelkonsummodells aus Beispiel IV 3-1:

$$\begin{array}{llll} \hat{Y}_t = 27,5939 - 0,6072X_{1t} + 0,0922X_{2t} + 0,2449X_{3t} & n = 40 \text{ (36)} \\ (0,1571) \quad (0,0399) \quad (0,0111) & \bar{R}^2 = 0,9896 \\ t = -3,86 \quad 2,31 \quad 22,07 & \text{RSS} = 143,07 \end{array}$$

⁸ Vgl. Akaike, H. (1981) und Schwarz, G. (1978).

$$\hat{Y}_t = 27,5415 - 0,4162X_{1t} + 0,2682X_{3t} \quad n = 40 \text{ (37)}$$

$$(0,1413) \quad (0,0049) \quad R^2 = 0,9884$$

$$t = -2,95 \quad 55,00 \quad RSS = 164,31$$

Unter Verwendung dieser Ergebnisse können wir für das korrekte Modell

$$AIC = \ln(143,07 / 40) + 2 \cdot (4) / 40 = 1,47$$

$$SC = \ln(143,07 / 40) + \ln(40) \cdot 4 / 40 = 1,64$$

berechnen, während wir für das Modell mit der vernachlässigten Variablen

$$AIC = \ln(164,31 / 40) + 2 \cdot (3) / 40 = 1,56$$

$$SC = \ln(164,31 / 40) + \ln(40) \cdot 3 / 40 = 1,69$$

erhalten. Eine solche Berechnung muss natürlich nicht "von Hand" durchgeführt werden. AIC und SC sind Standardbestandteile eines jeden EViews-Regressionsoutputs und können daher direkt abgelesen werden. In diesem Fall erkennen wir deutlich, dass AIC und SC beim korrekten Modell niedriger sind. Das Modell mit drei erklärenden Variablen ist also zu bevorzugen.

Wie dieses Beispiel zeigt, können wir im Vergleich zum Ramsey RESET zusammenfassend sagen, dass dieser die Existenz eines Spezifikationsfehlers feststellen kann, während AIC und SC dazu dienen, zwei alternative Spezifikationen zu vergleichen.

4.1.4 Verzögerte erklärende Variablen

Bisher sind wir davon ausgegangen, dass sich die erklärte Variable in einer Periode t durch die Ausprägungen der erklärenden Variablen in Periode t (*kontemporäre Variablen*) beschreiben lässt. Wir bezeichnen Modelle, die einen derartigen Zweck verfolgen, auch als **statische Modelle**. Was wir noch nicht berücksichtigt haben, ist die Tatsache, dass in der Regel auch Ausprägungen der erklärenden Variablen der Vorperiode t-1 Einfluss auf Y in Periode t haben können. Wir müssen also die Möglichkeit zulassen, dass eine gewisse Zeit verstreicht, bis die Veränderung einer erklärenden Variablen Einfluss auf die erklärte Variable hat. Wir bezeichnen diesen Zeitraum auch als Verzögerung (engl. lag). Viele ökonometrische Modelle, wie z.B.

$$Y_t = \beta_0 + \beta_1 X_{1,t-1} + \beta_2 X_{2,t} + \epsilon_t, \quad (\text{IV4: 8})$$

beinhalten **verzögerte erklärende Variablen**. Das Subskript t-1 bedeutet hier, dass die Beobachtung von X₁ aus einer Periode vor t stammt. Wir sagen, dass das Modell (IV4: 8) bzgl. der Variablen X₁ eine *Verzögerung von einer Periode* oder einen Lag von Eins hat. Bei derartigen Modellen sprechen wir auch von **einfachen Lag-Modellen**. Sie gehören zur Klasse der **dynamischen Modelle** (Modelle mit Variablen zu unterschiedlichen Zeitpunkten). Der Regressionskoeffizient einer Variablen X_k mit einer Verzögerung von einer Periode gibt (unter Konstanthaltung der anderen erklärenden Variablen) an, welche Veränderung von Y in t durch eine Erhöhung von X_k (um eine Einheit) in der Vorperiode hervorgerufen wird.

Beispiel IV 4-5

Verzögerte Variablen

Nehmen wir zum besseren Verständnis von (IV4: 8) an, dass Y_t das Angebot an Mais in Periode t, X_{1,t-1} den Preis von Mais in Periode t-1 und X_{2,t} den durchschnittlichen Lohn von

Hilfsarbeitern im landwirtschaftlichen Bereich in Periode t darstellen. Der Preis von Mais ist als verzögerte Variable ins Modell aufgenommen, da die Bauern bei Preiserhöhungen nicht sofort ihr Angebot erhöhen können. (Der Mais muss schließlich erst wachsen.) Das Angebot wird daher erst in der Folgeperiode zunehmen. β_1 würde in einem solchen Fall angeben, wieviele zusätzliche Einheiten (z.B. Tonnen) Mais in Periode t aufgrund einer Vorjahrespreiserhöhung von einer Einheit (z.B. Euro je Tonne) angeboten werden. Steigt das Lohnniveau für Hilfskräfte, können die Bauern relativ schnell darauf reagieren und Entlassungen veranlassen. Dieses Lohnniveau muss daher nicht als verzögerte Variable aufgenommen werden.

Zum besseren Verständnis des Begriffs "Verzögerung" betrachten wir folgende Tabelle, die hypothetisches Datenmateriel ($n = 10$) für unser Maisangebotsmodell darstellt:

t	1	2	3	4	5	6	7	8	9	10
Y_t	1.500	1.400	1.400	1.200	1.250	1.100	1.300	1.400	1.350	1.050
X_{1t}	150	160	130	120	115	110	105	100	120	140
X_{1t-1}		150	160	130	120	115	110	105	100	120
X_2	7,5	7,5	8,0	9,0	9,5	9,5	10,0	10,5	10,5	12,0

Wir erkennen, dass der Wert der verzögerten Variablen jeweils dem Wert aus der Vorperiode entspricht. Außerdem zeigt sich, dass das Datenmaterial, dass für unsere Schätzung zur Verfügung steht, von 10 auf 9 Beobachtungen reduziert wird, da uns für $t = 1$ keine Beobachtung von X_{1t-1} zur Verfügung steht. Je höher wir also die Verzögerung wählen, desto mehr Beobachtungen gehen verloren.

Modelle mit Verzögerungen von nur einer Periode der Art (IV4: 8), die die kontemporäre Variable nicht zusätzlich enthalten, können wie jedes andere lineare Regressionsmodell ohne Probleme mit OLS geschätzt werden. Hat ein Modell allerdings die Gestalt

$$Y_t = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_s X_{t-s} + \epsilon_t , \quad (\text{IV4: 9})$$

weist die Schätzgleichung also neben der Variablen X_t auch deren Verzögerungen um 1 bis s Perioden auf, ist dessen Schätzung manchmal problematisch (siehe unten). Wir sprechen bei derartigen **dynamischen Modellen** auch von **distributiven Lag-Modellen**, da Y hier eine Funktion gegenwärtiger und vergangener X -Werte ist, der Einfluss von X auf Y also über eine bestimmte Anzahl von Perioden verteilt wird. Die Koeffizienten β_0 bis β_s messen die Effekte der verzögerten Werte von X auf den gegenwärtigen Wert von Y . In vielen ökonomischen Anwendungen werden wir davon ausgehen, dass der Einfluss von X auf Y mit zunehmender Länge der Verzögerung (angezeigt durch Subskript der Koeffizienten) abnimmt. Die ersten Werte β_0 und β_1 sollten also absolut betrachtet größer sein als z.B. β_2 und β_s .

Beispiel IV 4-6

Distributive Lag-Modelle

Geldmenge: Typisches Beispiel für ein distributives Lag-Modell bietet die Analyse des Einflusses einer Geldmengenveränderung (ΔM) auf das Bruttoinlandsprodukt eines Landes. Aufgrund der Komplexität der Märkte dauert es nämlich eine Weile, bis die Wirtschaft vollständig auf ΔM reagiert hat.

Werbeausgaben: Auch die Tätigung von Werbeausgaben dürfte sich nicht unmittelbar auf den Umsatz einer Bank auswirken. Viele Banken investieren nämlich erhebliche Summen

in Marketingmaßnahmen gegenüber Kindern und Jugendlichen. Hintergrund derartiger Aktivitäten ist es, die Beworbenen bereits jetzt an die Bank zu binden, sodass sich mit dem Berufseinstieg der Beworbenen Geschäfte für die Bank ergeben. Ein heute vergebene Werbegeschenk zieht also erst mit zum Teil erheblicher Verzögerung einen Umsatz für die Bank nach sich.

Die Schätzung einer Gleichung wie (IV4: 9) mit OLS verursacht eine Reihe von Problemen. Zu diesen gehören beispielsweise folgende:

1. Es besteht die Gefahr, dass die verschiedenen verzögerten Werte von X stark multikollinear sind, was die Schätzungen der Koeffizienten unpräzise macht (vgl. Abschnitt IV 4.3).
2. Es kommt zu einer deutlichen Abnahme der Freiheitsgrade. Dies ist einerseits darauf zurückzuführen, dass für jedes verzögerte X ein Koeffizient geschätzt werden muss, was über ein steigendes K eine Abnahme der Freiheitsgrade $n - K - 1$ bedingt. Zusätzlich reduziert sich der Stichprobenumfang n für jede einzelne verzögerte Variable um Eins.

Aufgrund dieser Schwierigkeiten ist es Standard geworden, Gleichungen der Art (IV4: 9) unter gewissen Annahmen stark zu vereinfachen. Eine gängige Modifikation besteht in dem Versuch, alle verzögerten erklärenden Variablen durch den um eine Periode verzögerten Wert der erklärten Variablen zu ersetzen. Mit derartigen Formen *dynamischer Modelle* werden wir uns unter IV 4.7 noch näher befassen.

4.2 Modellspezifikation II: Funktionale Form

Nach Wahl der erklärenden Variablen gilt es die funktionale Form der Regressionsgleichung festzulegen. Nach Möglichkeit sollte diese nicht allein anhand der vorliegenden Stichprobe ausgewählt werden. Es könnte zwar so ein sehr guter "Modell-fit" erreicht werden, doch besteht dabei die Gefahr, dass zufallsbedingte Besonderheiten der Stichprobe modelliert werden, die für die allgemeinen Charakteristika der Grundgesamtheit unbedeutend sind. Resultat wäre ein geschätztes Modell, das für Prognosezwecke untauglich ist. Insbesondere sollten daher auch inhaltliche Überlegungen (z.B. das Gesetz des abnehmenden Grenzertrages) eine Rolle spielen.

Folge einer falschen funktionalen Form ist im Allgemeinen eine Verletzung der Annahme 2a und damit eine Verzerrung der Parameter- und Varianzschätzer. Selbst korrekte erklärende Variablen können so in einer Schätzung als nicht signifikant oder mit falschem Vorzeichen auftreten.

4.2.1 Bedeutung des konstanten Terms

Im linearen Regressionsmodell stellt die Konstante β_0 den Erwartungswert von Y für den Fall dar, dass alle erklärenden Variablen den Wert Null annehmen. Er besteht aus mindestens **drei Komponenten**:

1. dem wahren Wert der Konstanten, den wir im Folgenden mit β_0^w bezeichnen,
2. dem Erwartungswert des stochastischen Störterms der korrekt spezifizierten Gleichung, sofern dieser nicht Null ist,
3. und dem konstanten Einfluss von Spezifikationsfehlern (z.B. vernachlässigte Variablen, falsche funktionale Form, Messfehler bei der Ermittlung der Werte der erklärten Variablen).

Nehmen wir zur Veranschaulichung an, dass der Erwartungswert des stochastischen Störterms ξ_i eines korrekt spezifizierten Modells (wir blenden also den obigen Punkt 3 zunächst aus) der Form

$$Y_i = \beta_0^w + \beta_1 X_i + \xi_i \quad (\text{IV4: 10a})$$

den Wert a mit $a > 0$ annimmt, also $E(\xi_i) = a$ gilt. Damit wäre die Annahme verletzt, dass der Erwartungswert des Störterms Null sein muss. Das Modell ist also korrekt spezifiziert, aber der Erwartungswert des Störterms ist von Null verschieden. Subtrahieren wir nun a vom Störterm und Addieren a zu β_0^w , so erhalten wir

$$Y_i = \underbrace{(\beta_0^w + a)}_{\beta_0} + \beta_1 X_i + \underbrace{(\xi_i - a)}_{\epsilon_i}. \quad (\text{IV4: 10b})$$

Der neue Störterm $\epsilon_i = \xi_i - a$ besitzt nun einen Erwartungswert von Null, da $E(\epsilon_i) = E(\xi_i - a) = E(\xi_i) - a = a - a = 0$ gilt. Durch die Präsenz eines konstanten Terms, der einen Nicht-Null-Erwartungswert des stochastischen Störterms aufnimmt, wird also ein Erwartungswert des stochastischen Störterms von Null sichergestellt (solange alle anderen Annahmen erfüllt sind). Wir sehen auch, dass die Regressionskonstante β_0 mehr als nur eine Komponente aufweist. Im Falle eines fehlspezifizierten Modells (z.B. vernachlässigte Variable) würde der konstante Term zusätzlich den obi-

gen Punkt 3 aufnehmen, da auch hier der Erwartungswert des stochastischen Störterms von Null abweicht.

Unglücklicherweise lassen sich nun diese drei Komponenten der Konstanten β_0 nicht unterscheiden, da wir schließlich nur eine Schätzung für β_0 erhalten. Dies ist der Grund, warum wir β_0 anders als die anderen Regressionskoeffizienten analysieren müssen⁹ und wir β_0 bereits beim t-Test außer Acht gelassen haben. Grundsätzlich gilt, dass wir uns weder auf Schätzungen von β_0 beim Rückschluss auf β_0^w verlassen dürfen (außer bei korrekt spezifizierten Modellen mit erwartetem Störterm von Null), noch den Term aus der Regressionsgleichung ausschließen sollten.

Warum wir uns nicht auf **Schätzungen des konstanten Terms** für Rückschlüsse auf den wahren β_0^w -Wert der Grundgesamtheit verlassen sollten, hat folgenden Grund: β_0 beinhaltet, wie wir gesehen haben, nicht nur den wahren Wert β_0^w , sondern auch andere Komponenten. Der geschätzte Wert von β_0 würde also ggf. anders aussehen, wenn es diese nicht gäbe. Die Durchführung eines t-Tests für β_0 ist daher meist bedeutungslos.

Auch wenn wir uns nicht auf Schätzungen von β_0 verlassen können, ist dies kein Grund für eine **Vernachlässigung des konstanten Terms**. Ein Ausschluss von β_0 aus dem Regressionsmodell würde nämlich dazu führen, dass ein Nicht-Null-Erwartungswert des Störterms in die Schätzungen der anderen Koeffizienten "gezwungen" wird, was zu deren Verzerrung führt. Nehmen wir etwa an, die wahre Regressionsfunktion eines einfachen linearen Modells weist ein positives β_0 mit sehr geringer Steigung bzw. niedrigem Wert β_1 auf (vgl. Abbildung IV 15). Schätzen wir nun aber ohne Berücksichtigung von β_0 , so wird unsere geschätzte Regressionsgerade durch den Ursprung gezwungen und damit eine deutlich größere Steigung besitzen. Der OLS-Schätzer für β_1 ist also nach oben verzerrt und zeigt auch einen höheren t-Wert. Dies könnte aufgrund eines t-Tests zu dem Schluss führen, dass β_1 sehr hoch ist, obwohl dies nicht zutrifft.

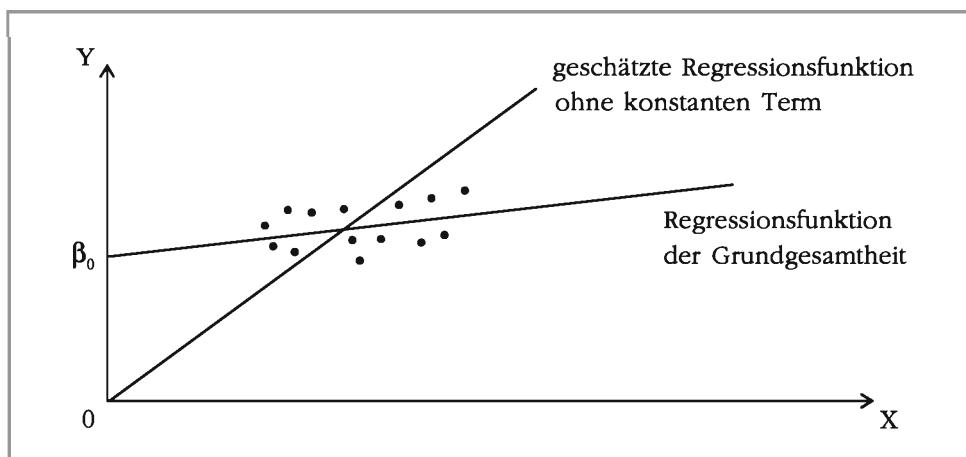


Abbildung IV 15: Vernachlässigung des konstanten Terms

⁹ Vgl. Allen, R.C., Stone, J.H. (2005).

Selbst wenn die Modelltheorie keinen konstanten Term vorsieht, sollte ein solcher trotzdem in die Gleichung aufgenommen werden, da wir uns prinzipiell nie wirklich sicher sein können, ob unser Modell korrekt spezifiziert ist, d.h. ob wir alle relevanten Variablen berücksichtigt, die korrekte funktionale Form gewählt und erklärte Variablen ohne Messfehler vorliegen haben.

4.2.2 Alternative funktionale Formen

Wie bereits mehrfach erwähnt wurde, ist es zur Anwendung linearer Regressions-techniken zwingende Voraussetzung, dass die Regressionsfunktion parameterlinear ist. Variablenlinearität ist nicht unbedingt notwendig, sodass die Regressionsfunktion lediglich die bereits unter Annahme 1 behandelte Form

$$f(Y_i) = \beta_0 + \beta_1 f(X_{1i}) + \dots + \beta_k f(X_{ki}) + \varepsilon_i$$

aufweisen muss. Auf einige besonders wichtige parameterlineare Funktionen wollen wir im Folgenden näher eingehen und uns dabei im Besonderen mit der Bedeutung ihrer Parameter und mit Anwendungsfällen beschäftigen, bei denen sich ihr Einsatz empfiehlt.

4.2.2.1 Lineare Form

Die Form, mit der wir uns bisher überwiegend beschäftigt haben, ist die **lineare Form**, die sich wie folgt darstellt:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (\text{IV4: 11a})$$

Besonderheit ist hier die Tatsache, dass die Veränderung der erklärten Variable Y bei Änderung einer erklärenden Variable X_k (bei Konstanthaltung aller anderen X_k) durch eine Konstante beschrieben wird, wie die partielle Ableitung von Y nach X_k zeigt:

$$\frac{\partial Y_i}{\partial X_{ki}} = \beta_k \quad (\text{IV4: 11b})$$

Verändert sich also X_k ceteris paribus um eine Einheit, so verändert sich Y um β_k Einheiten (*marginale Veränderung*). Mit derart konstanten Koeffizienten ergibt sich die (partielle) *Elastizität* von Y bezüglich X_k (prozentuale Veränderung von Y bei Zunahme der Variable X_k um 1 % und Konstanthaltung aller anderen X_k) nach

$$\text{Elastizität}_{Y_i, X_{ki}} = \frac{\Delta Y \text{ in \%}}{\Delta X_k \text{ in \%}} = \frac{\partial Y_i / Y_i}{\partial X_{ki} / X_{ki}} = \frac{\partial Y_i}{\partial X_{ki}} \frac{X_{ki}}{Y_i} = \beta_k \frac{X_{ki}}{Y_i}. \quad (\text{IV4: 11c})$$

Da es keine genauen Regeln für die Wahl der funktionalen Form gibt, empfiehlt es sich die lineare Form zu wählen, solange die Modelltheorie oder der gesunde Menschenverstand keine andere Form erfordert. Zur Rechtfertigung der linearen Form können im Zweifelsfall auch formale Tests herangezogen werden.¹⁰

¹⁰ MacKinnon, J., White, H., Davidson, R. (1983) und Bera, A. K., Jarque, C. M. (1982) schlagen beispielsweise Tests vor, die uns bei der Wahl zwischen einem linearen und einem Doppel-Log-Modell unterstützen können.

4.2.2.2 Doppel-Log-Form

Bei der **Doppel-Log-Form** ist der natürliche Logarithmus von Y die erklärte und sind die natürlichen Logarithmen der X_k die erklärenden Variablen:

$$\ln Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \dots + \beta_k \ln X_{ki} + \epsilon_i \quad (\text{IV4: 12a})$$

Eine derartige Form wird gewählt, wenn wir aufgrund unserer Modelltheorie annehmen, dass die Elastizitäten des Modells als konstant gelten.¹¹ In einem Doppel-Log-Modell können die Koeffizienten der erklärenden Variablen nämlich als Elastizitäten betrachtet werden, die außerdem konstant sind:

$$\text{Elastizität}_{Y_i, X_{ki}} = \frac{\Delta Y \text{ in \%}}{\Delta X_k \text{ in \%}} = \frac{\partial Y_i / Y_i}{\partial X_{ki} / X_{ki}} = \frac{\partial Y_i}{\partial X_{ki}} \frac{X_{ki}}{Y_i} = \frac{\partial(\ln Y_i)}{\partial(\ln X_{ki})} = \beta_k \quad (\text{IV4: 12b})$$

Die β_k im Doppel-Log-Modell sind daher wie folgt zu interpretieren: Nimmt X_k bei Konstanthaltung aller anderen X_k um 1 % zu, so verändert sich Y_i um β_k %. Interessieren wir uns in einem solchen Modell für die für die Veränderung von Y (in Einheiten), die ceteris paribus aus einer Erhöhung von X_k um eine Einheit ergibt, so erhalten wir diese als

$$\frac{\partial Y_i}{\partial X_{ki}} = \beta_k \frac{Y_i}{X_{ki}}. \quad (\text{IV4: 12c})$$

Ein Parameterwert von $\beta_k > 1$ impliziert, dass die Elastizität größer als Eins ist und die marginale Veränderung von Y mit steigendem Wert von X_k zunimmt. Für einen Parameterwert zwischen 0 und 1 nimmt die marginale Zunahme hingegen ab. Die Doppel-Log-Form ist daher eine funktionale Form zur Abbildung von zu- oder abnehmenden Grenzerträgen.

Zur Beantwortung der Frage, für welche konkreten Modelle sich das Doppel-Log-Modell damit eignet, betrachten wir Abbildung IV 16 näher. Abbildung IV 16 (oben) zeigt, dass die Doppel-Log-Form das ökonomische Konzept der Produktionsfunktion (oder von Indifferenzkurven) darstellen kann (vgl. auch Anwendungsbeispiel IV 3-5). Ist X_1 beispielsweise der Arbeitsinput und X_2 der Kapitalinput, so geben die sog. Isoquanten der Produktionsfunktion die verschiedenen Kombinationen von X_1 und X_2 an, die ein bestimmtes Outputlevel Y produzieren. Abbildung IV 16 (unten, links) zeigt die Beziehung zwischen Y und X_1 , die bei einer Doppel-Log-Form mit zwei erklärenden Variablen und Konstanthaltung von X_2 besteht. Für den Fall $\beta_1 < 0$ könnte die Doppel-Log-Form daher auch zur Modellierung einer typischen nicht-linearen Nachfragefunktion verwendet werden. Abbildung IV 16 (rechts) verdeutlicht die Möglichkeit der Modellierung nicht-linearer Angebotsfunktionen.

¹¹ Wichtig bei Verwendung der Doppel-Log Form ist außerdem, dass die Beobachtungswerte von Y und X_k stets größer als Null sein müssen, da sonst der natürliche Logarithmus \ln nicht definiert ist.

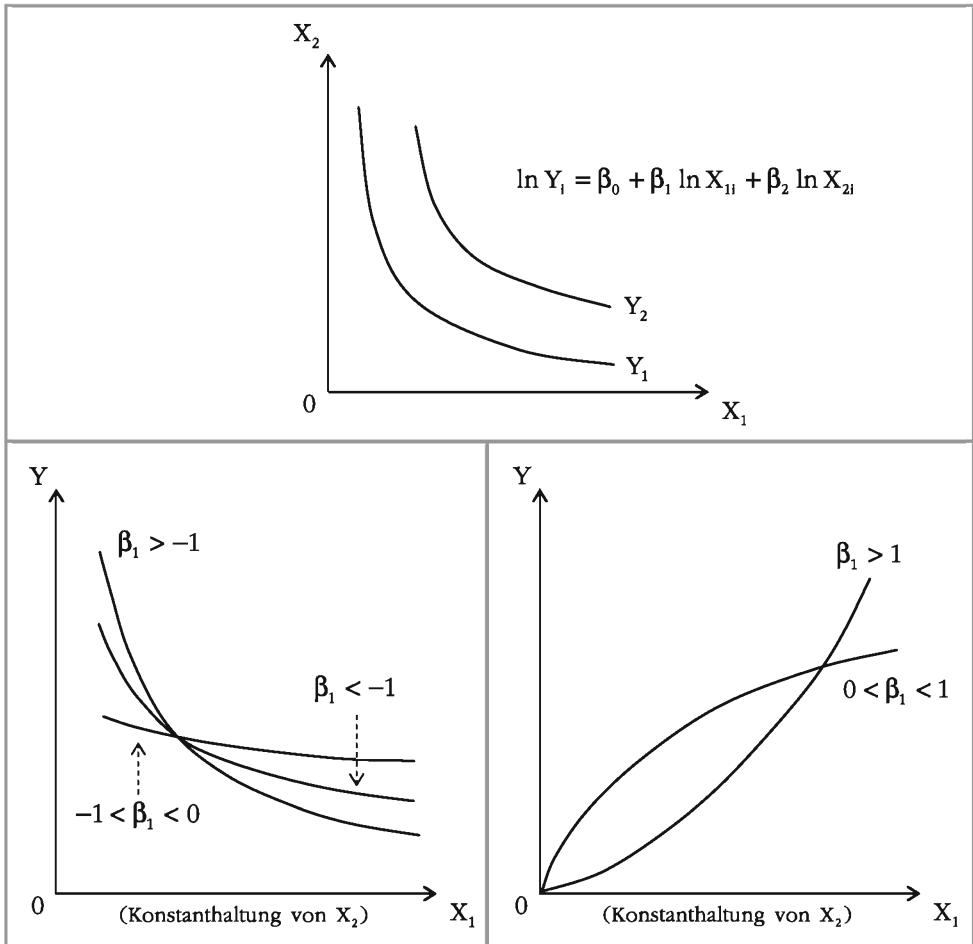


Abbildung IV 16: Doppel-Log-Form

4.2.2.3 Semi-Log-Form

Bezieht sich der natürliche Logarithmus nur auf die erklärenden Variablen und nicht auch auf die erklärte Variable, so liegt die sog. **rechtsseitige Semi-Log-Form**, als Variante der Doppel-Log-Form vor:

$$Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \dots + \beta_k \ln X_{ki} + \epsilon_i \quad (\text{IV4: 13a})$$

Anders als im Doppel-Log-Modell sind die Koeffizienten der erklärenden Variablen nun keine Elastizitäten mehr. Stattdessen gibt β_k nun an, um wie viele Einheiten sich Y bei einer 1 %igen Erhöhung von X_k (bei Konstanthaltung aller anderen erklärenden Variablen) verändert. Um dies zu zeigen betrachten wir zunächst die Veränderung von Y (in Einheiten) bei ceteris paribus Veränderung von X_k um eine Einheit.¹²

¹² Damit ergibt sich eine Elastizität von Y bzgl. X_k von β_k / Y_i .

$$\frac{\partial Y_i}{\partial X_{ki}} = \frac{\partial Y_i}{\partial \ln X_{ki}} \frac{\partial \ln X_{ki}}{\partial X_{ki}} = \beta_k \frac{1}{X_{ki}} \quad (\text{IV4: 13b})$$

Eine Umstellung von (IV4: 13b) nach β_k beweist seine geschilderte Eigenschaft

$$\beta_k = \frac{\partial Y_i}{\partial X_{ki} / X_{ki}}, \quad (\text{IV4: 13c})$$

da der Nenner von (IV4: 13c) gerade die prozentuale Veränderung der Variablen X_k darstellt, d.h. ΔX_k in %.

Von Form (IV4: 13a) gibt es wiederum eine Vielzahl von Ausprägungen. So müssen z.B. nicht alle X_k in Log-Form vorliegen. In einem Modell mit zwei erklärenden Variablen könnte X_2 etwa linear sein, während nur X_1 eine Log-Form annimmt. Ein solcher Fall ist in Abbildung IV 17 (links) unter Konstanthaltung von X_2 abgebildet. Ist $\beta_1 > 0$, so wird der Einfluss einer Veränderung von X_1 auf Y mit zunehmendem X_1 immer schwächer. Typische Anwendung findet ein solches Modell in Konsumfunktionen, da bei bestimmten Gütern davon ausgegangen wird, dass ihr Konsum ab einem bestimmten Einkommenslevel weniger stark als das verfügbare Einkommen zunimmt. Erklärt wird dies meist damit, dass ab einem bestimmten Einkommen mehr Luxusgüter konsumiert werden. Wir hätten also unser Geflügelkonsumbeispiel IV 3-1 auch in der Hinsicht spezifizieren können, dass wir den Logarithmus des verfügbaren Einkommens als erklärende Variable aufnehmen.

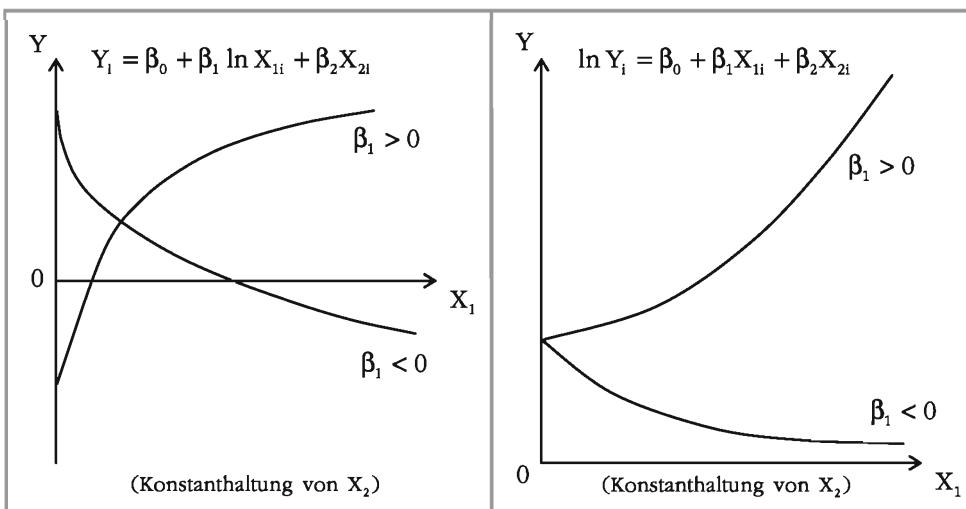


Abbildung IV 17: Semi-Log-Formen

Von einer sog. **linksseitigen Semi-Log-Form** (vgl. Abbildung IV 17, rechts) sprechen wir, wenn der natürliche Logarithmus der erklärenen Variablen im Modell auftaucht, die erklärenden Variablen aber keine Log-Form besitzen:

$$\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i \quad (\text{IV4: 14a})$$

Im Vergleich zu (IV4: 13a) ist sie die *praktisch relevantere Form*. In (IV4: 14a) gibt β_k an, dass sich Y bei Erhöhung von X_k um eine Einheit (bei Konstanthaltung aller anderen erklärenden Variablen) um β_k % verändert. Es gilt nämlich zunächst

$$\frac{\partial(\ln Y_i)}{\partial X_{ki}} = \beta_k,$$

was wir aufgrund von $\partial(\ln Y_i) / \partial Y_i = 1 / Y_i$ und damit $\partial(\ln Y_i) = \partial Y_i / Y_i$ als

$$\beta_k = \frac{\partial Y_i / Y_i}{\partial X_{ki}} \quad (\text{IV4: 14b})$$

ausdrücken können. Der Zähler von (IV4: 14b) ist dabei die prozentuale Veränderung der Variablen Y , d.h. ΔY in %.¹³

Gerade aufgrund der Eigenschaft (IV4: 14b) ist die Semi-Log-Form z.B. für Modelle geeignet, die das Einkommen von Individuen beschreiben. So könnte etwa das logarithmierte Einkommen eines Arbeiters linear von seiner Berufserfahrung und anderen Variablen abhängen. Sein Arbeitgeber wird nämlich in der Regel mit steigender Erfahrung prozentuale Gehaltserhöhungen vergeben. Wohl am häufigsten findet die linksseitige Semi-Log-Form aber bei der Modellierung von Wachstumsprozessen der Form

$$Y_t = Y_0 e^{\beta t} e^{\epsilon_t} \quad (\text{IV4: 15a})$$

Anwendung. Ein solcher Wachstumsprozess besagt, dass die Variable Y in Abhängigkeit von der fortschreitenden Zeit t bestimmt wird. Je höher der Wert Y_0 ist, desto höher ist der Ausgangswert von Y , bei dem das Wachstum zu $t = 0$ beginnt. Zur Schätzung der Parameter Y_0 und β derartiger Prozesse ist eine Logarithmierung von (IV4: 15a) erforderlich:

$$\ln Y_t = \alpha + \beta \cdot t + \epsilon_t \quad \text{mit} \quad \alpha = \ln Y_0 \quad (\text{IV4: 15b})$$

(IV4: 15b) stellt eine typische linksseitige Semi-Log-Form dar. Ihre Schätzung, die nun mit den Methoden der linearen Regression möglich ist, liefert die Werte $\hat{\beta}$ und $\hat{\alpha}$ sowie daraus $\hat{Y}_0 = e^{\hat{\alpha}}$. $\hat{\beta}$ kann dabei als geschätzte Wachstumsrate dieses exponentiellen Wachstumsprozesses interpretiert werden.¹⁴

4.2.2.4 Polynom-Form

Bei der **Polynom-Form** sind Variablen mehrfach, aber mit unterschiedlichen Exponenten als erklärende Variablen im Modell enthalten. Ein *quadratisches Polynom-Modell* hat beispielsweise die folgende Form:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i \quad (\text{IV4: 16a})$$

Anders als im linearen Modell ist die Änderung von Y , die durch eine Veränderung von X hervorgerufen wird, nicht länger konstant. Stattdessen hängt diese vom Wert von X selbst ab.¹⁵

$$\frac{dY_i}{dX_i} = \beta_1 + 2\beta_2 X_i \quad (\text{IV4: 16b})$$

¹³ Die Elastizität von Y bzgl. X_k erhalten wir damit als $\beta_k X_{ki}$ und ist nicht konstant.

¹⁴ Vgl. hierzu auch die Abschnitte IV 4.7.2.3 und IV 6.2.4.

¹⁵ Die Elastizität von Y bzgl. X ergibt sich damit als $(\beta_1 X_i + 2\beta_2 X_i^2) / Y_i$.

Gilt $\beta_1 < 0$ und $\beta_2 > 0$, so erhalten wir aus (IV4: 16a) grafisch eine typische nach oben geöffnete Parabel. Eine solche könnte zur Modellierung einer unternehmerischen Stückkostenfunktion eingesetzt werden. Während bei kleinen hergestellten Stückzahlen X_i noch die Fixkosten die Stückkosten dominieren, setzt mit zunehmendem X_i ein Kostendegressionseffekt ein. Die Gesamtkosten werden auf mehr Stück verteilt, was die Stückkosten irgendwann bei ihrem Minimum ankommen lässt, da der Fixkostendegressionseffekt immer schwächer wird. Der danach durch die Polynom-Form modellierte Anstieg der Stückkosten lässt sich damit erklären, dass mit zunehmender Stückzahl der Grenzkostenanstieg den Fixkostendegressionseffekt übersteigt.

Eine nach unten geöffnete Parabel ($\beta_1 > 0$ und $\beta_2 < 0$) könnte dazu verwendet werden, das Einkommen eines Arbeiters als Funktion seines Alters abzubilden (ggf. Erweiterung um andere Faktoren). Mit zunehmendem Lebensalter X_i wird sein Einkommen Y nämlich höchstwahrscheinlich aufgrund der steigenden Erfahrung zunehmen. Das positive β_1 aus (IV4: 16b) dominiert also in der Phase kleiner X_i die Einkommensentwicklung. Diese Einkommenssteigerungen werden aber im Zeitverlauf immer geringer ausfallen. Ab einem gewissen Alter kann das Einkommen sogar fallen. In dieser Phase fallenden Einkommens fällt der negative Term $2\beta_2 X_i$ aufgrund des hohen X_i nämlich stark ins Gewicht und bestimmt die Einkommensabnahme.

Der Einsatz von Polynom-Modellen ist nur bei vernünftiger theoretischer Rechtfertigung eines derartigen Verlaufs der Regressionsfunktion der Grundgesamtheit zu empfehlen. Wir wir bereits in unserem kleinen BIP-Regenmenge-Beispiel unter IV 3.1.3 gesehen haben, sollte z.B. ein quadratischer Term nicht nur ins Modell aufgenommen werden, um einen offensichtlich besseren "fit" zu erreichen.

4.2.2.5 Inverse Form

Bei einer **inversen Form** hängt die erklärte Variable vom Kehrwert mindestens einer erklärenden Variablen ab.¹⁶ Im Fall von nur einer sog. *inversen Variablen* ergibt sich beispielsweise

$$Y_i = \beta_0 + \beta_1 \frac{1}{X_{1i}} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i . \quad (\text{IV4: 17a})$$

Eine solche funktionale Form wird verwendet, wenn der Einfluss einer bestimmten erklärenden Variablen (hier X_i) auf Y mit zunehmendem Wert dieser Variablen abnimmt und sich Null annähert. Den Wert Null selbst darf sie aber nicht annehmen.

Die Veränderung von Y in Einheiten durch ceteris paribus Veränderung einer inversen Variablen X_k um eine Einheit kann hier wie folgt dargestellt werden:¹⁷

$$\frac{\partial Y_i}{\partial X_{ki}} = -\beta_k \frac{1}{X_{ki}^2} \quad (\text{IV4: 17b})$$

¹⁶ Das sog. Log-inverse Modell unterscheidet sich von (IV4:17a) dadurch, dass die Variable $\ln Y$ durch eine oder mehrere inverse Variablen erklärt. Für eine detaillierte Betrachtung dieses Modells vgl. z.B. von Auer, L. (2007).

¹⁷ Die Elastizität von Y bzgl. X_k ergibt sich damit als $-\beta_k / (X_{ki} Y_i)$.

Abbildung IV 18 zeigt den Fall eines Modells mit zwei erklärenden Variablen, wobei X_1 die inverse Variable ist. Unter Konstanthaltung von X_2 erkennen wir folgendes: Ist $\beta_1 > 0$, nähert sich Y mit zunehmendem X_1 dem (konstanten) Term $\beta_0 + \beta_2 X_2$ von oben her an. Dies ist der Fall, da $1/X_1$ gegen Null strebt. Gilt $\beta_1 < 0$, so findet diese Annäherung von unten her statt.

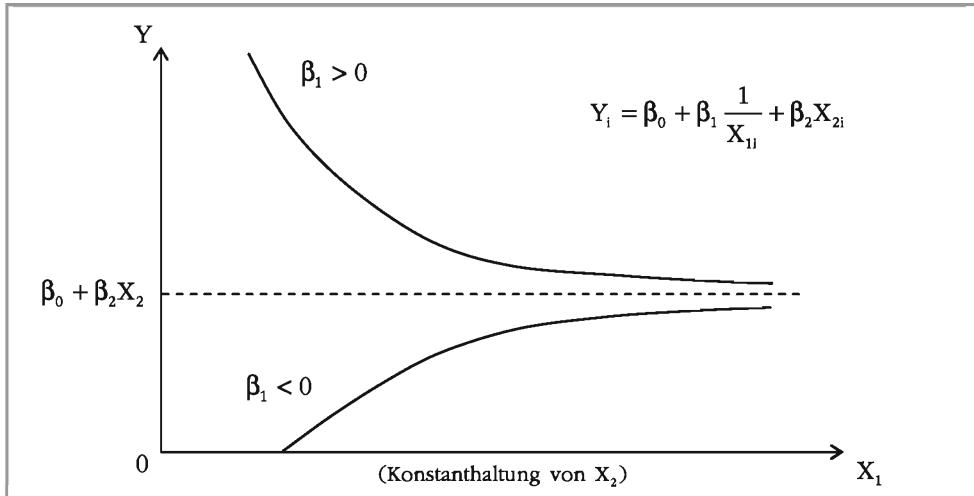


Abbildung IV 18: Inverse Form

Anwendung findet die inverse Form z.B. beim ökonomischen Konzept der Phillips-Kurve, die einen nichtlinearen, negativen Zusammenhang zwischen der Arbeitslosenquote und prozentualen Lohnänderungen herstellt. Jenseits einer gewissen Arbeitslosenquote führen Zunahmen der Arbeitslosenquote nach diesem Theorem aufgrund institutioneller und anderer Gründe jedoch nicht mehr zu einer Reduzierung von Lohnerhöhungen. Ein solches Modell könnte beispielsweise durch eine inverse Form

$$L_t = \beta_0 + \beta_1 (1/A_t) + \epsilon_t$$

mit L als prozentualer Lohnänderung und A als Arbeitslosenquote umgesetzt bzw. getestet werden.

4.2.2.6 Zusammenfassender Überblick

Die nachfolgende Tabelle bietet einen abschließenden Überblick über die wichtigsten funktionalen Formen und die Interpretation der Koeffizienten ihrer erklärenden Variablen. Wir haben uns bei der Darstellung mit Ausnahme der Polynomform auf Modelle mit einer erklärenden Variablen beschränkt. Für das Polynommodell und das inverse Modell können wir keine genaue Interpretation von β_1 angeben. Wir nennen daher nur eine Näherung für bestimmte Werte von X .

Form	Gleichung	Interpretation von β_1
linear	$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	Veränderung von Y (in Einheiten) bei Erhöhung von X um eine Einheit
Doppel-Log	$\ln Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	Prozentuale Veränderung von Y bei Erhöhung von X um ein Prozent
Semi-Log (rechtsseitig)	$Y_i = \beta_0 + \beta_1 \ln X_i + \epsilon_i$	Veränderung von Y (in Einheiten) bei einer Erhöhung von X um ein Prozent
Semi-Log (linksseitig)	$\ln Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$	Prozentuale Veränderung von Y bei Erhöhung von X um eine Einheit
Polynom	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$	Veränderung von Y (in Einheiten) bei Veränderung von X (in Einheiten), wenn X klein
invers	$Y_i = \beta_0 + \beta_1 \frac{1}{X_i} + \epsilon_i$	Entgegengesetzte Veränderung von Y (in Einheiten) bei Veränderung von X (in Einheiten), wenn X^2 ungefähr 1

4.2.3 Dummy-Variablen

4.2.3.1 Achsenabschnitts-Dummies

Bei Dummy-Variablen (binären Variablen) handelt es sich um Variablen, die abhängig von gewissen qualitativen Bedingungen nur die Werte 0 und 1 annehmen können. Führt eine Dummy-Variable zu Veränderungen des konstanten Terms eines Regressionsmodells, so spricht man von sog. **Achsenabschnitts-Dummies**. Diese werden folgendermaßen in ein einfaches Regressionsmodell integriert:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \epsilon_i$$

mit $D_i = \begin{cases} 1 & \text{i-te Beobachtung erfüllt bestimmte Bedingung} \\ 0 & \text{Bedingung nicht erfüllt} \end{cases}$ (IV4: 18)

Zum besseren Überblick bezeichnen wir Dummy-Variablen in diesem Abschnitt mit dem Buchstaben D. Abbildung IV 19 veranschaulicht, dass es durch die Verwendung von Achsenabschnitts-Dummies in der Tat zu einer Veränderung des Y-Achsenabschnitts der Regressionsfunktion abhängig von D kommt, wobei ihre Steigung unverändert bleibt. Nimmt die Dummy-Variable nämlich den Wert 1 an, so erhöht sich die Konstante des Modells bei $\beta_2 > 0$ um β_2 auf $\beta_0 + \beta_2$ (Verringerung bei $\beta_2 < 0$). Beim Wert 0 verbleibt die "dummylose" Funktion $\beta_0 + \beta_1 X_i$. Dies können wir durch Bildung des Erwartungswertes von (IV4: 18) bei deterministischem X und Gültigkeit der Annahme $E(\epsilon) = 0$ einfach zeigen:

$$E(Y_i) = \begin{cases} (\beta_0 + \beta_2) + \beta_1 X_i & \text{für } D_i = 1 \\ \beta_0 + \beta_1 X_i & \text{für } D_i = 0 \end{cases}$$

Die Regressionsfunktion wird also genau genommen durch zwei Funktionen beschrieben. Je nachdem welchen Wert D annimmt, wird der Erwartungswert von Y durch eine andere Funktion bestimmt. Es kommt also zu einem "Springen" von einer Funktion auf die andere.

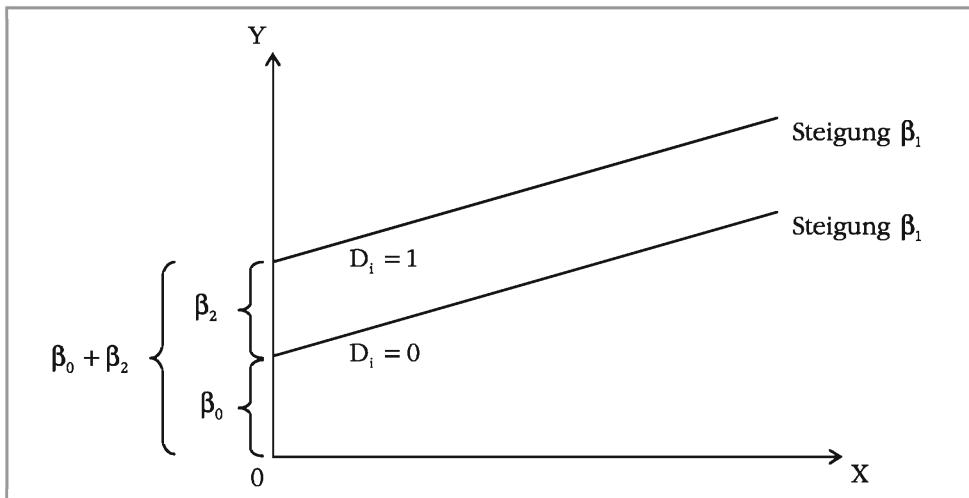


Abbildung IV 19: Achsenabschnitts-Dummy

Wichtig ist bei (IV4: 18) außerdem, dass wir hier trotz zweier Bedingungen (Erfüllung, Nicht-Erfüllung) nur eine Dummy-Variable benötigen. Dies liegt daran, dass immer ein Dummy weniger benötigt wird, als Bedingungen vorhanden sind. Die nicht direkt durch eine Dummy-Variable repräsentierte Bedingung, die vernachlässigte Bedingung, wird oft auch *Referenzkategorie* genannt. Die *Interpretation des Koeffizienten* der Dummy-Variablen wollen wir im folgenden Beispiel aufzeigen:

Beispiel IV 4-7

Achsenabschnitts-Dummy

Betrachten wir ein Modell $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 D_i + \varepsilon_i$ mit dem wir die Abhängigkeit des monatlichen Einkommens von Akademikern (Y) von ihrem Alter (X_1), ihrer Berufserfahrung (X_2) und ihrem Geschlecht (D) untersuchen wollen. Die Dummy-Variable D nimmt den Wert 1 an, wenn es sich beim Beobachtungssubjekt um einen Mann handelt. Andernfalls ist sie 0. Wir erkennen also, dass wir für die beiden Bedingungen "Mann" und "Frau" nur eine Dummy-Variable benötigen, wobei "Frau" die Referenzkategorie ist.

Nehmen wir an, die Schätzung des Modells liefert einen Koeffizienten $\hat{\beta}_3 = 512,32$. Ein solches Ergebnis würde bedeuten, dass bei gleichem Alter und gleicher Berufserfahrung das Einkommen der männlichen Akademiker im Durchschnitt 512,32 Euro höher ist, als das der weiblichen. Wäre das Vorzeichen negativ, wäre das Einkommen der Männer im Durchschnitt 512,32 Euro niedriger als das der Frauen.

Unsere bisherigen Dummy-Variablen waren solche, die qualitative Variablen mit nur zwei möglichen Ausprägungen (z.B. Mann/Frau) repräsentieren. Wie Dummy-Variablen in Modelle zu integrieren sind, wenn diese mehr als zwei mögliche Bedingungen vertreten sollen, zeigt folgendes einfache Beispiel:

Beispiel IV 4-8**Achsenabschnitts-Dummies**

Greifen wir unser vorheriges Akademikerbeispiel IV 4-7 noch einmal auf und modifizieren das Einkommensmodell zu $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 D_{1i} + \beta_4 D_{2i} + \varepsilon_i$. Wir interessieren uns nun nicht länger für das Geschlecht der Akademiker, sondern wollen die Art ihrer akademischen Qualifikation mit in die Gleichung aufnehmen. Wir vermuten nämlich, dass es einen Einfluss auf das Einkommen hat, ob der Akademiker einen Bachelor-, Master- oder einen Doktortitel vorweisen kann. Wir haben damit drei qualitative Bedingungen, die wir unterscheiden müssen und daher nur zwei Dummy-Variablen (eine weniger als Bedingungen) eingefügt. Es gilt:

$$D_{1i} = \begin{cases} 1 & \text{Master} \\ 0 & \text{sonst} \end{cases} \quad D_{2i} = \begin{cases} 1 & \text{Promotion} \\ 0 & \text{sonst} \end{cases}$$

Die vernachlässigte Bedingung "Bachelor" benötigt keine eigene Dummy-Variable, da sie durch ein gleichzeitiges "Nullwerden" beider Dummy-Variablen repräsentiert wird.

Eine besondere Anwendungsmöglichkeit von Achsenabschnitts-Dummies stellen sog. **saisonale Dummies** dar. Sie werden benutzt, um saisonalen Schwankungen in Zeitreihendaten Rechnung zu tragen. Würde auf sie verzichtet (vgl. vernachlässigte Variablen), ist es wahrscheinlich, dass Veränderungen von Y, die rein saisonbedingt sind, auf Veränderungen von X zurückgeführt werden, was zu einer Fehleinschätzung des Einflusses von X auf Y führt. Ein typisches *Quartalsmodell* hätte z.B. folgende Gestalt:

$$Y_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 X_t + \varepsilon_t \quad \text{mit} \quad (\text{IV4: 19})$$

$$D_{1t} = \begin{cases} 1 & \text{Quartal 1} \\ 0 & \text{sonst} \end{cases} \quad D_{2t} = \begin{cases} 1 & \text{Quartal 2} \\ 0 & \text{sonst} \end{cases} \quad D_{3t} = \begin{cases} 1 & \text{Quartal 3} \\ 0 & \text{sonst} \end{cases}$$

Es sind hier drei Dummy-Variablen zur Abbildung der vier möglichen Quartale enthalten. X ist eine "normale" erklärende Variable. t bezieht sich auf ein Quartal. β_1 beschreibt nun das Ausmaß, in dem sich der erwartete Wert von Y im ersten Quartal von dem des vierten Quartals unterscheidet. β_2 und β_3 sind analog zu interpretieren. Um die Hypothese signifikanter Saisonalität in den Daten zu testen, muss getestet werden, ob die Koeffizienten der Dummies gleichzeitig Null sind. Das Hypothesenpaar des notwendigen F-Tests muss also hier

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$	(keine Saisonalität)
$H_A : \text{mindestens ein } \beta \text{ ist ungleich Null}$	(Saisonalität)

(IV4: 20a)

lauten. Die "restriktive" Gleichung würde dann eine Gestalt

$$Y_t = \beta_0 + \beta_4 X_{4t} + \varepsilon_t \quad (\text{IV4: 20b})$$

annehmen. Um zu bestimmen, ob die Dummies wirklich zur Desaisonalisierung benötigt werden, wird der "fit" der "restriktiven" Gleichung mit der der originären Gleichung mittels des F-Tests verglichen. Kommen wir zu einer Ablehnung von H_0 , so werden alle Dummies beibehalten. Eine Nicht-Ablehnung würde dafür sprechen, alle Dummies aus der Regressionsgleichung zu entfernen.

Beispiel IV 4-9**Konsumtheorie nach Keynes**

Die Konsumtheorie nach John M. Keynes besagt, dass Individuen Ihren Konsum bei steigendem Einkommen erhöhen, diese Erhöhung aber nicht so stark ausfällt, wie die Einkommenserhöhung. Anders ausgedrückt wird eine marginale Konsumneigung (MKN) von größer als 0 und kleiner als 1 angenommen. Das zum Test dieser Theorie zu verwendende ökonometrische Modell kann z.B. die Form

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

besitzen, wobei Y die Konsumausgaben und X das verfügbare Einkommen beschreibt. β_1 misst dabei die MKN. Zur Schätzung dieses Modells für Deutschland bedienen wir uns nun zweier Zeitreihen (Quartalsdaten von 1999 bis 2007) aus dem Zeitreihenservice der Deutschen Bundesbank. Für die Konsumausgaben verwenden wir *saisonbereinigte* private Konsumausgaben aus der Volkswirtschaftlichen Gesamtrechnung (VGR) der Bundesrepublik Deutschland. Analog gewinnen wir auch das verfügbare Einkommen der privaten Haushalte aus der VGR. Das Ergebnis der Schätzung mit EViews zeigt sich wie folgt:

Dependent Variable: Y		Sample: 1999Q1 2007Q4		
Method: Least Squares		Included observations: 36		
	Coefficient	Std. Error	t-Statistic	Prob.
C	32.65175	4.505499	7.247089	0.0000
X	0.807545	0.012595	64.11577	0.0000
R-squared	0.991797	Mean dependent var		321.1092
Adjusted R-squared	0.991556	S.D. dependent var		15.78363
S.E. of regression	1.450399	Akaike info criterion		3.635506
Sum squared resid	71.52430	Schwarz criterion		3.723480
Log likelihood	-63.43912	F-statistic		4110.832
Durbin-Watson stat	1.151605	Prob(F-statistic)		0.000000

Für den Betrachtungszeitraum lag die MPC somit bei etwa 0,8, was bedeutet, dass eine Erhöhung des Einkommens um 1 Euro *im Durchschnitt* zu einer Erhöhung des Konsums um 80 Cent führte. Das sehr hohe R^2 , nahe Eins, zeigt einen sehr guten "fit" des Modells auf, allerdings weist der niedrige Wert der Durbin-Watson-Statistik auf mögliche Spezifikationsprobleme hin (vgl. Abschnitt IV 4.5 zu Autokorrelation und IV 4.7 zu nicht stationären Variablen).

Da wir mit saisonbereinigten Daten arbeiten, sollte das Ergebnis unserer Schätzung nicht durch (quartalsweise) saisonale Effekte in den Zeitreihendaten verfälscht worden sein. Wir wollen dies dennoch durch Schätzung des gleichen Modells mit Quartalsdummies überprüfen. Die neue Schätzung ist auf der Folgeseite zu finden. Nach (IV3: 6) können wir damit die Teststatistik des notwendigen F-Tests für das Hypothesenpaar (IV4: 20a) berechnen. Unter Verwendung der RSS-Werte der originären Regression und der Dummy-Regression erhalten wir folgendes:

$$F = \frac{(RSS_M - RSS) / M}{RSS / (n - K - 1)} = \frac{(71,5243 - 66,8571) / 3}{66,8571 / (36 - 1 - 1)} = 0,7912$$

Bei einem Signifikanzniveau von 5 % erhalten wir aus der F(3;34)-Verteilung den kritischen Wert 2,88. Da dieser deutlich über dem Wert der F-Statistik liegt, können wir H_0 nicht ablehnen. Es gibt also keine Hinweise auf Saisonalität.

Analog zu Beispiel IV 3-7 hätten wir diesen Test in EViews im Dummymodell auch viel schneller über View / Coefficient Tests / Wald - Coefficient Restrictions und Verwendung der Restriktion $c(2) = c(3) = c(4) = 0$ durchführen können.

Dependent Variable: Y		Sample: 1999Q1 2007Q4		
Method: Least Squares		Included observations: 36		
	Coefficient	Std. Error	t-Statistic	Prob.
C	32.99686	4.647134	7.100476	0.0000
D1	-0.378220	0.696507	-0.543024	0.5910
D2	0.093903	0.693680	0.135370	0.8932
D3	0.631693	0.692663	0.911978	0.3688
X	0.806335	0.012837	62.81278	0.0000
R-squared	0.992332	Mean dependent var		321.1092
Adjusted R-squared	0.991343	S.D. dependent var		15.78363
S.E. of regression	1.468564	Akaike info criterion		3.734693
Sum squared resid	66.85706	Schwarz criterion		3.954626
Log likelihood	-62.22447	F-statistic		1002.982
Durbin-Watson stat	1.077779	Prob(F-statistic)		0.000000

4.2.3.2 Steigungs-Dummies

Achsenabschnitts-Dummies haben im Regressionsmodell keinen Einfluss auf die Steigung der Regressionsgeraden. Anders ist dies bei sog. **Steigungs-Dummies**. Abhängig vom Wert des Dummies verändern solche nämlich deren Steigung. Ein Steigungs-Dummy wird in ein Modell durch Addition einer sog. *Interaktionsvariablen* integriert. Derartige Interaktionsterme werden verwendet, wenn die Veränderung von Y durch eine Variable X_k vom Wert einer anderen Variablen abhängt. Im Fall eines Steigungs-Dummies handelt es sich beim Interaktionsterm um das Produkt der Dummy-Variablen, die die Steigungsänderung hervorrufen soll, und der erklärenden Variablen, deren Steigungskoeffizient abhängig von der Dummy-Variablen verändert werden soll. Die einfachste Form eines Steigungs-Dummy-Modells ist daher

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i X_i + \varepsilon_i. \quad (\text{IV4: 21a})$$

Abbildung IV 20 (oben) und der Erwartungswert von (IV4: 21a), also

$$E(Y_i) = \begin{cases} \beta_0 + (\beta_1 + \beta_2)X_i & \text{für } D_i = 1 \\ \beta_0 + \beta_1 X_i & \text{für } D_i = 0, \end{cases}$$

veranschaulichen die Bedeutung dieses Modells. Für einen Wert der Dummy-Variable von Eins erhöht sich die Steigung bei $\beta_2 > 0$ um β_2 , wobei die Konstante β_0 unverändert bleibt (Steigungsverringerung bei $\beta_2 < 0$). Ein Wert von Null reduziert das Modell auf ein "dummyloses" Modell.

Typische Anwendung finden derartige Steigungs-Dummies z.B. bei der Modellierung von Konsumfunktionen, die über Zeiträume geschätzt werden sollen, in denen es zu Kriegen kam. Kriegszeiten würden nämlich die marginale Konsumneigung reduzieren und diese Änderung könnte durch eine Dummy-Variable "Krieg oder kein Krieg" modelliert werden.

Sollen sich in einem Modell in Abhängigkeit von einer Dummy-Variable sowohl Y-Achsenabschnitt als auch Steigung verändern, so ist dieses wie folgt zu formulieren:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i X_i + \varepsilon_i$$

(IV4: 21b)

Für den Erwartungswert von Y_i gilt hier

$$E(Y_i) = \begin{cases} (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_i & \text{für } D_i = 1 \\ \beta_0 + \beta_1 X_i & \text{für } D_i = 0. \end{cases}$$

Dies und auch Abbildung IV 20 (unten) zeigen, dass es für $\beta_2 > 0$ und $\beta_3 > 0$ bei einem Wert der Dummy-Variablen von Eins sowohl zu einer Erhöhung der Regressionskonstanten um β_2 , als auch der Steigung um β_3 kommt. Bei negativen Vorzeichen kommt es entsprechend zu einer Verringerung. Selbstverständlich können β_2 und β_3 unterschiedliche Vorzeichen annehmen und in gewissen Relationen zueinander und zu β_0 stehen, sodass eine Vielfalt unterschiedlicher Lagen von durch Dummies veränderten Regressionsgeraden denkbar ist.

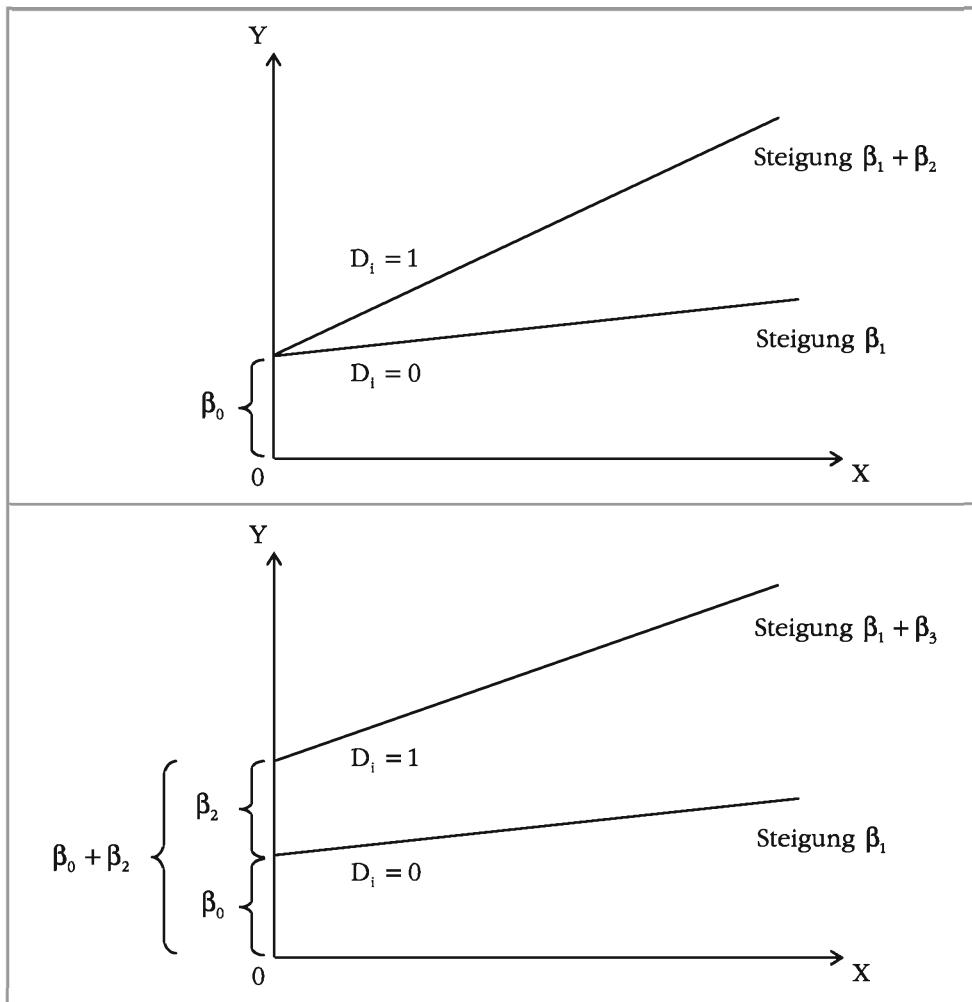


Abbildung IV 20: Steigungs-Dummies

Beispiel IV 4-10

Steigungs-Dummy

In unserem Einkommensmodell für Akademiker $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 D_i + \epsilon_i$ in Beispiel IV 4-7 sind wir davon ausgegangen, dass sexuelle Diskriminierung Einfluss auf das Einkommen von Akademikern hat. Wir haben daher den Steigungs-Dummy D = "Mann oder Frau" in unsere Modellgleichung aufgenommen. $\hat{\beta}_3$ war der Schätzer für die durchschnittliche Einkommensdifferenz zwischen Männern und Frauen (bei Konstanthaltung aller anderen erklärenden Variablen). Nehmen wir nun an, dass das Einkommen von Männern mit zunehmender Berufserfahrung (X_2) stärker zunimmt als das von Frauen, so können wir in ein solches Modell zusätzlich einen Steigungs-Dummy einbauen, sodass das neue Modell die Gestalt $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 D_i + \beta_4 D_i X_{2i} + \epsilon_i$ aufweist. $\hat{\beta}_4$ wäre nun eine Schätzung des unterschiedlichen Einflusses eines zusätzlichen Jahres an Erfahrung auf das Einkommen. Könnte in einem einseitigen t-Test festgestellt werden, dass β_4 signifikant positiv ist, wäre die These einer stärkeren Einkommenszunahme von Männern bei steigender Erfahrung bestätigt.

4.2.4 Folgen der Wahl einer falschen funktionalen Form

Nehmen wir an, die unserem Regressionsmodell zugrunde liegende Theorie verlangt eine nichtlineare Regressionsfunktion. So könnten wir z.B. eine Phillips-Kurve nicht nur über eine inverse Form, sondern auch als Doppel-Log-Form oder diverse Semi-Log-Formen modellieren. Wir könnten dabei in Versuchung geraten, die Form anhand des "fits" in der Stichprobe auszuwählen. Warum wir dies jedoch nicht tun sollten, können wir wie folgt begründen:

Wird eine falsche funktionale Form verwendet, wird im Allgemeinen die Annahme 2a verletzt und es steigt die Wahrscheinlichkeit falscher Rückschlüsse auf die wahren Parameter der Grundgesamtheit. Eine falsche funktionale Form ist nämlich ein mit einer vernachlässigten Variablen vergleichbarer Spezifikationsfehler. Betrachten wir dazu ein Polynommodell dritten Grades

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \epsilon_i, \quad (\text{IV4: 22a})$$

welches die wahren Zusammenhänge in der Grundgesamtheit beschreibt. Entscheiden wir uns stattdessen aber für die lineare Form

$$Y_i = \beta_0^* + \beta_1^* X_i + \epsilon_i^* \quad \text{mit} \quad \epsilon_i^* = \epsilon_i + \beta_2 X_i^2 + \beta_3 X_i^3, \quad (\text{IV4: 22b})$$

so ist klar, dass hier ein Fall vernachlässigter Variablen vorliegt. Aufgrund der Korrelation der erklärenden Variablen X^2 und X^3 mit X kommt es zu einer Verletzung der Annahme 2b und somit zur Verzerrung des verbleibenden Schätzers für β_1 (vgl. Abschnitt IV 4.1.1).

Zudem ist zu erwähnen, dass es durchaus sein kann, dass eine falsche funktionale Form zwar innerhalb einer Stichprobe, einen guten "fit" aufweist, aber zu schweren Prognosefehlern führt, wenn sie für Werte außerhalb der Stichprobe verwendet wird. Problematisch ist es insbesondere, wenn wir eine spezielle funktionale Form allein anhand von R^2 oder anderen Kriterien zur Messung des "fits" auswählen. Es besteht dabei die Gefahr, dass die Spezifikation an besondere Zufälligkeiten in der Stichprobe angepasst wird, die in der Grundgesamtheit keine Bedeutung haben. Die gewählte funktionale Form weist zwar dann innerhalb der Stichprobe einen

guten "fit" auf, beschreibt aber die wahren Beziehungen zwischen den Variablen nur schlecht. Außerhalb der Stichprobe kann dies deshalb zu schlechten Prognosen der erklärten Variable führen. Ist es möglich, die Entscheidung über die *funktionale Form* auch anhand *inhaltlicher Erwägungen* zu treffen, kann der Ökonometriker eher funktionale Fehlspezifikationen vermeiden als bei alleinigem Blick auf die vorliegenden Daten. Es gilt schließlich ein Modell zu finden, dass die Grundgesamtheit und nicht nur die vorliegende Stichprobe beschreibt.

4.3 Multikollinearität

4.3.1 Formen von Multikollinearität

Vollkommene Multikollinearität verletzt Annahme 6 des CLRM, die besagt, dass keine erklärende Variable eine lineare Funktion einer anderen erklärenden Variablen sein darf. Der Begriff "vollkommen" bezieht sich dabei auf die Tatsache, dass die Veränderung einer erklärenden Variablen vollständig durch die Variation einer oder mehrerer anderer Variablen erklärt werden kann. Im Fall eines multiplen linearen Regressionsmodells mit zwei erklärenden Variablen würde also vollkommene Multikollinearität vorliegen, wenn z.B.

$$X_{1i} = \alpha_0 + \alpha_1 X_{2i} \quad (\text{IV4: 23a})$$

gilt, der Wert von X_1 also bei gegebenem X_2 *exakt* mit einer linearen Funktion berechnet werden kann. Grafisch veranschaulicht Abbildung IV 21 (links) einen solchen Fall.

Beispiel IV 4-11

Vollkommene Multikollinearität

Studienleistungen: Versuchen wir die Studienleistungen Y eines Studenten in Abhängigkeit von X_1 : Einkommen der Eltern, X_2 : durchschnittliche Lernzeit pro Tag und X_3 : durchschnittliche Lernzeit pro Woche zu erklären, so wird schnell deutlich, dass X_2 und X_3 aufgrund von $X_3 = 7X_2$ vollkommen multikollinear sind.

Nominal- vs. Realzins: Nehmen wir an, wir integrieren sowohl den Nominalzins (i), als auch den Realzins (r) als erklärende Variablen in ein Modell. Beide Variablen sind nach der sog. Fisher-Identität in der Form $r_t = i_t - \pi_t^{\text{ew}}$ näherungsweise über die erwartete Inflationsrate (π^{ew}) verbunden. Ist die Inflationsrate (z.B. aufgrund strikter Preissetzung des Staates) konstant, so wären die beiden Variablen Nominal- und Realzins perfekt multikollinear.

Liegt vollkommene Multikollinearität vor, so ist OLS nicht in der Lage die Regressionskoeffizienten zu schätzen. In multiplen Regressionsmodellen messen diese nämlich die Veränderung von Y, die mit einer Veränderung einer erklärenden Variablen unter *Konstanthaltung aller anderen Variablen* verbunden ist. Aufgrund der Multikollinearität können nun aber nicht alle anderen Variablen der Regressionsgleichung konstant gehalten werden, da sich bei Veränderung einer Variablen auch eine andere ändert. Die meisten Softwarepakete liefern in einem solchen Fall eine Fehlermeldung, die auf die Unmöglichkeit der Schätzung hinweisen soll.¹⁸ Formal lässt sich die Unmöglichkeit der Schätzung z.B. bei Vorliegen von (IV4: 23a) mit $\alpha_0 = 0$ anhand des Nenners von $\hat{\beta}_1$ im Regressionsmodell mit zwei Regressoren, (IV1:18), aufzeigen. Würde dieser nämlich den Wert Null annehmen, käme es zu einer Division durch Null, wodurch der Schätzer nicht definiert wäre. Genau dies ist nun hier der Fall, wie die folgenden Umformungen zeigen:

¹⁸ EViews 6.0 zeigt hier die Fehlermeldung "Near singular matrix".

$$\begin{aligned}
 & (\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i}x_{2i})^2 \\
 &= (\sum (\alpha_1 x_{2i} - \alpha_1 \bar{X}_2)^2)(\sum (x_{2i} - \bar{X}_2)^2) - (\sum (\alpha_1 x_{2i} - \alpha_1 \bar{X}_2)(x_{2i} - \bar{X}_2))^2 \\
 &= \alpha_1^2 (\sum (x_{2i} - \bar{X}_2)^2)^2 - \alpha_1^2 (\sum (x_{2i} - \bar{X}_2)^2)^2 = 0
 \end{aligned}$$

Vollkommene Multikollinearität stellt jedoch in der Praxis kein allzu großes Problem dar. Sie ergibt sich häufig aufgrund logischer Fehler und ist einfach aufzudecken (Fehlermeldung der Statistiksoftware). Bei festgestellter vollkommener Multikollinearität können einfach entsprechende Variablen aus der Regressionsgleichung ausgeschlossen werden. Häufig stellen wir uns auch bereits im Vorfeld der Schätzung die Fragen, ob nicht eine Variable nur ein Vielfaches einer anderen ist, sie durch Addition einer Konstanten zu einer anderen Variablen oder eine lineare Kombination verschiedener anderer Variablen entsteht. Kann dies bejaht werden, kommt es sofort zum Ausschluss betroffener Variablen.

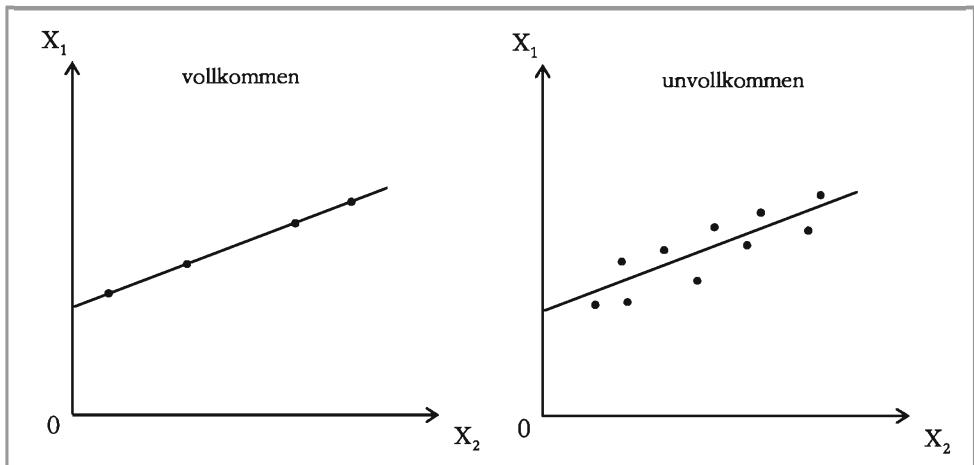


Abbildung IV 21: Vollkommene vs. unvollkommene Multikollinearität

Schwerer wiegt das Problem **unvollkommener Multikollinearität**. Die Beziehung zwischen zwei erklärenden Variablen hat hier nämlich z.B. die Form

$$X_{1i} = \alpha_0 + \alpha_1 X_{2i} + u_i . \quad (\text{IV4: 23b})$$

Der hier integrierte stochastische Störterm u_i und auch Abbildung IV 21 (rechts) zeigen, dass die Beziehung zwischen den erklärenden Variablen nicht mehr perfekt linear ist. Es verbleibt also ein Teil der Variation von X_1 , der sich nicht durch Veränderung von X_2 erklären lässt.

Ob erklärende Variablen unvollkommen multikollinear (starke lineare Beziehung zwischen ihnen) sind, hängt von ihrer theoretischen Beziehung zueinander und von der vorliegenden Stichprobe ab. So kann es z.B. vorkommen, dass zwei Variablen in einer Stichprobe nur schwach korreliert sind, während in einer anderen Stichprobe hohe Korrelation festgestellt und damit auf unvollkommene Multikollinearität geschlossen wird. Unvollkommene Multikollinearität kann also als ein *stichprobenabhängiges Phänomen* betrachtet werden.

4.3.2 Konsequenzen von Multikollinearität

Wie wir bereits erwähnt haben, hat *vollkommene Multikollinearität* die *Unmöglichkeit* der Schätzung von Regressionskoeffizienten zur Folge. Im Fall *unvollkommener Multikollinearität* ist eine *Schätzung möglich*, wobei aber folgende Punkte zu beachten sind:

1. Werden alle anderen Annahmen des CLRM eingehalten, so bleiben die *OLS-Schätzer* trotz Multikollinearität *unverzerrt*.
2. Da im Falle von Multikollinearität kaum zwischen dem Effekt einer Variablen auf Y und dem Effekt einer anderen Variablen unterschieden werden kann, steigt die Wahrscheinlichkeit geschätzte Betas zu erhalten, die weit vom wahren Koeffizienten der Grundgesamtheit entfernt sind. Auch unerwartete Vorzeichen werden wahrscheinlicher. Die *geschätzten Koeffizienten* stammen im Falle von Multikollinearität also aus Verteilungen mit größeren Varianzen und damit *größeren Standardfehlern*.¹⁹ Wir können dies leicht durch einen Blick auf die Varianzen (IV2: 26) und (IV2: 27) der Schätzer der Steigungskoeffizienten in einem Regressionsmodell mit zwei erklärenden Variablen zeigen:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1i}^2 (1 - r_{12}^2)} \quad \text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{12}^2)}$$

Sind die beiden Variablen X_1 und X_2 nicht miteinander korreliert ($r_{12} = 0$), so erreichen die Varianzen beider Schätzer ihren jeweiligen minimalen Wert. Unterschiede in den Varianzen der Schätzer sind nur durch Unterschiede der Varianzen der erklärenden Variablen bedingt. Liegt nun aber absolut betrachtet ein hoher Wert für r_{12} vor, so führt dies in beiden Fällen zu einem Anstieg der Varianzen der Schätzer und zeigt somit die zunehmende Ungenauigkeit der Schätzer an. Bei perfekter Korrelation ($r_{12} = \pm 1$) ist neben dem OLS-Schätzer (vgl. Abschnitt IV 4.3.1) auch die zugehörige Varianz nicht definiert.

3. Die Betrachtung der t-Statistik (IV3: 2a) bzw. (IV3: 2b)

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\hat{\sigma}_{\hat{\beta}_k}} \quad t_k = \frac{\hat{\beta}_k}{\hat{\sigma}_{\hat{\beta}_k}}$$

macht klar, dass größere Standardfehler der OLS-Schätzer auch zu *niedrigeren t-Werten* führen und damit Hypothesentests beeinträchtigen. Sie können uns dazu bewegen, fälschlicherweise wichtige erklärende Variablen aufgrund fehlender Signifikanz aus unserem Modell auszuschließen.

4. Liegt starke Multikollinearität vor, so können *Änderungen der Modellspezifikation* (Hinzufügen oder Ausschluss von erklärenden Variablen) *oder der Datengrundlage* zu *deutlichen Veränderungen der geschätzten Koeffizienten der multikollinearen Variablen* führen. Diese höhere Sensibilität der Schätzer beruht vor allem darauf, dass sich OLS z.B. bei zwei sich nahezu gleich entwickelnden Variablen zur Unterscheidung an Beobachtungen orientiert, bei

¹⁹ Dies beeinträchtigt allerdings nicht die Effizienz der OLS-Schätzer. Es gibt also trotz Multikollinearität keine andere unverzerrte Schätzmethode, welche eine geringere Schätzvarianz aufweist.

denen beide Variablen eine deutlich andere Entwicklungstendenz aufweisen. Führt eine Änderung des Stichprobenumfangs nun aber dazu, dass genau diese Beobachtungen eliminiert werden, die zur Unterscheidung der Variablen dienen, hat dies eine deutliche Änderung der geschätzten Koeffizienten zur Folge.

5. Auch wenn die individuellen t-Werte von multikollinearen Variablen oft sehr niedrig sind, kann es zu einem *relativ hohem R²* kommen oder die *Koeffizienten* dieser Variablen können *gemeinsam hoch signifikant* sein. Als erstes Anzeichen vorliegender Multikollinearität können wir daher die Kombination niedriger t-Werte bei hohem R² ansehen. Es kann auch vorkommen, dass die F-Statistik für den Ausschluss von mehreren multikollinearen Variablen hoch signifikant ist, wohingegen keine der individuellen t-Statistiken signifikant ist.

4.3.3 Aufdecken von Multikollinearität

Bei der Untersuchung eines geschätzten Regressionsmodells auf Multikollinearität müssen wir uns zweier wichtiger Punkte bewusst sein. Erstens tritt unvollkommene Multikollinearität in nahezu jeder Gleichung auf. In empirischen Daten ist es nämlich fast unmöglich, dass erklärende Variablen überhaupt nicht korreliert sind. Es gilt also nicht festzustellen, ob Multikollinearität vorliegt, sondern wie stark diese ist und wie hohe Multikollinearität möglicherweise die Varianzen der Schätzer erhöht hat. Bereits jetzt sei aber erwähnt, dass Multikollinearität nicht zwangsläufig die Varianzen der Schätzer erhöhen und damit die t-Werte verringern muss. Sie ist bei ökonometrischen Analysen also nicht immer ein Problem.

Grundsätzlich gibt es keine allgemeingültige Methode zur Messung der Stärke und der Folgen von Multikollinearität. Es existieren zwar eine Reihe formaler und auch weniger formaler Methoden, doch können diese nur als "Daumenregeln" betrachtet werden, die grobe Hinweise geben. Auch kommen sie nicht immer zum gleichen Ergebnis. Wir wollen nun zunächst einige dieser Methoden näher betrachten, wobei Regel Nr. 1 mehr ein klassisches Symptom von Multikollinearität als eine Methode zur Aufdeckung ist. Wir beschränken uns dabei auf für die Praxis wirklich interessante Verfahren.²⁰

1. Hohes R² bei wenig signifikanten t-Werten

Bei hohem R² wird der F-Test auf Gesamt signifikanz meistens die Nullhypothese ablehnen, dass alle Steigungskoeffizienten gleichzeitig Null sind, wobei individuelle t-Tests zeigen werden, dass keiner oder nur wenige der Koeffizienten signifikant von Null verschieden sind (Spezialfall von Punkt 5 aus Abschnitt IV 4.3.2).

2. Hohe paarweise Korrelation der erklärenden Variablen

In Modellen mit nur zwei erklärenden Variablen gilt der Korrelationskoeffizient zwischen den erklärenden Variablen als Indikator für Multikollinearität. Je höher dieser ist, desto schwerer wiegt das Problem der Multikollinearität. In Modellen mit mehr als zwei erklärenden Variablen ist dieser Indikator allerdings weniger aus-

²⁰ Einen umfassenden Überblick über weitere, jedoch nicht unbedingt praxisrelevante Verfahren bieten z.B. Gujarati, D.N. (2003) und Klien, L.R. (1962).

sagekräftig. Es ist nämlich möglich, dass Gruppen von erklärenden Variablen in einer multikollineararen Beziehung stehen, während nicht ein Korrelationskoeffizient hoch genug ist, um auf ernste Multikollinearität aufmerksam zu machen.

3. Hilfsregressionen und Varianzschwellungsfaktoren

Da Multikollinearität nichts anderes aussagt, als dass sich eine erklärende Variable als exakte oder approximative lineare Kombination anderer erklärender Variablen ergibt, besteht eine Möglichkeit diese aufzudecken darin, dass zunächst für jedes X_k ein Regressionsmodell geschätzt wird, welches diese Variable in Abhängigkeit von *allen* anderen erklärenden Variablen modelliert. Im Fall eines Modells mit drei erklärenden Variablen würden wir also die Modelle

$$\begin{aligned} X_{1i} &= \alpha_0 + \alpha_1 X_{2i} + \alpha_2 X_{3i} + u_i \\ X_{2i} &= \alpha'_0 + \alpha'_1 X_{1i} + \alpha'_2 X_{3i} + u_i \\ X_{3i} &= \alpha''_0 + \alpha''_1 X_{1i} + \alpha''_2 X_{2i} + u_i \end{aligned} \quad (\text{IV4: 24})$$

schätzen. Wir bezeichnen diese Regressionen auch als **Hilfsregressionen** (engl. auxiliary regressions). Nach der Schätzung der K Modelle kann für jede davon das Bestimmtheitsmaß R_k^2 erfasst werden, da ein hoher Wert des Bestimmtheitsmaßes als Hinweis auf Multikollinearität angesehen werden kann.²¹ Nachteil einer solchen Methode ist es jedoch, dass die Hilfsregressionen der X_k stets in Abhängigkeit *aller* anderen Variablen durchgeführt werden. Betrifft die Multikollinearität nur einige wenige Variablen, können die Hilfsregressionen natürlich wenig zur Diagnose beitragen. Um jeden möglichen Fall abzudecken, müssten weitere Hilfsregressionen geschätzt werden.

Die Bestimmtheitsmaße R_k^2 aus Hilfsregressionen der Art (IV4: 24) können zur Konstruktion sogenannter **VIFs** (engl. variance inflation factors) herangezogen werden. Sie können für jeden Parameterschätzer einer Regressionsgleichung berechnet werden und geben an, wie stark Multikollinearität die Varianz eines Parameterschätzers beeinflusst. Im Fall eines Regressionsmodells mit *zwei erklärenden Variablen* können wir den VIF für $\hat{\beta}_1$ aus der Varianz des Schätzers (IV2: 26) unter Beachtung der Tatsache herleiten, dass in den Hilfsregressionen $(r_{12})^2 = R_1^2$ gilt. Wir erhalten

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1i}^2(1 - r_{12}^2)} = \frac{\sigma^2}{\sum x_{1i}^2} \frac{1}{(1 - r_{12}^2)} = \frac{\sigma^2}{\sum x_{1i}^2} \frac{1}{(1 - R_1^2)}, \quad (\text{IV4: 25})$$

wobei der Term $1/(1 - R_1^2)$ den VIF für $\hat{\beta}_1$ darstellt. Dieses Ergebnis lässt sich auch auf den Fall eines Regressionsmodells mit K erklärenden Variablen erweitern, womit wir den Varianzschwellungsfaktor zum Schätzer $\hat{\beta}_k$ allgemein als

$$\text{VIF}(\hat{\beta}_k) = \frac{1}{(1 - R_k^2)} \quad (\text{IV4: 26})$$

definieren können. R_k^2 stellt dabei das Bestimmtheitsmaß einer Hilfsregression von X_k auf alle anderen erklärenden Variablen dar. Ein R_k^2 von Eins, was auf vollkom-

²¹ Die Signifikanz dieser Bestimmtheitsmaße kann durch einen klassischen F-Test mittels der Teststatistik (IV3: 11) geprüft werden.

mene Multikollinearität hindeutet, führt zu einem VIF der gegen unendlich strebt, wohingegen ein R_k^2 von Null auf die Abwesenheit von Multikollinearität hinweist und einen VIF von Eins liefert. Besonderheit bei diesem Faktor ist, dass im Falle eines Modells mit nur zwei erklärenden Variablen die beiden entstehenden VIFs identisch sind. Dies begründet sich dadurch, dass die beiden Hilfsregressionen das-selbe Bestimmtheitsmaß aufweisen.

Da es bei diesem Indikator keine formale Testentscheidung gibt, hat sich in der Praxis als grobe *Entscheidungsregel* eingebürgert, dass bei einem R_k^2 von über 0,90, also bei $VIF > 10$ von hoher Multikollinearität ausgegangen werden kann.²²

Die Varianzschwellungsfaktoren bleiben jedoch nicht ohne Kritik. Erstens sind sie nämlich gerade Resultat von Hilfsregressionen, die eine erklärende Variable in Abhängigkeit aller anderen darstellen, sodass auch für sie die im Zusammenhang mit den Hilfsregressionen aufgeführten Kritiken angebracht sind. Zweitens handelt es sich bei der Entscheidungsregel um eine recht willkürliche Vorgehensweise. Die sich daraus ergebenden Konsequenzen bleiben außerdem unklar, wie wir gleich sehen werden. Drittens ist ein hoher VIF weder notwendige noch hinreichende Bedingung für hohe Varianzen und Standardfehler der OLS-Schätzer. Wie wir nämlich in (IV4: 25) gesehen haben, hängt die Varianz eines OLS-Schätzers von drei Faktoren ab. Eine Verallgemeinerung dieses Ergebnisses liefert

$$\text{Var}(\hat{\beta}_k) = \frac{\sigma^2}{\sum x_{ki}^2} \cdot \text{VIF}(\hat{\beta}_k).$$

Ein hoher VIF kann also durch ein niedriges σ^2 oder ein hohes $\sum x_{ki}^2$ ausgeglichen werden. Hohe, durch einen VIF gemessene Multikollinearität, muss daher nicht unbedingt zu hohen Standardfehlern führen.

Zum Abschluss dieser Methodendiskussion wollen wir uns nun mit der Anwendung der beschriebenen Verfahren befassen. Wir veranschaulichen diese anhand eines hypothetischen Beispiels, das bewusst so konstruiert wurde, dass es die typischen Anzeichen von Multikollinearität aufweist. Im Anschluss daran untersuchen wir ein empirisches Beispiel und werden dabei feststellen, dass eine Multikollinearitätsanalyse mit "echten" Zahlen nicht mehr so eindeutig ist.

Beispiel IV 4-12

Hypothetischer Multikollinearitätsfall



Nehmen wir an, wir schätzen ein Modell $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ und erhalten das folgende Ergebnis:

$$\begin{aligned} \hat{Y}_i &= 35,8677 - 6,3265X_{1i} + 1,1790X_{2i} & n &= 25 (22) \\ &(33,7510) \quad (8,4383) & R^2 &= 0,7356 \\ t &= -0,19 \quad 0,21 \end{aligned}$$

Erstes deutliches Anzeichen für Multikollinearität sind die hohen Standardfehler und damit die niedrigen t-Werte, welche von einem relativ hohen R^2 begleitet werden. Ein Blick auf den Korrelationskoeffizienten zwischen X_1 und X_2 , der bei 0,999995 liegt, erhärtet diesen Verdacht. Die Durchführung der beiden Hilfsregressionen von X_1 auf X_2 und von X_2 auf X_1 blieben hier keinen besonderen Mehrwert, da wir deren identische Bestimmtheitsmaße

²² vgl. Kleinbaum, D.G., Kupper, L.L., Müller, K.E. (1998).

schließlich nach (IV1: 23) auch als Quadrat des Korrelationskoeffizienten zwischen X_1 und X_2 bestimmen können. Diese sind 0,99999, d.h. nahezu Eins. Die Variation von X_1 wird also zu fast 100 % durch die von X_2 erklärt und umgekehrt; ein unverkennbares Zeichen von Multikollinearität. Ganz klar kommt dann natürlich auch unsere VIF-Daumenregel aufgrund von $VIF(\hat{\beta}_1) = VIF(\hat{\beta}_2) = 1/(1-0,99999) = 100.000 > 10$ zu dem Schluss, dass deutliche Multikollinearität vorliegt.

Beispiel IV 4-13

Praktischer Multikollinearitätsfall



Zur Untersuchung empirischer Modelle auf Multikollinearität betrachten wir folgendes von Asteriou, D., Hall, S.D. (2007) geschätzte Modell zur Erklärung der Importe Großbritanniens:

$$\ln Y_t = \beta_0 + \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \varepsilon_t$$

Y sind dabei die Importe, X_1 ein BIP-Index und X_2 der britische Konsumentenpreisindex. Die Schätzung für den Zeitraum 1. Quartal 1990 bis 2. Quartal 1998 liefert das folgende Ergebnis:

$$\begin{aligned}\widehat{\ln Y_t} &= 0,6319 + 1,9269 \ln X_{1t} + 0,2743 \ln X_{2t} & n &= 34 (30) \\ &(0,1689) &(0,1374) & R^2 &= 0,9661 \\ t &= 11,41 & 2,00\end{aligned}$$

Eine neue Regression, die den Produzentenpreisindex als zusätzliche Variable $\ln X_3$ aufnimmt, ergibt Folgendes:

$$\begin{aligned}\widehat{\ln Y_t} &= 0,2139 + 1,9697 \ln X_{1t} + 1,0255 \ln X_{2t} - 0,7706 \ln X_{3t} & n &= 34 (30) \\ &(0,1568) &(0,3234) & R^2 &= 0,9660 \\ t &= 12,58 & 3,17 & -2,52\end{aligned}$$

Es fällt zunächst auf, dass sich der geschätzte Koeffizient von $\ln X_2$ stark ändert. Zudem erkennen wir, dass alle Koeffizienten in der neuen Schätzung hoch signifikant sind, wobei der Koeffizient von $\ln X_3$ jedoch ein völlig unerwartetes Vorzeichen aufweist. Bei zunehmenden Preisen für Produzenten im Inland sollten die Importe schließlich zu- und nicht abnehmen. Anhand der folgenden EViews-Korrelationsmatrix (Variablen markieren → "Open as Group" → View / Covariance Analysis → Ordinary Correlation) können wir außerdem zeigen, dass der Korrelationskoeffizient zwischen $\ln X_2$ und $\ln X_3$ bei 0,9836 liegt. Generell liegen ziemlich hohe Korrelationen zwischen den Variablen vor.

	Y	$\ln X_1$	$\ln X_2$	$\ln X_3$
$\ln Y$	1.000000	0.980660	0.907449	0.886093
$\ln X_1$	0.980660	1.000000	0.895348	0.889341
$\ln X_2$	0.907449	0.895348	1.000000	0.983597
$\ln X_3$	0.886093	0.889341	0.983597	1.000000

Während diese Korrelationen auf Multikollinearität hinweisen, liegen ihre klassischen Symptome (niedrige t-Werte bei hohem R^2) also nicht vor. Jemand der sich bei der Multikollinearitätsanalyse also lediglich auf diese verlassen hätte, würde davon ausgehen, dass keine Multikollinearität vorliegt. Vielmehr sind es hier die hohen paarweisen Korrelationen und das mit der Modelltheorie nicht im Einklang stehende Vorzeichen der neuen Preisvariablen, die verdächtig sind.

Da Korrelationskoeffizienten nur paarweise Korrelation untersuchen, uns aber ein Modell mit mehr als zwei erklärenden Variablen vorliegt, empfiehlt es sich mit den beschriebenen Hilfsregressionen zu arbeiten:

$$\widehat{\ln X_{1t}} = 1,0619 + 0,5150 \ln X_{2t} + 0,2103 \ln X_{3t} \quad n = 34 (31)$$

$$(0,3587) \quad (0,3476) \quad R^2 = 0,8040$$

$$t = \quad 1,44 \quad 0,61$$

$$\widehat{\ln X_{2t}} = 0,4601 + 0,1211 \ln X_{1t} + 0,8681 \ln X_{3t} \quad n = 34 (31)$$

$$(0,0843) \quad (0,0665) \quad R^2 = 0,9695$$

$$t = \quad 1,44 \quad 13,06$$

$$\widehat{\ln X_{3t}} = -0,5424 + 0,0555 \ln X_{1t} + 0,9748 \ln X_{2t} \quad n = 34 (31)$$

$$(0,0917) \quad (0,0746) \quad R^2 = 0,9678$$

$$t = \quad 0,61 \quad 13,06$$

Auf den ersten Blick ist hier auffällig, dass bei der Regression von $\ln X_1$ auf $\ln X_2$ und $\ln X_3$ die Koeffizienten nicht signifikant von Null verschieden sind (niedrige t-Werte). In der Hilfsregression für $\ln X_2$ ist $\ln X_3$ signifikant und in der für $\ln X_3$ ist es $\ln X_2$, wobei aber in beiden Fällen $\ln X_1$ nicht signifikant ist. Dies würde darauf hindeuten, dass ein Multikollinearitätsproblem hier nur zwischen den Variablen $\ln X_2$ und $\ln X_3$ besteht.

Betrachten wir nur die Bestimmtheitsmaße der Hilfsregressionen, kämen wir zu dem allgemeinen Schluss, dass Multikollinearität vorliegt. Die Variabilität von $\ln X_2$ und $\ln X_3$ wird hochgradig durch die anderen Variablen erklärt. Bei $\ln X_1$ können wir dies nicht sagen, da ein R^2 von 0,8040 bei Zeitreihenregressionen nicht als sehr hoch betrachtet werden kann. Ein konkretes Bild erhalten wir durch die Varianzschwellingsfaktoren $VIF(\hat{\beta}_1) = 5,1020$, $VIF(\hat{\beta}_2) = 32,7869$ und $VIF(\hat{\beta}_3) = 31,0559$. Sie legen nahe, dass $\ln X_2$ und $\ln X_3$ stark multikollinear sind.

Die Mehrheit der Daumenregeln weiß also darauf hin, dass $\ln X_2$ und $\ln X_3$ stark multikollinear sind, was auch zu erwarten war, da es sich bei beiden um Preisvariablen mit gleicher Entwicklungstendenz handelt.

4.3.4 Vorgehensweise bei festgestellter Multikollinearität

Im Folgenden beschreiben wir eine Auswahl alternativer Möglichkeiten, die unter gewissen Umständen geeignet sind, Multikollinearität zu begegnen. Da einige Möglichkeiten aber nicht immer realisierbar sind und neue Probleme mit sich bringen können, sind es in der *Praxis* häufig die *Punkte 1 und 2*, die bei der Problemlösung die größte Beachtung finden.

1. Untätig bleiben

Haben wir ernste Multikollinearität festgestellt, so ist es in vielen Fällen ratsam, diese Diagnose einfach zu ignorieren, insbesondere daher, da die im Folgenden noch beschriebenen Maßnahmen gewisse Risiken aufweisen. Darüber hinaus lassen sich generell zwei Gründe aufführen, die ein Untätigbleiben rechtfertigen:

Erstens reduziert Multikollinearität in vielen Fällen die t-Werte nicht so stark, dass die zugehörigen Parameter insignifikant werden, oder führt nicht zu derart starken Veränderungen der geschätzten Regressionskoeffizienten, dass diese vom erwarteten

ten Vorzeichen abweichen. *Zweitens* kann die Entfernung einer multikollineararen Variablen zu Verzerrungen der OLS-Schätzer führen, wenn die Variable aufgrund der Modelltheorie wichtig für das Modell ist. Eine derart bewusste Fehlspezifikation (vernachlässigte Variable) wird von erfahrenen Ökonometrikern gemieden. Selbst bei niedrigen t-Werten belassen sie multikollinare Variablen im Modell.

2. Ausschluss redundanter Variablen

Oftmals ist der Ausschluss einer multikollinearen Variablen eine gute Idee, vorausgesetzt es handelt sich bei dieser um eine sog. *redundante Variable* und nicht um eine solche, die für das Modell wichtig ist. Nehmen wir beispielsweise an, dass ein unerfahrener Ökonometriker zu viele Variablen in sein Modell aufgenommen hat, da er eine Verzerrung durch vernachlässigte Variablen fürchtet. Resultat einer solchen Vorgehensweise ist dann (neben einer drastischen Reduzierung der Freiheitsgrade) meistens, dass zwei oder mehr Variablen enthalten sind, die die selbe Größe beschreiben und damit multikollinear sind. Es macht z.B. wenig Sinn in eine Nachfragefunktion als erklärende Variablen das verfügbare Einkommen und das BIP aufzunehmen, da beide die Größe Einkommen messen. Da Einkommen für eine Nachfragefunktion eine wichtige erklärende Variable ist, können weder verfügbares Einkommen noch BIP als irrelevant bezeichnet werden. Stattdessen können wir von **Redundanz** sprechen, da nur eine Variable zur Messung des Einkommens benötigt wird. Eine dieser Variablen kann aus der Modellgleichung entfernt werden, ohne eine Verzerrung durch eine vernachlässigte Variable zu verursachen.

Beispiel IV 4-14

Multikollinearität und Variablenausschluss I

Greifen wir auf unser hypothetisches Beispiel IV 4-12 zurück. Unsere Schätzung, bei der wir starke Multikollinearität festgestellt haben, lautete wie folgt:

$$\hat{Y}_i = 35,8677 - 6,3265X_{1i} + 1,1790X_{2i} \quad n = 25 \text{ (22)} \\ (33,7510) \quad (8,4383) \quad R^2 = 0,7356 \\ t = -0,19 \quad 0,21$$

Betrachten wir nun zwei Modelle, in denen wir jeweils eine der multikollinearen Variablen (unter Annahme ihrer Redundanz) aus der Modellgleichung ausschließen:

$$\hat{Y}_i = 36,7186 + 0,8320X_{1i} \quad n = 25 \text{ (22)} \\ (0,1041) \quad R^2 = 0,7351 \\ t = 7,99$$

$$\hat{Y}_i = 36,6097 + 0,2080X_{2i} \quad n = 25 \text{ (22)} \\ (0,0260) \quad R^2 = 0,7352 \\ t = 7,99$$

Wir erkennen deutlich, dass in beiden Modellen die Multikollinearität beseitigt wurde, da die Standardfehler der OLS-Schätzer nun deutlich geringer sind und damit auch höhere t-Werte vorlegen.

In der Praxis stellt sich oft die Frage, welche redundante Variable aus der Gleichung ausgeschlossen werden soll. Im Falle ernster Multikollinearität macht es meist keinen statistischen Unterschied welche der multikollineararen Variablen ausgeschlossen wird. Dies verdeutlichen die R²- und t-Werte im Beispiel IV 4-14. Sie waren, egal welche der beiden Variablen ausgeschlossen wurde, nahezu identisch. Dies ist auch nicht verwunderlich, da sich die beiden Variablen X₁ und X₂ schließlich nahezu identisch verhalten. In praktischen Anwendungen ist es aber nicht immer egal, welche Variable ausgeschlossen wird.

Beispiel IV 4-15

Multikollinearität und Variablenausschluss II



Schätzen wir das Importmodell aus dem Beispiel IV 4-13 ohne ln X₂, so erhalten wir:

$$\widehat{\ln Y_t} = 0,6857 + 2,0938 \ln X_{1t} + 0,1196 \ln X_{3t} \quad n = 34 (30)$$

$$(0,1726) \quad (0,1361) \quad R^2 = 0,9602$$

$$t = 12,13 \quad 0,88$$

Wir stellen fest, dass der geschätzte Koeffizient von ln X₃ nun positiv und überhaupt nicht mehr signifikant ist. Die vorher festgestellte Signifikanz von ln X₃ ist also auf die bestehende Multikollinearität zwischen den beiden Preisvariablen zurückzuführen.

Die Schätzung des Modells ohne ln X₃ liefert hingegen:

$$\widehat{\ln Y_t} = 0,6319 + 1,9269 \ln X_{1t} + 0,2743 \ln X_{2t} \quad n = 34 (30)$$

$$(0,1689) \quad (0,1374) \quad R^2 = 0,9639$$

$$t = 11,41 \quad 2,00$$

Die geschätzten Koeffizienten beider erklärender Variablen sind positiv, wobei der Koeffizient von ln X₁ hoch signifikant und der von ln X₂ "gerade noch" signifikant ist. Wir würden uns also hier (da wir nur eine Preisvariable benötigen) für den Ausschluss von ln X₃ entscheiden. Auch wenn die Korrelation zwischen ln X₁ und ln X₂ (0,8953) nun darauf hindeutet, dass noch eine gewisse Multikollinearität vorliegt, würden wir nun nicht mehr so weit gehen, ln X₁ oder ln X₂ auszuschließen, da es sich bei diesen Variablen um für die Gleichung bedeutende Variablen handelt. Ein Ausschluss würde hier nur zur Verzerrung der OLS-Schätzer führen.

3. Erhöhung des Stichprobenumfangs

Wie wir wissen, kann Multikollinearität die Varianz der OLS-Schätzer erhöhen. Dies haben wir bereits für das Regressionsmodell mit zwei erklärenden Variablen über (IV2: 26) durch einen c.p. höheren Korrelationskoeffizienten veranschaulicht:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_{1t}^2 (1 - r_{12}^2)}$$

Es wird nun schnell klar, dass durch einen höheren Stichprobenumfang n die Varianz von $\hat{\beta}_1$ abnimmt, da $\sum x_{1t}^2$ im Nenner des obigen Terms gerade $n \cdot \text{Var}(X_1)$ ist und damit i.d.R. zunimmt. Das Multikollinearitätsproblem stellt also vor allem ein Datenproblem dar. Zusätzliche Daten können das Problem lösen, insbesondere dann, wenn diese die Multikollinearitätseigenschaften nicht so stark aufweisen. In der wirtschaftswissenschaftlichen Praxis ist jedoch eine Erhöhung des Stichprobenumfangs nicht ohne weiteres möglich. Die Daten, mit denen Modelle geschätzt werden, beziehen nämlich i.d.R. alle verfügbaren Informationen ein. Eine Erhö-

hung des Stichprobenumfangs n ist daher entweder unmöglich, da keine weiteren Daten vorhanden sind, oder mit enormen Kosten verbunden. In experimentellen Situationen wäre die Schaffung neuer Daten hingegen meist kein Problem.

4. Nutzung externer Informationen

Besitzen wir aus anderen Quellen (z.B. andere Studien) Informationen darüber, dass zwischen Modellparametern eine bestimmte lineare Beziehung besteht, so können wir diese für unsere eigene Schätzung nutzen. Nehmen wir z.B. an, wir wissen für ein Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i, \quad (\text{IV4: 27a})$$

dass die Variable X_2 irrelevant ist und der Parameter von X_3 dem doppelten Wert des Parameters von X_1 entspricht, also

$$\beta_2 = 0 \quad \text{und} \quad \beta_3 = 2\beta_1, \quad (\text{IV4: 27b})$$

gilt. Setzen wir nun (IV4: 27b) in (IV4: 27a) ein, so erhalten wir das neue Modell

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + 0 \cdot X_{2i} + 2\beta_1 X_{3i} + \beta_4 X_{4i} + \varepsilon_i \\ &= \beta_0 + \beta_1 X_{1i}^* + \beta_4 X_{4i} + \varepsilon_i \quad \text{mit} \quad X_{1i}^* = X_{1i} + 2X_{3i}, \end{aligned} \quad (\text{IV4: 28})$$

welches nur noch zwei zu schätzende Parameter enthält. Folglich erhöht sich die Zahl der Freiheitsgrade und die Parameter können mit geringerer Varianz geschätzt werden. Den Schätzer $\hat{\beta}_3$ können wir dann über (IV4: 27b) und seine Varianz nach den Rechenregeln für Varianzen als $\text{Var}(\hat{\beta}_3) = \text{Var}(2\hat{\beta}_1) = 4 \cdot \text{Var}(\hat{\beta}_1)$ bestimmen.

Offensichtliches Problem dieser auch als *restringierte OLS-Schätzung*²³ bezeichneten Methode ist, dass sie nur dann unverzerrt ist, wenn (IV4: 27b) korrekt ist. Bei Abweichungen liefert das restringierte Modell verzerrte Schätzer. Die geringere Schätzvarianz würde also mit einem erhöhten Verzerrungsrisiko erkauft. Um Anhaltspunkte über das Ausmaß der möglichen Verzerrung zu erhalten, sollte man in jedem Fall auf Basis des Modells (IV4: 27a) einen F-Test mit (IV4: 27b) als Nullhypothese durchführen.

5. Modelltransformation

In der Literatur wird auch häufig die Transformation unter Multikollinearität leidender Modelle vorgeschlagen. Eine solche Transformation ist aber mit Vorsicht zu betrachten, da sie die Verletzung anderer Annahmen des CLRM und andere Probleme mit sich bringen kann. Wir wollen im Folgenden zwei der gängigsten Transformationen näher betrachten:

Die erste Möglichkeit der Transformation besteht darin, aus multikollinearen Variablen eine *neue Variable* zu bilden und diese an Stelle der multikollinearen Variablen ins Modell aufzunehmen.²⁴ Nehmen wir beispielsweise an, dass in einem Modell mit zwei erklärenden Variablen die Variablen X_1 und X_2 multikollinear sind, so könnten sie durch eine neue Variable

$$X_3 = k_1 X_1 + k_2 X_2 \quad (\text{IV4: 29a})$$

²³ Eine ausführliche Darstellung der statistischen Eigenschaften der restringierten OLS-Schätzung findet sich z.B. in von Auer, L. (2007), Kapitel 21.

²⁴ Vgl. Studenmund, A.H. (2007), Kapitel 8.

ersetzt werden, sodass sich das neue Modell

$$Y_t = \beta_0 + \beta_3 X_3 + \varepsilon_t = \beta_0 + \beta_3 (k_1 X_1 + k_2 X_2) + \varepsilon_t \quad (\text{IV4: 29b})$$

ergibt, wobei die Konstanten k_1 und k_2 angeben, mit welchem Gewicht die ursprünglichen Variablen in die neuen eingehen. Nachteil einer solchen Methode ist allerdings, dass beiden "Teilen" der neuen Variablen der gleiche Koeffizient zugeordnet ist. Wir müssen daher äußerst vorsichtig sein, nicht Variablen mit unterschiedlichen erwarteten Koeffizienten (z.B. unterschiedliche Vorzeichen) oder deutlich unterschiedlichen Mittelwerten (z.B. unterschiedliche Größenordnungen) zusammenzufassen, ohne diesen Unterschieden durch entsprechende Wahl der Werte k_1 und k_2 Rechnung zu tragen. Auch ist eine solche Vorgehensweise nur zu wählen, wenn die neu entstehende Variable für sich selbst sinnvoll ist.

Eine zweite Möglichkeit der Transformation besteht bei Zeitreihendaten in der sog. *Differenzenbildung*. Um beispielsweise synchrone Entwicklungen von Variablen in einem Modell

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t \quad (\text{IV4: 30a})$$

zu bereinigen, kann von diesem das auch für die Vorperiode geltende Modell

$$Y_{t-1} = \beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t-1} + \varepsilon_{t-1} \quad (\text{IV4: 30b})$$

subtrahiert werden, sodass wir zunächst

$$Y_t - Y_{t-1} = \beta_1 (X_{1t} - X_{1t-1}) + \beta_2 (X_{2t} - X_{2t-1}) + (\varepsilon_t - \varepsilon_{t-1}) \quad (\text{IV4: 30c})$$

erhalten. Unter Verwendung einer alternativen Notation ergibt sich daraus ein in periodenweisen Veränderungen formuliertes Modell

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} \\ \Delta Y_t &= \beta_1 \Delta X_{1t} + \beta_2 \Delta X_{2t} + \Delta \varepsilon_t \quad \text{mit} \quad \Delta X_{kt} = X_{kt} - X_{kt-1}, \\ \Delta \varepsilon_t &= \varepsilon_t - \varepsilon_{t-1} \end{aligned} \quad (\text{IV4: 31})$$

welches wir auch als *erste Differenzenform* des Ausgangsmodells bezeichnen.²⁵ Unter ersten Differenzen verstehen wir hier also nichts anderes als die Differenz zwischen einer Variable zum Zeitpunkt t und ihrem Wert der Vorperiode $t - 1$.²⁶

Die Niveaus der Variablen mögen zwar multikollinear sein, aber dies bedeutet nicht, dass ihre Veränderungsbeträge im selben Ausmaß multikollinear sind. Der Grad der Multikollinearität ist im transformierten Modell des Öfteren geringer. Begründen lässt sich dies dadurch, dass eine derartige Variablentransformation (anders als einfache lineare Transformationen) die Struktur der Daten verändert und die ersten Differenzen der Variablen schwächer miteinander korreliert sind. Die Varianz der Schätzer dürfte damit abnehmen. Allerdings ist die Varianz der differenzierten Variablen häufig auch geringer, was für sich genommen die Varianz der Schätzer erhöht. Multikollinearität ist ein Phänomen, das größtenteils in Zeitreihen-

²⁵ Wir erkennen, dass im Modell (IV4: 31) die Regressionskonstante verlorengegangen ist. Bei der Schätzung des Modells wird daher wieder eine neue integriert, um die im Abschnitt IV 4.2.1 beschriebenen Nachteile zu vermeiden.

²⁶ Die Formulierung eines Modells in ersten Differenzen kommt nur bei Zeitreihendaten in Frage. In Querschnittsdaten existiert nämlich keine natürliche Reihenfolge in den Beobachtungen, was keine eindeutige Differenzenbildung zulässt.

datenmaterial auftritt. Hier ist die Wahrscheinlichkeit, dass die ersten Differenzen einem Trend folgen und damit mit anderen Variablen in multikollinearer Beziehung stehen, nämlich geringer als bei den Ursprungszeitreihen, aus denen sie bestimmt werden. Während nämlich z.B. das deutsche BIP von Jahr zu Jahr zunimmt, also einen klaren Aufwärtstrend zeigt, fluktuieren die Veränderungen des BIP (bzw. seine ersten Differenzen) stark.

Auch wenn erste Differenzen dazu beitragen können, die Folgen von Multikollinearität zu bekämpfen, führen diese neben dem Verlust an Freiheitsgraden (Reduzierung der Zahl der für die Schätzung zur Verfügung stehenden Beobachtungen um Eins) auch zu modelltheoretischen Komplikationen. So besteht z.B. ein Unterschied darin den Kapitalstock einer Volkswirtschaft oder die Veränderung des Kapitalstocks zu modellieren. Durch erste Differenzen würden wir hier nämlich Nettoinvestitionen abbilden. Ist dies aber nicht Ziel der empirischen Arbeit, so ist eine derartige Multikollinearitätsbekämpfung nicht zu empfehlen. Darüber hinaus ergibt sich ein schwerwiegendes Problem mit der neuen Störgröße $\Delta\epsilon_t$. Da $\Delta\epsilon_t = \epsilon_t - \epsilon_{t-1}$ und $\Delta\epsilon_{t-1} = \epsilon_{t-1} - \epsilon_{t-2}$ gilt, sind die beiden aufeinanderfolgenden Störgrößen $\Delta\epsilon_t$ und $\Delta\epsilon_{t-1}$ über ϵ_{t-1} korreliert und führen damit zu einer Verletzung von Annahme 4 des CLRM. Erneut erkaufen wir uns also einen scheinbaren Fortschritt durch neue Probleme an anderer Stelle (vgl. Abschnitt IV 4.5.2).

4.4 Heteroskedastizität

Annahme 3 des CLRM besagt, dass die Ausprägungen des Störterms aus Verteilungen mit gleicher Varianz stammen sollen. Wir sprechen hier von Homoskedastizität und formulieren $\text{Var}(\epsilon_i | X_i) = \sigma^2$ für $i = 1, 2, \dots, N$, wobei σ^2 eine Konstante ist. Abbildung IV 10 verdeutlichte diese Annahme.

In vielen praktischen Anwendungen ist die Annahme der Homoskedastizität nun jedoch unrealistisch. Häufig ist es nämlich so, dass sich die Verteilungen der ϵ_i in ihrer Varianz unterscheiden, d.h. die Varianzen $\text{Var}(\epsilon_i | X_i)$ nicht konstant sind. Wir sprechen dann von Heteroskedastizität und formulieren $\text{Var}(\epsilon_i | X_i) = \sigma_i^2$ für $i = 1, 2, \dots, N$. Die Varianz der Verteilung eines Störterms hängt dann davon ab, welche Beobachtung von X wir betrachten. Abbildung IV 11 veranschaulichte dies.

Heteroskedastizität tritt häufig in Querschnittsregressionen auf. Wir werden daher im Folgenden mit dem Subskript i für die jeweilige Beobachtung arbeiten. Dies bedeutet jedoch nicht, dass Heteroskedastizität in Zeitreihenregressionen unmöglich ist. Es hat sich nämlich herausgestellt, dass sie insbesondere auch bei Finanzmarktstudien häufig vorkommt.

4.4.1 Ursachen von Heteroskedastizität

Um das Wesen von Heteroskedastizität klarer zu machen, betrachten wir zunächst das einfache Modell

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

dessen Grundgesamtheitsregressionsfunktion zusammen mit einigen Stichprobenwerten im Falle von Homo- und Heteroskedastizität in Abbildung IV 22 dargestellt ist. Es ist dabei zu beachten, dass es sich darin bei (1) und (2a) um mit den Abbildungen IV 10 und IV 11 vergleichbare Darstellungen handelt. Die Abbildungen IV 10 und IV 11 betrachten nur drei X-Werte (X_1, X_2, X_3) und zeigen die Verteilung (Grundgesamtheit) des dazugehörigen Störterms. In Abbildung IV 22 hingegen ist nur eine Stichprobe von X-Y-Kombinationen zu finden, sodass es für X_1, X_2 und X_3 nur einzelne Y-Werte gibt. Dies hat zur Folge, dass wir nicht die vollständige Verteilung der Störterme abbilden, sondern nur einzelne Ausprägungen ablesen können. Wir haben diese zusätzliche Darstellung bewusst gewählt, da sich damit im Falle eines einfachen linearen Regressionsmodells Heteroskedastizität grafisch aufdecken lässt.

Grafik (1) in Abbildung IV 22 zeigt, dass die Abweichungen der Y von der Regressionsfunktion der Grundgesamtheit relativ gering ausfallen und mit zunehmendem X weder zu- noch abnehmen. Die Streuung (Varianz) der Abweichungen (Störterme) kann also als konstant betrachtet werden. Es liegt Homoskedastizität vor. Grafik (2a) hingegen zeigt, dass mit zunehmendem X auch die Abweichungen der Y von der Regressionsfunktion zunehmen. Die Streuung des Störterms steigt also mit X, sodass wir für jedes X_i eine andere Streuung feststellen und daher von Heteroskedastizität sprechen können. Grafik (2b) zeigt den umgekehrten Fall von (2a), also Heteroskedastizität mit fallender Varianz.

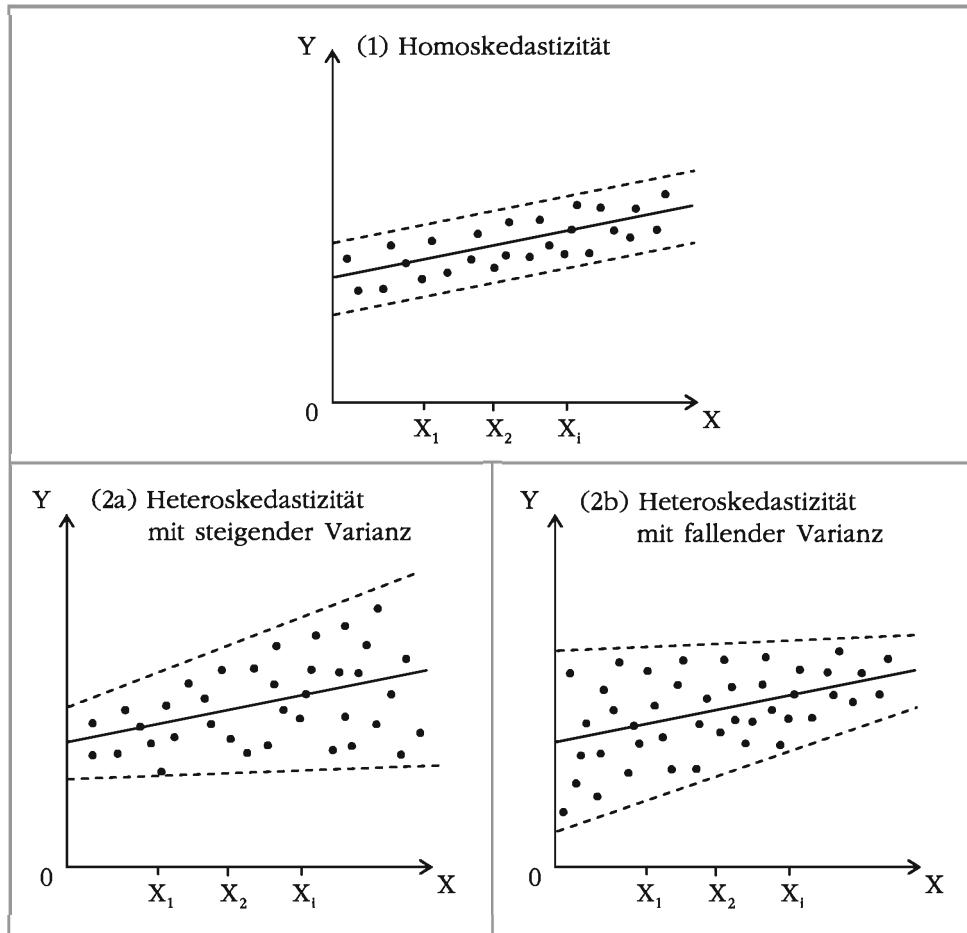


Abbildung IV 22: Homo- vs. Heteroskedastizität

Ein typisches Beispiel für *Heteroskedastizität mit steigender Varianz* kann anhand einer einfachen Konsumfunktion (Konsum in Abhängigkeit vom Einkommen) gegeben werden: Haushalte mit niedrigem Einkommen sind bei ihren Ausgabeentscheidungen in der Regel nicht sehr flexibel. Den Großteil ihres Einkommens werden sie für Lebensmittel, Kleidung und Fortbewegung ausgeben. Bei geringem Einkommen wird sich der Konsum der einzelnen Haushalte also kaum voneinander unterscheiden (geringe Streuung). Wohlhabende Haushalte hingegen sind flexibler und haben eine breitere Auswahl hinsichtlich ihrer Ausgaben. Einige konsumieren mehr und andere sparen mehr. Der individuelle Konsum kann sich also bei reichen Haushalten deutlich vom durchschnittlichen Konsum (gegeben durch die Regressionsfunktion der Grundgesamtheit) unterscheiden. Die Streuung bei hohen Einkommen wird also höher ausfallen als bei niedrigem.

Heteroskedastizität mit fallenden Varianzen können wir beispielsweise auf *evolutionäre Prozesse* zurückführen. So werden etwa im Zeitverlauf bessere Datenerhebungsmethoden entwickelt, die auftretende Messfehler verringern. Eine weitere Erklärung stellen *Lernprozesse* dar. Ist Y z.B. die Anzahl der Punkte in einem Test

und X die Anzahl der Versuche, mit denen ein Individuum den Test durchläuft, so nimmt die Variabilität von Y mit zunehmendem X ab.

Heteroskedastizität tritt häufig in Datensätzen auf, in denen ein *deutlicher Unterschied zwischen den größten und kleinsten beobachteten Werten der erklärenen Variablen* besteht. Je größer dieser Unterschied ist, desto größer ist die Wahrscheinlichkeit, dass die Störterme, die mit ihnen verbunden sind, heteroskedastisch sind. Es ist zu erwarten, dass die Verteilung der Störterme großer Beobachtungen hohe und die kleiner Beobachtung kleine Varianzen aufweisen. Besonders in *Querschnittsdaten* erhalten wir oft eine derartige Diskrepanz zwischen den niedrigsten und den höchsten Werten der erklärenen Variablen. So dürfte etwa in einer Studie zu Konsumausgaben in Abhängigkeit vom Einkommen in deutschen Bundesländern die Differenz zwischen Bayern und dem Saarland recht groß sein. Zudem werden wir annehmen können, dass die Ausgaben des Saarlandes absolut betrachtet nicht so variabel sind wie die von Bayern, da z.B. eine 5 %ige Ausgabenerhöhung in einem großen Bundesland einen höheren Betrag bedeutet, als in einem kleinen.

Nicht nur in Querschnittsdaten mit einer großen Variation der erklärenen Variable kann Heteroskedastizität auftreten. Sie ist auch in *Zeitreihendaten mit hohen Veränderungsraten* zu beobachten. Würden wir beispielsweise den Absatz an DVD-Playern zwischen 1980 und 2009 modellieren, so würden höchstwahrscheinlich heteroskedastische Störterme vorliegen. Mit Beginn des gigantischen Wachstums der Branche wird nämlich auch die Varianz des Störterms gestiegen sein. In Zeitreihen mit geringen Veränderungsraten ist Heteroskedastizität aber eher unwahrscheinlich.

Wie diese Beispiele zeigen, liegt Heteroskedastizität meist in der **Natur der Sache**. Die Grundgesamtheit verhält sich also oft einfach heteroskedastisch. Die Suche nach einer Spezifikation ohne Heteroskedastizität erübrigt sich daher. Vielmehr sind in einem solchen Fall Schätzverfahren heranzuziehen, die der Heteroskedastizität Rechnung tragen, oder Korrekturverfahren zu nutzen, die ihren Konsequenzen entgegenwirken (vgl. Abschnitt IV 4.4.4).

Heteroskedastizität kann ihre Ursache aber auch in **Spezifikationsfehlern** haben. Eine falsche funktionale Form führt in den seltensten Fällen zu Heteroskedastizität. Eine vernachlässigte Variable jedoch schon, da der Anteil des vernachlässigten Effekts auf Y, der nicht von einer der anderen erklärenden Variablen repräsentiert wird, vom Störterm absorbiert wird. Hat dieser Effekt eine heteroskedastische Komponente, kann der Störterm der fehlspezifizierten Gleichung heteroskedastisch sein, auch wenn es der Störterm der korrekten Gleichung nicht ist.

Eine Unterscheidung dieser beiden Heteroskedastizitätsursachen ist von großer Bedeutung. Einer durch Spezifikationsfehler verursachten Heteroskedastizität gilt es nämlich durch Beseitigung der Fehlspezifikation und nicht mit den Methoden unter IV 4.4.4 zu begegnen. Diese sind nur für den Fall vorgesehen, dass sich die Grundgesamtheit einfach heteroskedastisch verhält.

4.4.2 Konsequenzen von Heteroskedastizität

Ein heteroskedastischer Störterm hat für die Schätzung von Regressionsmodellen im Allgemeinen drei Folgen, die mit denen von Autokorrelation (vgl. Abschnitt IV 4.5.2) vergleichbar sind:

1. Keine Verzerrung der OLS-Schätzer

Im Falle von Heteroskedastizität kommt es zu keiner Verzerrung (und auch keiner Inkonsistenz) der OLS-Schätzer.²⁷ Ein Blick auf den Erwartungswert des Schätzers des Steigungskoeffizienten im einfachen Regressionsmodell, den wir in (IV2: 21) im Fall deterministischer X als

$$E(\hat{\beta}_1) = \beta_1 + \frac{\sum x_i E(\epsilon_i)}{\sum x_i^2} = \beta_1$$

hergeleitet hatten, kann uns helfen, dies klar zu machen. Wir erkennen nämlich deutlich, dass die Varianzen der Störterme keine Bedeutung bei der Erbringung dieses Beweises der Erwartungstreue haben. Auch bei Heteroskedastizität bleiben OLS-Schätzer also erwartungstreue.

2. OLS-Schätzer nicht länger effizient

Im Falle von Heteroskedastizität ist das Gauß-Markov-Theorem nicht länger gültig. Die OLS-Schätzer weisen nicht länger minimale Varianz auf (sind also nicht mehr effizient), da es andere lineare unverzerrte Schätzer gibt, die eine kleinere Varianz besitzen. Mit diesen (sog. WLS-Schätzer) werden wir uns unter IV 4.4.4.1 näher beschäftigen. Die höhere Varianz der OLS-Schätzer erklärt sich dadurch, dass ein heteroskedastischer Störterm Fluktuationen von Y verursacht, die OLS den erklärenden Variablen zuordnet. Insgesamt bleiben die Schätzer dabei jedoch unverzerrt, da Über- und Unterschätzungen gleichwahrscheinlich sind und sich daher gegenseitig ausgleichen.

3. Verzerrung der Schätzer der Standardfehler der OLS-Schätzer

Die Tatsache, dass Heteroskedastizität die Schätzer der OLS-Standardfehler verzerrt, können wir mit einer Herleitung der Varianz von $\hat{\beta}_1$ im einfachen Regressionsmodell veranschaulichen, auf die wir bisher verzichtet hatten. Wir gehen dazu zunächst von der allgemeinen Definition der Varianz von $\hat{\beta}_1$ aus, die aus der allgemeinen Definition der Varianz einer Zufallsvariablen hervorgeht:

$$\text{Var}(\hat{\beta}_1) = E\left[\left(\hat{\beta}_1 - E(\hat{\beta}_1)\right)^2\right]$$

Da $\hat{\beta}_1$ ein erwartungstreuer Schätzer für β_1 ist, also $E(\hat{\beta}_1) = \beta_1$ gilt, und außerdem aus (IV2: 20)

$$\hat{\beta}_1 - \beta_1 = \frac{\sum x_i \epsilon_i}{\sum x_i^2} = \sum k_i \epsilon_i \quad \text{mit} \quad k_i = \frac{x_i}{\sum x_i^2}$$

formuliert werden kann, erhalten wir aus dieser Definition

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= E\left[(\hat{\beta}_1 - \beta_1)^2\right] = E\left[\left(\sum k_i \epsilon_i\right)^2\right] \\ &= E\left[(k_1 \epsilon_1)^2 + 2k_1 k_2 \epsilon_1 \epsilon_2 + (k_2 \epsilon_2)^2 + \dots + (k_n \epsilon_n)^2\right]. \end{aligned}$$

²⁷ Liegt die Ursache der Heteroskedastizität in z.B. einer vernachlässigten Variablen kann es natürlich aufgrund der Fehlspezifikation zur Verzerrung kommen (vgl. z.B. IV 4.1.1). Die Heteroskedastizität selbst verursacht keine Verzerrung.

Durch die Annahme unkorrelierter Störterme

$$\text{Cov}(\epsilon_i, \epsilon_j) = E[(\epsilon_i - E(\epsilon_i))(\epsilon_j - E(\epsilon_j))] = E(\epsilon_i \epsilon_j) = 0 \quad \text{für } i \neq j \quad (\text{IV4: 32a})$$

erhalten wir

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= E[(k_1 \epsilon_1)^2] + E[(k_2 \epsilon_2)^2] + \dots + E[(k_n \epsilon_n)^2] \\ &= k_1^2 E(\epsilon_1^2) + k_2^2 E(\epsilon_2^2) + \dots + k_n^2 E(\epsilon_n^2) = \sum k_i^2 E(\epsilon_i^2). \end{aligned}$$

Resubstituieren wir k_i und bemerken, dass $E(\epsilon_i^2)$ nur ein alternativer Ausdruck für die Varianzen des stochastischen Störterms $\text{Var}(\epsilon_i) = \sigma_i^2$ ist, können wir allgemein festhalten, dass

$$\text{Var}(\hat{\beta}_1) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \quad (\text{IV4: 32b})$$

gilt. Im Falle von Heteroskedastizität hat also die Varianz des Schätzers $\hat{\beta}_1$ die Gestalt (IV4: 32b). Bei Annahme der Homoskedastizität, also $E(\epsilon_i^2) = \sigma^2$, erhalten wir unsere bisherige Formel

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2 \sum x_i^2}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}. \quad (\text{IV4: 33})$$

Diese stellt nur einen Sonderfall von (IV4: 32b) dar. Da OLS bei der Schätzung der Varianz der Schätzer nun auf die unter der Homoskedastizitätsannahme hergeleitete Standardformel für die Varianz der Schätzer (IV4: 33) und nicht auf (IV4: 32b) zurückgreift, führt ein Einsetzen von $\hat{\sigma}^2$ in die Standardformel natürlich zu falschen Schätzungen der Varianzen. OLS unterschätzt typischerweise in der Präsenz von Heteroskedastizität die Varianzen, da in der Praxis meist große σ_i^2 bei großen X_i auftreten. Die Folge sind daher unterschätzte Standardfehler, die wiederum zu erhöhten t-Werten führen, was Testentscheidungen verfälscht. Wir lehnen Nullhypotesen zu oft ab, sodass der Fehler 1. Art (Ablehnung einer korrekten Nullhypothese) deutlich zunimmt. Selbst wenn wir die korrekte Formel (IV4: 32b) zur Schätzung der Standardfehler heranziehen würden, hat dies keinen Effekt auf die Tatsache, dass Heteroskedastizität die Ineffizienz der OLS-Schätzer mit sich bringt. Die korrekten Varianzen wären nämlich größer als die der alternativen Schätzer unter IV 4.4.4.1.

4.4.3 Aufdecken von Heteroskedastizität

Es gibt generell zwei Möglichkeiten Heteroskedastizität aufzudecken. Zum einen können wir eine grafische Darstellung der Residuen auf gewisse Heteroskedastizitätsmuster untersuchen und zum anderen formale Tests zur Diagnose verwenden.

4.4.3.1 Grafische Methode

Im Fall von nur einer erklärenden Variablen, haben wir bereits im Zusammenhang mit Abbildung IV 22 gesehen, dass wir aus einer einfachen Grafik Hinweise auf Heteroskedastizität erhalten können. Im multiplen Regressionsmodell könnten wir für die einzelnen X_k ähnlich vorgehen. In der *Praxis* hat sich jedoch für *multiple*

Modelle die Methode eingebürgert, Informationen zu eventueller Heteroskedastizität aus zweidimensionalen Grafiken zu filtern, in denen e^2 auf der Y-Achse und \hat{Y} auf der X-Achse abgetragen wird.²⁸ Ziel ist es herauszufinden, ob es einen systematischen Zusammenhang zwischen \hat{Y} , d.h. einer *Linearkombination aller erklärenden Variablen*, und den quadrierten Residuen gibt. Besonders für große Stichproben können wir die e_i^2 nämlich als *Proxies für die Varianzen der Störterme* verwenden.²⁹ Insbesondere wenn wir mit dieser Methodik keinen systematischen Zusammenhang aufdecken können, also e^2 unabhängig von \hat{Y} auf einem bestimmten Niveau bleibt, empfiehlt es sich e^2 zusätzlich gegenüber den einzelnen erklärenden Variablen des Regressionsmodells abzutragen. Abbildung IV 23 zeigt exemplarische Resultate einer solchen Vorgehensweise. Hiermit sind wir in der Lage aufzudecken, ob die Varianz des stochastischen Störterms in irgendeiner funktionalen Beziehung zu *einer* der erklärenden Variablen steht. Für ein einfaches Regressionsmodell ist dies zweckmäßig. Auch für multiple Modelle können wir damit aufzeigen, ob eine Beziehung zwischen einzelnen erklärenden Variablen und der Störtermvarianz besteht. Problem ist jedoch, dass ein solches Verfahren nicht aufzudecken vermag, dass die Störtermvarianz ggf. in komplexeren funktionalen Strukturen von nur einigen der erklärenden Variablen abhängt. Für multiple Modelle ist daher ein Ausweichen auf formale Tests meist unvermeidlich.

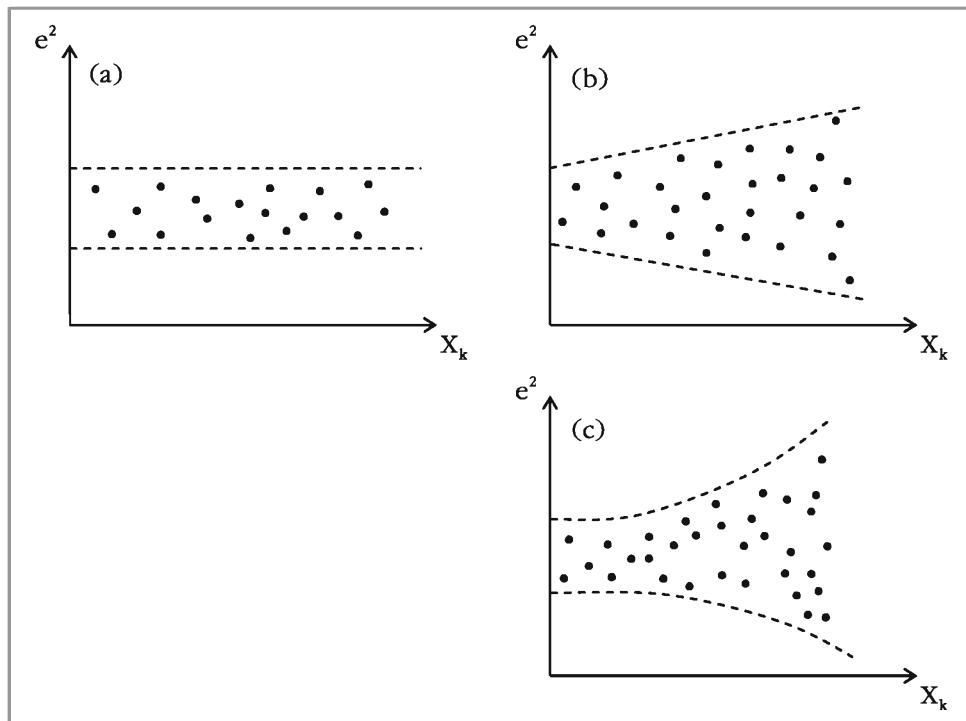


Abbildung IV 23: Grafische Aufdeckung von Heteroskedastizität

²⁸ Vgl. z.B. Gujarati, D.N. (2003), Kapitel 11.

²⁹ Eine detaillierte Behandlung des Zusammenhangs zwischen e_i und ϵ_i ist z.B. in Malinvaud, E. (1970) zu finden.

Nehmen wir an, wir betrachten ein Modell mit nur einer erklärenden Variable. Grafik (a) in Abbildung 23 lässt dann keinen systematischen Zusammenhang zwischen e^2 und X_k erkennen. Dies erlaubt die Aussage, dass wahrscheinlich keine Heteroskedastizität vorliegt. Die Grafiken (b) und (c) geben jedoch deutliche Hinweise auf Heteroskedastizität. (b) zeigt wie e^2 proportional zu X_k zunimmt. In (c) liegt ein quadratischer Zusammenhang zwischen e^2 und X_k vor, d.h. e^2 steigt proportional zu X_k^2 . Es besteht also eine klare Beziehung zwischen der betrachteten erklärenden Variable und der Varianz der Störterme (bzw. hier den quadrierten Residuen als deren Vertreter). Sind wir in der Lage die funktionale Beziehung zwischen der Variable X_k und der heteroskedastischen Varianz abzuleiten, so kann uns dies helfen, unsere Daten so zu transformieren, dass die Störtermvarianz in einem Regressionsmodell mit den transformierten Daten homoskedastisch ist. Mit solchen Transformationen werden wir uns unter IV 4.4.4.1 näher beschäftigen.

Beispiel IV 4-16

Eigentumswohnungen



Für eine deutsche Stadt sollen die Preise von Eigentumswohnungen in Euro Y durch die Anzahl der Wohnräume X_1 und der Wohnfläche in m^2 X_2 durch das einfache Modell

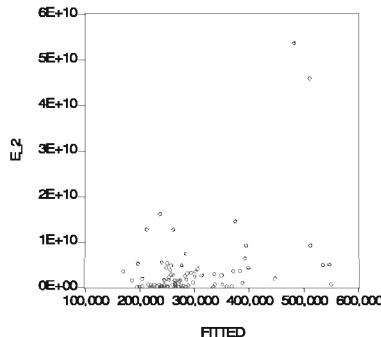
$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \epsilon_1$$

erklärt werden. EViews liefert dafür die folgende Schätzung, die uns zunächst zeigt, dass der Koeffizient von X_2 auf einem 5 %-Niveau signifikant ist, der Koeffizient von X_1 jedoch das Signifikanzniveau knapp verfehlt. Untersuchen wir nun, ob ggf. Heteroskedastizität im Modell vorliegt, die unsere Standardfehler und damit auch die t- und p-Werte verzerrt.

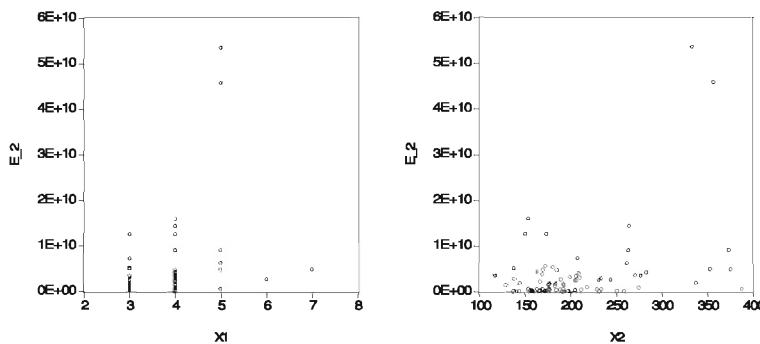
Dependent Variable: Y		Sample: 1 83		
Method: Least Squares		Included observations: 83		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-35194.85	33164.77	-1.061212	0.2918
X1	20044.33	10167.98	1.971318	0.0521
X2	1252.161	133.6639	9.367979	0.0000
R-squared	0.658418	Mean dependent var		294868.1
Adjusted R-squared	0.649878	S.D. dependent var		104056.6
S.E. of regression	61571.43	Akaike info criterion		24.92926
Sum squared resld	3.03E+11	Schwarz criterion		25.01669
Log likelihood	-1031.564	F-statistic		77.10216
Durbin-Watson stat	1.677034	Prob(F-statistic)		0.000000

Tragen wir zunächst die geschätzten Y-Werte gegenüber den quadrierten Residuen ab. In EViews können wir dazu über den Button "Genr" durch Eingabe von $e_2 = resid^2$ die Reihe der quadrierten Residuen erzeugen, da EViews nach Schätzung eines Modells in der Variable "resid" automatisch die Residuen ablegt. Die gefitteten Werte \hat{Y} können wir damit über den Befehl $fitted = y - resid$ bestimmen. Mit beiden neu erzeugten Reihen erhalten wir dann die auf der Folgeseite zu sehende Grafik. Anhand dieser wird schnell klar, dass wir nicht immer mustergültige Strukturen im Datenmaterial erhalten, die auf Heteroskedastizität hindeuten. Hier sind es vielmehr nur zwei Beobachtungen, die für ein auf Heteroskedastizität hindeutendes Bild sorgen. Gerade in einem solchen Fall gilt es den zugrunde liegenden Datensatz dahingehend zu überprüfen, dass es zu keinen Erfassungsfehlern gekommen ist und die Ausreißer tatsächlich für die zu untersuchende Grundge-

samtheit relevant sind. Dies sei in unserem Beispiel der Fall. Ein Ausschluss der Beobachtungen zur Heteroskedastizitätsbeseitigung ist daher nicht gerechtfertigt.



Untersuchen wir zusätzlich, ob eine Abhängigkeit zwischen den einzelnen erklärenden Variablen und den quadrierten Residuen bestehen könnte. Die beiden folgenden Grafiken bieten dazu keinen besonderen Mehrwert, da wir die Heteroskedastizität nicht eindeutig auf eine der Variablen zurückführen können. Wieder fallen nur die beiden Beobachtungen auf, die auf Heteroskedastizität hindeuten.



Aufgrund der geringen Aussagekraft dieser ersten grafischen Analyse, wenden wir uns besser statistischen Verfahren zu, die aussagekräftigere Ergebnisse liefern.

4.4.3.2 Breusch-Pagan LM-Test

Betrachten wir ein Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + \epsilon_i, \quad (\text{IV4: 34a})$$

in dem

$$\text{Var}(\epsilon_i) = \sigma_i^2 \quad (\text{IV4: 34b})$$

gilt, die Annahme der Homoskedastizität also verletzt ist. Für ein solches Modell sieht der **Breusch-Pagan LM-Test** folgende Vorgehensweise zur Aufdeckung der Heteroskedastizität vor:³⁰

³⁰ Diese Variante des ursprünglich von Breusch, T., Pagan, A. (1979) und Godfrey, L.G. (1978) entwickelten Tests wurde von Koenker, R. (1981) vorgeschlagen.

Zunächst wird das Modell (IV4: 34a) geschätzt, um die Residuen e_i zu erhalten. Diese werden dann verwendet um die Hilfsregression

$$e_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi} + u_i \quad (\text{IV4: 35})$$

zu schätzen, in der die Z_{pi} Variablen sind, von denen angenommen wird, dass sie die Varianz des Störterms bestimmen. Wir erkennen daran, dass die e_i^2 auch hier als Proxies für die unbekannten σ_i^2 dienen.

Auf Basis der Hilfsregression (IV4: 35) wird schließlich das Hypothesenpaar

$$\begin{aligned} H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0 & \quad (\text{Homoskedastizität}) \\ H_A : \text{mindestens ein } \alpha_p \text{ ungleich Null} & \quad (\text{Heteroskedastizität}) \end{aligned} \quad (\text{IV4: 36})$$

untersucht. Dahinter steckt die einfache Idee, dass wenn $\alpha_1 = \alpha_2 = \dots = \alpha_p = 0$ gilt, die Varianz der Störterme einer Konstanten α_0 entspricht. Es gilt nämlich dann $e_i^2 = \alpha_0 + u_i$ und die e_i^2 sind ja, wie bereits gesagt, die Proxies der σ_i^2 .

Für die Testentscheidung wird dann die LM-Statistik

$$LM = nR^2 \quad (\text{IV4: 37})$$

bestimmt, wobei n der Stichprobenumfang bei der Berechnung der Hilfsregression und R^2 das Bestimmtheitsmaß der Hilfsregression (IV4: 35) ist. LM folgt bei korrekter Nullhypothese *asymptotisch* einer Chi-Quadrat-Verteilung mit P Freiheitsgraden. Zu einer Ablehnung von H_0 und damit der Folgerung von Heteroskedastizität kommt es, wenn die LM-Statistik größer als der kritische Wert $\chi^2_{1-\alpha,p}$ bzw. der p-Wert geringer als das gewählte Signifikanzniveau ist.

(IV4: 35) lässt vermuten, dass wir mit dem Breusch-Pagan LM-Test nur lineare Formen von Heteroskedastizität berücksichtigen können. Breusch und Pagan haben jedoch gezeigt, dass der Test allgemein Heteroskedastizität in Bezug auf jede kontinuierliche Funktion $h(\alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi})$ entdeckt. Für die Durchführung des Tests ist es daher nicht notwendig die exakte Form der Heteroskedastizität zu formulieren. Wir erkennen auch, dass der Breusch-Pagan LM-Test die *Variablen Z_p nicht genauer spezifiziert*. In empirischen Anwendungen werden die Z_p daher meist durch alle oder eine Teilmenge der Regressoren ersetzt.

Von besonderer Bedeutung bei diesem Test ist, dass in der Literatur zwei verschiedene Varianten zu finden sind. Die von uns vorgestellte LM-Variante ist wie jeder **Lagrange-Multiplier-Test** (LM-Test) nur für *große Stichproben* geeignet, d.h. wir erhalten die genannte Verteilung der Teststatistik nur asymptotisch. Eine *Normalverteilungsannahme* für die Störterme des zu testenden Regressionsmodells ist hier *nicht erforderlich*. In seiner Urform wurden zur Testdurchführung in (IV4: 35) an Stelle der e_i^2 die Variablen $p_i = e_i^2 / \tilde{\sigma}^2$ mit $\tilde{\sigma}^2 = \sum e_i^2 / n$ verwendet. Die Teststatistik wurde als $0,5 \cdot ESS$ bestimmt, wobei ESS die erklärte Quadratesumme der Testregression war. Wichtige Annahme bei dieser Variante des Tests ist die der Normalverteilung der Störterme ϵ_i . Nur bei Homoskedastizität und Normalverteilung der ϵ_i folgt die Teststatistik $0,5 \cdot ESS$ asymptotisch einer Chi-Quadrat-Verteilung mit P Freiheitsgraden. Es ist daher besonders im Zusammenhang mit dieser ursprünglichen Testvariante empfehlenswert die Annahme der Normalverteilung der ϵ_i auf ihre Gültigkeit zu testen. In der Praxis wird dies aber meist vernachlässigt, da häufig Testverfahren verwendet werden, die die Annahme der Normalverteilung nicht

erfordern. Zu diesen zählt die vorgestellte LM-Variante des Breusch-Pagan-Tests oder der White-Test (vgl. Abschnitt IV 4.4.3.3).

Beispiel IV 4-17

Breusch-Pagan LM-Test

Im Beispiel IV 4-16 haben wir durch eine erste grafische Analyse bereits grobe Hinweise auf vorliegende Heteroskedastizität in unserem Eigentumswohnungsmodell bekommen. Versuchen wir nun diese Indizien mittels eines Breusch-Pagan LM-Tests statistisch zu untermauern. In EViews können wir diesen Test im Fenster der geschätzten Gleichung unter View / Residual Tests / Heteroskedasticity Tests / Breusch-Pagan-Godfrey durchführen. Standardmäßig wählt der Test die erklärenden Variablen des Modells als Erklärungsgrößen für die Varianz des Störterms. Benutzerdefiniert können jedoch auch andere Variablen angegeben werden. Mit den Standardeinstellungen erhalten wir folgendes Testergebnis:

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	8.925563	Prob. F(2,80)	0.0003
Obs*R-squared	15.14181	Prob. Chi-Square(2)	0.0005
Scaled explained SS	33.36539	Prob. Chi-Square(2)	0.0000

Test Equation:

Dependent Variable: RESID^2

Sample: 1 83

Method: Least Squares

Included observations: 83

	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.14E+10	3.95E+09	-2.879107	0.0051
X1	1.84E+09	1.21E+09	1.519835	0.1325
X2	40481554	15911959	2.544096	0.0129
R-squared	0.182431	Mean dependent var		3.65E+09
Adjusted R-squared	0.161992	S.D. dependent var		8.01E+09
S.E. of regression	7.33E+09	Akaike info criterion		48.30376
Sum squared resid	4.30E+21	Schwarz criterion		48.39119
Log likelihood	-2001.606	F-statistic		8.925563
Durbin-Watson stat	1.389305	Prob(F-statistic)		0.000317

Wir können hieraus unseren LM-Wert 15,1418 ablesen. Er ist größer als der kritische Wert 5,99, den wir bei einem Signifikanzniveau von 5 % aus einer Chi-Quadrat-Verteilung mit zwei Freiheitsgraden (entspricht Anzahl der erklärenden Variablen in der Hilfsregression) ablesen können. Wir können die Nullhypothese der Homoskedastizität ablehnen. Alternativ zeigt uns dies auch der deutlich unter 0,05 liegende p-Wert zur LM-Statistik von 0,0005. Wichtig ist außerdem, dass wir anhand der Testgleichung sehen, dass nur der geschätzte Parameter der Variable X₂ signifikant von Null verschieden ist. Dies deutet darauf hin, dass die Varianz des stochastischen Störterms nur von der Variable X₂ abhängt. Dies ist eine wertvolle Information, die bei der Bekämpfung der Heteroskedastizität hilfreich sein wird.

4.4.3.3 White-Test

Als Alternative zum Breusch-Pagan LM-Test bietet sich der **White-Test** an.³¹ Er ist ein für große Stichproben zu verwendender Lagrange-Multiplier-Test, der im Vergleich zum Breusch-Pagan LM-Test für die Z_p eine konkrete Belegung vorsieht.

³¹ Vgl. White, H. (1980).

Betrachten wir ein multiples Regressionsmodell der Form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \quad (\text{IV4: 38})$$

so sieht der White-Test zunächst die Schätzung dieses Modells zur Erlangung der e_i vor. Diese werden zur Schätzung der Hilfsregression

$$e_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 X_{1i} X_{2i} + u_i \quad (\text{IV4: 39})$$

herangezogen. Nach Formulierung des Hypothesenpaars

$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_5 = 0$	(Homoskedastizität)
$H_A : \text{mindestens ein } \alpha_p \text{ ungleich Null}$	(Heteroskedastizität) (IV4: 40)

wird die LM-Statistik

$$LM = nR^2 \quad (\text{IV4: 41})$$

berechnet, die bei korrekter Nullhypothese *asymptotisch* einer Chi-Quadrat-Verteilung mit Freiheitsgraden in Höhe der Anzahl der erklärenden Variablen P in der Hilfsregression (hier 5) folgt. n ist hierbei der Stichprobenumfang in der Hilfsregression und R^2 ihr Bestimmtheitsmaß. Ist LM größer als der kritische Wert aus der Chi-Quadrat-Verteilung oder der p-Wert kleiner als das gewählte Signifikanzniveau, so kann H_0 abgelehnt werden.

Die Anwendung des White-Tests ist trotz seiner Vorteile mit Vorsicht zu genießen. Weist ein Regressionsmodell nämlich viele erklärende Variablen auf, so können ihre Kreuzprodukte, die in der Hilfsregression (IV4: 39) entstehen, sehr schnell zu einer Aufzehrung von Freiheitsgraden führen. Im Fall von 3 erklärenden Variablen würden sich in der Hilfsregression schon 9 Terme ergeben. In einem solchen Fall empfiehlt sich folgender Ausweg:³²

An Stelle einer Hilfsregression des Typs (IV4: 39) schätzen wir

$$e_i^2 = \alpha_0 + \alpha_1 \hat{Y}_i + \alpha_2 \hat{Y}_i^2 + u_i, \quad (\text{IV4: 42})$$

wobei \hat{Y}_i wie gewohnt eine lineare Funktion aller X_k ist und \hat{Y}_i^2 als eine Funktion aller Quadrate und Kreuzprodukte der erklärenden Variablen interpretiert wird. Unabhängig von der Anzahl der erklärenden Variablen in der ursprünglichen Gleichung können wir die Nullhypothese der Homoskedastizität bzw. $H_0: \alpha_1 = \alpha_2 = 0$ dann ebenfalls unter Nutzung von (IV4: 41) prüfen, wobei hier dann natürlich das R^2 der Schätzung von (IV4: 42) zu verwenden ist und eine Chi-Quadrat-Verteilung mit zwei Freiheitsgraden zugrunde liegt.

Nun noch eine wichtige Anmerkung zu Spezifikationsfehlern: Ist das Ergebnis eines White-Tests die Ablehnung von H_0 , so muss dies nicht zwangsläufig ein Beweis für vorliegende Heteroskedastizität sein. Es wird nämlich häufig argumentiert, dass der White-Test sowohl als Test auf Heteroskedastizität als auch als Test auf Spezifikationsfehler (oder beides) betrachtet werden kann. Es könnte also genau so gut sein, dass ein Spezifikationsfehler für ein derartiges Testergebnis verantwortlich ist.³³

³² Vgl. Wooldridge, J.M. (2006), Kapitel 8.

³³ Thursby J.G. (1982) und Harris, R. (1995) betrachten dieses Problem im Detail.

Beispiel IV 4-18

White-Test



Führen wir unsere Prüfung auf Heteroskedastizität zu unserem Eigentumswohnungsmodell, nun mit dem White-Test durch. Wir können diesen in EViews analog zum Breusch-Pagan LM-Test finden und erhalten folgendes Ergebnis:

Heteroskedasticity Test: White

F-statistic	4.758280	Prob. F(5,77)	0.0008
Obs*R-squared	19.59181	Prob. Chi-Square(5)	0.0015
Scaled explained SS	43.17108	Prob. Chi-Square(5)	0.0000

Test Equation:

Dependent Variable: RESID^2

Sample: 1 83

Method: Least Squares

Included observations: 83

	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.29E+09	1.55E+10	-0.341745	0.7335
X1	1.00E+10	6.74E+09	1.483567	0.1420
X1^2	-1.65E+09	9.28E+08	-1.779285	0.0791
X1*X2	24471340	17615717	1.389177	0.1688
X2	-1.58E+08	98453842	-1.609082	0.1117
X2^2	197901.6	205855.2	0.961363	0.3394
R-squared	0.236046	Mean dependent var		3.65E+09
Adjusted R-squared	0.186439	S.D. dependent var		8.01E+09
S.E. of regression	7.22E+09	Akaike info criterion		48.30823
Sum squared resid	4.02E+21	Schwarz criterion		48.48308
Log likelihood	-1998.791	F-statistic		4.758280
Durbin-Watson stat	1.349238	Prob(F-statistic)		0.000768

Der p-Wert unserer LM-Statistik liegt bei 0,0015, was bei einem Signifikanzniveau von 5 % wieder deutlich zu einer Ablehnung der Nullhypothese der Homoskedastizität führt.

Unsere alternative Testmöglichkeit bzw. die Hilfsregression (IV4: 42) ist im Folgenden geschätzt. Wir können aus den Ergebnissen $LM = n \cdot R^2 = 83 \cdot 0,196 = 16,27$ berechnen. Bei einem Signifikanzniveau von 5 % können wir dieser den kritischen Wert 5,99 aus der Chi-Quadrat-Verteilung mit 2 Freiheitsgraden gegenüberstellen. Da der LM-Wert größer ist als der kritische Wert, können wir auch hier die Nullhypothese Homoskedastizität ablehnen.

Dependent Variable: E_2

Sample: 1 83

Method: Least Squares

Included observations: 83

	Coefficient	Std. Error	t-Statistic	Prob.
C	6.95E+09	1.08E+10	0.646420	0.5199
FITTED	-52334.41	64240.76	-0.814660	0.4177
FITTED^2	0.129107	0.089197	1.447440	0.1517
R-squared	0.195676	Mean dependent var		3.65E+09
Adjusted R-squared	0.175568	S.D. dependent var		8.01E+09
S.E. of regression	7.27E+09	Akaike info criterion		48.28743
Sum squared resid	4.23E+21	Schwarz criterion		48.37486
Log likelihood	-2000.928	F-statistic		9.731192
Durbin-Watson stat	1.331788	Prob(F-statistic)		0.000165

4.4.4 Vorgehensweise bei festgestellter Heteroskedastizität

Geben unsere Prüfverfahren Hinweise auf Heteroskedastizität, besteht der nächste Schritt darin, das vorliegende Modell noch einmal gründlich zu betrachten. Es ist sicherzustellen, dass keine klaren Spezifikationsfehler (v.a. vernachlässigte Variablen) vorliegen, die die festgestellte Heteroskedastizität verursacht haben könnten. Können derartige *Spezifikationsfehler ausgeschlossen* werden und liegen *große Stichproben* vor, so können die ersten beiden der folgenden Verfahren zur Bekämpfung der nachteiligen Effekte von Heteroskedastizität angewendet werden. Im Falle von durch Spezifikationsfehler verursachter Heteroskedastizität gilt es diesen Spezifikationsfehlern zu begegnen.

4.4.4.1 Gewichtetes Prinzip der kleinsten Quadrate (WLS)

Die Anwendung des *gewichteten Prinzips der kleinsten Quadrate* (WLS, engl. weighted least squares) führt im Vergleich zur originären Schätzung zu neuen, effizient geschätzten Regressionsparametern, korrekten Standardfehlern und damit auch richtigen t-Werten. WLS sorgt nämlich durch seine Methodik dafür, dass die OLS-Schätzer wieder BLUE sind. Zum besseren Verständnis betrachten wir dafür das multiple Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i, \quad (\text{IV4: 43a})$$

in dem

$$\text{Var}(\epsilon_i) = \sigma_i^2 \quad (\text{IV4: 43b})$$

gilt, sonst jedoch alle weiteren Annahmen des CLRM erfüllt sind. Dividieren wir nun beide Gleichungsseiten von (IV4: 43a) durch die Standardabweichungen der stochastischen Störterme σ_i bzw. gewichten jeden Term mit $1/\sigma_i$ so erhalten wir

$$\frac{Y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{X_{1i}}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \dots + \beta_k \frac{X_{ki}}{\sigma_i} + \frac{\epsilon_i}{\sigma_i} \quad (\text{IV4: 44})$$

oder mit anderer Symbolik $\tilde{Y}_i = Y_i / \sigma_i$, $\tilde{X}_{0i} = 1 / \sigma_i$, $\tilde{X}_{ki} = X_{ki} / \sigma_i$ und $\tilde{\epsilon}_i = \epsilon_i / \sigma_i$

$$\tilde{Y}_i = \beta_0 \tilde{X}_{0i} + \beta_1 \tilde{X}_{1i} + \beta_2 \tilde{X}_{2i} + \dots + \beta_k \tilde{X}_{ki} + \tilde{\epsilon}_i. \quad (\text{IV4: 45})$$

Für dieses modifizierte Modell gilt nun

$$\text{Var}(\tilde{\epsilon}_i) = \text{Var}\left(\frac{\epsilon_i}{\sigma_i}\right) = \frac{\text{Var}(\epsilon_i)}{\sigma_i^2} = 1. \quad (\text{IV4: 46})$$

Die Varianzen der stochastischen Störterme des Modells (IV4: 45) sind also konstant (gleich Eins), d.h. homoskedastisch. Wir können daher mit OLS schätzen, wenn wir einfach die transformierten Variablen verwenden. Die im Zuge der Schätzung von (IV4: 45) verwendeten OLS-Schätzer werden auch als *WLS-Schätzer* bezeichnet. Sie sind aufgrund von (IV4: 46) wieder BLUE.

Das Prinzip von WLS ist auf Anhieb einleuchtend, doch hat es ein bedeutendes praktisches Problem. Die Werte σ_i^2 und damit auch die $1/\sigma_i$ sind nämlich unbekannt, sodass eine Schätzung von (IV4: 45) nicht ohne explizite Annahmen zur Struktur der σ_i^2 durchgeführt werden kann. Nehmen wir beispielsweise an, dass sich die Struktur der σ_i^2 wie folgt darstellt:

$$\text{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 Z_i \quad (\text{IV4: 47})$$

Z ist dabei eine Variable, von der wir glauben, dass sie als einzige σ_i^2 beeinflusst und deren Werte für alle i bekannt sind. Wäre Z eine der erklärenden Variablen X_k , so würden wir durch (IV4: 47) eine Struktur wie in Abbildung IV 23 (b) unterstellen.³⁴ Dividieren wir nun jeden Term in (IV4: 43a) durch $\sqrt{Z_i}$, so erhalten wir

$$\frac{Y_i}{\sqrt{Z_i}} = \beta_0 \frac{1}{\sqrt{Z_i}} + \beta_1 \frac{X_{1i}}{\sqrt{Z_i}} + \beta_2 \frac{X_{2i}}{\sqrt{Z_i}} + \dots + \beta_K \frac{X_{Ki}}{\sqrt{Z_i}} + \frac{\varepsilon_i}{\sqrt{Z_i}} \quad (\text{IV4: 48})$$

bzw. wieder in anderer Symbolik eine Form (IV4: 45). Wir können dann in diesem Fall für die Varianzen der ε_i^* Folgendes feststellen:

$$\text{Var}(\varepsilon_i^*) = \text{Var}\left(\frac{\varepsilon_i}{\sqrt{Z_i}}\right) = \frac{\text{Var}(\varepsilon_i)}{Z_i} = \frac{\sigma^2 Z_i}{Z_i} = \sigma^2 \quad (\text{IV4: 49})$$

Dies zeigt, dass das Problem der Heteroskedastizität im modifizierten (gewichteten) Modell nicht mehr auftritt. Wir können also (IV4: 48) mit OLS schätzen und unsere Schätzer sind wieder effizient, vorausgesetzt unsere Strukturannahme (IV4: 47) ist korrekt. Auch unsere t-Werte sind dann wieder verwendbar. Das R^2 aus der Regression (IV4: 48) kann jedoch *nicht mehr als Bestimmtheitsmaß für Y interpretiert* werden. Es gibt nun den Anteil der Variation von Y / \sqrt{Z} an, der durch die X_k^* erklärt wird. Es handelt sich dabei aber um nur schwer interpretierbare Variablen.

Bei der Betrachtung von (IV4: 48) fällt auf, dass die Modellgleichung keinen konstanten Term mehr aufweist, da β_0 zum Koeffizienten von $1 / \sqrt{Z_i}$ wird. An der Interpretation der Koeffizienten im Vergleich zum Ursprungsmodell ändert sich jedoch nichts. Nehmen wir z.B. $\text{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 X_{2i}$ also $Z_i = X_{2i}$ an. Wir erhalten dann aus (IV4: 48)

$$\frac{Y_i}{\sqrt{X_{2i}}} = \beta_0 \frac{1}{\sqrt{X_{2i}}} + \beta_1 \frac{X_{1i}}{\sqrt{X_{2i}}} + \beta_2 \frac{X_{2i}}{\sqrt{X_{2i}}} + \dots + \beta_K \frac{X_{Ki}}{\sqrt{X_{2i}}} + \frac{\varepsilon_i}{\sqrt{X_{2i}}} \quad (\text{IV4: 50})$$

In (IV4: 50) geben die β_k ($k = 1, 2, \dots, K$) weiterhin den partiellen Einfluss von X_k auf Y an. Schätzen wir das neue Modell, so erhalten wir $\hat{\beta}_0$ und $\hat{\beta}_1, \dots, \hat{\beta}_K$. Interessieren wir uns also für den heteroskedastizitätsfrei geschätzten Parameter von X_2 , so erhalten wir diesen als Schätzung des Koeffizienten von $X_{2i} / \sqrt{X_{2i}}$ in (IV4: 50). Benötigen wir die Schätzung des konstanten Terms β_0 , so finden wir diese als Schätzung des Koeffizienten von $1 / \sqrt{X_{2i}}$ in (IV4: 50).

Beispiel IV 4-19

WLS



Nehmen wir in unserem Eigentumswohnungsmodell aus Beispiel IV 4-16 an, dass es sich bei dem dortigen Modell um ein korrekt spezifiziertes handelt und dass es allein die erklärende Variable X_2 (Wohnfläche) ist, an die die Varianz des Störterms gebunden ist. Der Breusch-Pagan LM-Test gab im Beispiel IV 4-17 Hinweise darauf. Konkret ausgedrückt,

³⁴ Selbstverständlich sind auch andere Strukturannahmen bzw. funktionale Beziehungen zwischen der Varianz von ε_i und Z denkbar. Abbildung IV 24 (c) würden wir etwa über $\text{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 Z_i^2$ modellieren und im Folgenden mit $1/Z_i$ transformieren.

gehen wir also davon aus, dass $\text{Var}(\epsilon_i) = \sigma_i^2 = \sigma^2 X_{2i}$, also eine Struktur wie in Abbildung IV 23 (b) gilt. Die WLS-Schätzung erfolgt daher mit den Gewichten $1/\sqrt{X_{2i}}$. In EViews können wir eine WLS-Schätzung durchführen, indem wir bei der Gleichungsformulierung unter Quick / Estimate Equation / Options die Option "Weighted LS/TSLS" aktivieren und das Gewicht als $X2^{(-.5)}$ eingeben. Die WLS-Schätzung liefert dann folgendes Ergebnis:

Dependent Variable: Y		Sample: 1 83		
Method: Least Squares		Included observations: 83		
Weighting series: X2^(-.5)				
	Coefficient	Std. Error	t-Statistic	Prob.
C	-16731.43	32710.55	-0.511499	0.6104
X1	17517.87	9471.463	1.849542	0.0681
X2	1207.120	136.3049	8.856030	0.0000
Weighted Statistics				
R-squared	0.601611	Mean dependent var		285302.0
Adjusted R-squared	0.591651	S.D. dependent var		66636.36
S.E. of regression	56628.98	Akaike info criterion		24.76190
Sum squared resid	2.57E+11	Schwarz criterion		24.84933
Log likelihood	-1024.619	F-statistic		60.40436
Durbin-Watson stat	1.742107	Prob(F-statistic)		0.000000
Unweighted Statistics				
R-squared	0.656856	Mean dependent var		294868.1
Adjusted R-squared	0.648277	S.D. dependent var		104056.6
S.E. of regression	61712.07	Sum squared resid		3.05E+11
Durbin-Watson stat	1.663748			

Vorteil des Regressionsoutputs in EViews ist es, dass in diesem die transformierten Variablen X_k' nicht auftauchen. EViews ordnet die neuen geschätzten Koeffizienten direkt den originären Variablen zu. Im Vergleich zu den Regressionsergebnissen des originären Modells ohne Berücksichtigung der Heteroskedastizität

$$\hat{Y}_i = -35.194,85 + 20.044,33 X_{1i} + 1.252,16 X_{2i} \quad n = 83 (80)$$

$$(10.167,98) \quad (133,66) \quad R^2 = 0,6499$$

$$t = \quad 1,97 \quad 9,37$$

fällt auf, dass die geschätzten Regressionskoeffizienten deutlich gesunken sind. Auch die Standardfehler und damit die t-Statistiken und p-Werte haben sich geändert. Der Parameter von X_2 ist nach wie vor hoch signifikant, derjenige von X_1 wurde mit einem p-Wert von nun 0,0681 noch insignifikanter.

Kritisch ist im Zusammenhang mit WLS anzumerken, dass zum einen die Identifikation des Einflussfaktors Z ziemlich schwierig ist. Zum anderen stellt auch die funktionale Form, die den Einfluss von Z auf die Varianz der stochastischen Störterme beschreibt, ein Problem dar. Ohne ihre Kenntnis oder zumindest plausible Annahme kann WLS nicht angewendet werden.

Darüber hinaus ist zu bedenken, dass die Varianz von ϵ_i durchaus auch von mehreren Variablen Z bzw. vor allem in multiplen Regressionsmodellen von allen oder

Teilmengen der erklärenden Variablen in verschiedenen funktionalen Formen abhängen kann. In einem Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

wäre es z.B. denkbar, dass

$$\text{Var}(\varepsilon_i) = \sigma^2 = \sigma^2 \cdot \exp(\alpha_1 Z_{1i} + \alpha_2 Z_{2i})$$

gilt, die Varianz also in einer exponentiellen Form von zwei Einflussgrößen abhängt, die auch die erklärenden Variablen des Modells sein können. Wir nutzen hier ausnahmsweise $\exp(\dots)$ zur Darstellung der Exponentialfunktion um Verwechslungen mit den Residuen e_i zu vermeiden. Bei der Durchführung der WLS-Schätzung stehen wir nun vor dem Problem, dass wir zunächst die unbekannten Parameter α_1 und α_2 schätzen müssen. Konsistente Schätzer dafür können wir aus den OLS-Residuen e_i des Ursprungsmodells durch Schätzung von

$$\ln e_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + u_i$$

gewinnen.³⁵ Von dieser Schätzung benötigen wir die prognostizierten Werte (engl. fitted values). Damit berechnen wir $\hat{h}_i = \sqrt{\exp(\text{fitted value } i)}$ und verwenden diese Größen zur Transformation

$$\frac{Y_i}{\hat{h}_i} = \beta_0 \frac{1}{\hat{h}_i} + \beta_1 \frac{X_{1i}}{\hat{h}_i} + \beta_2 \frac{X_{2i}}{\hat{h}_i} + \frac{\varepsilon_i}{\hat{h}_i}.$$

Der Störterm dieses Modells erfüllt schließlich *asymptotisch* die Annahmen des CLRM. Bei einem solchen Verfahren, das zunächst Schätzungen für die Struktur der Heteroskedastizität (im Beispiel $\hat{\alpha}_0$, $\hat{\alpha}_1$ und $\hat{\alpha}_2$) erfordert, sprechen wir auch von **FWLS** (engl. feasible weighted least squares). Ein solches liefert jedoch nur für *große Stichproben* und bei *korrekter Strukturannahme* effiziente Schätzungen der Modellparameter des Ursprungsmodells, da wir die Parameter α_1 und α_2 nur konsistent schätzen können. In kleinen Stichproben sind die Abweichungen von $\hat{\alpha}_1$ und $\hat{\alpha}_2$ von ihren wahren Werten noch zu groß. FWLS ist daher in kleinen Stichproben nicht unbedingt effizienter als OLS.

FWLS-Schätzungen erfordern die korrekte Erfassung der Heteroskedastizitätsstruktur der Grundgesamtheit. Die von uns vorgestellten Testverfahren auf Heteroskedastizität sind zu allgemein. Sie testen nicht auf spezielle Strukturen, was aber gerade für die FWLS-Prozedur von enormer Bedeutung ist. Es sind daher hier speziellere Testverfahren erforderlich, mit denen auf konkrete funktionale Formen getestet werden kann.³⁶

4.4.4.2 White Standardfehler

Da im Falle von Heteroskedastizität die OLS-Schätzer der Regressionsparameter erwartungstreu bleiben, jedoch die Schätzer für ihre Varianzen bzw. Standardfehler verzerrt und inkonsistent sind, schlug White eine Methode vor, mit der sich konsistente Schätzer der Standardfehler gewinnen lassen.³⁷ Für *große Stichproben* und bei Heteroskedastizität liefern die White-Schätzer genauere Schätzungen der Standardfehler als die OLS-Methode sie ergibt.

³⁵ Für eine detaillierte Beschreibung dieses Verfahrens vgl. Verbeek, M. (2008), Kapitel 4.

³⁶ Vgl. dazu Verbeek, M. (2008), Kapitel 4.

³⁷ Vgl. White, H. (1980).

OLS schätzt den Standardfehler des Steigungsparameters aufgrund der Homoskedastizitätsannahme im einfachen Regressionsmodell, wie wir unter IV 2.3.1 aufgeführt und unter IV 4.4.2 hergeleitet haben, nach

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2} \quad \text{mit} \quad \sigma^2 = \frac{\sum e_i^2}{n - 2}.$$

In der Präsenz von Heteroskedastizität wäre aber ein Schätzer notwendig, der sich des Zusammenhangs (IV4: 32b), also

$$\text{Var}(\hat{\beta}_1) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}$$

bedient. Genau hier setzt White an und schlägt vor, in dieser Formel die σ_i^2 durch die quadrierten Residuen e_i^2 zu ersetzen. Wir erhalten damit einen konsistenten Schätzer der Varianz von $\hat{\beta}_1$. Die Wurzel davon wird als **White oder heteroskedastiekonsistenter Standardfehler** bezeichnet. In Hypothesentests empfiehlt es sich bei Heteroskedastizität in großen Stichproben auf diese und nicht auf die von OLS gelieferten Standardfehler zurückzugreifen. Die Ausgabe von White Standardfehlern ist mittlerweile in jedem Ökonometricssoftwarepaket möglich. Wie wir im folgenden Beispiel sehen werden, verändert die Verwendung von White Standardfehlern natürlich weder das R^2 einer geschätzten Gleichung noch die geschätzten Koeffizienten. EViews berechnet lediglich andere Standardfehler, t- und p-Werte. Im Vergleich zu WLS werden also nicht die effizientesten Schätzungen der Koeffizienten geliefert. In der Praxis werden aber dennoch korrigierte Standardfehler bevorzugt, da sich damit das Fehlerpotenzial durch ggf. fehlerhafte Heteroskedastizitätsstrukturannahmen bei WLS bzw. FWLS vermeiden lässt.

Beispiel IV 4-20

White Standardfehler



In Beispiel IV 4-19 begegneten wir der vorliegenden Heteroskedastizität durch WLS. Wir haben dabei festgestellt, dass sich durch WLS sowohl die geschätzten Regressionskoeffizienten als auch die Standardfehler der geschätzten Koeffizienten (und daraus resultierende Größen) verändern. Betrachten wir nun das Schätzergebnis unter Verwendung von White Standardfehlern. In EViews ist die Berechnung der White Standardfehler eine Option bei der Eingabe der Regressionsgleichung (Quick / Estimate Equation / Options - Heteroskedasticity consistent coefficient covariance - White). Im Vergleich zum originären Regressionsoutput des untersuchten Modells im Beispiel IV 4-16 haben sich anders als bei WLS die geschätzten Koeffizienten nicht verändert. Lediglich die Standardfehler wurden korrigiert, was für die Testentscheidung verlässlichere t- und p-Werte liefert.

Wie wir nun erkennen, ist auch der Parameter von X_1 bei einem Signifikanzniveau von 5 % signifikant von Null verschieden. Durch die WLS-Prozedur hatten wir ein solches Resultat jedoch nicht erhalten. Als Ursache für die unterschiedlichen Ergebnisse können wir zum einen die Tatsache sehen, dass der hier vorliegende Stichprobenumfang nicht sonderlich groß ist, d.h. die White Standardfehler noch nicht ihre optimalen Werte annehmen. Auf der anderen Seite könnte jedoch auch die im Rahmen der WLS-Schätzung getroffene Strukturannahme für die Varianz des stochastischen Störterms unzureichend sein. Möglicherweise hängt diese von mehr als einer Variablen ab und ggf. auch von solchen, die nicht im Modell enthalten sind.

Dependent Variable: Y		Sample: 1 83		
Method: Least Squares		Included observations: 83		
White Heteroskedasticity-Consistent Standard Errors & Covariance				
	Coefficient	Std. Error	t-Statistic	Prob.
C	-35194.85	44758.26	-0.786332	0.4340
X1	20044.33	9980.110	2.008428	0.0480
X2	1252.161	181.3212	6.905758	0.0000
R-squared	0.658418	Mean dependent var		294868.1
Adjusted R-squared	0.649878	S.D. dependent var		104056.6
S.E. of regression	61571.43	Akaike info criterion		24.92926
Sum squared resid	3.03E+11	Schwarz criterion		25.01669
Log likelihood	-1031.564	F-statistic		77.10216
Durbin-Watson stat	1.677034	Prob(F-statistic)		0.000000

4.4.4.3 Variablenredefinition

Neben den formalen Verfahren der beiden vorangegangenen Abschnitte besteht auch die Möglichkeit die Variablen eines Regressionsmodells so umzudefinieren, dass damit Heteroskedastizität vermieden wird. Dabei ist jedoch Vorsicht geboten, da damit die Modellspezifikation verändert wird, wodurch sich die Bedeutung der Regressionsgleichung ändern kann.

In vielen Fällen ist es zur Beseitigung von Heteroskedastizität ausreichend von einer *linearen funktionalen Form auf eine Doppel-Log-Form zu wechseln*. Diese unterliegt nämlich aufgrund ihres Wesens geringerer Variation als die lineare Form, sodass die Wahrscheinlichkeit auftretender Heteroskedastizität geringer ist. Zusätzlich ist die Verwendung einer Doppel-Log-Form in vielen Studien ebenso gerechtfertigt, wie die der linearen Form. Dies gilt insbesondere, wenn die lineare Form nur aufgrund der Tatsache gewählt wurde, dass sie eine Standardform darstellt.

Manchmal hilft es auch, die Variablen nicht absolut, sondern in *Pro-Kopf* zu definieren. Analysieren wir beispielsweise die Ausgaben verschiedener Gemeinden in Abhängigkeit vom jeweiligen Gesamteinkommen der Bevölkerung, so verwenden wir besser Pro-Kopf-Ausgaben und Pro-Kopf-Einkommen.

4.5 Autokorrelation

Thema dieses Abschnitts ist die Korrelation des stochastischen Störterms bzw. die Verletzung von Annahme 4 des CLRM. Sie tritt häufig in Zeitreihendaten auf, sodass wir uns in den Formeln dieses Abschnitts für die Verwendung des Variablen-subsripts t entschieden haben. Nehmen wir z.B. an, eine Regressionsfunktion erkläre die Geldnachfrage durch verschiedene Variablen. In einem solchen Fall kann eine unerwartete Änderung der Geldpolitik der Zentralbanken einen Schock auslösen, der sich im Störterm niederschlägt und gewöhnlich mehrere Perioden Auswirkungen auf die Geldnachfrage hat. Dies bedeutet, dass der Störterm in jeder Periode nicht nur Schocks aus dieser Periode, sondern auch Nachwirkungen von früheren Schocks beinhaltet. Dies führt dazu, dass die Störterme der verschiedenen Perioden miteinander korreliert sind. Wir sprechen dann von Autokorrelation (engl. serial correlation) des Störterms.

4.5.1 Formen von Autokorrelation

Wir sprechen allgemein von Autokorrelation, wenn Annahme 4 des CLRM, also die der Unkorreliertheit des stochastischen Störterms, verletzt wird. Sie liegt also vor, wenn Folgendes gilt:

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-s} | X_t, X_{t-s}) \neq 0 \quad \text{für } s \neq 0 \quad (\text{IV4: 51})$$

Der einfachste Prozess, der zu Autokorrelation führt, ist ein autoregressiver Prozess erster Ordnung oder kurz *AR(1)-Prozess*. Hier hängt der Wert des Störterms in der Periode t vom Störterm der Vorperiode $t-1$ ab, d.h. es gilt

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad \text{mit } |\rho| < 1. \quad (\text{IV4: 52})$$

ρ ist dabei ein Parameter und u_t ein klassischer (nicht autokorrelierter) Störterm, d.h. es liegt $u_t \sim \text{NID}(0, \sigma_u^2)$ vor (NID, engl. normally and independently distributed). Wie zu erkennen ist, wird der Störterm ε_t in zwei Komponenten zerlegt. Die erste Komponente besteht in der Störung, die in der Periode $t-1$ bereits wirksam war und nun in abgeschwächter Form $|\rho| < 1$ auch in Periode t wirksam wird. Die zweite Komponente ist die neue Störung u_t , die erst in Periode t aufgetreten ist. Wir beschränken die möglichen Werte von ρ bewusst auf ein Intervall zwischen -1 und 1 , da für Werte außerhalb davon schwerwiegende Folgen auftreten würden. Nehmen wir beispielsweise an, dass $\varepsilon_{t-1} = 0$ und $u_t \neq 0$ gilt. Selbst wenn in den Folgeperioden keine weiteren Störungen auftreten ($u_{t+1} = u_{t+2} = \dots = 0$), nähern sich die Störgrößen ε_{t+1} , ε_{t+2} usw. nicht mehr dem Wert Null an, sondern verharren auf u_t , falls $\rho = 1$. Für $|\rho| > 1$ würden die Störgrößen in diesem Fall sogar über die Zeit ins Unendliche wachsen. Ist bei einem AR(1)-Prozess $|\rho| < 1$ und sind damit diese Eigenschaften ausgeschlossen, wird er auch als **stationär** bezeichnet. Bei einem stationären Prozess sind u.a. sein *Erwartungswert* und seine *Varianz konstant*. Mit weiteren Eigenschaften stationärer Prozesse werden wir uns unter IV 4.7.2.1 noch näher beschäftigen.

Analysieren wir nun den autokorrelierten Störterm (IV4: 52) etwas näher. Wir können zunächst festhalten, dass für den *Erwartungswert* von ϵ_t trotz AR(1)-Prozess weiterhin $E(\epsilon_t) = 0$ gilt. Es gilt nämlich zunächst

$$E(\epsilon_t) = \rho E(\epsilon_{t-1}) + E(u_t).$$

Unterstellen wir Stationarität und damit $|\rho| < 1$ für den AR(1)-Prozess, so folgt daraus $E(\epsilon_t) = E(\epsilon_{t-1})$ (konstanter Erwartungswert) und wir können

$$E(\epsilon_t) = \rho E(\epsilon_{t-1}) + E(u_t)$$

schreiben. Da $|\rho| < 1$ gilt, folgt daraus

$$E(\epsilon_t) = \frac{E(u_t)}{1 - \rho}. \quad (\text{IV4: 53})$$

Da $E(u_t) = 0$ ist, folgt unmittelbar $E(\epsilon_t) = 0$. Wäre $\rho = 1$, so wäre die rechte Seite von (IV4: 53) nicht definiert.

Die *Varianz* von ϵ_t erhalten wir als

$$\text{Var}(\epsilon_t) = \rho^2 \text{Var}(\epsilon_{t-1}) + \sigma_u^2,$$

woraus sich unter Annahme der Stationarität und damit $\text{Var}(\epsilon_t) = \text{Var}(\epsilon_{t-1})$ (konstante Varianz)

$$\text{Var}(\epsilon_t) = \rho^2 \text{Var}(\epsilon_t) + \sigma_u^2$$

bzw.

$$\text{Var}(\epsilon_t) = \frac{\sigma_u^2}{1 - \rho^2} = \sigma^2 \quad (\text{IV4: 54})$$

ergibt. Neben dem Erwartungswert ist also auch die Varianz für alle ϵ_t identisch. Sie ist konstant und damit ist ϵ_t homoskedastisch. Bisher sind also trotz AR(1)-Störterm weder Annahme 2a, 2b oder 3 verletzt.

Bevor wir mit unseren Betrachtungen fortfahren, müssen wir an dieser Stelle die Begriffe **Autokovarianz** und **Autokorrelationskoeffizient** näher erläutern. Es handelt sich dabei um Maße, die uns den linearen Zusammenhang aktueller und vergangener Variablenwerte aufzeigen. Stellen wir den Realisationen einer Variablen X in Periode t jeweils die Realisation der Vorperiode $t-1$ gegenüber und berechnen mit diesen Wertepaaren die Kovarianz bzw. den Korrelationskoeffizienten, so sprechen wir von einer Autokovarianz bzw. einem Autokorrelationskoeffizienten **erster Ordnung**. Wir schreiben

$$\text{Cov}(X_t, X_{t-1}) = E[(X_t - E(X_t))(X_{t-1} - E(X_{t-1}))] \quad (\text{IV4: 55a})$$

und

$$\text{Corr}(X_t, X_{t-1}) = \frac{\text{Cov}(X_t, X_{t-1})}{\sigma(X_t) \cdot \sigma(X_{t-1})}, \quad (\text{IV4: 55b})$$

wobei $\sigma(\dots)$ die Standardabweichung der jeweiligen Größe darstellt. Stellen wir jeweils die Realisation vor s Perioden gegenüber, sprechen wir von **s -ter Ordnung** und schreiben analog zu (IV4: 55a) und (IV4: 55b) $\text{Cov}(X_t, X_{t-s})$ bzw. $\text{Corr}(X_t, X_{t-s})$. Es gilt $-1 \leq \text{Corr}(X_t, X_{t-s}) \leq 1$.

Die *Kovarianz* zweier aufeinanderfolgender Störgrößen bzw. die Autokovarianz erster Ordnung des Störterms berechnen wir im Falle eines AR(1)-Prozesses gemäß

$$\begin{aligned}\text{Cov}(\varepsilon_t, \varepsilon_{t-1}) &= E[(\varepsilon_t - E(\varepsilon_t))(\varepsilon_{t-1} - E(\varepsilon_{t-1}))] \\ &= E(\varepsilon_t \varepsilon_{t-1}) = E((\rho \varepsilon_{t-1} + u_t) \varepsilon_{t-1}) \\ &= \rho E(\varepsilon_{t-1}^2) + E(u_t \varepsilon_{t-1}).\end{aligned}$$

Setzen wir hier $E(\varepsilon_{t-1}^2) = \text{Var}(\varepsilon_t) = \sigma^2$ ein und berücksichtigen $E(u_t \varepsilon_{t-1}) = 0$, da die Störgröße u_t annahmegemäß nicht mit früheren Störungen korreliert ist, ergibt sich

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-1}) = \rho \sigma^2 \neq 0,$$

was klar Annahme 4 verletzt und uns zum Autokorrelationskoeffizienten

$$\text{Corr}(\varepsilon_t, \varepsilon_{t-1}) = \frac{\text{Cov}(\varepsilon_t, \varepsilon_{t-1})}{\sigma(\varepsilon_t) \cdot \sigma(\varepsilon_{t-1})} = \frac{\rho \sigma^2}{\sigma \sigma} = \rho \quad (\text{IV4: 56})$$

führt. Die Verwendung des Buchstabens ρ in (IV4: 52) war also kein Zufall. Er misst nämlich die Stärke der **Autokorrelation erster Ordnung** bzw. die Korrelation zwischen ε_t und ε_{t-1} .³⁸ Es lässt sich für einen AR(1)-Prozess allgemein zeigen, dass $\text{Cov}(\varepsilon_t, \varepsilon_{t-s}) = \rho^s \sigma^2 \neq 0$ und $\text{Corr}(\varepsilon_t, \varepsilon_{t-s}) = \rho^s$ gilt. Da $-1 < \rho < 1$ gilt, fällt der Wert der Autokovarianz und des Autokorrelationskoeffizienten also mit der zeitlichen Distanz der Störgrößen.

Ist ρ gleich Null, liegt keine Autokorrelation vor. Es würde nämlich dann in (IV4: 52) der Zusammengang $\varepsilon_t = u_t$ gelten. Nähert sich ρ dem Wert 1 oder -1 an, wird die Ausprägung des Störterms der Vorperiode für die Bestimmung des Störterms der aktuellen Periode immer wichtiger und es existiert hohe Autokorrelation. Das Vorzeichen von ρ gibt die Art der Autokorrelation an. Ist ρ positiv, bedeutet dies, dass der Störterm dazu tendiert von einer Periode auf die andere das gleiche Vorzeichen zu haben. Auf einen positiven Störterm folgt also wieder ein positiver und auf einen negativen ein negativer. Wir sprechen in einem solchen Fall von *positiver Autokorrelation*. Abbildung IV 24 zeigt zwei Beispiele positiver Autokorrelation. Die Ausprägungen der Störterme sind hier in chronologischer Reihenfolge angeordnet. Selbstverständlich wäre auch eine Darstellung möglich, wie wir sie in Abbildung IV 12 gewählt haben. *Negative Autokorrelation* liegt vor, wenn ρ negativ ist. Der Störterm hat hier die Tendenz in aufeinanderfolgenden Beobachtungen sein Vorzeichen zu wechseln. Auf positive Störterme folgen also negative, auf negative positive. Beispiele dafür sind ebenfalls in Abbildung IV 24 zu finden. Dort ist außerdem ein typischer Fall fehlender Autokorrelation (d.h. $\rho = 0$) enthalten.

Positive Autokorrelation ist in Zeitreihenmodellen häufig auf externe Schocks zurückzuführen, die Auswirkungen auf mehrere Folgeperioden haben. Der Störterm wird je nach Art des Schocks von mehreren Perioden positiver oder negativer Werte gekennzeichnet sein. Wir begegnen positiver Autokorrelation in der Regel häufiger als negativer. Negative Autokorrelation tritt oft bei Störterminen von Gleichungen auf, die in ersten Differenzen formuliert sind (vgl. Abschnitt IV 4.3.4), da die Regressionsbeziehung für die Beobachtungen in t die Größe $\Delta \varepsilon_t = \varepsilon_t - \varepsilon_{t-1}$ und in $t-1$ die Größe $\Delta \varepsilon_{t-1} = \varepsilon_{t-1} - \varepsilon_{t-2}$ als Störterm enthält.

³⁸ Wie wir diesen Autokorrelationskoeffizienten schätzen können, werden wir im Abschnitt IV 4.5.3.1 im Detail behandeln.

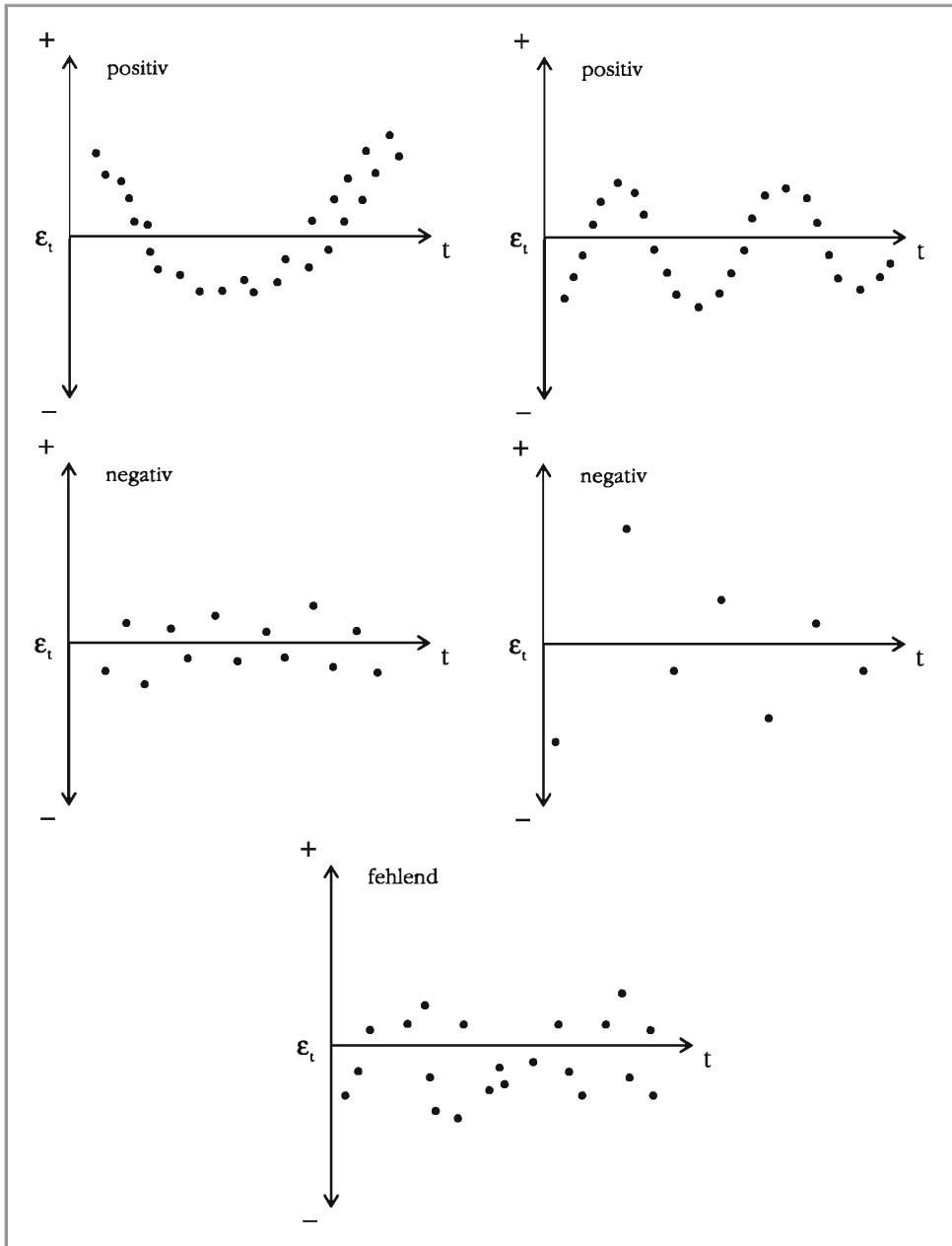


Abbildung IV 24: Positive, negative und fehlende Autokorrelation

Abgesehen von einer Form erster Ordnung kann Autokorrelation noch viele andere Formen annehmen. So könnte z.B. in einem Quartalsmodell der im aktuellen Quartal beobachtete Störterm von dem des Vorjahresquartals abhängen:

$$\varepsilon_t = \rho \varepsilon_{t-4} + u_t \quad \text{mit} \quad |\rho| < 1 \quad (\text{IV4: 57})$$

Es besteht ebenfalls die Möglichkeit, dass die aktuelle Störtermbeobachtung von mehr als einer vergangenen Störtermbeobachtung abhängt:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_s \varepsilon_{t-s} + u_t \quad (\text{IV4: 58})$$

Wir sprechen in einem solchen Fall auch von *Autokorrelation s-ter Ordnung* oder davon, dass der Störterm einem *AR(s)-Prozess* folgt.³⁹ Autokorrelationen höherer Ordnung sind jedoch in den meisten praktischen Anwendungen nicht so wahrscheinlich wie Autokorrelation erster oder zweiter Ordnung.

Bisher haben wir in diesem Lehrbuch von einem korrekt spezifiziertem Modell gesprochen, wenn es keine relevanten Variablen vernachlässigt und die korrekte funktionale Form aufweist. Nach dieser Auffassung wäre auch ein Modell

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (\text{IV4: 59a})$$

mit

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t, \quad (\text{IV4: 59b})$$

in dem ε_t autokorriert und u_t wieder ein klassischer nicht autokorrelierter Störterm ist, korrekt spezifiziert. Streng genommen darf es aber bei einem *wirklich korrekt spezifizierten Modell*, das dadurch gekennzeichnet ist, dass es auch die komplette Dynamik richtig abbildet, *keine Autokorrelation* des Störterms geben. Die komplette Dynamik ist ja spezifiziert. Das Modell (IV4: 59a) wäre demnach nicht wirklich korrekt spezifiziert, da die Dynamik nicht komplett berücksichtigt wurde. Wir können daraus aber ein wirklich korrekt spezifiziertes Modell generieren. Lösen wir (IV4: 59a) und das um eine Periode verzögerte (IV4: 59a) nach dem Störterm auf, so erhalten wir Folgendes:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \varepsilon_t & \Leftrightarrow \varepsilon_t &= Y_t - \beta_0 - \beta_1 X_t \\ Y_{t-1} &= \beta_0 + \beta_1 X_{t-1} + \varepsilon_{t-1} & \Leftrightarrow \varepsilon_{t-1} &= Y_{t-1} - \beta_0 - \beta_1 X_{t-1} \end{aligned}$$

Setzen wir dies in (IV4: 59b) ein, so erhalten wir

$$Y_t - \beta_0 - \beta_1 X_t = \rho(Y_{t-1} - \beta_0 - \beta_1 X_{t-1}) + u_t$$

und nach einigen Umformungen

$$Y_t = \beta_0(1 - \rho) + \beta_1 X_t - \rho \beta_1 X_{t-1} + \rho Y_{t-1} + u_t. \quad (\text{IV4: 59c})$$

Wir erhalten damit ein wirklich korrekt spezifiziertes *dynamisches Modell*⁴⁰ mit **unkorreliertem Störterm** u_t .⁴¹ Schreiben wir (IV4: 59c) in der Form

$$Y_t = \delta_0 + \delta_1 X_t + \delta_2 X_{t-1} + \gamma_1 Y_{t-1} + u_t, \quad (\text{IV4: 59d})$$

so erhalten wir ein Modell mit den gleichen Variablen wie in (IV4: 59c), jedoch anderen Koeffizienten. Diese stehen zueinander in den Beziehungen $\delta_0 = \beta_0(1 - \rho)$, $\delta_1 = \beta_1$, $\delta_2 = -\rho \beta_1$ und $\gamma_1 = \rho$. (IV4: 59c) ist eine restriktivere Form des mit OLS schätzbaren Modells (IV4: 59d). Gilt für das Modell (IV4: 59d) die testbare Restriktion $\delta_2 = -\delta_1 \gamma_1$ nicht, so wäre das Modell aus (IV4: 59a) und (IV4: 59b) sogar fehl spezifiziert. Daher empfiehlt es sich, nach Möglichkeit immer die Dynamik im Mo-

³⁹ Johnston, J., Dinardo, J. (1997) erläutern die Stationaritätsbedingung für einen AR(2)-Prozess, Enders, W. (2004) für einen allgemeinen AR(s)-Prozess.

⁴⁰ Den Themenkomplex dynamischer Modelle werden wir unter IV 4.7.1 noch vertiefen.

⁴¹ Das entstehende Modell (IV4: 59c) ist nicht mehr parameterlinear, d.h. nicht mehr mit OLS schätzbar. Eine Schätzung der Modellparameter kann nur mit iterativen Verfahren erfolgen.

dell und nicht im Störterm abzubilden. Durch das Hinzufügen von verzögerten X und Y kann im Allgemeinen die Autokorrelation eliminiert werden.⁴² Eine solch vollständige Spezifikation von allen empirischen Modellen zu verlangen, geht in der Praxis jedoch zu weit, da wir im Allgemeinen über die Dynamik nicht allzu viel wissen. Außerdem ist es bei vielen Fragestellungen nicht möglich und notwendig, ein in allen Details vollständig spezifiziertes Modell zu haben. Wir können daher unsere bisherige Auffassung korrekt spezifizierter Modelle aufrechterhalten. Wichtig ist nämlich nur, dass es zu keiner schwerwiegenden Verletzung der Annahme 2b kommt. Darauf kann Autokorrelation jedoch ein Hinweis sein:

Die *Vernachlässigung* einer für das Modell *bedeutsamen Variable*, die mit den verbleibenden Variablen korreliert ist, führt neben einer Verletzung von Annahme 2b und damit einer Verzerrung der OLS-Schätzer i.d.R. auch zur Autokorrelation des Störterms. Betrachten wir z.B. das korrekte Modell

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t, \quad (\text{IV4: 60a})$$

in dem wir eine Variable vernachlässigen und damit ein Modell

$$Y_t = \beta_0^* + \beta_1^* X_{1t} + \varepsilon_t^* \quad \text{mit} \quad \varepsilon_t^* = \varepsilon_t + \beta_2 X_{2t} \quad (\text{IV4: 60b})$$

erhalten. Wird eine Variable vernachlässigt, muss der Anteil dieses vernachlässigten Effekts auf die erklärte Variable, der nicht durch die einbezogenen Variablen repräsentiert werden kann, vom stochastischen Störterm absorbiert werden (vgl. IV 4.1.1). Dies ist in (IV4: 60b) geschehen. Anders als in (IV4: 60a) ist der Störterm nun also auch eine Funktion der vernachlässigten Variablen. Resultat ist, dass der neue Störterm ε^* auch dann autokorriert sein kann, wenn dies für den wahren Störterm ε nicht gilt. Der neue Störterm wird insbesondere dann autokorriert sein, wenn die *vernachlässigte Variable selbst autokorriert* ist (bei Zeitreihen sehr wahrscheinlich). Betrachten wir dies mit Hilfe folgender Gleichungen näher:

$$\varepsilon_t^* = \varepsilon_t + \beta_2 X_{2t} \quad \varepsilon_{t-1}^* = \varepsilon_{t-1} + \beta_2 X_{2t-1}$$

Wie wir sehen, ist ε_t^* abhängig von der Ausprägung der Variable X_2 in t. Dies gilt analog auch in Periode t-1. Nehmen wir nun an, die Variable X_2 ist positiv autokorriert. Einem großen Wert von X_2 in t-1 folgt also tendenziell ein großer Wert von X_2 in t. Überwiegt nun im Durchschnitt dieser Einfluss auf ε^* den Einfluss von ε , wird ε^* größtenteils von X_2 bestimmt und ist damit ebenfalls positiv autokorriert.

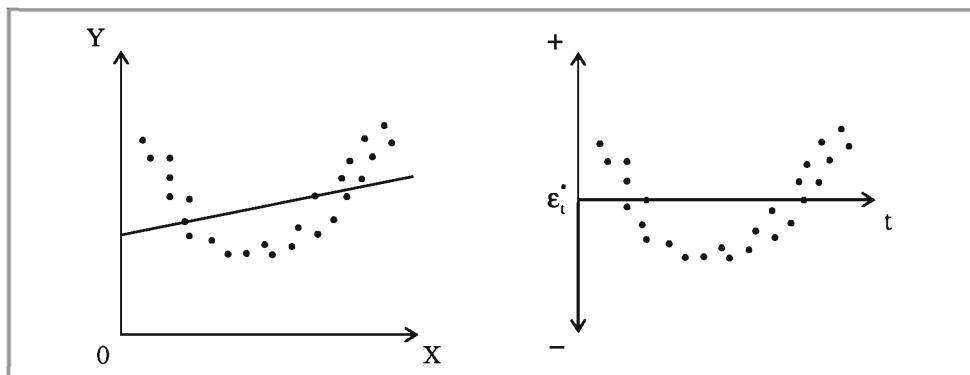


Abbildung IV 25: Autokorrelation des Störterms und funktionale Form

⁴² Für eine praktische Anwendung vgl. Abschnitt IV 6.2.3.

Auch eine *falsche funktionale Form* kann Auslöser für Autokorrelation sein. Nehmen wir beispielsweise an, das korrekte Modell hat die polynomiale Form

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \epsilon_t, \quad (\text{IV4: 61a})$$

wohingegen aber stattdessen ein lineares Modell

$$Y_t = \beta_0^* + \beta_1^* X_t + \epsilon_t^* \quad \text{mit} \quad \epsilon_t^* = \epsilon_t + \beta_2 X_t^2 \quad (\text{IV4: 61b})$$

zur Schätzung verwendet wurde. Der neue Störterm ϵ^* ist nun eine Funktion des wahren Störterms und X^2 . Auch hier liegt genau genommen lediglich ein Fall einer vernachlässigten Variablen vor. Wie Abbildung IV 25 skizziert, ergibt sich dadurch eine typische positive Autokorrelation des neuen Störterms.

4.5.2 Konsequenzen von Autokorrelation

Wir wir bereits unter IV 4.4.2 erwähnt haben, sind die Folgen von Autokorrelation mit denen von Heteroskedastizität vergleichbar. Wir wollen im Folgenden die drei wichtigsten näher betrachten:

1. Keine Verzerrung der OLS-Schätzer

Ist der stochastische Störterm autokorriert, so ist Annahme 4 des CLRM verletzt, doch führt dies zu keiner Verzerrung (und auch keiner Inkonsistenz) der OLS-Schätzer.⁴³ Dies gilt unabhängig davon, ob die Autokorrelation positiv, negativ, erster oder höherer Ordnung ist. Ein Blick auf den Erwartungswert des Schätzers $\hat{\beta}_1$ im einfachen Regressionsmodell, den wir in (IV2: 21) im Fall deterministischer X als

$$E(\hat{\beta}_1) = \beta_1 + \frac{\sum x_i E(\epsilon_i)}{\sum x_i^2} = \beta_1$$

hergeleitet hatten, kann uns hier wieder helfen, dies klar zu machen. Wir sehen deutlich, dass diese Beziehung bzw. die Erwartungstreue auch im Fall von Autokorrelation erhalten bleibt, da keinerlei Kovarianzen der stochastischen Störterme bei der Bestimmung des Erwartungswertes eine Rolle spielen.⁴⁴

2. OLS-Schätzer nicht länger effizient

Wird Annahme 4 des CLRM verletzt, so hat die Verteilung der OLS-Schätzer (unter allen linearen unverzerrten Schätzern) nicht mehr die minimale Varianz. OLS-Schätzer sind also im Fall von Autokorrelation nicht mehr BLUE, da es andere Schätzer (vgl. GLS-Schätzer unter IV 4.5.4.1) gibt, die eine geringere Varianz aufweisen. Die höhere Varianz der OLS-Schätzer lässt sich damit begründen, dass der autokorrierte Störterm Fluktuationen von Y verursacht, die von OLS den erklä-

⁴³ Im Falle von Autokorrelation, die durch eine Fehlspezifikation verursacht wird, kann jedoch eine Verzerrung auftreten. Die Autokorrelation selbst bewirkt in statischen Modellen keine Verzerrung. Eine Verzerrung durch Autokorrelation entsteht in dynamischen Modellen, in denen die verzögerte endogene Variable als erklärende Variable im Modell auftaucht (vgl. Abschnitt IV 4.7.1.2).

⁴⁴ Ist der Regressor stochastisch, so kann gezeigt werden, dass OLS erwartungstreu ist, falls X strikt exogen ist. Vgl. dazu z.B. Wooldridge, J.M. (2006), Kapitel 11. Für strikte Exogenität ist notwendig, dass die X aus *allen* Perioden (X_1, X_2, \dots, X_T) unkorreliert mit ϵ_t sind.

renden Variablen zugeordnet werden. Es ist also in der Präsenz von Autokorrelation wahrscheinlicher, dass die OLS-Schätzungen vom wahren Koeffizienten abweichen. Die Varianz der OLS-Schätzer steigt also. Die Schätzer bleiben aber unverzerrt, da Über- und Unterschätzungen sich die Waage halten.

3. Verzerrung der Schätzer der Standardfehler der OLS-Schätzer

Wie bei der durch Heteroskedastizität verursachten Verzerrung der geschätzten Standardfehler, können wir diese auch bei Autokorrelation anhand der bereits unter IV 4.4.2 gezeigten Herleitung veranschaulichen. Damit die OLS-Standardfehler korrekt sind, ist nämlich nicht nur die Annahme der Homoskedastizität, sondern auch das Fehlen von Autokorrelation ($\text{Cov}(\epsilon_i, \epsilon_j) = 0$) erforderlich (vgl. (IV4: 32a)). Nehmen wir nun an, dass Annahme 3 gültig, jedoch Annahme 4 des CLRM nicht erfüllt ist. Wir sind dann lediglich in der Lage

$$\text{Var}(\hat{\beta}_1) = E[(k_1 \epsilon_1)^2 + 2k_1 k_2 \epsilon_1 \epsilon_2 + (k_2 \epsilon_2)^2 + \dots + (k_n \epsilon_n)^2]$$

zu

$$\begin{aligned}\text{Var}(\hat{\beta}_1) &= \sum_{i \neq j} k_i^2 E(\epsilon_i^2) + \sum_{i \neq j} k_i k_j E(\epsilon_i \epsilon_j) \\ &= \sigma^2 \sum_{i \neq j} k_i^2 + \sigma^2 \sum_{i \neq j} k_i k_j \rho^s \quad \text{mit } s = |i - j|\end{aligned}$$

zu vereinfachen.⁴⁵ Resubstitution von k und weitere Umformungen liefern

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2} \left(1 + \frac{1}{\sum x_i^2} \sum_{i \neq j} \sum x_i x_j \rho^s \right). \quad (\text{IV4: 62})$$

Sind die Störterme nicht autokorriert ($\rho = 0$), so erhalten wir aus (IV4: 62) unsere klassische Formel für die Varianz von $\hat{\beta}_1$. Je höher ρ jedoch ist, desto stärker weicht (IV4: 62) davon ab. Positive Autokorrelation führt c.p. zu einer höheren Varianz der OLS-Schätzer. Negative Autokorrelation kann die Varianz c.p. sogar verringern.

Da OLS bei Schätzung des Standardfehlers des Steigungskoeffizientenschätzers (im einfachen Regressionsmodell) auf die falsche Formel (IV4: 33) zurückgreift, liefert OLS falsche Schätzungen der Standardfehler. Auch der Schätzer von σ^2 , also konkret (IV2: 25), ist verzerrt, da dieser unter den Annahmen 2, 3 und 4 hergeleitet wird und eine davon natürlich bei Autokorrelation verletzt ist.⁴⁶ Im Falle positiver Autokorrelation wird OLS die Standardfehler der OLS-Schätzer unterschätzen im Falle negativer Autokorrelation überschätzen, falls X selbst positiv autokorriert ist. Gewöhnlich weisen aber die Zeitreihen in den Wirtschaftswissenschaften positive Autokorrelation auf. Liegt also der wahrscheinlichere Fall positiver Autokorrelation der Störterme vor, so führt dies auch zur Überhöhung der t-Statistik, was wiederum die Wahrscheinlichkeit der Ablehnung einer korrekten Nullhypothese (Fehler 1. Art) erhöht.

⁴⁵ Es gilt $E(\epsilon_i \epsilon_j) = \text{Cov}(\epsilon_i, \epsilon_j) = \rho^s \sigma_i \sigma_j = \rho^s \sigma^2$.

⁴⁶ Vgl. Goldfeld, S.M., Quandt, R.E. (1972).

4.5.3 Aufdecken von Autokorrelation

4.5.3.1 Grafische Methode

Ein einfacher Weg, Anhaltspunkte für Autokorrelation zu gewinnen, ist die Untersuchung des Graphen der Residuen auf Muster wie in Abbildung IV 24 bzw. IV 12.

Beispiel IV 4-21

Naive Geldnachfragefunktion



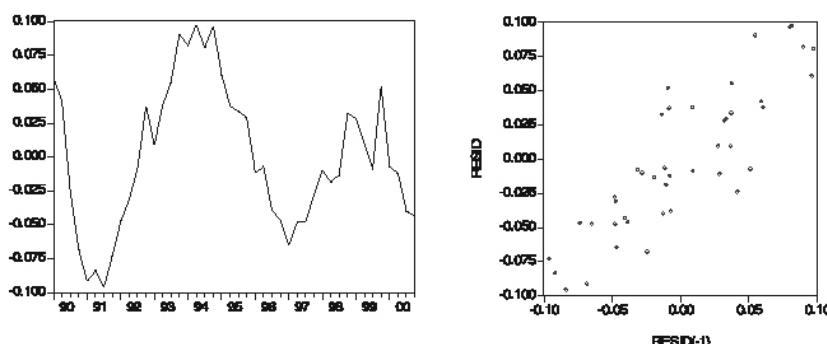
Greene, W.H. (2008) schätzt zur Verdeutlichung von Autokorrelation, die durch Fehlspezifikation verursacht wird, eine Geldnachfragefunktion für die USA in der Form

$$\ln Y_t = \beta_0 + \beta_1 \ln X_{1t} + \beta_2 \ln X_{2t} + \varepsilon_t.$$

Dabei ist Y die Geldmenge $M1$, X_1 das reale Bruttoinlandsprodukt und X_2 der Konsumtenpreisindex. Die Nachfrage nach Geld bzw. die Geldmenge $M1$ ist schließlich umso größer, je größer das reale BIP und das Preisniveau sind. Hier wird die Autokorrelation durch eine vernachlässigte Variable verursacht. Die moderne Geldnachfragetheorie kommt nämlich nicht ohne eine Variable aus, die die Opportunitätskosten der Geldhaltung beschreibt. Diese Opportunitätskosten werden meist durch eine Zinsvariable implementiert, was hier nicht getan wurde.

Eine Schätzung des Modells mit Datenmaterial vom 1. Quartal 1990 bis 4. Quartal 2000 ergab die folgenden Ergebnisse und auch eine grafische Darstellung der Residuen.

Dependent Variable: LOG(Y)		Sample: 1990Q1 2000Q4		
Method: Least Squares		Included observations: 44		
	Coefficient	Std. Error	t-Statistic	Prob.
C	3.406346	0.722046	4.717629	0.0000
LOG(X1)	-2.049544	0.330010	-6.210545	0.0000
LOG(X2)	3.573927	0.426378	8.382064	0.0000
R-squared	0.778173	Mean dependent var		6.945513
Adjusted R-squared	0.767352	S.D. dependent var		0.111688
S.E. of regression	0.053871	Akaike info criterion		-2.938692
Sum squared resid	0.118987	Schwarz criterion		-2.817043
Log likelihood	67.65123	F-statistic		71.91438
Durbin-Watson stat	0.269749	Prob(F-statistic)		0.000000



Bei Betrachtung der Schätzergebnisse fällt sofort das völlig implausible Vorzeichen der BIP-Variable X_1 auf. Das hier resultierende negative Vorzeichen würde bedeuten, dass mit zunehmendem BIP, d.h. steigendem Produktionsniveau und wachsender Wirtschaft, die Geldnachfrage sinkt, was völlig unsinnig wäre. Uns liegt damit ein deutliches Indiz dafür vor, dass eine wichtige Variable vernachlässigt wurde, die die verbleibenden OLS-Schätzer nun verzerrt.

Die Darstellung der Residuen im Zeitverlauf (links) lässt erkennen, dass auf negative (positive) Residuen in der Regel wieder negative (positive) folgen. Auch im Streudiagramm (rechts), in dem die um eine Periode verzögerten Residuen auf der X-Achse und kontemporäre Residuen auf der Y-Achse abgetragen sind, lässt deutlich positive Korrelation erkennen. Wir können also vermuten, dass der stochastische Störterm positiv auto-korreliert ist.

Im Zusammenhang mit dieser grafischen Methode bzw. um deren Aussagekraft zu verstärken, wird in der Praxis häufig zusätzlich eine Schätzung des Autokorrelationskoeffizienten als Beurteilungshilfe herangezogen. Interessieren wir uns also z.B. für eine Schätzung des Autokorrelationskoeffizienten erster Ordnung, so müssen wir lediglich das Pendant zu (IV4: 52), also $e_t = \rho e_{t-1} + u_t$, mittels OLS schätzen. Der Schätzer $\hat{\rho}$ hat damit die Gestalt

$$\hat{\rho} = \frac{\sum_{t=2}^n e_{t-1} e_t}{\sum_{t=2}^n e_{t-1}^2}. \quad (\text{IV4: 63})$$

Da für $t = 1$ kein Residuum e_{t-1} vorliegt, erfolgt die Schätzung nur mit Datenmaterial von $t = 2$ bis $t = n$. Alternativ können wir zur Gewinnung von $\hat{\rho}$ natürlich auch einfach die klassische Formel zur Berechnung des Korrelationskoeffizienten für die Variable e_t und die verzögerte Variable e_{t-1} anwenden. Weicht $\hat{\rho}$ nun deutlich von Null ab, so ist dies ein Indiz für Autokorrelation erster Ordnung. Da jedoch nicht klar ist, ab wann wir von einem signifikant von Null verschiedenen Koeffizienten sprechen können, sollten wir unsere Entscheidung anhand der folgenden Hypothesentestverfahren treffen. Eine grafische Analyse und ein erster Blick auf den Autokorrelationskoeffizienten sollten nur als erste Anhaltspunkte und nicht als alleinige Entscheidungsgrundlage dienen.

4.5.3.2 Durbin-Watson d-Test

Der **Durbin-Watson d-Test** ist eines der bekanntesten Verfahren zum Test auf Autokorrelation.⁴⁷ Er bedient sich der Residuen eines geschätzten Modells, um festzustellen, ob *Autokorrelation erster Ordnung* vorliegt. Autokorrelation höherer Ordnung vermag er nicht aufzudecken. Voraussetzungen für seine Anwendung sind, dass das Regressionsmodell eine Konstante β_0 beinhaltet, der stochastische Störterm normalverteilt ist und in der Modellgleichung keine verzögerte erklärte Variable als erklärende Variable auftaucht.⁴⁸

⁴⁷ Vgl. Durbin, J., Watson, G.S. (1951).

⁴⁸ Konkret dürfen nur deterministische oder strikt exogene Variablen als Regressoren vorkommen. Zum Begriff der strikten Exogenität vgl. Fußnote 44 dieses Abschnitts.

Gehen wir von einem Modell

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_K X_{Kt} + \varepsilon_t, \quad (\text{IV4: 64a})$$

mit

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad \text{mit} \quad |\rho| < 1 \quad (\text{IV4: 64b})$$

aus, wobei u_t ein nicht autokorrelierter Störterm ist. Um die Hypothese $H_0: \rho = 0$ (keine Autokorrelation erster Ordnung) zu testen, ist zunächst (IV4: 64a) zu schätzen, um die Residuen zu erhalten. Damit berechnen wir die Teststatistik

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad (\text{IV4: 65a})$$

Wir beachten dabei, dass der Zähler eine Beobachtung weniger aufweist als der Nenner, da ähnlich wie bei (IV4: 63) eine Beobachtung dazu benutzt werden muss, um e_{t-1} zu bestimmen. Die Durbin-Watson d-Statistik nimmt im Falle extremer positiver Autokorrelation ($\rho = 1$) genau den Wert Null, bei extremer negativer Autokorrelation ($\rho = -1$) einen Wert nahe Vier und bei fehlender Autokorrelation ($\rho = 0$) einen Wert nahe Zwei an. Dies lässt sich anhand einer einfachen Umformung von (IV4: 65a)

$$d = \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} = \frac{\sum_{t=2}^n e_t^2}{\sum_{t=1}^n e_t^2} + \frac{\sum_{t=2}^n e_{t-1}^2}{\sum_{t=1}^n e_t^2} - 2 \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \approx 1 + 1 - 2\hat{\rho} \quad (\text{IV4: 65b})$$

veranschaulichen, die zu

$$d \approx 2(1 - \hat{\rho}) \quad (\text{IV4: 65c})$$

führt.⁴⁹ Wie wir deutlich sehen, ist diese Beziehung zwischen d und $\hat{\rho}$ nur näherungsweise gültig. Aufgrund der unterschiedlichen Summationsgrenzen im Nenner und Zähler von (IV4: 65b) entstehen nämlich nicht exakt die beiden Einsen und zudem stimmt der letzte Bruch in (IV4: 65b) nicht genau mit (IV4: 63) überein. Dennoch können wir mit (IV4: 65c) unsere obigen Argumente untermauern. Geht $\hat{\rho}$ gegen 1, so tendiert d zu 0, geht $\hat{\rho}$ gegen -1 tendiert d zu 4 und für $\hat{\rho}$ gegen 0 nähert sich d dem Wert 2. In der Praxis ist meistens positive Autokorrelation ein Problem. Deswegen werden uns meist sehr kleine Werte der d-Statistik begegnen. Als *Daumenregel* können wir uns merken, dass wir beim Auftreten eines *d nahe 2* davon ausgehen können, dass *weder positive noch negative Autokorrelation erster Ordnung* vorliegt.

Zur exakten Testdurchführung sind unter Berücksichtigung des Stichprobenumfangs n und der Anzahl K der erklärenden Variablen der untere und obere kritische Wert d_u und d_o aus den Tabellen der d-Statistik abzulesen. Diese werden dazu verwendet, um die in Abbildung IV 26 dargestellte Schematik aufzustellen.

⁴⁹ Wir können den letzten Bruchterm in (IV4: 65b) als Schätzer von ρ verwenden, da $\bar{e} = 0$.

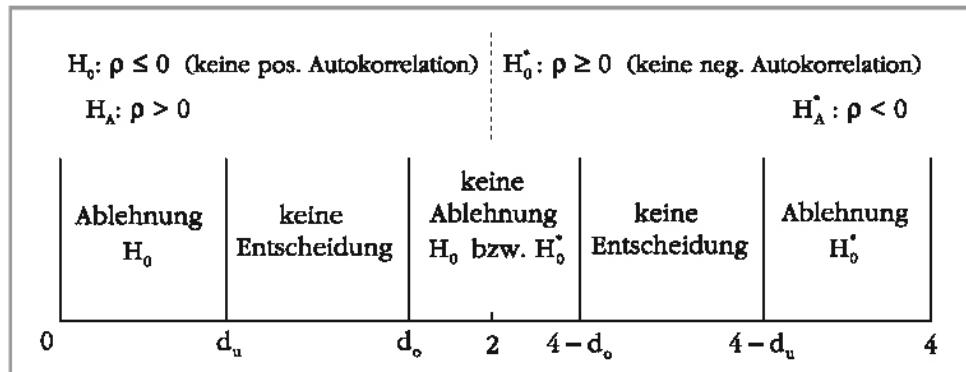


Abbildung IV 26: Durbin-Watson d-Test

Mit diesem Schema können wir einen Autokorrelationstest durchführen. Wollen wir beispielsweise **positive Autokorrelation** nachweisen, also

$$\begin{array}{ll} H_0: \rho \leq 0 & \text{(keine positive Autokorrelation)} \\ H_A: \rho > 0 & \text{(positive Autokorrelation)} \end{array} \quad (\text{IV4: 66a})$$

testen, bedienen wir uns der Entscheidungsregel

$$\begin{array}{ll} \text{Ablehnung } H_0, \text{ wenn} & 0 < d < d_u \\ \text{Nicht-Ablehnung } H_0, \text{ wenn} & d_o < d < 4 - d_o \\ \text{keine Entscheidung, wenn} & d_u \leq d \leq d_o \end{array} \quad (\text{IV4: 66b})$$

Die Entscheidungsregel für einen Test auf negative Autokorrelation kann analog aus Abbildung IV 26 abgelesen werden. Wie wir schnell erkennen, gibt es beim Durbin-Watson d-Test, anders als z.B. beim t- oder F-Test nicht nur einen Annahme- und einen Ablehnungsbereich, sondern auch einen Bereich, in dem der Test kein Ergebnis liefern kann. Gerade aus diesem Grund und den eingangs geschilderten Einschränkungen des Tests, weichen wir in der Praxis häufig auf den Breusch-Godfrey LM-Test (vgl. Abschnitt IV 4.5.3.3) aus.

Beispiel IV 4-22

Durbin-Watson-Test

Versuchen wir unsere Geldnachfragefunktion aus Beispiel IV 4-21 auf Autokorrelation zu testen. Vorteil der Arbeit mit EViews ist hierbei, dass der Wert der Durbin-Watson d-Statistik automatisch berechnet wird und als Teil des Standard-Regressionsoutputs ausgegeben wird. Wir können also in diesem Fall direkt den Wert 0,2697 ablesen. Bei einem so niedrigen d-Wert erübrigt sich eigentlich die Aufstellung der Entscheidungsregel. Wir können sofort feststellen, dass ein Autokorrelationsproblem vorliegt. Wir wollen den Test aber trotzdem durchführen: Wir vermuten positive Autokorrelation und verwenden daher das Hypothesenpaar (IV4: 66a) und die Entscheidungsregel (IV4: 66b). Bei einem Stichprobenumfang von $n = 44$, einer Anzahl von erklärenden Variablen $K = 2$ und dem von uns gewähltem Signifikanzniveau von 5 % erhalten wir durch Nachschlagen im Tabellenwerk $d_u = 1,43$ und $d_o = 1,62$. Da wir nur für $n = 40$ und $n = 45$ Tabellenwerte finden, haben wir als Approximation für $n = 44$ den Wert von $n = 45$ verwendet. Dies führt schließlich zu folgender konkreter Entscheidungsregel:

Ablehnung H_0 , wenn	$0 < d < 1,43$
Nicht-Ablehnung H_0 , wenn	$1,62 < d < 2,38$
keine Entscheidung, wenn	$1,43 \leq d \leq 1,62$

Mit einem d -Wert von 0,2697 können wir H_0 hier deutlich ablehnen, sodass dieses Testergebnis auf vorliegende positive Autokorrelation der Störterme schließen lässt.

4.5.3.3 Breusch-Godfrey LM-Test

Wie wir bereits angesprochen haben, weist der Durbin-Watson d -Test eine Reihe gewichtiger Nachteile auf. So führt er nicht immer zu einer Testentscheidung und kann nicht verwendet werden, wenn eine verzögerte erklärte Variable im Modell enthalten ist. Außerdem ist er nicht in der Lage auf Autokorrelation höherer Ordnung zu testen. Der **Breusch-Godfrey LM-Test** weist keines dieser Probleme auf.⁵⁰ Zur Veranschaulichung seiner Vorgehensweise betrachten wir das Modell

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_K X_{Kt} + \epsilon_t, \quad (\text{IV4: 67a})$$

in dem

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \dots + \rho_s \epsilon_{t-s} + u_t \quad (\text{IV4: 67b})$$

gilt. Wir testen, ob Autokorrelation vorliegt, über das Hypothesenpaar

$H_0 : \rho_1 = \rho_2 = \dots = \rho_s = 0$	(keine Autokorrelation)
$H_A : \text{mindestens ein } \rho \text{ ist ungleich Null}$	(Autokorrelation)

. (IV4: 68)

Um H_0 zu testen, wird zunächst Gleichung (IV4: 67a) mit OLS geschätzt. Die sich ergebenden e_t werden dann dazu verwendet, um eine Hilfsregression

$$e_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \dots + \alpha_K X_{Kt} + \alpha_{K+1} e_{t-1} + \dots + \alpha_{K+s} e_{t-s} + v_t \quad (\text{IV4: 69})$$

zu schätzen. v_t ist hier ein klassischer Störterm. Das R^2 der Schätzung von (IV4: 69) wird dazu verwendet, die LM-Statistik

$$LM = (n - s) \cdot R^2 \quad (\text{IV4: 70})$$

zu berechnen. Sie ist bei korrekter Nullhypothese *asymptotisch* mit s Freiheitsgraden Chi-Quadrat-verteilt. Ist der Wert der LM-Statistik größer als der kritische Wert bei gewähltem Signifikanzniveau, kann H_0 abgelehnt werden.

Welche Anzahl s von verzögerten Residuen in (IV4: 69) verwendet wird, richtet sich nach dem Grad der vermuteten Autokorrelation. Meistens wird sich aber an der Periodizität (Monate, Quartale, Jahre, etc.) des vorliegenden Datenmaterials orientiert. Bei Jahresdaten wählen wir häufig 1 oder 2, bei Quartalsdaten 4 oder 5.

⁵⁰ Vgl. Breusch, T.S. (1978) und Godfrey, L.G. (1978).

Beispiel IV 4-23**Breusch-Godfrey-LM-Test**

In unserem Geldnachfragebeispiel IV 4-21 wollen wir das verwendete Modell nun auch auf Autokorrelation höherer Ordnung untersuchen. Da uns Quartalsdaten vorliegen, verwenden wir $s = 4$. Der Test kann in EViews im Schätzfenster Ober View / Residual Tests / Serial Correlation LM Test unter Angabe von s durchgeführt werden. Wir erhalten damit die folgenden Ergebnisse:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	26.43188	Prob. F(4,37)	0.0000
Obs*R-squared	32.59365	Prob. Chi-Square(4)	0.0000

Test Equation:

Dependent Variable: RESID
 Method: Least Squares
 Presample missing value lagged residuals set to zero.

Sample: 1990Q1 2000Q4
 Included observations: 44

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.056236	0.387410	0.145159	0.8854
LOG(X1)	0.035286	0.178939	0.197199	0.8448
LOG(X2)	-0.060862	0.231945	-0.262397	0.7945
RESID(-1)	0.903795	0.163865	5.515498	0.0000
RESID(-2)	0.117065	0.220804	0.530176	0.5992
RESID(-3)	-0.160807	0.221518	-0.725931	0.4724
RESID(-4)	-0.068652	0.186647	-0.411958	0.6827
R-squared	0.740765	Mean dependent var		-4.52E-15
Adjusted R-squared	0.698727	S.D. dependent var		0.052604
S.E. of regression	0.028873	Akaike info criterion		-4.106894
Sum squared resid	0.030846	Schwarz criterion		-3.823045
Log likelihood	97.35166	F-statistic		17.62125
Durbin-Watson stat	1.909418	Prob(F-statistic)		0.000000

Wir erkennen, dass der Wert der LM-Statistik bei 32,5937 liegt. Bei einem Signifikanzniveau von 5 % erhalten wir aus der Tabelle der Chi-Quadrat-Verteilung mit 4 Freiheitsgraden den kritischen Wert 9,49. Dieser ist deutlich kleiner als der Wert der LM-Statistik, so dass die Nullhypothese (keine Autokorrelation) abgelehnt werden kann. Dies bestätigt auch der niedrige p-Wert des LM-Tests (0,0000). Es liegt also definitiv Autokorrelation vor.

Ein Blick auf die t-Statistiken der verzögerten Residuen zeigt aber, dass nur der erste geschätzte Koeffizient signifikant von Null verschieden ist. Dies deutet darauf hin, dass nur Autokorrelation erster Ordnung vorliegt. Dies bestätigt eine weitere Durchführung des Breusch-Godfrey LM-Tests für $s = 1$ (siehe Folgeseite). Der geschätzte Koeffizient des verzögerten Residiums ist nun sogar noch signifikanter (höherer t-Wert) als vorher.

Gerade im Zusammenhang mit diesem Beispiel stellt sich der Leser sicher die Frage, warum der Durbin-Watson d-Test überhaupt benötigt wird, wenn der Breusch-Godfrey LM-Test doch auch auf Autokorrelation erster Ordnung testen kann. Die Antwort darauf ist recht einfach. Der Durbin-Watson d-Test war der erste formale Test auf Autokorrelation. Breusch und Godfrey bauten auf den Arbeiten von Durbin und Watson auf. Zudem ist der d-Test Bestandteil eines jeden Regressionsoutputs und damit ein erster wichtiger Autokorrelationsindikator.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	104.9522	Prob. F(1, 40)	0.0000
Obs*R-squared	31.85807	Prob. Chi-Square(1)	0.0000

Test Equation:

Dependent Variable: RESID Sample: 1990Q1 2000Q4

Method: Least Squares Included observations: 44

Presample missing value lagged residuals set to zero.

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.052854	0.384046	0.137624	0.8912
LOG(X1)	0.074428	0.175662	0.423700	0.6741
LOG(X2)	-0.117558	0.227054	-0.517754	0.6075
RESID(-1)	0.859466	0.083894	10.24462	0.0000
R-squared	0.724047	Mean dependent var		-4.52E-15
Adjusted R-squared	0.703351	S.D. dependent var		0.052604
S.E. of regression	0.028651	Akaike info criterion		-4.180763
Sum squared resid	0.032835	Schwarz criterion		-4.018564
Log likelihood	95.97678	F-statistic		34.98408
Durbin-Watson stat	1.686806	Prob(F-statistic)		0.000000

4.5.4 Vorgehensweise bei festgestellter Autokorrelation

Nach Diagnose von Autokorrelation ist der wichtigste Schritt zunächst, die Modellspezifikation noch einmal zu überdenken. Es ist sicherzustellen, dass alle relevanten Variablen im Modell enthalten sind und die funktionale Form als korrekt betrachtet werden kann. Ist dies nämlich nicht der Fall, kann die vorliegende Autokorrelation nur durch Korrektur der Spezifikationsfehler beseitigt werden. Weist ein Modell danach immer noch autokorrelierte Störterme auf, so hat der Ökonometriker einen Teil des systematischen Verhaltens nicht modelliert, sondern dem Störterm zugeschlagen. Erstrebenswert ist es aber, ein möglichst umfassendes Modell aller systematischen Effekte zu erhalten und dabei den Störterm auf "White-Noise"-Eigenschaften (Erwartungswert Null und konstante Varianz) zu reduzieren. Wie wir unter IV 4.5.1 gesehen haben, ist dies theoretisch möglich. Praktisch stehen wir jedoch häufig vor dem Problem, dass wir die Dynamik des Modells nicht vollständig kennen. Wir sind daher meist gezwungen zur Beseitigung der Autokorrelation der Störterme bzw. ihrer nachteiligen Effekte die Verfahren dieses Abschnitts heranzuziehen. Sie sind jedoch nur zulässig, wenn folgende **Voraussetzungen** erfüllt sind:

1. Das Modell weist keine Mängel wie vernachlässigte Variablen oder falsche funktionale Form auf.
2. Der Stichprobenumfang ist sehr groß.
3. Es darf keine verzögerte erklärte Variable als erklärende Variable im Modell enthalten sein,⁵¹ da sonst durch die folgenden Verfahren sogar inkonsistente Schätzer resultieren würden.

⁵¹ Genauer müssten wir sagen, dass im Modell nur strikt exogene Regressoren enthalten sein dürfen. Vgl. dazu Fußnote 44 dieses Abschnitts.

4.5.4.1 Verallgemeinertes Prinzip der kleinsten Quadrate (GLS)

Das *verallgemeinerte Prinzip der kleinsten Quadrate* (GLS, engl. generalised least squares) verfolgt das Ziel, eine Modellgleichung von Autokorrelation zu befreien. GLS transformiert dazu die Gleichung, die aufgrund von Autokorrelation die Annahme 4 des CLRM verletzt, in eine Gleichung, die diese Verletzung nicht hervorruft. Zur Veranschaulichung dieser Transformation und der Schätzung der transformierten Gleichung (sog. GLS-Gleichung) beschränken wir uns im Folgenden auf Autokorrelation erster Ordnung. Betrachten wir dazu das Modell

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_K X_{Kt} + \varepsilon_t, \quad (\text{IV4: 71a})$$

in dem Autokorrelation erster Ordnung vorliegt, sodass

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad \text{mit } |\rho| < 1 \quad (\text{IV4: 71b})$$

gilt, wobei u_t wieder ein klassischer (nicht autokorrelierter) Störterm ist. Ist nun das Modell (IV4: 71a) in Periode t gültig, so besitzt es diese Gültigkeit auch in $t-1$, sodass wir alternativ auch

$$Y_{t-1} = \beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t-1} + \dots + \beta_K X_{Kt-1} + \varepsilon_{t-1} \quad (\text{IV4: 72})$$

schreiben können. Die Multiplikation beider Seiten von (IV4: 72) mit ρ liefert

$$\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{1t-1} + \rho \beta_2 X_{2t-1} + \dots + \rho \beta_K X_{Kt-1} + \rho \varepsilon_{t-1}. \quad (\text{IV4: 73})$$

Subtrahieren wir nun (IV4: 73) von (IV4: 71a), so erhalten wir

$$\begin{aligned} Y_t - \rho Y_{t-1} &= \beta_0(1-\rho) + \beta_1(X_{1t} - \rho X_{1t-1}) + \beta_2(X_{2t} - \rho X_{2t-1}) + \dots \\ &\quad + \beta_K(X_{Kt} - \rho X_{Kt-1}) + (\varepsilon_t - \rho \varepsilon_{t-1}) \end{aligned} \quad (\text{IV4: 74})$$

oder

$$\begin{aligned} Y_t^* &= Y_t - \rho Y_{t-1} \\ Y_t^* &= \beta_0^* + \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \dots + \beta_K X_{Kt}^* + u_t \quad \text{mit} \\ \beta_0^* &= \beta_0(1-\rho) \\ u_t &= \varepsilon_t - \rho \varepsilon_{t-1} \end{aligned} \quad (\text{IV4: 75})$$

Wir sprechen bei Y_t^* und X_{kt}^* auch von *quasi differenzierten Daten*. Wäre $\rho = 1$, würden wir differenzierte Daten erhalten. Wir nehmen aber $|\rho| < 1$ an. Gleichung (IV4: 75) ist nur für die Perioden $t = 2, 3, \dots, T$ definiert, da im Rahmen der Transformation eine Beobachtung verloren geht. Besonders bei hohem Stichprobenumfang wird in der Praxis ein derartiger Informationsverlust meist toleriert. In kleinen Stichproben sollte eine Berücksichtigung von $t = 1$ über die spezielle Transformation $Y_1^* = \sqrt{1-\rho^2}Y_1$ bzw. $X_{k1}^* = \sqrt{1-\rho^2}X_{k1}$ erfolgen.⁵²

Der Störterm u_t der quasi differenzierten Gleichung (IV4: 75) entstammt aus (IV4: 71b), ist daher nicht autokorreliert und erfüllt alle Gauss-Markov-Annahmen. Die GLS-Gleichung (IV4: 75) kann daher mit OLS geschätzt werden, wobei wir die Schätzer dann als *GLS-Schätzer* bezeichnen. Sie weisen im Vergleich zu den Schätzern des Ursprungsmodells eine geringere Varianz auf, sind also *BLUE*, sofern auch die Beobachtung in $t = 1$ entsprechend berücksichtigt wird und ρ bekannt ist.

⁵² Zur Rechtfertigung dieser Vorgehensweise vgl. z.B. Hill, R.C., Griffiths, W.E., Judge, G.G. (2001), Kapitel 12.

Die Koeffizienten β_k der GLS-Gleichung unterscheiden sich in ihrer Interpretation nicht von denen der originären (autokorrelierten) Gleichung. Da es durch die GLS-Transformation zu einer Veränderung der erklärten Variable kommt, ist jedoch ein GLS-R² nicht als Bestimmtheitsmaß von Y interpretierbar.

Mit der Schätzung von GLS-Gleichungen ist nun das Problem verbunden, dass wir dazu nicht direkt OLS verwenden können, da ρ unbekannt ist. Neben den üblichen Regressionskoeffizienten ist also zudem ρ zu schätzen. Die wohl bekannteste Methode zur Schätzung von GLS-Gleichungen ist die mehrstufige **Cochrane-Orcutt Methode**.⁵³ In einem ersten Schritt wird dabei das autokorrelierte Modell (IV4: 71a) mit OLS geschätzt und die Residuen e_t berechnet. Im zweiten Schritt werden diese zur Schätzung des unbekannten ρ herangezogen. Dies kann z.B. durch OLS-Schätzung des Modells

$$e_t = \rho e_{t-1} + u_t \quad (\text{IV4: 76})$$

erfolgen. Im dritten Schritt wird die Schätzung $\hat{\rho}$ dann zur Durchführung der GLS-Transformation in Gleichung (IV4: 74) verwendet und die so erhaltene neue Gleichung mit OLS geschätzt. Als Ergebnis erhalten wir Werte für $\hat{\beta}_0$ und die $\hat{\beta}_k$. In der Praxis wird dann häufig ein iteratives Verfahren angewendet. Die Parameterschätzungen werden dabei in die originäre Gleichung (IV4: 71a) eingesetzt und die "neuen" Residuen

$$e_t^{\text{neu}} = Y_t - \hat{\beta}_0 + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_k X_{kt} \quad (\text{IV4: 77})$$

berechnet. Verwenden wir diese neuen Residuen erneut in (IV4: 76), so erhalten wir eine neue Schätzung von ρ . Diese Prozedur können wir beliebig oft wiederholen. Standard ist es aber, dass der Iterationsprozess gestoppt wird, wenn sich die neue Schätzung von ρ um weniger als 0,01 oder 0,005 von der vorhergehenden unterscheidet. Die letzte Schätzung von ρ und die daraus resultierenden Ergebnisse der GLS-Schätzung können dann als finale Ergebnisse betrachtet werden. Es kann aber theoretisch nicht gesagt werden, dass diese Iterationsvorgehensweise zu effizienteren Schätzungen führt als nur ein Durchlauf des Verfahrens.

Wir sprechen bei derartigen Iterationsverfahren, in denen nicht ρ sondern $\hat{\rho}$ zur Datentransformation verwendet wird, auch von **FGLS** (engl. feasible generalized least squares). Wir können bei Verwendung von FGLS insbesondere in *kleinen Stichproben* nicht generell sagen, dass wir die Modellparameter effizient schätzen, da dies nur bei bekanntem ρ der Fall ist oder wenn $\hat{\rho}$ sehr nahe beim wahren ρ liegt. Wir können aber sagen, dass wir konsistent schätzen, da wir uns mit zunehmendem Stichprobenumfang dem wahren Wert ρ immer mehr annähern. Auch unsere Testverfahren besitzen im FGLS-Kontext streng genommen nur asymptotisch Gültigkeit. Wir sollten daher FGLS nur bei *großen Stichproben* einsetzen. In großen Stichproben führt FGLS dann nämlich zu effizienteren Schätzern als OLS, falls die Modellannahmen erfüllt sind.⁵⁴

⁵³ Vgl. Cochrane, D., Orcutt, G.H. (1949).

⁵⁴ Man beachte außerdem: (1) Bei nicht strikt exogenen Regressoren ist OLS konsistent, so lange $\text{Cov}(e_t X) = 0$ gilt. FGLS ist in einem solchen Fall jedoch inkonsistent. (2) Ist X nicht stationär, so ist die OLS-Schätzung problematisch (vgl. Abschnitt IV 4.7.2.4). GLS ist jedoch als Quasi-Differenzierung etwas besser, da durch Differenzieren einer Nicht-Stationarität begegnet werden kann.

Die Cochrane-Orcutt Methode ist in EViews nicht direkt implementiert. Stattdessen erfolgt die GLS-Schätzung durch das Hinzufügen eines AR(1)-Terms in die Schätzgleichung, welcher darauf hinweist, dass von Autokorrelation erster Ordnung im Störterm ausgegangen wird.⁵⁵ EViews schätzt die Parameter β_0 , β_1 , ..., β_k und ρ dann simultan über gewisse iterative nichtlineare Regressionstechniken, auf die wir hier nicht näher eingehen wollen. Damit werden dieselben Schätzungen der Koeffizienten wie bei der Cochrane-Orcutt Methode erzielt. Wir wollen die Einfachheit der Anwendung dieser Methodik im nachfolgenden Beispiel veranschaulichen.

Beispiel IV 4-24

GLS-Schätzung in EViews



Wir haben bereits festgestellt, dass im Geldnachfragebeispiel IV 4-21 Autokorrelation erster Ordnung vorliegt. GLS kommt hier aber nicht als Gegenmaßnahme in Frage, da die Autokorrelation durch eine offensichtliche Fehlspezifikation verursacht wird. Stattdessen muss die vorliegende Autokorrelation durch Korrektur des Modells (z.B. Hinzufügen der offensichtlich fehlenden Zinsvariablen) beseitigt werden.

In unserem Anwendungsbeispiel IV 3-1 zum Geflügelkonsum in den USA lässt sich zeigen, dass auch hier Autokorrelation vorliegt. Die entsprechenden Tests möge der Leser selbst nachvollziehen. Da wir hier davon ausgehen wollen, dass keine Variablen vernachlässigt wurden und auch die funktionale Form des Modells korrekt gewählt wurde, ist eine GLS-Schätzung gerechtfertigt. Um in EViews die Resultate einer FGLS-Schätzung zu erzielen, müssen wir bei der Modellformulierung unter Quick / Estimate Equation / Equation Specification lediglich den Ausdruck AR(1) an die Modellgleichung anhängen. Im Falle von Autokorrelation s-ter Ordnung müssten zu deren Korrektur einfach die Ausdrücke AR(1), AR(2), ..., AR(s) in die zu schätzende Gleichung aufgenommen werden. In diesem Anwendungsbeispiel erhalten wir durch Eingabe der konkreten Gleichung Y C X1 X2 X3 AR(1) folgendes Schätzergebnis:

Dependent Variable: Y	Sample (adjusted): 1961 1999			
Method: Least Squares	Included observations: 39 after adjustments			
	Convergence achieved after 12 Iterations			
C	23.51656	2.478523	9.488133	0.0000
X1	-0.092925	0.103985	-0.893642	0.3778
X2	0.089972	0.045913	1.959608	0.0583
X3	0.243144	0.015859	15.33207	0.0000
AR(1)	0.801227	0.093037	8.611962	0.0000
R-squared	0.995894	Mean dependent var	51.26077	
Adjusted R-squared	0.995411	S.D. dependent var	19.28900	
S.E. of regression	1.306643	Akaike info criterion	3.492009	
Sum squared resid	58.04875	Schwarz criterion	3.705286	
Log likelihood	-63.09417	F-statistic	2061.780	
Durbin-Watson stat	2.056647	Prob(F-statistic)	0.000000	
Inverted AR Roots	.80			

⁵⁵ Vgl. z.B. Asteriou, D., Hall, S.G. (2007), Kapitel 8.

Wir erkennen, dass 12 Iterationsläufe notwendig waren, bis der letzte geschätzte Wert von $\hat{\rho}$ gefunden war. Im Regressionsoutput taucht $\hat{\rho}$ als Koeffizient der Größe AR(1) auf. Es gilt also hier $\hat{\rho} = 0,8012$. Wir sehen außerdem, dass eine Beobachtung aufgrund der GLS-Transformation verlorengegangen ist, da im Vergleich zu den ursprünglichen 40 nur noch 39 Beobachtungen in die Schätzung eingehen.

Wichtiger ist nun aber ein Vergleich der neuen geschätzten Koeffizienten mit denen der ursprünglichen OLS-Schätzung ohne Berücksichtigung der Autokorrelation. Diese Schätzung lieferte nämlich folgendes Ergebnis:

$$\begin{aligned}\hat{Y}_t &= 27,4939 - 0,6072X_{1t} + 0,0922X_{2t} + 0,2449X_{3t} & n &= 40 \quad (36) \\ (0,1571) && (0,0399) & R^2 = 0,9896 \\ t = -3,86 && 2,31 & d = 0,8978 \\ && 22,06\end{aligned}$$

Es fällt auf, dass trotz der Tatsache, dass Autokorrelation keine Verzerrung der OLS-Schätzer verursacht, die mit OLS geschätzten Koeffizienten nicht mit denen der GLS-Schätzung identisch sind. Bei X_1 führt die Beseitigung der Autokorrelation zu einer deutlichen Änderung des geschätzten Koeffizienten. Außerdem ist er nun nicht mehr signifikant. Auch der geschätzte Koeffizient von X_2 ist nur noch schwach signifikant. Die bestehende Autokorrelation hat also dazu geführt, dass die t-Werte bisher überschätzt wurden und unsere bisherigen Testentscheidungen falsch gewesen sein könnten. Dies verdeutlicht die besondere Wichtigkeit der Überprüfung der Erfüllung aller Annahmen des CLRM (und ggf. Korrekturmaßnahmen) bevor ein geschätztes Modell für Rückschlüsse auf die Grundgesamtheit verwendet wird. Kritisch ist dabei jedoch, dass der hier vorliegende Stichprobenumfang von 40 Beobachtungen relativ gering ist, die GLS-Prozedur aber nur für große Stichprobenumfänge angewendet werden sollte. Die neuen Parameterschätzungen und Schätzungen der Standardfehler sind daher auch mit Vorsicht zu genießen.

4.5.4.2 Newey-West Standardfehler

Newey-West Standardfehler sind ähnlich, wie die unter IV 4.4.4.2 behandelten White Standardfehler, auch *konsistente Schätzer für die Standardfehler der OLS-Schätzer*.⁵⁶ Da OLS zur Schätzung der Standardfehler des Steigungskoeffizienten im einfachen Regressionsmodell auf die Varianz (IV2: 23) zugreift, ist es offensichtlich, dass OLS verzerrte Schätzungen der Standardfehler liefert, falls Autokorrelation oder Heteroskedastizität vorliegt. Im Falle positiver Autokorrelation sind die OLS-Schätzer häufig nach unten, im Falle negativer nach oben verzerrt. Newey-West Standardfehler tragen nun dem Fakt der *Autokorrelation beliebiger Ordnung* Rechnung. Sie sind bei *großen Stichproben* und vorliegender Autokorrelation genauer als die OLS-Standardfehler, sind gewöhnlich größer (bei positiver Autokorrelation) als diese und führen damit zu niedrigeren t-Werten. Im Vergleich zu GLS hat eine Verwendung von Newey-West Standardfehlern keine Veränderung der geschätzten Koeffizienten des betrachteten Modells zur Folge, da die Regressionsparameter weiterhin mit OLS und lediglich die Standardfehler anders geschätzt werden.

Mit der genauen Berechnung der Newey-West Standardfehler wollen wir uns an dieser Stelle nicht beschäftigen. Es sei aber noch erwähnt, dass diese nicht nur der Autokorrelation sondern *auch der Heteroskedastizität* Rechnung tragen. Sie stellen damit eine Alternative zu den behandelten White Standardfehlern dar.

⁵⁶ Vgl. Newey, W.K., West, K.D. (1987).

Beispiel IV 4-25**Newey-West Standardfehler**

In Beispiel IV 4-24 sind wir mittels GLS der Autokorrelation in unserem Geflügelkonsummodell aus Beispiel IV 3-1 begegnet. Betrachten wir nun die Ergebnisse, die wir mit Newey-West Standardfehlern erzielen können. In EViews ist die Berechnung der Newey-West Standardfehler wie auch die der White Standardfehler eine Option bei der Eingabe der Regressionsgleichung. Eine Autokorrelationskorrektur mittels dieses Verfahrens ist daher relativ einfach. Wir erhalten folgendes Resultat:

Dependent Variable: Y		Sample: 1960 1999		
Method: Least Squares		Included observations: 40		
Newey-West HAC Standard Errors & Covariance (lag truncation=3)				
	Coefficient	Std. Error	t-Statistic	Prob.
C	27.59394	1.896189	14.55231	0.0000
X1	-0.607160	0.166734	-3.641477	0.0008
X2	0.092188	0.041867	2.201917	0.0342
X3	0.244860	0.009930	24.65786	0.0000
R-squared	0.990391	Mean dependent var		50.56725
Adjusted R-squared	0.989590	S.D. dependent var		19.53879
S.E. of regression	1.993549	Akaike info criterion		4.312350
Sum squared resid	143.0726	Schwarz criterion		4.481238
Log likelihood	-82.24700	F-statistic		1236.776
Durbin-Watson stat	0.897776	Prob(F-statistic)		0.000000

Wir sehen, dass wir keine Beobachtung unserer Stichprobe verlieren und sich im Vergleich zum Ursprungsmodell die geschätzten Koeffizienten nicht verändert haben. Lediglich ihre Standardfehler und damit ihre t- und p-Werte sind anders. Die Standardfehler der Schätzer $\hat{\beta}_1$ und $\hat{\beta}_2$ sind höher. Bei $\hat{\beta}_3$ ist dies aber nicht der Fall. Hier ist sogar ein niedrigerer Standardfehler zu beobachten. Im Vergleich zur GLS-Schätzung fällt auf, dass nun wieder alle Koeffizienten signifikant sind. Auch hier ist jedoch kritisch anzumerken, dass auch für die Newey-West Standardfehler der Stichprobenumfang relativ klein ist. Auch die Resultate dieses Verfahrens sind daher kritisch zu betrachten.

4.6 Korrelation zwischen erklärenden Variablen und stochastischem Störterm

In diesem Abschnitt wollen wir uns näher mit der Annahme 2b des CLRM befassen. Sie wird verletzt, wenn eine Korrelation zwischen erklärenden Variablen und dem stochastischen Störterm vorliegt. Dies hat, wie wir unter IV 4.6.1 sehen werden, bei stochastischem X die Verzerrung und Inkonsistenz der OLS-Schätzer zur Folge. Wie wir bereits unter IV 4.1.1 gesehen haben, kann eine Ursache für die Verletzung von Annahme 2b in *vernachlässigten Variablen* liegen. Als weitere mögliche Ursachen sind *Messfehler* bei der Erfassung von Variablen, die Verwendung *verzögterer erklärender Variablen* als erklärende Variablen und *Simultanität* zu betrachten. Wir wollen diese Ursachen im weiteren Verlauf dieses Abschnitts noch näher beleuchten. Besonderes Augenmerk legen wir auch auf die sog. *Instrumentenvariabenschätzung*, die einer Verletzung der Annahme 2b begegnen kann.

4.6.1 Konsequenzen

Zur Veranschaulichung der Probleme durch eine Verletzung von Annahme 2b des CLRM, betrachten wir ein einfaches Regressionsmodell, in dem nach (IV2: 20)

$$\hat{\beta}_1 = \beta_1 + \frac{\sum x_i \epsilon_i}{\sum x_i^2}$$

gilt. Der Schätzer $\hat{\beta}_1$ ist unverzerrt, wenn $E(\hat{\beta}_1) = \beta_1$ gilt, also der Erwartungswert des Bruchterms auf der rechten Seite von (IV2: 20) gleich Null ist. Wie wir unter IV 2.3.1 gesehen haben, ist dies bei *deterministischem* X und $E(\epsilon_i) = 0$ der Fall. Bei deterministischem X haben wir bei unseren Ausführungen zu Annahme 2b unter IV 2.2.2 außerdem festgestellt, dass Annahme 2b automatisch erfüllt ist. Nun wollen wir aber von dem realistischeren Fall ausgehen, dass X selbst *stochastisch* ist:

Bei genauerer Betrachtung obiger Gleichung fällt auf, dass der Bruchterm auf der rechten Seite nichts anderes als der Quotient aus Stichprobenkovarianz zwischen X und ϵ und Stichprobenvarianz von X ist. Wir können also alternativ auch

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n-1} \sum x_i \epsilon_i}{\frac{1}{n-1} \sum x_i^2} = \beta_1 + \frac{\widehat{\text{Cov}}(X, \epsilon)}{\widehat{\text{Var}}(X)} \quad (\text{IV4: 78})$$

schreiben. Mit zunehmendem Stichprobenumfang konvergieren diese Schätzer gegen ihre wahren Grundgesamtheitswerte. Ist Annahme 2b erfüllt, konvergiert der Schätzer der Kovarianz also gegen Null, sodass auch der Schätzer $\hat{\beta}_1$ gegen seinen wahren Wert β_1 konvergiert. Die Schätzung liegt also immer näher am wahren Wert der Grundgesamtheit.

Besteht nun aber eine Korrelation zwischen erklärender Variable und stochastischem Störterm, so ist aufgrund von (IV2: 16) auch Annahme 2a verletzt. Der OLS-Schätzer ist dann *verzerrt*, da bei Erwartungswertbildung von (IV4: 78) der Erwartungswert des Bruchterms nicht Null wird. Auch eine *Konsistenz* des OLS-Schätzers ist *nicht gegeben*, da die Stichprobenkovarianz zwischen X und ϵ nicht gegen Null strebt.

4.6.2 Ursachen

4.6.2.1 Vernachlässigte Variablen

Im Abschnitt IV 4.1.1 haben wir gesehen, dass die Vernachlässigung relevanter Variablen im Fall deterministischer X zu einer Verzerrung der OLS-Schätzer führt, falls die vernachlässigten Variablen mit den im Modell enthaltenen erklärenden Variablen korreliert sind. Die Verletzung von Annahme 2b führt deshalb auch im Falle stochastischer X zur *Verzerrung*. Selbst die Konsistenz ist nicht mehr gegeben, d.h. auch bei noch so großen Stichproben kommt es zur Verzerrung. Warum eine vernachlässigte Variable eine Verletzung von Annahme 2b zur Folge hat, wollen wir im Folgenden noch einmal kurz begründen: Vernachlässigen wir in einem korrekt spezifizierten Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

die Variable X_2 , so führt dies zu

$$Y_i = \beta_0^* + \beta_1^* X_{1i} + \varepsilon_i^* \quad \text{mit} \quad \varepsilon_i^* = \varepsilon_i + \beta_2 X_{2i},$$

da der stochastische Störterm die Effekte jeglicher vernachlässigter Variablen auf Y beinhaltet. Da nun in nahezu jedem Regressionsmodell eine Korrelation (wenn auch nur eine geringe) zwischen den erklärenden Variablen besteht, kommt es im neuen Modell bei Veränderung von X_2 sowohl zu einer Veränderung von X_1 als auch von ε^* . Wir können also sagen, dass X_1 und ε^* korreliert sind und Annahme 2b verletzt wird. Besteünde keine Korrelation zwischen X_1 und X_2 , käme es natürlich auch zu keiner Annahmenverletzung und keiner Verzerrung.

4.6.2.2 Messfehler

Im Folgenden wollen wir die Folgen von Messfehlern bei der erklärten und bei den erklärenden Variablen aufzeigen. Wir bedienen uns dabei zur Veranschaulichung jeweils des einfachen Regressionsmodells.

1. Messfehler bei der erklärten Variablen

Betrachten wir die Regressionsfunktion einer Grundgesamtheit der Gestalt

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i. \quad (\text{IV4: 79})$$

Bei der Messung der Variablen Y entstehen nun Messfehler, sodass an Stelle der wahren Werte Y_i die Werte

$$Y_i^* = Y_i + u_i \quad \rightarrow \quad Y_i = Y_i^* - u_i \quad (\text{IV4: 80})$$

erfasst werden. u_i stellt dabei den auftretenden Messfehler dar. Setzen wir (IV4: 80) in (IV4: 79) ein und nehmen noch eine kleine Umformung vor, so erhalten wir

$$Y_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i^* \quad \text{mit} \quad \varepsilon_i^* = \varepsilon_i + u_i. \quad (\text{IV4: 81})$$

Wir erkennen deutlich, dass solange X_i und u_i nicht korreliert sind, also formal $\text{Cov}(u_i, X_i) = 0$ vorliegt, die Gleichung ohne Probleme mit OLS geschätzt werden kann. Die Gültigkeit von $\text{Cov}(u_i, X_i) = 0$ ist unmittelbar einleuchtend. Warum sollte auch der Messfehler bei Y mit der Variable X korreliert sein? Die OLS-Schätzer bleiben *unverzerrt* und *konsistent*. Messfehler bei der erklärten Variablen erhöhen jedoch die Varianz der Störterme. Dem wird allerdings bei der Schätzung des Standardfehlers der Regression Rechnung getragen, indem die Standardabweichung der

$\hat{\epsilon}_i$ und nicht der ϵ_i geschätzt wird. Damit kommt es zu keiner Beeinträchtigung der t-Werte und Hypothesentests. Kommt es zu systematisch positiven oder negativen Messfehlern, so gilt $E(\hat{\epsilon}_i) \neq 0$. Wie wir aber unter IV 4.2.1 gesehen haben, wird ein solcher von Null verschiedener Erwartungswert von der Modellkonstante absorbiert und verursacht somit kaum Probleme. Einzig und allein der "fit" (R^2) wird aufgrund der höheren Varianz des Störterms schlechter.

2. Messfehler bei den erklärenden Variablen

Während Messfehler bei der erklärten Variablen weniger problematisch sind, sind sie bei erklärenden Variablen kritisch. Nehmen wir an, für das wahre Modell

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (\text{IV4: 82})$$

können aufgrund von Messfehlern nur die Werte

$$\tilde{X}_i = X_i + v_i \quad \rightarrow \quad X_i = \tilde{X}_i - v_i \quad (\text{IV4: 83})$$

erfasst werden. Das Modell hat nach Einsetzen von (IV4: 83) in (IV4: 82) die Form

$$Y_i = \beta_0 + \beta_1 \tilde{X}_i + \hat{\epsilon}_i \quad \text{mit} \quad \hat{\epsilon}_i = \epsilon_i - \beta_1 v_i. \quad (\text{IV4: 84})$$

Selbst wenn wir nun annehmen, dass der Messfehler bezüglich der Variable X einen Erwartungswert von Null aufweist, nicht autokorreliert und vom Störterm der korrekten Gleichung unabhängig ist, entstehen Probleme bei der Anwendung von OLS als Schätzmethode. Am einfachsten erkennen wir dies daran, dass der Störterm $\hat{\epsilon}_i$ und die Variable X_i in (IV4: 83) korreliert sind, denn $\epsilon_i - \beta_1 v_i$ und $X_i + v_i$ sind korreliert, da v_i sowohl in $\hat{\epsilon}_i$ als auch in X_i enthalten ist. Sie sind hier negativ korreliert (unterschiedliche Vorzeichen von v). Deswegen kommt es bei Messfehlern im Allgemeinen dazu, dass der Parameter β_1 nach unten verzerrt geschätzt wird.

Insgesamt können wir also sagen, dass Messfehler bei den erklärenden Variablen zu *verzerrten* und *inkonsistenten* OLS-Schätzern führen. Der Grad der Verzerrung und Inkonsistenz hängt dabei von der Varianz des Messfehlers ab.

4.6.2.3 Verzögerte endogene Variable

Verzögerte endogene Variablen als erklärende Variablen sind zwar erst Thema des Abschnitts IV 4.7.1.1, doch aufgrund der Tatsache, dass sie eine Verletzung der Annahme 2b verursachen können, wollen wir bereits hier kurz auf sie eingehen. Betrachten wir dazu ein einfaches Modell mit autokorreliertem Störterm

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \epsilon_t \quad \text{mit} \quad \epsilon_t = \rho \epsilon_{t-1} + u_t.$$

Dieses Modell hat auch in Periode $t-1$ Gültigkeit, sodass wir auch

$$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + \beta_2 Y_{t-2} + \epsilon_{t-1}$$

schreiben können. Untersuchen wir nun, was passiert, wenn sich ϵ_{t-1} bei positiver Autokorrelation erhöht. Dies führt zunächst zu einem steigenden Wert Y_{t-1} und aufgrund der Autokorrelation auch zu einem höheren ϵ_t . In unserem originären Modell sind also Y_{t-1} und ϵ_t korreliert, da sich beide in Reaktion auf ϵ_{t-1} verändern. Dies stellt eine klare Verletzung von Annahme 2b dar.⁵⁷

⁵⁷ Anders ausgedrückt, könnten wir ϵ_{t-1} auch als vernachlässigte Variable betrachten, die mit Y_{t-1} korreliert ist. Wie wir wissen, hat auch eine solche Vernachlässigung eine Verzerrung zur Folge, da auch sie zur Verletzung von Annahme 2b führt.

Während also Autokorrelation in "normalen" Modellen zu keiner Verzerrung der OLS-Schätzer führt, bewirkt Autokorrelation in Modellen mit verzögerten endogenen Variablen als erklärenden Variablen eine Verzerrung. Deswegen führen auch die verschiedenen FGLS-Verfahren (engl. feasible generalized least squares, vgl. Abschnitt 4.5.4.1) in diesem Fall zu inkonsistenten Schätzungen.

4.6.2.4 Simultanität

Bisher sind wir davon ausgegangen, dass in unseren Modellen die erklärte Variable keinen Einfluss auf die erklärenden Variablen besitzt. Wir waren daher in der Lage unsere Modelle mit nur einer Gleichung abzubilden. Hat nun aber die erklärte Variable einen eindeutigen Effekt auf zumindest eine erklärende Variable, so reicht eine Gleichung nicht länger zur Modellbildung aus. Wir stellen daher sog. **simultane Systeme**⁵⁸ auf, in denen wir häufig nicht länger zwischen erklärter und erklärenden Variablen, sondern zwischen Variablen unterscheiden, die simultan bestimmt werden (Y , *endogene Variablen*), und denen, für die dies nicht gilt (X , *exogene Variablen*). (IV4: 85) zeigt ein typisches Beispiel eines solchen Systems.

Menge	(1)	$Y_{1i} = \beta_0 + \beta_1 Y_{2i} + \beta_2 X_{1i} + \beta_3 X_{2i} + \varepsilon_{1i}$	(IV4: 85)
Preis	(2)	$Y_{2i} = \alpha_0 + \alpha_1 Y_{1i} + \alpha_2 X_{1i} + \alpha_3 X_{3i} + \varepsilon_{2i}$	

Nehmen wir an, die beiden Gleichungen in (IV4: 85) beschreiben den Markt für Schweinefleisch in Deutschland. Y_1 und Y_2 beschreiben die Menge und den Preis von Schweinefleisch, X_1 den Preis von Rindfleisch (Substitutionsgut), X_2 das durchschnittliche deutsche Haushaltseinkommen und X_3 den Preis von Futtermittel für Schweine. Gleichung 1 modelliert damit die Nachfrage-, Gleichung 2 die Angebotsseite. Kommt es nun in einem solchen System zu einer Veränderung der von den Konsumenten nachgefragten Menge Y_1 , so führt dies in Gleichung 2 zu einer Erhöhung des Preises Y_2 , was wiederum Y_1 beeinflusst. Ein solches Verhalten (sog. Feedback) kennzeichnet endogene Variablen. Eine exogene Variable wie z.B. X_1 kann zwar die Veränderung einer endogenen Variablen hervorrufen, die Veränderung der endogenen Variablen führt aber dann zu keiner Rückkopplung auf X_1 .

Wie kann nun aber ein simultanes System zu einer Verletzung von Annahme 2b des CLRM und damit zur Verzerrung (und Inkonsistenz) der OLS-Schätzer führen? Betrachten wir dazu (IV4: 85). Nimmt ε_1 zu, so führt dies zu einer Erhöhung von Y_1 in Gleichung 1, was wiederum zu einer Erhöhung von Y_2 über Gleichung 2 führt, da Y_1 hier eine erklärende Variable ist. Diese Erhöhung von Y_2 geht dann wieder in Gleichung 1 ein, da Y_2 hier eine erklärende Variable ist. Eine Erhöhung des Störterms ε_1 führt also in Gleichung 1 auch zu einer Erhöhung der erklärenden Variable Y_2 . ε_1 und Y_2 sind also korreliert und verletzen Annahme 2b und damit auch 2a. Damit ist der Schätzer für β_1 verzerrt. Auch die Schätzer der Koeffizienten der exogenen Variablen werden verzerrt. Nehmen wir β_2 im obigen Modell als Beispiel. β_2 soll den Effekt der Veränderung von X_1 auf Y_1 bei *Konstanthaltung aller anderen erklärenden Variablen* beschreiben. Eine Veränderung von X_1 führt nun aber über die Veränderung von Y_1 in Gleichung 2 zu einer Veränderung von Y_2 . Es ist also nicht möglich die anderen Variablen konstant zu halten. Der Schätzer für β_2 wird verzerrt.

⁵⁸ Für eine ausführlichere Behandlung dieses Themas vgl. z.B. Studenmund, A.H. (2007).

4.6.3 Instrumentenvariablen schätzung

Nachdem wir uns mit den Ursachen der Verletzung von Annahme 2b befasst haben, wollen wir nun mögliche Korrekturmaßnahmen betrachten. Die *Instrumentenvariablen schätzung* gehört zu den am häufigsten vorgeschlagenen Verfahren. Sie kann bei Messfehlern, verzögerten endogenen Variablen und Simultanität eingesetzt werden. Im Falle vernachlässigter Variablen empfiehlt sich hingegen die Korrektur des Spezifikationsfehlers durch Einbeziehen der vernachlässigten Variablen. Ist dies nicht möglich, da z.B. für die Variablen keine Daten verfügbar sind, wird mit sog. *Proxy-Variablen* gearbeitet, die möglichst hoch mit den vernachlässigten Variablen korreliert sind und an ihrer Stelle in ein Modell aufgenommen werden.⁵⁹

4.6.3.1 Instrumentenvariablen

Nehmen wir an, im einfachen linearen Regressionsmodell

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

ist die Annahme 2b des CLRM verletzt. Die Variable X, die wir dann als *endogen* bezeichnen, ist mit dem Störterm korreliert, was dazu führt, dass OLS verzerrt und inkonsistent ist. Wichtigster Schritt bei der Instrumentenvariablen schätzung ist nun die Suche nach einer neuen Variable Z, die hochgradig mit der Variable X korreliert und gleichzeitig nicht mit dem stochastischen Störterm des Modells korreliert ist. Eine Zufallsvariable Z, wird daher dann als **Instrument** bezeichnet, wenn⁶⁰

- die Korrelation zwischen Z und ϵ , bei Null liegt (*Instrument-Exogenität*)

$$\text{Cov}(\epsilon_i, Z_i) = 0 \quad (\text{IV4: 86a})$$

- und die Korrelation zwischen Z und X mit zunehmendem Stichprobenumfang deutlich von Null unterscheidet (*Instrument-Relevanz*)

$$\text{Cov}(X_i, Z_i) \neq 0 . \quad (\text{IV4: 86a})$$

Haben wir eine solche Variable Z gefunden, so können wir die Schätzer so modifizieren, dass wir zwar nicht erwartungstreue jedoch *konsistente Schätzer* erhalten:

Ist X die mit dem Störterm korrelierte Variable und Z das für sie gefundene Instrument, so hat der **Instrumentenvariablen schätzer** (IV-Schätzer) die Gestalt⁶¹

$$\hat{\beta}_1^{\text{IV}} = \frac{\sum y_i z_i}{\sum x_i z_i} = \frac{\widehat{\text{Cov}}(Y, Z)}{\widehat{\text{Cov}}(X, Z)} . \quad (\text{IV4: 87})$$

Das stellt nichts anderes als die Stichprobenkovarianz zwischen Y und Z dividiert durch die Stichprobenkovarianz zwischen X und Z dar. Dies verdeutlicht, dass der IV-Schätzer nur dann Ergebnisse liefern kann, wenn Z und X korreliert sind. An-

⁵⁹ Die notwendigen Bedingungen, damit diese Vorgehensweise zu konsistenten Schätzern führt, erläutert z.B. Wooldridge, J.M. (2006), Kapitel 9.

⁶⁰ Es reicht aus, wenn diese Bedingungen mit zunehmendem Stichprobenumfang erfüllt sind. Zur Vereinfachung verzichten wir aber im Folgenden auf eine Limes-Schreibweise.

⁶¹ Eine Herleitung findet sich z.B. in Hill, R.C., Griffiths, W.E., Judge, G.G. (2001), Kapitel 13.

dernfalls ist er nicht definiert. Es ist kein Zufall, dass wir aus (IV4: 87) für den Fall $Z = X$ den OLS-Schätzer erhalten. Ist X nämlich exogen, so kann X sich selbst instrumentieren, sodass der IV- dem OLS-Schätzer entspricht.

In ähnlicher Weise, in der wir auch (IV2: 20) hergeleitet haben, können wir nun auch die Beziehung zwischen dem Instrumentenvariablen-Schätzer und dem wahren Regressionsparameter der Grundgesamtheit β_1 aufstellen. Es gilt

$$\hat{\beta}_1^{\text{IV}} = \beta_1 + \frac{\sum z_i \epsilon_i}{\sum x_i z_i}. \quad (\text{IV4: 88})$$

Analog zu unseren Ausführungen unter IV 4.6.1 zeigt diese Darstellung, dass sich bei Gültigkeit von (IV4: 86a) $\hat{\beta}_1^{\text{IV}}$ mit zunehmendem Stichprobenumfang n dem wahren Wert β_1 annähert (Stichprobenkovarianz zwischen Z und ϵ geht gegen Null). $\hat{\beta}_1^{\text{IV}}$ ist daher ein konsistenter Schätzer für β_1 . In großen Stichproben ist $\hat{\beta}_1^{\text{IV}}$ *approximativ* normalverteilt mit

$$E(\hat{\beta}_1^{\text{IV}}) = \beta_1 \quad (\text{IV4: 89})$$

$$\text{Var}(\hat{\beta}_1^{\text{IV}}) = \frac{\sigma^2}{\sum x_i^2 r_{zx}^2}, \quad (\text{IV4: 90})$$

wobei σ^2 die Varianz des Störterms ist und r_{zx}^2 für den quadrierten Korrelationskoeffizienten zwischen X und Z in der Stichprobe steht. Da $\sum x_i^2$ nichts anderes als $n \cdot \text{Var}(X)$ ist, nimmt die Varianz des Instrumentenvariablen-Schätzers i.d.R. mit zunehmendem Stichprobenumfang ab. Je höher die Korrelation zwischen X und seinem Instrument Z ist, desto geringer ist die Varianz und desto effizienter ist der IV-Schätzer. Gerade deswegen ist eine möglichst hohe Korrelation zwischen X und Z so wichtig. Geht diese gegen Null, strebt die Varianz des Schätzers gegen unendlich. Hieran erkennen wir auch die Kosten des IV-Schätzers, die wir in Kauf nehmen müssen, wenn wir vermuten, dass eine erklärende Variable endogen ist. Sind X und ϵ nämlich doch unkorreliert, dann hat der IV-Schätzer eine höhere Varianz als der OLS-Schätzer, solange $r_{zx}^2 < 1$.

Den Schätzwert für σ^2 erhalten wir ähnlich wie für OLS als

$$\hat{\sigma}_{\text{IV}}^2 = \frac{\sum (e_i^{\text{IV}})^2}{n - 2} = \frac{\sum (Y_i - \hat{\beta}_0^{\text{IV}} - \hat{\beta}_1^{\text{IV}} X_i)^2}{n - 2}. \quad (\text{IV4: 91})$$

Es werden also in (IV2: 25) lediglich die OLS-Schätzer durch die IV-Schätzer ersetzt.

Nutzen wir die Residuen der Instrumentenvariablen-Schätzung e_i^{IV} zur Berechnung des Bestimmtheitsmaßes nach der klassischen Formel $R^2 = 1 - \text{RSS/TSS}$, so ergeben sich diverse Probleme. So kann R^2 dann auch negative Werte annehmen, besitzt keine natürliche Interpretation und kann nicht im Rahmen eines F-Tests zur Berechnung der F-Statistik herangezogen werden. R^2 ist nun aber im Rahmen unserer Analyse nicht die entscheidende Größe. Was nutzt uns nämlich z.B. ein hohes OLS- R^2 , wenn wir β_1 nicht konsistent schätzen können. Unser Ziel bei der IV-Schätzung ist anders als bei OLS keine R^2 -Maximierung, sondern bessere Schätzungen des partiellen Effekts von X auf Y , wenn X mit ϵ korreliert ist.

In der Literatur wird zum Teil recht ungenau formuliert, dass die Variable X im Zuge der Instrumentenvariablen-Schätzung durch Z ersetzt werden soll. Würden wir dieser Anleitung strikt Folge leisten, so erhielten wir nicht (IV4: 87) sondern

$$\hat{\beta}_1^* = \frac{\sum y_i z_i}{\sum z_i^2},$$

dessen Umformung

$$\hat{\beta}_1^* = \beta_1 \frac{\sum x_i z_i}{\sum z_i^2} + \frac{\sum z_i \epsilon_i}{\sum z_i^2}$$

liefert. Mit zunehmendem Stichprobenumfang nähert sich auch hier der zweite Term Null an. Der erste wird jedoch nur β_1 erreichen, wenn der dahinterstehende Bruch 1 erreicht. Das wäre aber nur der Fall, wenn $X = Z$ gilt. Da dann wieder der letzte Term nicht gegen Null konvergiert, ist ein solcher Schätzer nicht konsistent.

Zusammenfassend liegt uns mit der Instrumentenvariablen schätzung eine Technik vor, die bei Vorhandensein eines geeigneten Instruments konsistente Schätzer liefert. Gerade die Tatsache, dass es nicht leicht ist, geeignete Instrumente zu finden, erschwert aber die praktische Umsetzung. Besonders schwierig wird es bei multiplen Regressionen, bei der mehrere Variablen mit dem Störterm korreliert sind. Für jede endogene Variable ist dann nämlich mindestens ein eigenes Instrument nötig.

4.6.3.2 Zweistufige Methode der kleinsten Quadrate (TSLS)

Bei der praktischen Umsetzung der Instrumentenvariablen schätzung wird auf die *zweistufige Methode der kleinsten Quadrate* (TSLS, engl. two stage least squares) zurückgegriffen. Sie ist vor allem dann relevant, wenn es *mehr potentielle Instrumente als endogene Regressoren* gibt. Wir verwenden in einem solchen Fall nämlich nicht das Z mit der höchsten Korrelation mit X , sondern berücksichtigen alle Z in geeigneter Weise.

Betrachten wir ein **einfaches Regressionsmodell**

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (\text{IV4: 92})$$

mit einer endogenen erklärenden Variablen X und zwei für sie in Frage kommenden Instrumenten Z_1 und Z_2 , für die $\text{Cov}(\epsilon_i, Z_{1i}) = 0$ und $\text{Cov}(\epsilon_i, Z_{2i}) = 0$ sowie $\text{Cov}(X_i, Z_{1i}) \neq 0$ und $\text{Cov}(X_i, Z_{2i}) \neq 0$ gilt. Wir interessieren uns nun für die Instrumentenvariablen schätzung der Parameter β_0 und β_1 . In einem solchen Fall folgt TSLS dem folgenden 2-Stufen-Schema:

Stufe 1: Es wird die Variable, die die Verletzung von Annahme 2b hervorruft (hier also X), auf die exogenen Instrumentenvariablen regressiert:

$$X_i = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + v_i \quad (\text{IV4: 93})$$

Mit dem Ergebnis der Schätzung dieses Modells werden dann die gefiteten Werte $\hat{X}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_{1i} + \hat{\alpha}_2 Z_{2i}$ berechnet. Da diese von den Instrumentenvariablen abhängen bzw. *Linearkombinationen der Instrumente* sind, sind sie asymptotisch nicht mit dem Störterm ϵ des UrsprungsmodeLLS korreliert.⁶²

⁶² Warnung: Das Hinzufügen einer relativ großen Anzahl von zulässigen Instrumenten kann im Fall endlicher Stichproben in Stufe 2 zur Inkonsistenz der Schätzer führen. Mit diesem Problem befassen sich Bound, J., Jaeger, D.A., Baker, R.M. (1995) näher.

Stufe 2: Die gefitteten Werte \hat{X}_i des 1. Schritts werden dann zur Schätzung von

$$Y_i = \gamma_0 + \gamma_1 \hat{X}_i + u_i \quad (\text{IV4: 94})$$

verwendet. Die OLS-Schätzung dieser Gleichung liefert die gesuchten Instrumentenvariablenabschätzungen als $\hat{\gamma}_0$ für β_0 und $\hat{\gamma}_1$ für β_1 .

Im Falle nur eines Instruments entspricht $\hat{\gamma}_1$ aus der Schätzung von (IV4: 94) genau dem Schätzwert, den wir auch über (IV4: 87) erhalten hätten. Der Standardfehler aus (IV4: 94) stimmt jedoch nicht mit dem des IV-Schätzers überein, was jedoch bei der TSLS-Prozedur in allen gängigen Softwarepaketen (so auch in EViews) korrigiert wird.⁶³ Generell greifen wir daher in der Praxis *auch bei Modellen mit nur einer Instrumentenvariablen* auf TSLS zurück.

Die TSLS-Methode ist einfach auf die **multiple Regression** übertragbar. Es ist aber zu beachten, dass dann mindestens so viele Instrumente zur Verfügung stehen müssen, wie endogene erklärende Variablen vorhanden sind und diese Instrumente Variablen sein müssen, die *nicht bereits im Modell enthalten* sind. Liegen genau so viele Instrumente wie endogene erklärende Variablen vor, so sprechen wir von *exakter Identifikation*, bei zu vielen Instrumenten von *Überidentifikation* und bei zu wenigen von *Unteridentifikation*. Im Falle von Unteridentifikation ist eine Schätzung unmöglich.

Nehmen wir beispielsweise an, wir verfügen in einem Modell der Form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \quad (\text{IV4: 95})$$

in dem **eine erklärende Variable** (konkret X_1) **endogen** ist, über zwei exogene Instrumente Z_1 und Z_2 mit $\text{Cov}(\varepsilon_i, Z_{1i}) = 0$ und $\text{Cov}(\varepsilon_i, Z_{2i}) = 0$ sowie $\text{Cov}(X_{1i}, Z_{1i}) \neq 0$ und $\text{Cov}(X_{1i}, Z_{2i}) \neq 0$. Wir sagen hier auch, dass eine Überidentifikation vom Grad 1 vorliegt, da wir ein Instrument mehr zur Verfügung haben, als wir benötigen. Wir können nun mit TSLS in den folgenden 2 Stufen schätzen:

Stufe 1: Wir regressieren die endogene erklärende Variable auf die exogenen Modellvariablen und die verfügbaren Instrumente, also

$$X_{1i} = \alpha_0 + \alpha_1 X_{2i} + \alpha_2 Z_{1i} + \alpha_3 Z_{2i} + v_i, \quad (\text{IV4: 96})$$

wobei der Störterm v den Erwartungswert $E(v) = 0$ besitzt und nicht mit den erklärenden Variablen korreliert ist, d.h. $\text{Cov}(v_i, X_{2i}) = \text{Cov}(v_i, Z_{1i}) = \text{Cov}(v_i, Z_{2i}) = 0$. Die Schätzergebnisse nutzen wir zur Bestimmung der $\hat{X}_{1i} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{2i} + \hat{\alpha}_2 Z_{1i} + \hat{\alpha}_3 Z_{2i}$, die als *Linearkombinationen exogener Variablen* sicher nicht mit dem Störterm ε des Ursprungsmodeells korreliert sind.

Stufe 2: Wir nutzen die \hat{X}_{1i} aus Stufe 1 für die OLS-Schätzung von

$$Y_i = \gamma_0 + \gamma_1 \hat{X}_{1i} + \gamma_2 X_{2i} + u_i \quad (\text{IV4: 97})$$

und erhalten daraus unsere Instrumentenvariablenabschätzungen.

⁶³ Generell kann jedoch TSLS genau wie OLS unter Heteroskedastizität und Autokorrelation leiden, sodass dennoch falsche Standardfehler ausgewiesen werden. Wooldridge, J.M. (2006) gibt daher in Kapitel 15 einen Überblick über mögliche Gegenmaßnahmen. Bewährt hat sich u.a. die Verwendung von Newey-West Standardfehlern.

Wären in Stufe 1 $\alpha_2 = 0$ und $\alpha_3 = 0$, so sind die beiden Instrumente *nicht relevant* und es liegt Unteridentifikation vor. Als Folge wäre \hat{X}_1 linear abhängig von X_2 , weshalb die Schätzung der Stufe 2 perfekt multikollinear und damit unmöglich wäre. Es muss daher in Stufe 1 $\alpha_2 \neq 0$ oder $\alpha_3 \neq 0$ gelten, damit wir schätzen können. Die Signifikanz der Instrumente in Stufe 1 kann als Indikator für ihre Relevanz betrachtet werden. Ein hoher Wert der F-Statistik für gemeinsame Signifikanz aller Instrumentenvariablen (Z_1 und Z_2 , also ohne X_2) in Stufe 1 ist notwendig. Ein kleiner Wert der F-Statistik ist ein Indiz für schwache (wenig Erklärung der Veränderung von X_1) Instrumente.

Beispiel IV 4-26

Lohnmodell und TSLS



Wooldridge, J.M. (2006) schätzt mit einem Datensatz von 428 Beobachtungen ein typisches Modell zur Erklärung der Löhne von arbeitstätigen Frauen

$$\ln(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \varepsilon_i,$$

in dem Y den Stundenlohn (in US-Dollar), X_1 die Schulbildung (in Jahren) und X_2 die Berufserfahrung (in Jahren) darstellen, mit Hilfe der Instrumente Schulbildung der Mutter Z_1 und Schulbildung des Vaters Z_2 , da er eine Endogenität der Variable X_1 vermutet.

Schätzen wir das Modell zunächst ohne Berücksichtigung einer evtl. Endogenität mit OLS, also wie Gleichung (IV4: 95), so erhalten wir folgendes Ergebnis:

Dependent Variable: LOG(Y)		Sample: 1 428		
Method: Least Squares		Included observations: 428		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.522041	0.198632	-2.628179	0.0089
X1	0.107490	0.014146	7.598332	0.0000
X2	0.041567	0.013175	3.154906	0.0017
X2^2	-0.000811	0.000393	-2.062834	0.0397
R-squared	0.156820	Mean dependent var		1.190173
Adjusted R-squared	0.150854	S.D. dependent var		0.723198
S.E. of regression	0.666420	Akaike info criterion		2.035509
Sum squared resid	188.3051	Schwarz criterion		2.073445
Log likelihood	-431.5990	F-statistic		26.28615
Durbin-Watson stat	1.960988	Prob(F-statistic)		0.000000

Zur TSLS-Schätzung in EViews (Quick / Estimate Equation - Method: TSLS) sind lediglich die originäre Modellgleichung und eine Liste der Instrumente anzugeben. Dabei ist jedoch zu beachten, dass in die Liste nicht nur die "externen" Instrumente aufgenommen werden. Es müssen auch die exogenen Variablen des Modells vorkommen, da diese als Instrumente für sich selbst dienen. Die zwei TSLS-Schritte laufen dann automatisch ab. Auf der Folgeseite sind die IV-Schätzungen der zweiten Stufe von TSLS, d.h. Ergebnisse der Regression analog zu (IV4: 97), wiedergegeben. Ebenfalls sind die Ergebnisse der ersten Stufe aufgeführt, die von EViews nicht ausgewiesen werden. Dazu wird eine Regression von X_1 auf X_2 , X_2^2 , Z_1 und Z_2 durchgeführt.

Die von EViews im Rahmen der TSLS-Prozedur berechneten Statistiken, wie z.B. das R^2 von 0,1357, beziehen sich auf die Residuen, die aus dem ursprünglichen Modell unter Verwendung der Instrumentenvariablenabschätzungen des letzten TSLS-Schritts entstehen. Dazu werden ins Analogon zu Gleichung (IV4: 95) an Stelle der β_k die geschätzten γ_k aus dem Analogon zu (IV4: 97) eingesetzt. Sie dürfen nicht mit den Residuen des zweiten TSLS-Schritts, d.h. von Gleichung (IV4: 97), verwechselt werden.

Dependent Variable: X1 Method: Least Squares		Stufe 1		Sample: 1 428 Included observations: 428	
		Coefficient	Std. Error	t-Statistic	Prob.
C	9.102640	0.426561	21.33958	0.0000	
X2	0.045225	0.040251	1.123593	0.2618	
X2^2	-0.001009	0.001203	-0.838572	0.4022	
Z1	0.157597	0.035894	4.390609	0.0000	
Z2	0.189548	0.033756	5.615173	0.0000	
R-squared	0.211471	Mean dependent var		12.65888	
Adjusted R-squared	0.204014	S.D. dependent var		2.285376	
S.E. of regression	2.038967	Akaike info criterion		4.274378	
Sum squared resid	1758.575	Schwarz criterion		4.321797	
Log likelihood	-909.7168	F-statistic		28.36041	
Durbin-Watson stat	1.939888	Prob(F-statistic)		0.000000	

Dependent Variable: LOG(Y) Method: Two-Stage Least Squares Instrument list: Z1 Z2 X2 X2^2		Stufe 2		Sample: 1 428 Included observations: 428	
		Coefficient	Std. Error	t-Statistic	Prob.
C	0.048100	0.400328	0.120152	0.9044	
X1	0.061397	0.031437	1.953024	0.0515	
X2	0.044170	0.013432	3.288329	0.0011	
X2^2	-0.000899	0.000402	-2.237993	0.0257	
R-squared	0.135708	Mean dependent var		1.190173	
Adjusted R-squared	0.129593	S.D. dependent var		0.723198	
S.E. of regression	0.674712	Sum squared resid		193.0200	
F-statistic	8.140709	Durbin-Watson stat		1.945659	
Prob(F-statistic)	0.000028	Second-Stage SSR		212.2096	

Beim Vergleich der beiden Regressionsergebnisse (OLS-Schätzung des UrsprungsmodeLLS und zweite Stufe von TSLS) fällt sofort die deutlich veränderte Schätzung des Regressionskoeffizienten der Variable \hat{X}_1 (aus dem Analog zu (IV4: 97)) im Vergleich zu X_1 (aus dem Analogon zu (IV4: 95)) bei nun höherem Standardfehler und niedrigerem t-Wert auf. Endogenität scheint hier also einen Einfluss auf die Schätzergebnisse zu haben.

Bisher hatten wir lediglich eine Vermutung, dass X_1 Annahme 2b verletzt. Vor Durchführung der Instrumentenvariablen-Schätzung ist es aber empfehlenswert durch ein geeignetes Testverfahren (vgl. Abschnitt 4.6.3.3) zu prüfen, ob dies auch wirklich der Fall ist. Führen wir die Instrumentenvariablen-Schätzung nämlich durch, wenn keine Verletzung vorliegt, führt dies zwar zu keinem verzerrten Schätzer, jedoch zu einem Schätzer mit höherer Varianz (im Vergleich zu OLS).

Wir können TSLS auch in Modellen mit **mehreren endogenen erklärenden Variablen** einsetzen.⁶⁴ Betrachten wir beispielsweise das Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \epsilon_i,$$

in dem X_1 und X_2 endogen und X_3 , X_4 sowie X_5 exogen sind. Um dieses Modell mit TSLS zu schätzen, benötigen wir **mindestens zwei** exogene Variablen, die nicht im

⁶⁴ Für eine detaillierte Beschreibung vgl. Wooldridge, J.M. (2006), Kapitel 15.

Modell enthalten und hochgradig mit X_1 und X_2 korreliert sind. Haben wir mindestens zwei Instrumente gefunden, können wir in der ersten Stufe von TSLS dann jeweils X_1 und X_2 auf alle Instrumente und exogenen erklärenden Variablen des Ursprungsmodells regressieren und die damit bestimmbaren \hat{X}_{1i} und \hat{X}_{2i} in der zweiten TSLS-Stufe als Variablen verwenden.

Da die Suche geeigneter Instrumentenvariablen oft schwer fällt, geben wir dem Leser abschließend folgende **Tipps zur praktischen Instrumentensuche** an die Hand:

1. Manchmal ist $\text{Cov}(\epsilon_t, X_t) \neq 0$, aber $\text{Cov}(\epsilon_t, X_{t-1}) = 0$. Zudem sind X_t und X_{t-1} meist korreliert. Wir können dann X_{t-1} als Instrument für X_t verwenden. Typischerweise ist dies der Fall, wenn bei X Messfehler entstehen und wir annehmen, dass der Messfehler nicht autokorreliert ist.
2. In einem simultanen System $Y_{1t} = \beta_0 + \beta_1 Y_{2t} + \beta_2 X_t + \epsilon_t$ (zweite Gleichung unbekannt, Y_2 aber sicher endogen) eignen sich z.B. X_{t-1} oder X_{t-2} als Instrumente für Y_2 , falls eine hohe Korrelation besteht. Problematisch ist dabei jedoch, dass X_{t-1} oder X_{t-2} in einem dynamischen Modell (vgl. Abschnitt IV 4.7.1) bereits enthalten sein können. Wir haben dann derartige Instrumente nicht zur Verfügung.
3. Vor allem bei Messfehlern können sich die Rangziffern der Variablen als Instrumente eignen. Wichtig ist dabei jedoch, dass das Ranking nicht durch die Messfehler beeinflusst werden darf.

4.6.3.3 Hausman-Test und Verletzung von Annahme 2b

Sind die erklärenden Variablen exogen, so ist der OLS-Schätzer effizienter als der IV-Schätzer. Es ist daher sinnvoll zu prüfen, ob eine erklärende Variable, von der wir vermuten, dass sie endogen ist, dies auch tatsächlich ist. Andernfalls ist eine Anwendung von TSLS nämlich gar nicht erforderlich und wir verwenden den effizienteren OLS-Schätzer. Dafür bietet sich der sog. Hausman-Test an.⁶⁵ Da die Residuen e_i nicht zum Test einer Korrelation zwischen einer erklärenden Variablen und dem Störterm verwendet werden können,⁶⁶ schlägt Hausman ein anderes Testverfahren vor. Die Grundidee des Tests besteht im direkten Vergleich von OLS- und IV-Schätzer. Unter der Nullhypothese der Exogenität ist der OLS-Schätzer konsistent (und für große Stichproben) effizient, wohingegen er inkonsistent ist, wenn die Alternativhypothese Endogenität gilt. Der IV-Schätzer hingegen ist konsistent, egal ob die Nullhypothese korrekt ist oder nicht, jedoch ineffizient, wenn sie gilt.

Ein großer Unterschied zwischen OLS- und IV-Schätzung wird von Hausman als Evidenz zugunsten der Alternativhypothese gedeutet. Unter Gültigkeit der Nullhypothese ist im *einfachen Regressionsmodell* die Statistik

$$H = \frac{\left(\hat{\beta}_1^{\text{IV}} - \hat{\beta}_1^{\text{OLS}} \right)^2}{\text{Var}(\hat{\beta}_1^{\text{IV}}) - \text{Var}(\hat{\beta}_1^{\text{OLS}})} \quad (\text{IV4: 98})$$

mit einem Freiheitsgrad asymptotisch Chi-Quadrat-verteilt. Zur Schätzung beider Varianzen soll dabei die gleiche Schätzung für σ^2 herangezogen werden. Die Null-

⁶⁵ Vgl. Hausman, J. (1978).

⁶⁶ Diese sind per Konstruktion in OLS nie mit der betroffenen erklärenden Variablen korreliert. Vgl. dazu (IV2: 13e).

hypothesen kann verworfen werden, wenn H_0 den kritischen Wert übersteigt. Schlussfolgerung ist dann, dass die zum Parameter β_i gehörende Variable endogen sein muss (unter der Voraussetzung, dass die zur IV-Schätzung verwendeten Instrumente wirklich exogen sind). Diese Methode kann einfach auf ein *multiples Regressionsmodell* erweitert werden.⁶⁷

Hausman-Tests kommen vor allem bei der Erkennung von Messfehlern, vernachlässigten Variablen und Simultanitätsverzerrungen zum Einsatz. Ein von Davidson und MacKinnon vorgeschlagener **Spezialfall des Hausman-Tests** wird besonders häufig angewendet.⁶⁸ Bei diesem wird im Falle nur einer endogener erklärenden Variablen zunächst die als endogen vermutete erklärende Variable X_1 auf alle exogenen Variablen des Modells und ein oder mehrere Instrumente regressiert. Die Residuen dieser Hilfsregression werden dann als zusätzliche Variable im originären Regressionsmodell berücksichtigt und das neue Modell geschätzt. Liefert ein t-Test für den Koeffizienten der Residuenvariable Signifikanz, kann dies als ein Hinweis dafür interpretiert werden, dass die Korrelation zwischen X_1 und ϵ einen Einfluss auf das Schätzergebnis hat. Im Falle mehrerer endogener erklärender Variablen wäre hier natürlich ein F-Test heranzuziehen, da für jede als endogen vermutete erklärende Variable eine Residuenvariable entsteht.

Nehmen wir an, wir vermuten, dass im Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i \quad (\text{IV4: 99})$$

eine erklärende Variable (konkret X_1) endogen ist. Haben wir ein Instrument Z für X_1 gefunden, so gehen wir in den folgenden zwei Schritten vor:

Schritt 1: Wir schätzen zunächst in einem ersten Schritt die Gleichung

$$X_{1i} = \alpha_0 + \alpha_1 X_{2i} + \alpha_2 X_{3i} + \alpha_3 Z_i + v_i, \quad (\text{IV4: 100})$$

d.h. wir regressieren die als endogen vermutete erklärende Variable auf alle exogenen Variablen und Z und bestimmen die Residuen \hat{v}_i . Da jede der in (IV4: 100) integrierten exogenen Variablen mit ϵ unkorreliert ist, ist X_1 nur dann mit ϵ unkorreliert, wenn v und ϵ unkorreliert sind. Dies ist es, was wir nun testen wollen. Schreiben wir $\epsilon_i = \delta_1 v_i + u_i$, wobei u ein klassischer Störterm ist, dann sind ϵ und v unkorreliert, wenn $\delta_1 = 0$ ist.

Schritt 2: Der einfachste Weg $\delta_1 = 0$ zu testen, ist die Durchführung eines t-Tests nach Aufnahme von v in (IV4: 99). Da v aber nicht beobachtbar ist, verwenden wir stattdessen die im ersten Schritt berechneten \hat{v}_i . Wir schätzen also das neue Modell

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \delta_1 \hat{v}_i + u_i \quad (\text{IV4: 101})$$

mit OLS und testen $H_0: \delta_1 = 0$ mittels der t-Statistik. Sind wir in der Lage H_0 abzulehnen bzw. ist δ_1 signifikant von Null verschieden, so können wir schließen, dass X_1 tatsächlich endogen ist, da ϵ und v korreliert sind, und OLS verzerrte Ergebnisse liefert.

⁶⁷ Vgl. z.B. Johnston, J., Dinardo, J. (1997), Kapitel 8.

⁶⁸ Vgl. Davidson, R., MacKinnon, J. G. (1989).

Interessante Eigenschaft der Schätzung von (IV4: 101) ist, dass die geschätzten Koeffizienten der erklärenden Variablen (abgesehen von \hat{v} natürlich) genau den Ergebnissen einer TSLS-Schätzung von (IV4: 99) entsprechen. Dieses Verfahren liefert uns also sowohl die TSLS-Schätzergebnisse als auch eine gleichzeitige Möglichkeit der Prüfung, ob die Anwendung von TSLS überhaupt gerechtfertigt war. Wir können daher das vorliegende Testverfahren wie folgt interpretieren: Das Hinzufügen von \hat{v} zum Ursprungsmodell (IV4: 99) und die anschließende OLS-Schätzung bestätigt die Endogenität von X_1 . Schätzen wir also zunächst (IV4: 99) mit OLS, können wir sehen wie stark sich die Schätzung $\hat{\beta}_1$ durch Hinzufügen von \hat{v} ändert und damit den Einfluss einer Endogenität von X_1 erkennen.

Beispiel IV 4-27

Hausman-Test



Prüfen wir für das Lohnmodell aus Beispiel IV 4-26, ob die Anwendung von TSLS tatsächlich gerechtfertigt war, d.h. ob die Variable X_1 Annahme 2b wirklich verletzt. In einem ersten Schritt ist zunächst die Variable X_1 in Abhängigkeit von ihren Instrumenten Z_1 und Z_2 sowie den exogenen Variablen X_2 und X_2^2 des Ursprungsmodells zu schätzen. Da diese Regression genau der ersten Stufe der TSLS-Prozedur entspricht, haben wir sie bereits unter Beispiel IV 4-26 durchgeführt. Die Residuen dieser Stufe, die sich aufgrund des im Beispiel IV 4-26 angegebenen Regressionsergebnisses leicht über die Gleichung

$$\hat{v}_1 = X_{11} - 9,1026 - 0,0452X_{21} + 0,0010X_{21}^2 - 0,1576Z_{11} - 0,1895Z_{21}$$

bestimmen lassen, werden nun als zusätzliche Variable in die ursprüngliche Modellgleichung integriert und das so entstehende neue Modell geschätzt. Dies liefert den folgenden Regressionsoutput:

Dependent Variable: LOG(Y)		Sample: 1 428		
Method: Least Squares		Included observations: 428		
	Coefficient	Std. Error	t-Statistic	Prob.
C	0.048100	0.394575	0.121904	0.9030
X1	0.061397	0.030985	1.981499	0.0482
X2	0.044170	0.013239	3.336272	0.0009
X2^2	-0.000899	0.000396	-2.270623	0.0237
V	0.058167	0.034807	1.671105	0.0954
R-squared	0.162350	Mean dependent var		1.190173
Adjusted R-squared	0.154429	S.D. dependent var		0.723198
S.E. of regression	0.665016	Akaike info criterion		2.033602
Sum squared resid	187.0701	Schwarz criterion		2.081022
Log likelihood	-430.1908	F-statistic		20.49611
Durbin-Watson stat	1.931343	Prob(F-statistic)		0.000000

Es wird dadurch unsere Behauptung bestätigt, dass die durch das Integrieren der Residuenvariable resultierenden Parameterschätzungen mit denen der TSLS-Prozedur identisch sind.⁶⁹ Wir erkennen jedoch, dass der p-Wert für den Koeffizienten der integrierten Residuenvariable (in E-Views V) bei 0,0954 liegt. Ein zweiseitiger t-Test der Nullhypothese der Exogenität (Koeffizient von V gleich Null) würde daher bei einem Signifikanzniveau von 5 % nicht zu einer Ablehnung von H_0 führen. Bei einem Signifikanzniveau von 10 % würde

⁶⁹ Die Unterschiede bei den Standardfehlern ergeben sich aus der verschiedenen Anzahl von Freiheitsgraden bei der jeweiligen Berechnung der Standardfehler.

sich jedoch die Alternativhypothese der Endogenität bestätigen lassen. Wir haben damit nur einen moderaten Hinweis auf Korrelation zwischen ϵ und X_1 . Es empfiehlt sich daher aufgrund der Tatsache, dass die TSLS-Schätzung von 0,0614 doch deutlich unter der OLS-Schätzung von 0,1074 liegt, beide Ergebnisse festzuhalten.

Dieser Spezialfall des Hausman-Tests kann einfach für **mehrere endogene erklärende Variablen** verallgemeinert werden. Dazu wird allerdings für jede als endogen vermutete Variable mindestens ein Instrument benötigt. Zur Testdurchführung berechnen wir lediglich für jede als endogen vermutete erklärende Variable die Residuen der Regression des Typs (IV4: 100), verwenden diese als zusätzliche erklärende Variablen im ursprünglichen Modell und testen deren gemeinsame Signifikanz mittels eines F-Tests. Gemeinsame Signifikanz deutet an, dass mindestens eine der als endogen vermuteten erklärenden Variablen dies auch ist.

4.6.3.4 Sargan-Test und Güte von Instrumenten

Wie wir gesehen haben, müssen Instrumente relevant (Korrelation mit der endogenen Variable) und exogen (keine Korrelation mit dem Störterm) sein. Während ein F-Test auf gemeinsame Signifikanz aller Instrumente in Stufe 1 von TSLS Hinweise auf die Relevanz der Instrumente liefert, kann die Korrelation zwischen Instrument und dem unbeobachtbaren Störterm der Grundgesamtheit nicht unmittelbar getestet werden. Um die Güte von Instrumenten trotzdem beurteilen zu können, eignet sich im Falle von *Überidentifikation* (Anzahl der Instrumente größer als Anzahl der endogenen erklärenden Variablen) der sog. **Sargan-Test**.⁷⁰

Nehmen wir zu dessen Veranschaulichung an, unser Modell sei

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \quad (\text{IV4: 102})$$

und wir verfügen für die endogene erklärende Variable X_i über zwei potentielle Instrumente Z_1 und Z_2 . Die Testdurchführung erfolgt nun in zwei Schritten:

Zunächst wird die überidentifizierte Gleichung mit TSLS geschätzt und die dabei gewonnenen Parameterschätzungen, d.h. die γ_k aus Gleichung (IV4: 97) als Schätzungen für die β_k zur Bestimmung der Residuen e_i im Modell (IV4: 102) eingesetzt. Anschließend wird dann eine Regression von e_i auf *alle exogenen Variablen*

$$e_i = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \alpha_3 X_{2i} + v_i \quad (\text{IV4: 103})$$

durchgeführt. Falls Z_1 , Z_2 und X exogen sind, sollten sie nicht mit e korreliert sein und die Schätzung von (IV4: 103) ein niedriges Bestimmtheitsmaß R^2 liefern. Unter der Nullhypothese, dass *alle* Instrumente exogen sind, ist die Teststatistik

$$\text{SARG} = n \cdot R^2, \quad (\text{IV4: 104})$$

bei der n den Stichprobenumfang und R^2 das Bestimmtheitsmaß der Schätzung von (IV4: 103) darstellt, mit q Freiheitsgraden *asymptotisch* Chi-Quadrat-verteilt. q ist dabei die Anzahl der Instrumentenvariablen Z abzüglich der Anzahl der endogenen erklärenden Variablen, d.h. der Grad der Überidentifikation (hier $q = 2 - 1 = 1$).

⁷⁰ Vgl. Sargan, J. D. (1964). Voraussetzung der Anwendung des Tests ist Homoskedastizität im Bezug auf die Instrumentenvariablen, d.h. $\text{Var}(\epsilon_i | Z_{1i}, Z_{2i}, \dots) = \sigma^2$.

Ist SARG größer als der entsprechende kritische Wert der Chi-Quadrat-Verteilung, muss die Nullhypothese, dass alle Instrumente exogen sind, verworfen werden. Eine Ablehnung von H_0 bedeutet, dass mindestens ein Instrument mit dem Störterm ε korreliert ist und daher die IV-Schätzungen, die auf den gewählten Instrumenten basieren, nicht gültig sind.

Wichtig ist, dass der Sargan-Test nur benutzt werden kann, wenn mehr Instrumente vorliegen, als wir unbedingt brauchen. Haben wir genauso viele Instrumente wie endogene erklärende Variablen, wäre das R^2 , das wir im Zuge des Sargan-Tests erhalten gleich Null. Wir können die *Exogenität der Instrumente* im Falle *exakter Identifikation* also *nicht testen*.

Beispiel IV 4-28

Sargan-Test



Für das Lohnmodell aus Beispiel IV 4-26 hatten wir die Instrumente Schulbildung der Mutter Z_1 und Schulbildung des Vaters Z_2 zur TSLS-Schätzung herangezogen. Die Relevanz der Instrumente können wir durch einen F-Test für ihre gemeinsame Signifikanz in der ersten TSLS-Stufe, d.h. in der Regression von X_i auf X_2, X_2^2, Z_1 und Z_2 , untersuchen. Diese erste Stufe haben wir bereits im Beispiel IV 4-26 geschätzt. Um diesen F-Test durchzuführen dürfen wir nun nicht einfach den F-Wert des EViews-Outputs von 28,3604 heranziehen, da dieser die Gesamtsignifikanz aller Modellvariablen testet. Wir benötigen nur einen Test für die Parameter von Z_1 und Z_2 . Diesen können wir in EViews im Schätzfensster dieser Regression über View / Coefficient Tests / Wald - Coefficient Restrictions und Eingabe der Restriktion $c(4) = c(5) = 0$ durchführen (vgl. dazu auch Beispiel IV 3-7). Das Resultat eines solchen F-Tests ist $F = 55,4003$. Die Variable X_i ist also klar (partiell) korreliert mit Z_1 und Z_2 .

Die Instrument-Exogenität können wir mit dem eben beschriebenen Sargan-Test überprüfen. Dazu benötigen wir zunächst die im Anschluss an Gleichung (IV4: 102) erläuterten Residuen e_i . Diese werden von E-Views nach der TSLS-Schätzung in der Variable "resid" abgelegt und dürfen nicht mit den Residuen des zweiten TSLS-Schritts, \hat{u}_i , verwechselt werden. Wir legen für die relevanten Residuen eine Variable E1 an und regressieren sie auf alle exogenen Variablen, was folgendes Ergebnis liefert:⁷¹

Dependent Variable: E1		Sample: 1 428		
Method: Least Squares		Included observations: 428		
	Coefficient	Std. Error	t-Statistic	Prob.
C	0.010964	0.141257	0.077618	0.9382
Z1	-0.006607	0.011886	-0.555804	0.5786
Z2	0.005782	0.011179	0.517263	0.6052
X2	-1.83E-05	0.013329	-0.001376	0.9989
X2^2	7.34E-07	0.000398	0.001842	0.9985
R-squared	0.000883	Mean dependent var		3.18E-16
Adjusted R-squared	-0.008565	S.D. dependent var		0.672337
S.E. of regression	0.675210	Akaike info criterion		2.064029
Sum squared resid	192.8495	Schwarz criterion		2.111448
Log likelihood	-436.7021	F-statistic		0.093496
Durbin-Watson stat	1.946859	Prob(F-statistic)		0.984495

⁷¹ Dies ist erforderlich, da die Variable "resid" in EViews nicht direkt für weiterführende Berechnungen verwendet werden kann.

Mit $R^2 = 0,0009$ und dem Stichprobenumfang $n = 428$ können wir nun den konkreten Wert $SARG = 428 \cdot 0,0009 = 0,3852$ ermitteln, was in einer Chi-Quadrat-Verteilung mit einem Freiheitsgrad einen sehr kleinen Wert darstellt, sodass die Instrumentenvariablen den Sargan-Test bestehen. Fügen wir Z_3 , die Schulbildung des Ehemannes, in unsere Liste von Instrumentenvariablen ein, so erhalten wir eine Überidentifikation vom Grad $q = 2$ und $SARG = 428 \cdot 0,0026 = 1,1128$. Auch damit bestehen die Instrumente den Test.

Dependent Variable: E2		Sample: 1 428		
Method: Least Squares		Included observations: 428		
	Coefficient	Std. Error	t-Statistic	Prob.
C	0.008606	0.177272	0.048549	0.9613
Z1	-0.010385	0.011869	-0.874997	0.3821
Z2	0.000673	0.011380	0.059179	0.9528
Z3	0.006781	0.011426	0.593484	0.5532
X2	5.60E-05	0.013229	0.004236	0.9966
X2^2	-8.88E-06	0.000396	-0.022452	0.9821
R-squared	0.002605	Mean dependent var		-4.75E-16
Adjusted R-squared	-0.009212	S.D. dependent var		0.666942
S.E. of regression	0.670007	Akaike info criterion		2.050863
Sum squared resid	189.4399	Schwarz criterion		2.107767
Log likelihood	-432.8847	F-statistic		0.220457
Durbin-Watson stat	1.938177	Prob(F-statistic)		0.953682

Die Aufnahme des zusätzlichen Instruments führt im Vergleich zur Schätzung mit nur zwei Instrumenten ($\hat{\beta}_1 = 0,0614$ bei einem Standardfehler von 0,0314) zu $\hat{\beta}_1 = 0,0804$ bei einem niedrigeren Standardfehler von 0,0218, wodurch der Parameter von X_1 an Signifikanz gewinnt.

Dependent Variable: LOG(Y)		Sample: 1 428		
Method: Two-Stage Least Squares		Included observations: 428		
Instrument list: Z1 Z2 Z3 X2 X2^2				
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.186857	0.285396	-0.654730	0.5130
X1	0.080392	0.021774	3.692104	0.0003
X2	0.043097	0.013265	3.248981	0.0013
X2^2	-0.000863	0.000396	-2.177745	0.0300
R-squared	0.149524	Mean dependent var		1.190173
Adjusted R-squared	0.143506	S.D. dependent var		0.723198
S.E. of regression	0.669298	Sum squared resid		189.9347
F-statistic	11.52474	Durbin-Watson stat		1.952048
Prob(F-statistic)	0.000000	Second-Stage SSR		207.8396

4.7 Besonderheiten bei der Arbeit mit Zeitreihen

Die Durchführung von Zeitreihenregressionen verursacht oft Komplikationen, die in Querschnittsregressionen nicht auftauchen können. So haben wir bereits gesehen, dass distributive Lag-Modelle häufig unter Multikollinearität leiden, was ihre Schätzung bzw. die Interpretation von Schätzergebnissen erschwert. Wir wollen uns daher in diesem Kapitel näher mit der Modellklasse der *dynamischen Modelle* beschäftigen. Des Weiteren befassen wir uns näher mit dem Konzept der *Stationarität* von Zeitreihen. Verändern sich die Eigenschaften einer Variablen (z.B. Erwartungswert oder Varianz) im Zeitverlauf drastisch, wird sie als nicht stationär bezeichnet. Ein besonderes Problem mit nicht stationären Variablen ist, dass sie t-Werte und die Maße für den "Gesamtfit" von Modellen in die Höhe treiben. Zudem werden wir uns im Kontext von nicht stationären Variablen mit dem Konzept der Kointegration und den Fehlerkorrekturmodellen beschäftigen. Dieses Konzept führte zu einer Revolution in der Ökonometrie. Engle und Granger erhielten für ihre Arbeiten auf diesem Gebiet im Jahr 2003 den Nobelpreis für Ökonomie.

4.7.1 Dynamische Modelle

4.7.1.1 Grundlagen

In Abschnitt IV 4.1.4 haben wir bereits zwei Arten dynamischer Modelle kennengelernt: die einfachen und die distributiven Lag-Modelle. Eine weitere Art ergibt sich in Folge von Problemen, die mit distributiven Lag-Modellen verbunden sind. Wir haben gesehen, dass mit distributiven Lag-Modellen der Form

$$Y_t = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_s X_{t-s} + \varepsilon_t \quad (\text{IV4: 105})$$

insbesondere Multikollinearität und der Verlust von Freiheitsgraden einhergehen. Um diesen Problemen zu begegnen, werden in der Praxis unter recht restriktiven Annahmen Modifikationen derartiger Modelle vorgenommen. Bekannteste Umformung ist dabei wohl die **Koyck-Transformation**.⁷² Unter den Annahmen, dass alle β_k das gleiche Vorzeichen besitzen, s gegen ∞ geht und die β_k mit zunehmendem Lag geometrisch abnehmen,⁷³ lässt sich (IV4: 105) in ein Modell der Form

$$Y_t = \gamma_0 + \beta_0 X_t + \gamma_1 Y_{t-1} + u_t \quad (\text{IV4: 106})$$

überführen.⁷⁴ Der sofortige Effekt einer Veränderung von X auf Y ist nach wie vor durch β_0 gegeben. Der verzögerte Effekt kommt nun aber durch γ_1 zum Ausdruck. Unter Gültigkeit der aufgeführten Annahmen können wir also in (IV4: 105) die verzögerten erklärenden Variablen durch die um eine Periode verzögerte erklärte Variable ersetzen und so den Problemen distributiver Lag-Modelle begegnen.

⁷² Vgl. Koyck, L.M. (1954).

⁷³ Dies bewirkt, dass Einflüsse der jüngsten Vergangenheit einen stärkeren Effekt auf Y haben, als weit in der Vergangenheit liegende, was durchaus plausibel ist.

⁷⁴ Hill, R.C., Griffiths, W.E., Judge, G.G. (2001) zeigen in Kapitel 15 diese Überführung formal auf. Sie bieten außerdem einen Überblick über andere Transformationsmöglichkeiten unter anderen Annahmen bezüglich der β_k .

Häufig werden dynamische Modelle mit verzögerten erklärten Variablen aber nicht so restriktiv formuliert wie bei der Koyck-Transformation. Ein typisches Beispiel dafür haben wir in (IV4: 59d) gesehen. Hier war nämlich zudem auch eine Variable X_{t-1} enthalten. (IV4: 59d) zeigt uns außerdem, dass nicht nur distributive Lag-Modelle dynamische Modelle mit verzögerten endogenen Variablen erforderlich machen können. Hier war es nämlich eine Autokorrelation des Störterms, die durch Hinzufügen von Y_{t-1} (und X_{t-1}) ins Modell berücksichtigt werden konnte.

Ein dynamisches Modell, das *nur* verzögerte erklärte Variablen beinhaltet, wird auch als **autoregressives Modell** bezeichnet. Typische Beispiele hierfür haben wir bereits im Kapitel zur Autokorrelation des Störterms bzw. als (IV4: 52), (IV4: 57) und (IV4: 58) kennengelernt. Wird ein autoregressives Modell noch um andere erklärende Variablen (oder deren Verzögerungen) ergänzt, sprechen wir von **autoregressiven distributiven Lag-Modellen** (ARDL-Modellen).

4.7.1.2 Problem der Autokorrelation in ARDL-Modellen

Zwar beseitigen dynamische Modelle mit verzögerten erklärten Variablen die Probleme distributiver Lag-Modelle oder berücksichtigen eine Autokorrelation des stochastischen Störterms, doch bringen sie eigene Schwierigkeiten mit sich, mit denen generell alle autoregressiven Modelle zu kämpfen haben. So führt Autokorrelation in ARDL-Modellen auch bei großen Stichproben zu einer *Verzerrung der OLS-Schätzer*, während die Schätzer in statischen Regressionsmodellen bei Autokorrelation unverzerrt bleiben. Diese Verzerrung beruht, wie wir unter IV 4.6.2.3 gesehen haben, auf einer durch die Autokorrelation verursachten Verletzung der Annahme 2b, da Y_{t-1} von u_{t-1} abhängt und u_t mit u_{t-1} korreliert ist.

Neben der Verzerrung der Schätzer der Regressionskoeffizienten führt Autokorrelation in einem ARDL-Modell außerdem zu *verzerrten Schätzungen der Standardfehler* und zu *"verzerrten" Residuen*. Ersteres führt zu Beeinträchtigung von Hypothesentests und Weiteres zu einer Untauglichkeit des Durbin-Watson d-Tests zur Aufdeckung von Autokorrelation, da dieser auf den Residuen basiert. Als Testverfahren für Autokorrelation in ARDL-Modellen ist daher z.B. der *Breusch-Godfrey LM-Test* (vgl. Abschnitt IV 4.5.3.3) zu verwenden.

Wurde Autokorrelation in einem ARDL-Modell diagnostiziert, haben wir allgemein zwei Möglichkeiten der weiteren Vorgehensweise. Zunächst empfiehlt es sich die *Modellspezifikation* noch einmal zu analysieren. Es könnte nämlich ein Fall von Autokorrelation vorliegen, die z.B. durch eine vernachlässigte Variable verursacht wird. Wie wir aber bereits gesehen haben, handelt es sich dabei um ein recht schwieriges Unterfangen. Können wir von einer korrekten Modellspezifikation ausgehen, so kann die unter IV 4.6.3 behandelte *Instrumentenvariablen schätzung* herangezogen werden. Können wir nämlich ein Instrument finden, das mit Y_{t-1} korreliert und gleichzeitig nicht mit u_t korreliert ist, so erhalten wir durch dieses im Rahmen der IV-Schätzung zumindest konsistente Schätzer. Wie wir gesehen haben, ist aber das Finden eines solchen Instruments ebenfalls nicht einfach.⁷⁵ Die unter

⁷⁵ In der Literatur wird für ein Modell der Form (IV4: 106), welches aus einer Koyck-Transformation resultiert, häufig die Verwendung von X_{t-1} als Instrument für Y_{t-1} vorgeschlagen. Vgl. z.B. Hill, R.C., Griffiths, W.E., Judge, G.G. (2001), Kapitel 15.

IV 4.5.4.1 behandelte *GLS-Schätzung* als Gegenmaßnahme bei Autokorrelation heranzuziehen führt zu inkonsistenten Schätzern.⁷⁶

4.7.2 Nichtstationäre Zeitreihen und Kointegration

Ein weiteres Problem bei der Arbeit mit Zeitreihendaten besteht darin, dass in Zeitreihenregressionen Koeffizienten signifikanter erscheinen können, als sie es eigentlich sind, wenn die zugehörigen erklärenden Variablen dem gleichen Trend wie die erklärte Variable unterliegen. Liegt eine derart starke, nicht kausale Beziehung zwischen Variablen vor, sprechen wir auch von **Scheinkorrelation**. Die Schätzung eines Regressionsmodells mit scheinkorrelierten Variablen wird auch als **Scheinregression** (engl. spurious regression) bezeichnet, deren Charakteristika überhöhte t-Werte und ein überschätzter Gesamtfit sind.⁷⁷ Ursache von Scheinkorrelation bei Zeitreihendaten sind meist sog. *nicht-stationäre Zeitreihen*.

4.7.2.1 Stationarität vs. Nicht-Stationarität

Zeitreihen werden allgemein als stationär bezeichnet, wenn sich ihre grundlegenden Eigenschaften wie z.B. Erwartungswert und Varianz im Zeitverlauf nicht verändern. Bei nicht-stationären Zeitreihen kommt es hingegen zu derartigen Veränderungen. Betrachten wir z.B. die Pro-Kopf-Produktion einer Volkswirtschaft, so werden wir hier von einer nicht-stationären Zeitreihe sprechen können, da die Pro-Kopf-Produktion typischerweise im Zeitverlauf zunimmt. Die durchschnittliche Wachstumsrate der Pro-Kopf-Produktion verändert sich jedoch häufig kaum im Zeitverlauf, kann also als stationär angesehen werden.

Formal können wir sagen, dass eine Zeitreihenvariable Y_t dann **stationär** ist, wenn für alle t und s

- der Erwartungswert der Variablen im Zeitverlauf konstant ist,

$$E(Y_t) = \mu \quad (\text{IV4: 107a})$$

- ihre Varianz im Zeitverlauf konstant ist

$$\text{Var}(Y_t) = \sigma_Y^2 \quad (\text{IV4: 107b})$$

- und die Kovarianz zwischen zwei Perioden t und $t-s$ nur vom Abstand s der Perioden und nicht vom Zeitpunkt t abhängt, zu dem die Kovarianz berechnet wurde.

$$\text{Cov}(Y_t, Y_{t-s}) = \gamma_s \quad (\text{IV4: 107c})$$

Einfach formuliert ist also eine Zeitreihe dann stationär, wenn ihr Erwartungswert, ihre Varianz und ihre Autokovarianz (für verschiedene Lags) gleich bleiben, egal zu welchem Zeitpunkt wir sie messen. Wird eine der Voraussetzungen nicht erfüllt, ist die Zeitreihenvariable Y_t **nicht stationär**. Anders ausgedrückt verändern sich also für eine nicht stationäre Zeitreihe Erwartungswert und Varianz im Zeitverlauf.

Auch wenn sich unsere Definitionen von Stationarität und Nicht-Stationarität auf Variablen beziehen, sei angemerkt, dass sie natürlich auch für Störerme gelten. So

⁷⁶ Vgl. Betancourt, R., Kelejian, H. (1981).

⁷⁷ Yule, G.U. (1926) entdeckte dieses Phänomen erstmals. Granger, C.W.J., Newbold, P. (1974) befassen sich intensiv mit dem Thema Scheinregression.

ist der stochastische Störterm des klassischen Regressionsmodells stationär, da er einen konstanten Erwartungswert von 0 und eine konstante Varianz σ^2 besitzt sowie nicht autokorreliert ist, d.h. die Autokovarianzen für alle Lags gleich 0 sind. Besitzt eine Variable allgemein diese drei spezifischen Eigenschaften, sagen wir, sie ist **Weißes Rauschen** (White Noise). Im Falle von Heteroskedastizität haben wir gesehen, dass der stochastische Störterm seine White Noise Eigenschaft und damit seine Stationarität verlieren kann. Hier haben wir nämlich festgestellt, dass sich die Varianz des Störterms im Zeitverlauf verändern kann. Ein heteroskedastischer Störterm ist also ein nicht stationärer Störterm.

4.7.2.2 Random Walks und Unit Roots

In der Empirie begegnet uns häufig ein ganz spezieller Typ nicht-stationärer Zeitreihen, der Random Walk. Wir unterscheiden allgemein zwei Typen von Random Walks: den Random Walk ohne Drift (ohne konstanten Term) und den Random Walk mit Drift (mit konstantem Term).

1. Random Walk ohne Drift

Nehmen wir an, u_t ist ein White Noise Störterm mit Erwartungswert 0 und Varianz σ^2 . Eine Zeitreihe Y_t folgt dann einem *Random Walk ohne Drift*, wenn

$$Y_t = Y_{t-1} + u_t \quad (\text{IV4: 108})$$

gilt. Der Wert von Y zum Zeitpunkt t hängt also von seinem Vorperiodenwert und einer rein zufälligen Komponente ab. Es liegt also ein AR(1)-Modell ohne Modellkonstante und Steigungsparameter von Eins vor. Da die Reihe zum Zeitpunkt 0 mit einem bestimmten Wert Y_0 beginnt, können wir aus (IV4: 108) auf

$$\begin{aligned} Y_1 &= Y_0 + u_1 \\ Y_2 &= Y_1 + u_2 = Y_0 + u_1 + u_2 \\ Y_3 &= Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3 \end{aligned}$$

und damit auf

$$Y_t = Y_0 + \sum u_t \quad (\text{IV4: 109})$$

schließen. Es gilt daher

$$E(Y_t) = E(Y_0 + \sum u_t) = Y_0 \quad (\text{IV4: 110})$$

und aufgrund der Unabhängigkeit der u_t

$$\text{Var}(Y_t) = \text{Var}(Y_0 + \sum u_t) = \text{Var}(\sum u_t) = \sum \text{Var}(u_t) = \sum \sigma^2 = t\sigma^2 \quad .78 \quad (\text{IV4: 111})$$

(IV4: 110) und (IV4: 111) zeigen klar, dass ein Random Walk ohne Drift zwar einen konstanten Erwartungswert, jedoch eine von t abhängende Varianz besitzt und damit *nicht stationär* ist.

Eine interessante Eigenschaft von Random Walks ist die Tatsache, dass sie ein *unbegrenztes Gedächtnis* bezüglich zufälliger Schocks besitzen. Wie (IV4: 109) zeigt, ist Y_t nämlich die Summe des Startwertes Y_0 plus der Summe zufälliger Einflüsse,

⁷⁸ Häufig wird Y_0 gleich Null gesetzt, sodass $E(Y_t) = 0$ gilt.

die sich bis t ereignet haben. Ist also z.B. $u_2 = 3$ an Stelle von $u_2 = 0$, so sind alle Y_t von Y_2 an 3 Einheiten höher. Wir sagen daher, dass der Schock dem Random Walk unbegrenzt im Gedächtnis bleibt.

Drücken wir nun (IV4: 108) als

$$Y_t - Y_{t-1} = \Delta Y_t = u_t \quad (\text{IV4: 112})$$

aus, wobei Δ die erste Differenz symbolisiert. Obwohl Y_t nicht stationär ist, lässt sich damit einfach zeigen, dass die *ersten Differenzen stationär* sind. Sie entsprechen ja gerade der White Noise Störgröße u_t und sind damit stationär.

2. Random Walk mit Drift

Fügen wir zu (IV4: 108) den sog. *Driftparameter* δ hinzu, so erhalten wir den sog. *Random Walk mit Drift* als

$$Y_t = \delta + Y_{t-1} + u_t . \quad (\text{IV4: 113})$$

Wir betrachten also wieder ein AR(1)-Modell, diesmal jedoch mit Konstante und einem Steigungsparameter von Eins. Daraus folgt wie bei (IV4: 109)

$$Y_t = Y_0 + t \cdot \delta + \sum u_t , \quad (\text{IV4: 114})$$

woraus wir

$$E(Y_t) = Y_0 + t \cdot \delta \quad (\text{IV4: 115})$$

$$\text{Var}(Y_t) = t \sigma^2 \quad (\text{IV4: 116})$$

erhalten. Dies erklärt die Bezeichnung "Drift" für die Konstante δ . Mit fortschreitender Zeit driftet nämlich der Erwartungswert nach oben ($\delta > 0$) oder nach unten ($\delta < 0$) vom Startwert Y_0 weg. Wir erkennen außerdem, dass nun neben der Varianz auch der Erwartungswert von Y_t von t abhängt und damit auch ein Random Walk mit Drift *nicht stationär* ist.

Wie auch beim Random Walk ohne Drift können wir beim Random Walk mit Drift durch die Bildung der *ersten Differenz* eine *stationäre Reihe* generieren. Es gilt nämlich

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t . \quad (\text{IV4: 117})$$

Da δ eine Konstante ist, ist der Erwartungswert von (IV4: 117) gleich δ und seine Varianz ist konstant. Es ist also Stationarität gegeben.

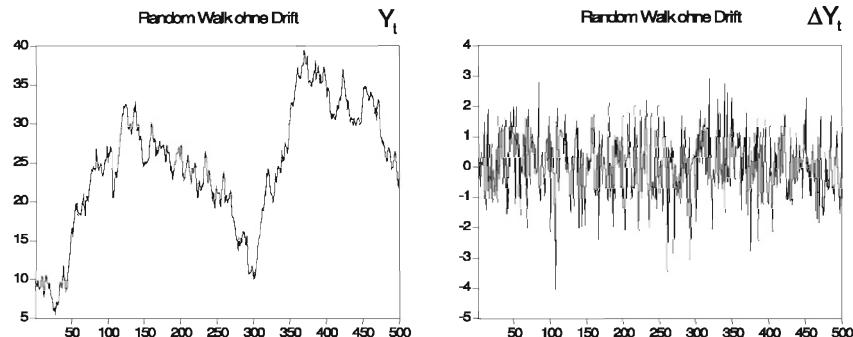
Beispiel IV 4-29

Random Walk Simulation

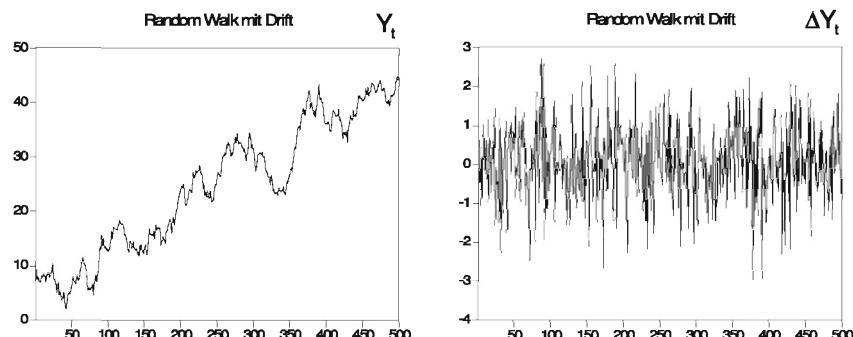


Zur besseren Veranschaulichung der beiden vorgestellten Random Walks führen wir zwei Simulationen durch und stellen die Ergebnisse grafisch dar. Wir lassen für den Störterm u_t von einem $N(0;1)$ -verteilten Zufallsgenerator 500 Beobachtungen erzeugen und bestimmen damit bei einem Startwert $Y_0 = 10$ nach (IV4: 108) die Werte Y_t des Random Walk ohne Drift. Für den Random Walk mit Drift gehen wir analog vor, wobei wir in (IV4: 113) einen Drift von $\delta = 0,1$ ansetzen. Wir erhalten damit die auf der Folgeseite dargestellten Reihen (Y_t links, ΔY_t rechts).

Wie sehr schön zu erkennen ist, fluktuieren die Veränderungen von Y_t beim Random Walk in nahezu konstanter Schwankungsbreite um den Wert Null, was auf einen Erwartungswert der Veränderungen von Null und eine konstante Varianz hindeutet. Dies zeigt, dass aus einem nicht-stationären Random Walk durch Differenzenbildung eine stationäre Reihe generiert werden kann. Zum dazugehörigen Verlauf von Y_t ist zu sagen, dass dieser durch die Kumulation der zufällig erzeugten Veränderungen entsteht und damit selbst ein Zufallsprozess ist. Das Bild, das sich hier zeigt, erinnert an typische Kursverläufe, wie wir sie aus Börsenzeitschriften oder den Nachrichten kennen. In der Tat wird von einer Vielzahl von Wissenschaftlern angenommen, dass Aktienkurse ungefähr Random Walks folgen. Dies wird damit begründet, dass der aktuelle Kurs stets alle auf dem Markt verfügbaren Informationen beinhaltet, da die Aktionäre schließlich auf deren Basis handeln und so den Kurs beeinflussen. Eine Kursveränderung kann damit nur durch neue, unerwartete Informationen entstehen, die in ihrer Natur stochastisch sind. Gerade daher ist die grafische Analyse der vergangenen Kursentwicklung (sog. Chart-Analyse), die in der Praxis betrieben wird, gänzlich untauglich, um daraus Schlüsse für die künftige Kursentwicklung zu ziehen. Ein Chart-Analyst würde bei der Betrachtung unseres Verlaufs von Y_t wohl am äußeren Rand einen Abwärtstrend (oder andere Muster, wie z.B. die Schulterformation) in die Grafik hineininterpretieren, der aber definitiv nicht vorhanden ist, da wir ja wissen, dass die Reihe rein zufällig erzeugt wurde.



Bei den ersten Differenzen des Random Walk mit Drift zeigt sich ein ähnliches Bild wie bei denen des Random Walk ohne Drift. Es ist eine konstante Varianz der Veränderungen zu erkennen. Diese fluktuieren jedoch um den Wert 0,1. Dies deutet auf die geltende erwartete Veränderung von $\delta = 0,1$ hin und erklärt auch, warum Y_t im Zeitverlauf ansteigt.



In der Literatur wird der Random Walk häufig *fälschlicherweise* alternativ mit dem Begriff **Unit Root** (dt. Einheitswurzel) bezeichnet. Genau genommen ist dieser aber ein Spezialfall einer Unit Root. Diesen Spezialfall haben wir durch Einsetzen von $\rho = 1$ (bzw. auch $\delta = 0$ beim Random Walk ohne Drift) ins AR(1)-Modell

$$Y_t = \delta + \rho Y_{t-1} + u_t$$

mit *White Noise Störterm* erhalten. Eine allgemeinere Klasse von Unit Roots erhalten wir, wenn der Störterm u_t bei ebenfalls $\rho = 1$ *nicht* White Noise ist. So wäre es z.B. denkbar, dass u_t z.B. selbst einem AR(1)-Modell folgt. In einem solchen Fall sind nicht alle Eigenschaften der Random Walks gültig, die wir in diesem Abschnitt hergeleitet haben. Die wichtigste Eigenschaft, dass der Wert von Y hochgradig mit vergangenen Werten von Y korreliert ist, bleibt jedoch erhalten. Dies ist die Kern-eigenschaft einer Unit Root.

4.7.2.3 Differenzstationarität vs. Trendstationarität

Bei der Analyse von Zeitreihen müssen wir zwischen zwei Arten von Trends unterscheiden: stochastischen Trends und deterministischen Trends. Ist der Trend einer Zeitreihe vollständig prognostizierbar und nicht variabel, so sprechen wir von einem *deterministischen Trend*. Beim deterministischen Trend sind die Abweichungen vom Trend stationär, d.h. es gibt immer wieder eine Tendenz zurück zum Trend. Bei einem *stochastischen Trend* gilt dies jedoch nicht. Hier sind die Abweichungen vom Trend nicht stationär.

Beim nicht stationären *Random Walk ohne Drift*, der einen *stochastischen Trend* aufweist, können wir feststellen, dass seine Veränderungen (IV4: 112) bzw.

$$Y_t - Y_{t-1} = \Delta Y_t = u_t$$

einen Erwartungswert von Null aufweisen und eine stationäre Reihe liefern, da u_t einen White Noise Prozess darstellt. Wir bezeichnen Zeitreihen, deren Differenzen stationär sind, auch als *differenzstationär*.

Auch der nicht stationäre *Random Walk mit Drift* weist einen *stochastischen Trend* auf. Seine Veränderungen (IV4: 117) bzw.

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t$$

sind im Erwartungswert von Null verschieden bzw. konkret δ . Dies bedeutet, dass die Variable Y_t einen positiven ($\delta > 0$) oder negativen ($\delta < 0$) Drift aufweist. Auch der Random Walk mit Drift ist *differenzstationär*, da Erwartungswert und Varianz der ΔY_t konstant sind.

Wird der Wert einer Variablen direkt von der fortschreitenden Zeit bestimmt, d.h. steigt oder fällt sie im Zeitverlauf, so sagen wir, sie besitzt einen *deterministischen Trend*. Eine solche Variable können wir durch ein Modell

$$Y_t = \delta + \gamma \cdot t + u_t \quad (\text{IV4: 118})$$

beschreiben, wobei u_t ein stationärer Störterm (z.B. White Noise oder AR(1)-Prozess mit $|\rho| < 1$) und t eine Variable ist, die aus den fortlaufenden natürlichen Zahlen besteht, die den Zeitbezug von Y beschreiben. Dies bedeutet, dass wir ein solches Modell typischerweise aus einem Datensatz der folgenden Struktur schätzen:

Realisation der erklären Variable Y	Y_1	Y_2	...	Y_n
Realisation der erklärenden Variable t	1	2	...	n

Der Erwartungswert von (IV4: 118) ist mit $E(Y_t) = \delta + \gamma \cdot t$ klar nicht konstant, die Varianz $\text{Var}(Y_t) = \sigma^2$ hingegen schon. Die Zeitreihe ist also *nicht stationär*. Subtrahieren wir $E(Y_t)$ von Y_t , ist die resultierende Reihe *stationär*, da sie nur noch durch den White Noise u_t bestimmt wird. Diese Prozedur der Beseitigung eines deterministischen Trends bezeichnen wir auch als *Trendbereinigung*. Da durch Trendbereinigung eine stationäre Reihe generiert werden kann, bezeichnen wir (IV4: 118) als *trendstationär*.

Differenzen- und Trendstationarität sind klar voneinander abzugrenzen. Es gilt nämlich folgendes:

1. Im Falle einer differenzstationären Zeitreihe führt eine Trendbereinigung nicht zu einer stationären Reihe, da in (IV4: 114) der Term Σu_t erhalten bleibt.
2. Im Falle eines deterministischen Trends führt Differenzenbildung (sog. Überdifferenzierung) ebenfalls zu schwerwiegenden Problemen.⁷⁹

Beispiel IV 4-30

Deterministischer vs. stochastischer Trend



Zur Veranschaulichung des Unterschieds zwischen deterministischen und stochastischen Trends betrachten wir die folgende Abbildung:



Die Reihe mit stochastischem Trend haben wir durch einen Random Walk mit Drift bzw. konkret $Y_t = 0,2 + Y_{t-1} + u_t$ unter Verwendung 100 zufällig aus einer Standardnormalverteilung gezogener Werte von u_t und eines Startwertes $Y_0 = 1$ erzeugt. Die Reihe mit deterministischem Trend entstammt der Gleichung $Y_t = 0,2t + u_t$ mit $t = 1, 2, \dots, 100$ und ebenfalls zufällig erzeugten Werten von u_t .

Wir zu erkennen ist, sind im Fall des *deterministischen Trends* die Abweichungen von der Trendlinie, die den *zeitvariablen Erwartungswert* repräsentiert, rein zufällig und entfernen sich nicht weit bzw. nur in einer gewissen Schwankungsbreite (*konstante Varianz*) von ihr. Wir sagen auch, dass sie keinen Einfluss auf die langfristige Entwicklung der Zeitreihe haben, die von der Trendkomponente 0,2t bestimmt wird. Im Falle des *stochastischen Trends* hingegen, ist es gerade die zufällige Komponente u_t , die den langfristigen Kurs der Zeitreihe Y_t mitbestimmt. Wir erkennen nicht die Tendenz, dass die Zeitreihe zu einem bestimmten Trend zurückkehrt. Was wir sehen, ist neben dem *zeitvariablen Erwartungswert* auch eine *zeitvariable Varianz*, da keine gleich bleibende Schwankungsbreite um eine Trendlinie zu beobachten ist.

⁷⁹ Der differenzierte Störterm stellt einen Moving-Average-Prozess (MA-Prozess) mit einer Unit Root dar. Zu MA-Prozessen vgl. z.B. Pindyck, R.S., Rubinfeld, D.L. (1998), Kapitel 17.

Kombinieren wir einen Random Walk mit einem deterministischen Trend, so erhalten wir ein Modell der Form

$$Y_t = \delta + Y_{t-1} + \gamma \cdot t + u_t . \quad (\text{IV4: 119})$$

Y_t ist klar *nicht stationär*. Selbst die ersten Differenzen

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + \gamma \cdot t + u_t \quad (\text{IV4: 120})$$

bilden hier keine stationäre Reihe, da ihr Erwartungswert von t abhängt. Eine *stationäre Reihe* können wir hier nur erhalten, wenn wir zusätzlich zur *Differenzierung* eine *Trendbereinigung* vornehmen.

Bevor wir im nächsten Abschnitt die Auswirkungen der Verwendung nicht stationärer Variablen in Regressionsmodellen aufzeigen, wollen wir noch einige Begrifflichkeiten klären, die im Zusammenhang mit Zeitreihenregressionen bzw. der Stationarität von Bedeutung sind: Wir nennen eine Zeitreihe allgemein **integriert mit Grad 1** oder kurz I(1), wenn sie nicht stationär ist, ihre ersten Differenzen jedoch stationär sind. Typische Beispiele für I(1)-Reihen sind daher der Random Walk mit und ohne Drift. Muss eine Zeitreihe zweimal differenziert werden bzw. müssen die ersten Differenzen nochmals differenziert werden, um eine stationäre Reihe zu erhalten, nennen wir sie **integriert mit Grad 2** bzw. I(2).⁸⁰ Führt also allgemein das d -malige Differenzieren einer nicht stationären Reihe Y_t zu einer stationären Reihe, nennen wir sie I(d) und schreiben $Y_t \sim I(d)$. Ist eine Zeitreihe Y_t bereits in ihrer Urform stationär (z.B. typischer White Noise), schreiben wir $Y_t \sim I(0)$.

4.7.2.4 Scheinregression und ihre Bekämpfung

Besonders bei makroökonomischen Zeitreihen stellt fehlende Stationarität bei Regressionen ein ernstes Problem dar.⁸¹ Führen wir eine Regression nämlich mit nicht stationären Variablen durch, so werden wir höchstwahrscheinlich hoch signifikante Parameterschätzungen erhalten, obwohl zwischen erklärter und erklärender Variable möglicherweise überhaupt kein Zusammenhang besteht.⁸² Wir sprechen dann von einer *Scheinregression*. Folgendes Beispiel zeigt dies auf:

Beispiel IV 4-31

Scheinregressionen



Simulieren wir die beiden Random Walks

$$X_t = X_{t-1} + u_t \quad Y_t = Y_{t-1} + v_t$$

indem wir zufällig jeweils 500 Beobachtungen von u_t und v_t aus einer Standardnormalverteilung ziehen und jeweils Startwerte von Null verwenden. Die resultierenden Zeitreihen sind nicht stationär bzw. genauer I(1). Da Y_t und X_t unabhängig voneinander zufällig erzeugt wurden, würden wir nun in einer Regression von Y_t auf X_t erwarten, dass das R^2 der Regression nicht signifikant ist. X_t sollte nämlich nicht in der Lage sein, Y_t zu erklären. Auch sollte der Parameter von X_t nicht signifikant sein. Wir beobachten nun aber folgende Ergebnisse:

⁸⁰ Ist Y_t also I(2), dann ist $\Delta \Delta Y_t = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$ stationär. Man beachte dabei, dass $\Delta \Delta Y_t = \Delta^2 Y_t \neq Y_t - Y_{t-2}$.

⁸¹ Vgl. z.B. Nelson, C.R., Plosser, C.I. (1982).

⁸² Außerdem sind die OLS-Schätzer der Koeffizienten autoregressiver Variablen nach unten verzerrt und die t-Statistiken selbst bei Gültigkeit von H_0 nicht t-verteil.

Dependent Variable: Y		Sample: 1 500		
Method: Least Squares		Included observations: 500		
	Coefficient	Std. Error	t-Statistic	Prob.
C	1.489786	0.392653	3.794156	0.0002
X	0.514560	0.020317	25.32610	0.0000
R-squared	0.562932	Mean dependent var		10.32061
Adjusted R-squared	0.562055	S.D. dependent var		6.100306
S.E. of regression	4.037027	Akaike info criterion		5.632886
Sum squared resid	8116.198	Schwarz criterion		5.649745
Log likelihood	-1406.222	F-statistic		641.4113
Durbin-Watson stat	0.069547	Prob(F-statistic)		0.000000

Der Koeffizient von X_t ist hoch signifikant und das nicht geringe R^2 ist signifikant von Null verschieden (p-Wert von Null bei der F-Statistik). Als typisch bei *stochastischen Trends* können wir außerdem einen sehr niedrigen Wert der Durbin-Watson d-Statistik feststellen. Wüssten wir nun nicht, dass die beiden Variablen in keinerlei Beziehung zueinander stehen, könnten wir in Versuchung geraten, diese Ergebnisse dahingehend zu interpretieren, dass eine signifikante statistische Beziehung zwischen Y_t und X_t besteht.

Führen wir eine analoge Simulation und Regression für die Random Walks mit Drift

$$X_t = 0,1 + X_{t-1} + u_t \quad Y_t = 0,1 + Y_{t-1} + v_t$$

durch, so erhalten wir das folgende Ergebnis:

Dependent Variable: Y		Sample: 1 500		
Method: Least Squares		Included observations: 500		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.794763	0.368401	-23.87277	0.0000
X	0.254990	0.006073	41.98535	0.0000
R-squared	0.779721	Mean dependent var		4.135194
Adjusted R-squared	0.779279	S.D. dependent var		9.622947
S.E. of regression	4.520955	Akaike info criterion		5.859316
Sum squared resid	10178.64	Schwarz criterion		5.876174
Log likelihood	-1462.829	F-statistic		1762.769
Durbin-Watson stat	0.050948	Prob(F-statistic)		0.000000

Wieder ist der Parameter der erklärenden Variablen hoch signifikant, obwohl zwischen den Variablen kein Zusammenhang besteht. Zudem erhalten wir nun auch ein hohes und signifikantes R^2 . Dies lässt sich damit begründen, dass durch den stochastischen Trend mit Drift beide Variablen eine ähnliche Entwicklungstendenz aufweisen. Wir stellen außerdem wieder einen sehr niedrigen Wert der Durbin-Watson d-Statistik fest.

Würden wir zwei Variablen mit ähnlichen deterministischen Trends aufeinander regressieren, ist R^2 sogar noch höher. Am besten erkennen wir dies durch Simulation der Variablen

$$X_t = 0,1t + u_t \quad Y_t = 0,2t + v_t$$

nach der bisher verwendeten Methodik. Regressieren wir Y_t auf X_t , erhalten wir das folgende Ergebnis, welche unsere R^2 -Behauptung bestätigt. Signifikanz ergibt sich hier nur aufgrund des ähnlichen deterministischen Trends. Wieder hängen die Variablen nämlich in keiner Weise voneinander ab. Die Betrachtung der Durbin-Watson d-Statistik zeigt nun ein anderes Bild. Bei deterministischen Trends zeigen sich keine derart niedrigen Werte, wie

wir sie bei den stochastischen Trends beobachtet hatten. u_t und v_t sind schließlich anders als bei den Random Walks nicht autokorriktiert.

Dependent Variable: Y		Sample: 1 500		
Method: Least Squares		Included observations: 500		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.076253	0.198653	-0.383853	0.7013
X	1.998653	0.006859	291.3911	0.0000
R-squared	0.994169	Mean dependent var		50.05015
Adjusted R-squared	0.994157	S.D. dependent var		29.06362
S.E. of regression	2.221539	Akaike info criterion		4.438269
Sum squared resid	2457.747	Schwarz criterion		4.455128
Log likelihood	-1107.567	F-statistic		84908.75
Durbin-Watson stat	1.982894	Prob(F-statistic)		0.000000

Nachdem wir diese fatalen Auswirkungen nicht stationärer Zeitreihen auf die Interpretation von Regressionsergebnissen kennengelernt haben, stellt sich nun die Frage, wie wir diesen begegnen können. Wir haben diese bereits indirekt in den vorhergehenden Kapiteln beantwortet. Differenzstationäre Reihen können wir durch Differenzieren in stationäre Reihen überführen⁸³ und trendstationäre durch Trendbereinigung. Im ersten Fall können wir das Modell dann in bestimmten Fällen einfach mit den Differenzreihen neu schätzen.⁸⁴ Der Vorgang der Trendbereinigung ist dagegen noch erklärbungsbedürftig:

Nehmen wir an, wir wollen ein Modell

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (\text{IV4: 121})$$

schätzen, in dem beide Variablen einem deterministischen Trend folgen. Wir haben gesehen, dass wir einen deterministischen Trend bei Y_t beseitigen können, wenn wir von $Y_t = \delta + \gamma \cdot t + u_t$ den zugehörigen Erwartungswert $E(Y_t) = \delta + \gamma \cdot t$ subtrahieren. In der praktischen Umsetzung bedeutet dies, dass wir trendbereinigte Werte \tilde{Y}_t dadurch erhalten, dass wir das Modell

$$Y_t = \delta + \gamma \cdot t + u_t \quad (\text{IV4: 122})$$

mit OLS schätzen und dann

$$\tilde{Y}_t = Y_t - \hat{\delta} - \hat{\gamma} \cdot t \quad (\text{IV4: 123})$$

berechnen. Da (IV4: 123) gerade den Residuen der OLS-Schätzung entspricht, können wir sagen, dass die Residuen trendbereinigte Größen darstellen, d.h. $\hat{u}_t = \tilde{Y}_t$.

Führen wir eine analoge Trendbereinigung auch für die erklärende Variable (bzw. mehrere erklärende Variablen) durch, können wir schließlich \tilde{Y}_t auf \tilde{X}_t regressieren, da diese ja stationär sind und wir keine verfälschten Ergebnisse erhalten.

⁸³ Da die GLS-Schätzung eine Quasidifferenzierung von im Modell verwendeten Zeitreihen darstellt, kann auch sie den Folgen einer Scheinregression etwas entgegenwirken.

⁸⁴ Eine solche Methodik ist bei ökonomischen Reihen nur zweckmäßig, wenn die Variablen nicht kointegriert sind, d.h. in keiner langfristigen Beziehung stehen (vgl. IV 4.7.2.6).

Alternativ zu dieser Vorgehensweise können wir in das Modell (IV4: 121) auch einfach die Trendvariable t in der Form

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 t + v_t \quad (\text{IV4: 124})$$

integrieren.⁸⁵ Die Schätzung des Parameters β_1 aus (IV4: 124) stimmt mit der Parameterschätzung überein, die wir aus der Regression von \tilde{Y}_t auf \tilde{X}_t erhalten. Eine Trendbereinigung (Bereinigung eines linearen deterministischen Trends) kann also einfach durch Hinzufügen einer Trendvariablen t in das Modell erfolgen.

Beispiel IV 4-32

Scheinregressionen - Gegenmaßnahmen



Regressieren wir für die Random Walks des vorhergehenden Beispiels IV 4-31 die ersten Differenzen ΔY_t auf die ersten Differenzen ΔX_t , so erhalten wir die im Folgenden abgedruckten Ergebnisse. Wir haben dabei in EViews die Funktion D(...) genutzt, die automatisch die ersten Differenzen der jeweiligen Reihe bildet.

Es sticht sofort ins Auge, dass die erklärenden Variablen in beiden Modellen nun nicht mehr signifikant sind, wie es auch sein sollte. Außerdem sind auch die Bestimmtheitsmaße der Regressionen nun sehr klein und nicht mehr signifikant. Die Werte der Durbin-Watson d-Statistik liegen nun nahe 2, da der stochastische Trend beseitigt ist.

Dependent Variable: D(Y)		Sample (adjusted): 2 500		
Method: Least Squares		Included observations: 499 after adjustments		
<i>Random Walk ohne Drift</i>	Coefficient	Std. Error	t-Statistic	Prob.
C	0.036594	0.044461	0.823063	0.4109
D(X)	-0.028704	0.043117	-0.665739	0.5059
R-squared	0.000891	Mean dependent var		0.036661
Adjusted R-squared	-0.001119	S.D. dependent var		0.992630
S.E. of regression	0.993186	Akaike info criterion		2.828202
Sum squared resid	490.2497	Schwarz criterion		2.845086
Log likelihood	-703.6364	F-statistic		0.443209
Durbin-Watson stat	1.982643	Prob(F-statistic)		0.505887

Dependent Variable: D(Y)		Sample (adjusted): 2 500		
Method: Least Squares		Included observations: 499 after adjustments		
<i>Random Walk mit Drift</i>	Coefficient	Std. Error	t-Statistic	Prob.
C	0.066019	0.044397	1.487006	0.1376
D(X)	-0.067548	0.042508	-1.589072	0.1127
R-squared	0.005055	Mean dependent var		0.050344
Adjusted R-squared	0.003053	S.D. dependent var		0.968445
S.E. of regression	0.966966	Akaike info criterion		2.774693
Sum squared resid	464.7063	Schwarz criterion		2.791577
Log likelihood	-690.2858	F-statistic		2.525149
Durbin-Watson stat	2.036095	Prob(F-statistic)		0.112680

⁸⁵ Genauso können wir auch bei Modellen mit mehreren erklärenden Variablen vorgehen.

Das Modell mit den zwei Variablen mit deterministischen Trends, welches wir im Beispiel IV 4-31 geschätzt hatten, wollen wir nun trendbereinigt schätzen. Wir fügen dazu eine Trendvariable mit den Werten 1,2,...,500 ins Modell ein und erhalten die folgenden "korrekten" Ergebnisse, die uns klar zeigen, dass der Koeffizient von X nicht signifikant ist und anders als die Variable t keinen Erklärungsbeitrag für Y leistet.

Dependent Variable: Y		Sample: 1 500		
Method: Least Squares		Included observations: 500		
Deterministischer Trend	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.307592	0.093363	-3.294586	0.0011
X	-0.022670	0.048214	-0.470202	0.6384
t	0.203299	0.004838	42.01758	0.0000
R-squared	0.998719	Mean dependent var		50.05015
Adjusted R-squared	0.998714	S.D. dependent var		29.06362
S.E. of regression	1.042261	Akaike info criterion		2.926644
Sum squared resid	539.8951	Schwarz criterion		2.951931
Log likelihood	-728.6610	F-statistic		193758.4
Durbin-Watson stat	1.950665	Prob(F-statistic)		0.000000

Würden wir eine derartige Trendbereinigung in einem Modell vornehmen, in dem die Variablen einem stochastischen und keinem deterministischen Trend unterliegen, würde trotzdem weiterhin ein signifikanter Zusammenhang zwischen den Variablen festgestellt werden. Auch wäre R^2 weiterhin hoch und der d-Wert niedrig. Dies können wir zeigen, indem wir in unser Modell aus den beiden Variablen mit Random Walks mit Drift eine Trendvariable mit den Werten 1,2,...,500 einfügen:

Dependent Variable: Y		Sample: 1 500		
Method: Least Squares		Included observations: 500		
Random Walk mit Drift	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.593064	0.432621	-12.92832	0.0000
X	0.670867	0.036966	18.14830	0.0000
t	-0.096966	0.008526	-11.37302	0.0000
R-squared	0.825210	Mean dependent var		4.135194
Adjusted R-squared	0.824507	S.D. dependent var		9.622947
S.E. of regression	4.031233	Akaike info criterion		5.632003
Sum squared resid	8076.666	Schwarz criterion		5.657291
Log likelihood	-1405.001	F-statistic		1173.210
Durbin-Watson stat	0.092475	Prob(F-statistic)		0.000000

4.7.2.5 Prüfung auf Stationarität

Nachdem wir nun die Auswirkungen fehlender Stationarität auf Regressionsergebnisse und auch bereits erste Ansätze kennengelernt haben, diesen zu begegnen, stellt sich die Frage, wie wir prüfen können, ob eine Zeitreihe stationär ist. Nur dann dürfen wir sie schließlich in unseren klassischen Regressionsmodellen verwenden.⁸⁶

⁸⁶ Im Abschnitt IV 4.7.2.5 werden wir sehen, dass wir unter gewissen Voraussetzungen auch mit nicht stationären Reihen schätzen können.

Wir haben bereits gesehen, dass uns eine grafische Darstellung der Zeitreihe erste Hinweise auf Nicht-Stationarität geben kann. Trends zeigen zeitvariable Erwartungswerte auf und Schwankungen der Zeitreihe in nicht konstanten Intervallen deuten auf zeitvariable Varianz hin. Handfestere Hinweise liefern uns die Autokorrelationsfunktion und sog. Dickey-Fuller-Tests.

1. Autokorrelationsfunktion und Korrelogramm

Ein einfacher Test auf Stationarität beruht auf der sog. *Autokorrelationsfunktion*. Sie ist definiert als

$$p_s = \frac{\text{Cov}(Y_t, Y_{t-s})}{\text{Var}(Y_t)}, \quad (\text{IV4: 125})$$

gibt also nichts anderes als den Autokorrelationskoeffizienten in Abhängigkeit der Verzögerung s an. Tragen wir s auf der x-Achse und p_s auf der Y-Achse ab, so erhalten wir das sog. *Korrelogramm*. Da wir die Autokorrelationskoeffizienten nicht kennen, müssen wir sie schätzen. Wir erhalten sie unter Verwendung erwartungstreuer Schätzer für Varianz und Autokovarianz zu

$$\hat{p}_s = \frac{\sum (Y_t - \bar{Y})(Y_{t-s} - \bar{Y})}{\frac{n-s}{n-1} \sum (Y_t - \bar{Y})^2}. \quad (\text{IV4: 126})$$

Die Berechnung der Autokorrelationsfunktion nimmt uns EViews ab. Wie wir sie konkret dazu benutzen können, eine Zeitreihe auf Stationarität zu überprüfen, wollen wir direkt anhand eines Beispiels veranschaulichen.

Beispiel IV 4-33

Autokorrelationsfunktion



In Beispiel IV-29 hatten wir einen Random Walk ohne Drift simuliert und diesen zusammen mit den aus ihm gewonnenen ersten Differenzen grafisch dargestellt. Wir wissen bereits, dass der Random Walk ohne Drift nicht stationär ist, seine ersten Differenzen aber schon. Versuchen wir dies nun anhand der Autokorrelationsfunktion aufzuzeigen.

Wählen wir die Zeitreihen in EViews an, können wir unter Views / Correlogram nach Angabe der Verzögerung s jeweils die Autokorrelationsfunktion von 1 bis s erzeugen. Wählen wir $s = 20$, so erhalten wir die auf der Folgeseite dargestellten Ergebnisse. In den Spalten AC ist die jeweilige Autokorrelationsfunktion bzw. sind die Autokorrelationskoeffizienten für die verschiedenen Lags enthalten. Die Balkengrafiken veranschaulichen die Autokorrelationsstruktur, wobei die vertikale Linie die Nullachse repräsentiert. Balken rechts davon veranschaulichen demnach positive, Balken links davon negative Autokorrelationskoeffizienten. Die Länge der Balken symbolisiert die Höhe der Autokorrelation.

Betrachten wir zunächst das Korrelogramm der ersten Differenzen. Wir erkennen für den simulierten White Noise Prozess, der sich dahinter verbirgt, ausschließlich Autokorrelationskoeffizienten um Null. Dies ist das typische Bild des Korrelogramms eines White Noise Prozesses. Ähnelt das Korrelogramm einer empirisch beobachteten Zeitreihe also dem einer White Noise Reihe, kann sie als stationär betrachtet werden.

Das Korrelogramm des Random Walk zeigt ein völlig entgegengesetztes Bild. Hier haben wir sehr hohe Autokorrelationskoeffizienten. Selbst bei hohen Lags ist die Autokorrelation noch sehr hoch. Das hier zu sehende Bild ist typisch für eine nicht stationäre Zeitreihe: Die Autokorrelationskoeffizienten beginnen bei einem hohen Wert und sinken kaum mit zunehmendem Lag.

Random Walk ohne Drift Y_t :

Sample: 1 500
Included observations: 500

Autocorrelation	AC	Q-Stat	Prob
	1 0.988	490.73	0.000
	2 0.974	969.06	0.000
	3 0.961	1435.2	0.000
	4 0.949	1890.7	0.000
	5 0.936	2334.6	0.000
	6 0.923	2767.7	0.000
	7 0.910	3189.6	0.000
	8 0.896	3599.5	0.000
	9 0.884	3998.8	0.000
	10 0.870	4386.7	0.000
	11 0.856	4762.8	0.000
	12 0.842	5127.2	0.000
	13 0.830	5481.9	0.000
	14 0.819	5828.2	0.000
	15 0.807	6165.6	0.000
	16 0.796	6493.8	0.000
	17 0.783	6812.2	0.000
	18 0.770	7121.0	0.000
	19 0.758	7420.5	0.000
	20 0.744	7710.3	0.000

Zugehörige erste Differenzen ΔY_t :

Sample: 1 500
Included observations: 500

Autocorrelation	AC	Q-Stat	Prob
b	1 0.069	2.4085	0.121
	2 -0.017	2.5548	0.279
	3 -0.054	4.0400	0.257
	4 0.040	4.8426	0.304
	5 -0.047	5.9625	0.310
	6 0.030	6.4329	0.376
	7 0.062	8.4098	0.298
	8 -0.041	9.2557	0.321
	9 0.044	10.231	0.332
	10 0.009	10.271	0.417
	11 -0.035	10.892	0.452
	12 -0.088	14.888	0.248
	13 -0.042	15.788	0.261
	14 0.027	16.173	0.303
	15 0.046	17.256	0.304
	16 0.016	17.391	0.361
	17 -0.008	17.428	0.426
	18 -0.032	17.948	0.459
	19 0.065	20.122	0.387
p	20 0.081	23.541	0.263

Die Betrachtung geschätzter Autokorrelationskoeffizienten ist nun aber nicht ausreichend, um die White Noise Eigenschaft einer Zeitreihe zu analysieren. Wir müssen wissen, ob sie auch signifikant sind. Dazu können wir uns der Tatsache bedienen, dass wenn eine Zeitreihe White Noise ist, die Schätzungen der Autokorrelationskoeffizienten

$$\hat{\rho}_k \stackrel{d}{\sim} N(0; 1/n) \quad (\text{IV4: 127})$$

sind.⁸⁷ Dies bedeutet, dass die Autokorrelationskoeffizienten in *großen Stichproben* mit einem Erwartungswert von Null und einer Varianz von 1 dividiert durch den Stichprobenumfang n approximativ normalverteilt sind. Damit können wir eine Z-Statistik

$$Z = \frac{\hat{\rho}_k - 0}{\sqrt{1/n}} \quad (\text{IV4: 128})$$

berechnen und den resultierenden Wert mit dem kritischen Wert aus der Standardnormalverteilung vergleichen. Ist der Absolutbetrag der Teststatistik größer als der kritische Wert, können wir $H_0: \rho_k = 0$ ablehnen, andernfalls nicht.

Anstatt die Signifikanz einzelner Autokorrelationskoeffizienten zu testen, können wir die Hypothese testen, dass alle ρ_k bis zu einem bestimmten Lag k gleichzeitig gleich Null sind. Dazu können wir die Q-Statistik

$$Q = n \sum_{i=1}^k \hat{\rho}_i^2 \quad (\text{IV4: 129})$$

⁸⁷ Vgl. Bartlett, M.S. (1964).

nutzen,⁸⁸ in der n wieder den Stichprobenumfang und k die Anzahl der Autokorrelationskoeffizienten darstellt, die einbezogen werden sollen. Die Q-Statistik wird verwendet, um zu testen, ob eine Zeitreihe White Noise ist. In *großen Stichproben* ist sie mit k Freiheitsgraden approximativ Chi-Quadrat-verteilt. Überschreitet Q den kritischen Wert aus der Chi-Quadrat-Verteilung, so kann die Nullhypothese $H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$ gegenüber $H_A: "mindestens ein \rho_s von Null verschieden"$ abgelehnt werden.

Eine Variante der Q-Statistik ist die sog. **Lijung-Box-Statistik**, die als

$$LB = n(n+2) \sum_{s=1}^k \frac{\hat{\rho}_s^2}{n-k} \quad (\text{IV4: 130})$$

definiert ist⁸⁹ und in großen Stichproben ebenfalls approximativ einer Chi-Quadrat-Verteilung mit k Freiheitsgraden folgt. Es hat sich gezeigt, dass diese auch in *kleinen Stichproben* vorteilhafte Eigenschaften aufweist, weshalb sie in einem solchen Fall der Q-Statistik vorgezogen wird.

Beispiel IV 4-34

Signifikanz von Autokorrelationskoeffizienten

Betrachten wir das Korrelogramm des Random Walk ohne Drift aus unserem letzten Beispiel. Nehmen wir an, wir interessieren uns für die Signifikanz des Autokorrelationskoeffizienten der Ordnung 20. Wir können bei einem Signifikanzniveau von 5 % mit resultierendem kritischen Wert $z_{0,975} = 1,96$ (zweiseitiger Test) feststellen, dass

$$\left(\frac{0,744 - 0}{\sqrt{1/500}} = 16,64 \right) > 1,96$$

gilt. Wir können also $H_0: \rho_{20} = 0$ klar ablehnen.

Wollen wir jedoch wissen, ob alle Autokorrelationskoeffizienten bis Lag 20 gemeinsam signifikant von Null verschieden sind, bedienen wir uns der sehr hohen Q-Statistik 7.710,3 (vgl. Tabelle in Beispiel IV 4-33), die eine Ablehnung von H_0 offensichtlich macht. Alternativ könnten wir auch den zugehörigen p-Wert verwenden. Er liegt bei Null, sodass wir bei einem Signifikanzniveau von 5 % die Nullhypothese der Form $H_0: \rho_1 = \dots = \rho_{20} = 0$ klar ablehnen können.

Der Random Walk ohne Drift des letzten Beispiels ist durchgehend durch p-Werte zur Q-Statistik von Null gekennzeichnet. Bei seinen ersten Differenzen zeigen sich relativ hohe p-Werte, die es uns nicht erlauben H_0 abzulehnen, sodass wir nicht auf signifikante Autokorrelationen schließen können.

2. Dickey-Fuller-Tests

Ausgangspunkt des Dickey-Fuller-Tests⁹⁰ ist die Gleichung

$$Y_t = \rho Y_{t-1} + u_t, \quad (\text{IV4: 131})$$

in der u_t ein White Noise Störterm ist. Wie wir wissen, erhalten wir für $\rho = 1$ einen Random Walk ohne Drift, also eine nicht stationäre Reihe. Ist jedoch $|\rho| < 1$, so

⁸⁸ Vgl. Box, G.E.P., Pierce, D.A. (1970).

⁸⁹ Vgl. Lijung, G.M., Box, G.E.P. (1978).

⁹⁰ Vgl. Dickey, D.A., Fuller, W.A. (1979, 1981). Für einen Überblick über weitere Testverfahren vgl. z.B. Campbell, J.Y., Perron, P. (1991) sowie Wolters, J., Hassler, U. (2006).

erhalten wir eine stationäre Zeitreihe.⁹¹ Die Nullhypothese der Nicht-Stationarität kann also als $H_0: \rho = 1$ formuliert werden. Die Alternativhypothese der Stationarität lautet in der Praxis meist $H_A: \rho < 1$.⁹²

Aus Vereinfachungsgründen wird der Dickey-Fuller-Test in der Praxis jedoch nicht direkt mittels (IV4: 131) durchgeführt, sondern diese Gleichung zunächst über

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\ \Delta Y_t &= (\rho - 1)Y_{t-1} + u_t \end{aligned}$$

modifiziert, was mit $\vartheta = (\rho - 1)$ folgende Darstellung zulässt:

$$\Delta Y_t = \vartheta Y_{t-1} + u_t \quad (\text{IV4: 132})$$

$\vartheta = 0$ gilt nun, wenn $\rho = 1$, d.h. Y_t eine nicht stationäre Reihe ist. $\vartheta < 0$ erhalten wir, wenn $\rho < 1$, d.h. Y_t stationär ist. Um die Stationarität von Y_t zu überprüfen, können wir also Gleichung (IV4: 132) mit OLS schätzen und das Hypothesenpaar

$H_0: \vartheta = 0$	(Nicht-Stationarität)
$H_A: \vartheta < 0$	(Stationarität)

(IV4: 133)

testen. Eine Ablehnung von H_0 bedeutet, dass die Zeitreihe mit $E(Y_t) = 0$ *stationär* ist. Da gerade bei ökonomischen Zeitreihen ein Erwartungswert von Null unrealistisch ist, macht eine solche Alternativhypothese nicht immer Sinn. Deswegen haben Dickey und Fuller noch weitere Testgleichungen vorgeschlagen:

Erweitern wir (IV4: 132) um eine Driftkomponente δ bzw. betrachten als Ausgangspunkt in (IV4: 131) einen Random Walk mit Drift, erhalten wir die Testgleichung

$$\Delta Y_t = \delta + \vartheta Y_{t-1} + u_t. \quad (\text{IV4: 134})$$

Prüfen wir mittels dieser Gleichung das Hypothesenpaar (IV4: 133), so bedeutet eine Ablehnung von H_0 nun, dass die Zeitreihe mit dem von Null verschiedenen Erwartungswert $E(Y_t) = \delta/(1 - \rho)$ *stationär* ist.

Wird in (IV4: 134) nun noch ein deterministischer Trend integriert, erhalten wir als letzte Testgleichung

$$\Delta Y_t = \delta + \vartheta Y_{t-1} + \gamma t + u_t. \quad (\text{IV4: 135})$$

Bei einer Ablehnung von H_0 sagen wir hier, dass die Zeitreihe *trendstationär* ist. Dies bedeutet, dass wir sie erst nach Bereinigung des deterministischen Trends als wirklich stationär bezeichnen können.

Die konkrete Durchführung eines Dickey-Fuller-Tests ähnelt nun einem t-Test des Koeffizienten ϑ der verzögerten Variablen Y_{t-1} . Die Teststatistik des Dickey-Fuller-Tests, die auch als Tau- bzw. t -Statistik bezeichnet wird, ist also die des t-Tests, d.h.

⁹¹ Dies können wir uns klar machen, wenn wir unter der Bedingung $|\rho| < 1$ analog zu unserer Vorgehensweise in Abschnitt IV 4.5.1 den Erwartungswert und die Varianz von (IV4: 131) bestimmen. Wir erhalten $E(Y_t) = 0$ und $\text{Var}(Y_t) = 1/(1-\rho^2)$. Da diese Konstanten sind und die Autokovarianz nur von s abhängt, ist Y_t stationär.

⁹² Die Alternative $\rho > 1$ wird in der Regel nicht betrachtet, da diese bedeuten würde, dass Y_t "explosiv" wächst.

$$\tau = \frac{\hat{\vartheta}}{\hat{\sigma}_{\hat{\vartheta}}}. \quad (\text{IV4: 136})$$

Bei korrekter Nullhypothese folgt τ nun aber keiner konventionellen t-Verteilung mehr, sodass wir für unsere Testentscheidung nicht auf die klassischen kritischen t-Werte zurückgreifen können. Stattdessen sind abhängig von der verwendeten Testgleichung (IV4: 132), (IV4: 134) und (IV4: 135) spezielle von Dickey und Fuller bestimmte kritische Werte zu verwenden. Die nachfolgende Tabelle zeigt typische kritische Werte des Dickey-Fuller Tests für *große Stichproben* im Vergleich zu den klassischen t-Werten.⁹³ In der praktischen Anwendung liefern Ökonometriesoftwarepakete die notwendigen kritischen Werte automatisch, weshalb wir hier auf weitere Tabellierung verzichten.

Modell	Interpretation, wenn $\vartheta = 0$	α		
		1 %	5 %	10 %
$\Delta Y_t = \vartheta Y_{t-1} + u_t$	Y_t folgt einem Random Walk ohne Drift	-2,58	-1,95	-1,62
$\Delta Y_t = \delta + \vartheta Y_{t-1} + u_t$	Y_t folgt einem Random Walk mit Drift	-3,43	-2,86	-2,57
$\Delta Y_t = \delta + \vartheta Y_{t-1} + \gamma t + u_t$	Y_t folgt einem Random Walk mit Drift und deterministischem Trend	-3,96	-3,41	-3,12
Klassische t-Werte		-2,33	-1,65	-1,28

Ist der Wert der t -Statistik kleiner als der kritische Wert, können wir H_0 ablehnen und damit folgern, dass Y_t in den Fällen (IV4: 132) und (IV4: 134) stationär und im Fall (IV4: 135) trendstationär ist.

Besonders wichtig ist bei der Durchführung des Dickey-Fuller-Tests die korrekte Wahl der Testgleichung. Wählen wir die falsche, verursachen wir eine Fehlspezifikation, deren Konsequenzen uns bereits aus IV 4.1 und 4.2 bekannt sind. Wir haben bereits erwähnt, dass (IV4: 132) aufgrund des fehlenden Drifts für ökonomische Zeitreihen meist unrealistisch ist. In der Praxis sind daher eher (IV4: 134) und (IV4: 135) relevant. Welchen Einfluss die Wahl der Testgleichung auf die Testergebnisse hat, wollen wir im nachfolgenden Beispiel im Detail betrachten.

Beispiel IV 4-35

Dickey-Fuller-Test

Simulieren wir in der gewohnten Art und Weise durch zufälliges Ziehen von Störtermrealisierungen aus einer $N(0;1)$ -Verteilung die Variable Y_t , die einem deterministischen Trend der Form

$$Y_t = 0,1 + 0,2t + u_t$$

folgt. Ein Test auf Stationarität müsste nun Trendstationarität der entstandenen Reihe feststellen, d.h. H_0 ablehnen. Wir wissen bereits, dass aufgrund unserer Simulation nur Glei-

⁹³ Modifiziert entnommen aus MacKinnon, J.G. (1991).

chung (IV4: 135) dazu in der Lage sein kann. Testen wir nun dennoch mittels aller drei möglichen Testgleichungen auf Stationarität. Es ist dazu anzumerken, dass der Dickey-Fuller-Test in EViews als Spezialfall des sog. Augmented Dickey-Fuller-Tests (vgl. Ausführungen der Folgeseite) zu behandeln ist. Es ist also die Zeitreihe anzuwählen, unter View / Unit Root Test der Augmented Dickey-Fuller-Test zu selektieren und die Lag-Länge benutzerdefiniert auf Null zu setzen. Nach Auswahl der Form der Testgleichung erhalten wir dann die folgenden (gekürzt dargestellten) Ergebnisse:

Null Hypothesis: Y has a unit root
Exogenous: **None** Lag Length: 0 (Fixed)

	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	2.177890	0.9933
Test critical values:		
1% level	-2.569604	
5% level	-1.941459	
10% level	-1.616273	

Null Hypothesis: Y has a unit root
Exogenous: **Constant** Lag Length: 0 (Fixed)

	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	-0.628667	0.8612
Test critical values:		
1% level	-3.443228	
5% level	-2.867112	
10% level	-2.569800	

Null Hypothesis: Y has a unit root
Exogenous: **Constant, Linear Trend** Lag Length: 0 (Fixed)

	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	-23.00531	0.0000
Test critical values:		
1% level	-3.976554	
5% level	-3.418852	
10% level	-3.131965	

Wie zu erwarten war, sind die Tests auf Basis von (IV4: 132) und (IV4: 134) nicht in der Lage die Trendstationarität aufzuzeigen. Die sich aus den speziellen kritischen Werten ergebenden p-Werte sind sehr hoch, d.h. wir können H_0 bei einem Signifikanzniveau von 5 % nicht ablehnen. Der Test auf Basis von (IV4: 135) liefert jedoch einen p-Wert von Null. Wir können H_0 klar ablehnen, was auf Trendstationarität hinweist.

Besonders wichtig ist, dass die Test-Gleichungen und kritischen Werte des Dickey-Fuller-Tests der Annahme unterliegen, dass die Störterme nicht autokorreliert sind. Für den Fall *autokorrelierter Störterme* haben Dickey und Fuller die Testgleichungen durch Einbeziehung verzögerter erster Differenzen der erklärenen Variablen modifiziert. Werden diese modifizierten Gleichungen verwendet, sprechen wir auch vom **ADF-Test** (engl. augmented Dickey-Fuller-test). Die drei möglichen Formen des ADF-Tests sind

$$\Delta Y_t = \delta Y_{t-1} + \sum_{i=1}^s \phi_i \Delta Y_{t-i} + u_t \quad (\text{IV4: 137})$$

$$\Delta Y_t = \delta + \delta Y_{t-1} + \sum_{i=1}^s \phi_i \Delta Y_{t-i} + u_t \quad (\text{IV4: 138})$$

$$\Delta Y_t = \delta + \gamma t + \delta Y_{t-1} + \sum_{i=1}^s \phi_i \Delta Y_{t-i} + u_t, \quad (\text{IV4: 139})$$

wobei $\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$, $\Delta Y_{t-2} = Y_{t-2} - Y_{t-3}$, etc. gilt. Die Lag-Anzahl s wird dabei in der Regel über das Akaike- oder das Schwarz-Informationskriterium bestimmt. Sinnvoll ist es zudem für die gewählte Lag-Anzahl zu prüfen, ob die Residuen der ADF-Regression autokorriert sind oder nicht (z.B. mit dem Breusch-Godfrey LM-Test). Sind sie dies nicht, ist die richtige Lag-Zahl gefunden. Zur Testentscheidung werden auch hier der Wert der t -Statistik und die speziellen kritischen Werte herangezogen.

Beispiel IV 4-36

Ist der Goldpreis stationär?

Testen wir die Zeitreihe des Feingoldpreises in Euro je kg von Juni 1986 bis April 2009 mit insgesamt 491 Preisbeobachtungen aus der Zeitreihendatenbank der Deutschen Bundesbank auf Stationarität. Bereits ein erster Blick auf die grafische Darstellung weist auf fehlende Stationarität hin, da weder Erwartungswert noch Varianz im Zeitverlauf konstant zu sein scheinen.



In der Literatur wird häufig vorgeschlagen, Rohstoffpreise mittels der Testgleichungen auf Stationarität zu testen, die auf dem Random Walk mit Drift oder dem Random Walk mit Drift und deterministischem Trend beruhen.⁹⁴ Wir entscheiden uns zunächst für die erstgenannte Möglichkeit und berücksichtigen außerdem eine evtl. Autokorrelation des Störterms, d.h. wir verwenden Testgleichung (IV4: 138) des ADF-Tests. Zur Vereinfachung bedienen wir uns einer automatischen Lag-Bestimmung mittels des Akaike-Informationskriteriums durch EViews. Wir erhalten damit das auf der Folgeseite (nun vollständig) abgedruckte Testergebnis.

Der Regressionsoutput verdeutlicht, dass die ADF-Teststatistik genau den Wert der t -Statistik für den Koeffizienten von Y_{t-1} in der ADF-Testgleichung annimmt. Durch das Akaike-Informationskriterium hat EViews für die verzögerten Differenzen eine Lag-Anzahl s von 15 bestimmt und damit der Autokorrelation des Störterms Rechnung getragen. Aufgrund des hohen p -Werts von 0,7081 können wir die Nullhypothese der Nicht-Stationarität

⁹⁴ Vgl. z.B. Pindyck, R.S., Rubinfeld, D.L. (1998), Kapitel 16.

bei einem Signifikanzniveau von 5 % nicht ablehnen. Es lässt sich also keine Stationarität vermuten. Selbst unter Einbeziehung eines deterministischen Trends erhalten wir einen p-Wert von 0,8165. Es kann also auch nicht auf Trendstationarität geschlossen werden.

Null Hypothesis: Y has a unit root

Exogenous: Constant

Lag Length: 14 (Automatic based on AIC, MAXLAG=15)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.122794	0.7081
Test critical values:		
1% level	-3.443863	
5% level	-2.867392	
10% level	-2.569950	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(Y)

Method: Least Squares

Sample (adjusted): 1969M09 2009M04

Included observations: 476 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	-0.007905	0.007040	-1.122794	0.2621
D(Y(-1))	-0.034874	0.046520	-0.749653	0.4538
D(Y(-2))	-0.044773	0.046840	-0.955890	0.3396
D(Y(-3))	-0.074971	0.047013	-1.594673	0.1115
D(Y(-4))	0.057077	0.048084	1.187040	0.2358
D(Y(-5))	0.099042	0.048138	2.057455	0.0402
D(Y(-6))	-0.074008	0.048737	-1.518507	0.1296
D(Y(-7))	0.039359	0.049157	0.800684	0.4237
D(Y(-8))	0.072273	0.049399	1.463035	0.1441
D(Y(-9))	0.106529	0.049411	2.155982	0.0316
D(Y(-10))	0.009844	0.049532	0.198739	0.8426
D(Y(-11))	0.100366	0.049514	2.027039	0.0432
D(Y(-12))	0.019737	0.049653	0.397502	0.6912
D(Y(-13))	-0.030692	0.049897	-0.615114	0.5388
D(Y(-14))	-0.119526	0.050275	-2.377419	0.0178
C	120.2412	78.35573	1.534555	0.1256
R-squared	0.064094	Mean dependent var	39.86970	
Adjusted R-squared	0.033576	S.D. dependent var	675.4479	
S.E. of regression	664.0117	Akaike info criterion	15.86751	
Sum squared resid	2.03E+08	Schwarz criterion	16.00753	
Log likelihood	-3760.468	F-statistic	2.100174	
Durbin-Watson stat	1.999847	Prob(F-statistic)	0.009066	

EViews bietet uns die komfortable Möglichkeit den ADF-Test auch für die ersten (und auch zweiten) Differenzen einer Variablen durch einfachen Mausklick durchzuführen. Führen wir diesen Test mit der auf dem Random Walk mit Drift basierenden Testgleichung und automatischer Lag-Bestimmung durch EViews durch, liefert der ADF-Test einen p-Wert von Null, mit dem wir die Stationarität der ersten Differenzen des Goldpreises bei jedem Signifikanzniveau aufzeigen können.

4.7.2.6 Kointegration und Fehlerkorrekturmodell

In unseren bisherigen Ausführungen haben wir gesehen, dass nicht stationäre Zeitreihen nicht (in ihrer Urform) in Regressionen verwendet werden sollten, um das Problem der Scheinregression zu vermeiden. Zu dieser Regel gibt es jedoch eine Ausnahme. Betrachten wir dazu das Regressionsmodell

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad (\text{IV4: 140})$$

in dem wir mittels des Dickey-Fuller-Tests oder des ADF-Tests festgestellt haben, dass sowohl $Y_t \sim I(1)$ als auch $X_t \sim I(1)$ gilt. Obwohl eine Linearkombination von Y und X,⁹⁵ wie z.B.

$$\varepsilon_t = Y_t - \beta_0 - \beta_1 X_t, \quad (\text{IV4: 141})$$

in der Regel ebenfalls $I(1)$ ist, besteht die Möglichkeit, dass (IV4: 141) $I(0)$ ist. Trotz der Tatsache, dass Y und X stochastische Trends aufweisen, kann ihre Linearkombination (IV4: 141) also stationär sein. Durch die Linearkombination heben sich die stochastischen Trends der Reihen sozusagen auf. Nehmen wir z.B. an, uns liegen die $I(1)$ -Variablen Konsum und Einkommen vor, so könnte die Ersparnis (Einkommen abzüglich Konsum) eine stationäre Variable sein. In einem solchen Fall ist eine Regression von Konsum auf Einkommen sinnvoll und würde zu keiner Scheinregression führen. Eine stationäre Linearkombination impliziert nämlich auch, dass die Variablen Y und X einem ähnlichen stochastischen Trend folgen, sich also nie all zu weit voneinander entfernen. Ökonomisch bedeutet dies, dass zwischen den Variablen eine *langfristige Beziehung* (oder Gleichgewichtsbeziehung) besteht.

In der hier geschilderten Konstellation nennen wir Y und X **kointegriert**,⁹⁶ bezeichnen (IV4: 140) als *Kointegrationsregression* und β_1 als *Kointegrationsparameter*. Die einfachste Möglichkeit zu testen, ob die Variablen Y und X kointegriert sind, ist die Störterme (IV4: 141) der Kointegrationsregression mittels eines Dickey-Fuller-Tests oder ADF-Tests auf Stationarität zu testen.⁹⁷ Da wir diese jedoch nicht beobachten können, prüfen wir stattdessen die Residuen $e_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t$. Sind die Residuen stationär, sind Y und X kointegriert. Sind sie nicht stationär, sind Y und X nicht kointegriert und ihre Regression aufeinander wäre eine typische Scheinregression.

Der Test auf Stationarität der Residuen basiert auf der Testgleichung

$$\Delta e_t = \vartheta e_{t-1} + u_t, \quad (\text{IV4: 142})$$

⁹⁵ Eine Linearkombination von Y und X ist eine neue Variable $Z = \alpha_0 + \alpha_1 X + \alpha_2 Y$. Hier erhalten wir (IV4: 141), indem wir $\alpha_0 = -\beta_0$, $\alpha_1 = -\beta_1$ und $\alpha_2 = 1$ setzen und die Reihe Z mit ε bezeichnen.

⁹⁶ Wäre eine der Variablen $I(1)$ und die andere $I(0)$, könnten sie nicht kointegriert sein. Eine stationäre Variable lässt sich schließlich nicht durch eine nicht stationäre erklären und umgekehrt, weshalb es auch keine langfristige Beziehung zwischen ihnen geben kann.

⁹⁷ Einen Überblick über Kointegrationstests bei Modellen mit mehreren erklärenden Variablen bietet z.B. Kennedy, P. (2003). Zu den wichtigsten modernen Verfahren zählt der Test von Johansen, S. (1991, 1995). Er besitzt anders als das im Folgenden vorgestellte Verfahren nach Engle, R.F., Granger, C.W.J. (1987) den Vorteil, auch auf mehr als eine Kointegrationsbeziehung unter den Variablen testen zu können.

die wieder Terme der Art Δe_{t-1} , Δe_{t-2} , etc. auf der rechten Gleichungsseite enthalten kann, um Autokorrelation von e_t zu begegnen. Von besonderer Bedeutung ist, dass (IV4: 142) *keine Modellkonstante* enthalten darf, da der Mittelwert der Residuen bei Null liegt. Unser Hauptaugenmerk bei der Testdurchführung gilt wieder dem Wert der t -Statistik für den Parameter ϑ . Da der Test auf den Residuen der Kointegrationsregression beruht, sind der Dickey-Fuller Test und der ADF-Test in Richtung eines Auffindens von Kointegration verzerrt. Um dieses Problem zu umgehen, sind nicht die von Dickey und Fuller gelieferten kritischen Werte des Dickey-Fuller-Tests, sondern *spezielle von Engle und Granger tabellierte kritische Werte* für die Testentscheidung zu verwenden.⁹⁸ Die nachfolgende Tabelle zeigt typische kritische Werte für *große Stichproben*.⁹⁹ Je nachdem, welche Form das Ursprungsmodell (IV4: 140) aufweist, aus dessen Schätzung die Residuen stammen, sind also andere kritische Werte erforderlich.

Modell	Beschreibung	α		
		1 %	5 %	10 %
$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$	Standardmodell	-3,96	-3,37	-3,07
$Y_t = \beta_0 + \beta_1 X_t + \beta_2 t + \varepsilon_t$	Standardmodell mit deterministischem Trend	-3,98	-3,42	-3,13
Klassische t-Werte		-2,33	-1,65	-1,28

Beispiel IV 4-37

Langfristiger Einfluss der EZB

Im Rahmen der Hauptfinanzierungsgeschäfte tauschen Geschäftsbanken zeitlich begrenzt notenbankfähige Sicherheiten in Zentralbankgeld um, d.h. sie beschaffen sich Mittel für ihre geschäftlichen Operationen. Der diesen Wertpapierpensionsgeschäften zugrunde liegende Zinssatz wird als Hauptfinanzierungssatz bezeichnet und ist der wichtigste Leitzins der Europäischen Zentralbank (EZB). Da die Geschäftsbanken den Hauptfinanzierungssatz durch Zinsen auf Wertpapiere und Kredite finanzieren, kann die EZB damit indirekt die Zinsen am Geld- und Kapitalmarkt beeinflussen.



⁹⁸ Vgl. Engle, R.F., Granger, C.W.J. (1987). Vgl. auch Phillips, P.C.B., Ouliaris, S. (1990).

⁹⁹ Modifiziert entnommen aus Hamilton, J.D. (1999), Stichprobengröße 500.

In diesem Beispiel interessieren wir uns dafür, ob eine langfristige Beziehung zwischen dem Hauptfinanzierungssatz und der Rendite börsennotierter Bundeswertpapiere mit 2 Jahren Restlaufzeit besteht. Wir wollen also untersuchen, ob beide Variablen kointegriert sind. Um dies zu prüfen, steht uns Zahlenmaterial der Deutschen Bundesbank (Monatsdaten von Januar 1999 bis April 2009) zur Verfügung. Genau genommen sind für die Untersuchung von Kointegrationsbeziehungen relativ *lange Zeiträume erforderlich* (vgl. dazu Aufgabe ZR 2). Das ist *nicht durch eine hohe Periodizität zu kompensieren*. Wir können oftmals keine Kointegration feststellen, da die betrachteten Zeiträume einfach zu kurz sind. Auch das hier vorliegende Zeitfenster ist relativ eng. Aufgrund der anschaulichen Interpretation dieses Zinsbeispiels führen wir die Untersuchung aber dennoch durch.

Die uns vorliegenden Reihen (X: Hauptfinanzierungssatz, Y: Rendite Bundeswertpapire) müssen wir zunächst auf Stationarität testen. Der ADF-Test (Lag-Wahl durch Akaike, Grundmodell Random Walk *mit Drift*) liefert für X einen p-Wert von 0,2302 und für Y einen p-Wert von 0,4339 und damit keinen Hinweis auf Stationarität. Ein entsprechender Test für die ersten Differenzen von X liefert einen p-Wert von 0,0156 und bei Y einen p-Wert von 0,0000. Wir können also von zwei I(1)-Reihen ausgehen. Mit diesen können wir nun die Kointegrationsregression $Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$ schätzen. Wir erhalten das folgende Ergebnis, mit dem wir die benötigten Residuen für unseren Kointegrationstest bestimmen können:

Dependent Variable: Y		Sample: 1999M01 2009M04		
Method: Least Squares		Included observations: 124		
	Coefficient	Std. Error	t-Statistic	Prob.
C	0.873099	0.162164	5.384041	0.0000
X	0.822446	0.050890	16.16128	0.0000
R-squared	0.681618	Mean dependent var		3.383548
Adjusted R-squared	0.679008	S.D. dependent var		0.915061
S.E. of regression	0.518439	Akaike info criterion		1.540007
Sum squared resid	32.79099	Schwarz criterion		1.585496
Log likelihood	-93.48045	F-statistic		261.1869
Durbin-Watson stat	0.227406	Prob(F-statistic)		0.000000

Wir können nun die Hypothesen

$$\begin{aligned} H_0 : & \text{ "Reihen nicht kointegriert"} \leftrightarrow \text{"Residuen nicht stationär"} \\ H_A : & \text{ "Reihen kointegriert"} \leftrightarrow \text{"Residuen stationär"} \end{aligned}$$

testen. Der ADF-Test für die Residuen auf Basis von (IV4: 142) liefert die Ergebnisse der Folgeseite. Um Autokorrelation Rechnung zu tragen, wurden von EViews zwei Differenzterme in die Testgleichung integriert. Das Problem bei der Testentscheidung in EViews ist nun, dass nicht die benötigten kritischen Werte von Engle und Granger ausgegeben werden, sondern die speziellen kritischen Werte von Dickey und Fuller, auf denen auch der p-Wert basiert. Wir können diesen daher nicht zur Entscheidungsfindung verwenden. Greifen wir stattdessen (näherungsweise) auf die Werte aus der vorhergehenden Tabelle zu, können wir hier feststellen, dass bei einem Signifikanzniveau von 5 % die Beziehung $-3,4154 < -3,37$ gilt, wir also H_0 ablehnen können. Die Residuen sind also stationär und X und Y damit kointegriert.

Welche ökonomische Bedeutung hat nun eine Kointegration von Hauptfinanzierungssatz und der Rendite der Bundeswertpapiere? Wir können zunächst sagen, dass zwischen beiden Variablen eine fundamentale Beziehung besteht und die geschätzte (langfristige) Beziehung zwischen ihnen nicht als Scheinregression anzusehen ist. Verändert die EZB den Hauptfinanzierungssatz, wird sich die Rendite von Bundeswertpapieren ebenfalls ändern, wodurch sichergestellt wird, dass die Geldpolitik der Zentralbank auf die gesamte

Volkswirtschaft wirkt. Wir sprechen dabei auch von der Transmission geldpolitischer Impulse. Im Kontrast dazu würden geldpolitische Maßnahmen erheblich behindert, wenn nur eine "Scheinbeziehung" zwischen beiden Größen besteht. Dies würde nämlich bedeuten, dass Veränderungen beider Größen fundamental nichts miteinander zu tun haben.

Null Hypothesis: E has a unit root

Exogenous: None

Lag Length: 2 (Automatic based on AIC, MAXLAG=15)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.415398	0.0008
Test critical values:		
1% level	-2.584214	
5% level	-1.943494	
10% level	-1.614970	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(E)

Method: Least Squares

Sample (adjusted): 1999M04 2009M04

Included observations: 121 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
E(-1)	-0.154399	0.045207	-3.415398	0.0009
D(E(-1))	0.162624	0.090820	1.790621	0.0759
D(E(-2))	0.150085	0.092537	1.621884	0.1075
R-squared	0.099885	Mean dependent var		-0.000667
Adjusted R-squared	0.084629	S.D. dependent var		0.247708
S.E. of regression	0.236995	Akaike info criterion		-0.017076
Sum squared resid	6.627656	Schwarz criterion		0.052242
Log likelihood	4.033069	Durbin-Watson stat		2.008154

Sind $Y_t \sim I(1)$ und $X_t \sim I(1)$ in einem Modell (IV4: 140) *nicht kointegriert*, so weisen sie keine langfristige Beziehung zueinander auf. Ihre *kurzfristige Beziehung* können wir durch Modellschätzung in Differenzen beschreiben (vgl. Abschnitt IV 4.7.2.4). Es ist dabei zu beachten, dass wir dadurch zwar einer Scheinregression begegnen können, sich die Modellbedeutung aber ändern kann (vgl. Ausführungen im Abschnitt IV 4.3.4, Punkt 5). Würden wir diese Vorgehensweise auch bei kointegrierten Niveauvariablen wählen, würden wir den Langfristzusammenhang zwischen ihnen vernachlässigen. Schätzen wir das Modell mit den kointegrierten Variablen, bilden wir den *Langfristzusammenhang* ab,¹⁰⁰ jedoch nicht den kurzfristigen. Für kointegrierte Variablen wurde daher eine fundamentale Darstellung des wahren Modells, das sog. **Fehlerkorrekturmodell** nach Engle und Granger, entwickelt.¹⁰¹ Dieses wollen wir im Folgenden kurz vorstellen:

¹⁰⁰ Die Schätzung der Kointegrationsparameter kann dabei mit OLS oder dem dynamischen OLS-Schätzer von Stock, J.H., Watson, M.W. (1993) erfolgen.

¹⁰¹ Das Modell wurde erstmalig verwendet von Sargan, J.D. (1984) und später von Engle und Granger popularisiert.

Sind $Y_t \sim I(1)$ und $X_t \sim I(1)$ in einem Modell (IV4: 140) *kointegriert*, so besteht zwischen ihnen eine langfristige Beziehung, die kurzfristig aber verletzt sein kann. Wir können den Störterm (IV4: 141) als Fehlerwert betrachten, um den Y_t kurzfristig von seinem Langfristwert abweichen. Nach dem Granger-Repräsentations-Theorem lassen sich zwei kointegrierte Variablen X und Y unter Einbeziehung dieses Fehlers in der Art

$$\Delta Y_t = \alpha_0 + \alpha_1 \Delta X_t + \alpha_2 \varepsilon_{t-1} + u_t, \quad (\text{IV4: 143})$$

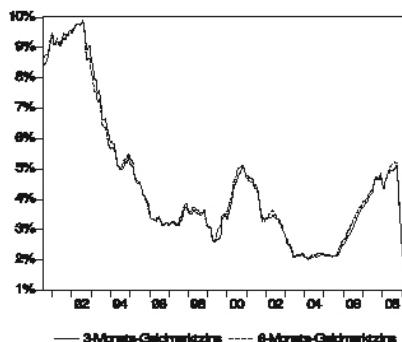
darstellen, in der u_t ein klassischer Störterm, ε_{t-1} der um eine Periode verzögerte Fehler aus der Kointegrationsregression ist und ΔY_t und ΔX_t die ersten Differenzen der Variablen X und Y sind. Wir sprechen hierbei vom sog. *Fehlerkorrekturmodell*. Es ist dadurch gekennzeichnet, dass es *nur I(0)-Variablen* beinhaltet und damit unsere Teststatistiken wieder den üblichen Verteilungen folgen. Zur Schätzung von (IV4: 143) ziehen wir in der Praxis aufgrund der unbekannten ε_{t-1} die Residuen e_{t-1} der Kointegrationsregression heran. Außerdem wird das Fehlerkorrekturmodell häufig auch noch um dynamische Elemente wie z.B. ΔY_{t-1} und ΔX_{t-1} erweitert.

(IV4: 143) besagt, dass ΔY_t neben ΔX_t auch vom Fehlerkorrekturterm ε_{t-1} abhängt. Ist dieser von Null verschieden, ist das Modell aus dem Gleichgewicht. Nehmen wir z.B. an, ΔX_t ist Null und ε_{t-1} ist positiv. Dies bedeutet, dass Y_{t-1} zu hoch ist, um im Gleichgewicht zu sein. Anders ausgedrückt liegt Y_{t-1} über seinem Gleichgewichtswert von $\beta_0 + \beta_1 X_{t-1}$. Da wir erwarten, dass α_2 negativ ist, ist auch $\alpha_2 \varepsilon_{t-1}$ negativ und damit ΔY_t zur Wiederherstellung des Gleichgewichts ebenfalls negativ. Ist Y also über seinem Gleichgewichtswert, wird Y in der Folgeperiode fallen um den Gleichgewichtsfehler zu korrigieren (daher der Name Fehlerkorrektummodell). Analog ist bei negativem ε_{t-1} der Term $\alpha_2 \varepsilon_{t-1}$ positiv, was zu positivem ΔY_t führt. Wir sehen daran, dass es der Wert von α_2 ist, der bestimmt, wie schnell das Gleichgewicht wiederhergestellt wird.

Beispiel IV 4-38

Fehlerkorrektummodell und Geldmarktsätze

Seit Einführung des Euro ist der EURIBOR (Euro Interbank Offered Rate) der Zinssatz, zu dem sich Geschäftsbanken im Eurosysteem gegenseitig Geld leihen. Vorher war der für Deutschland relevante Zins der FIBOR (Frankfurt Interbank Offered Rate). Die verknüpften 3- (X) und 6-Monatsgeldmarktzinsen (Y) aus dem Datenbestand der Deutschen Bundesbank sind für den Zeitraum von Juli 1990 bis April 2009 in folgender Abbildung zu sehen:



Aus der volkswirtschaftlichen Theorie und auch aufgrund des Verlaufs beider Zinsen würden wir erwarten, dass eine langfristige Beziehung zwischen den beiden Zinssätzen besteht (diese also kointegriert sind), da andernfalls Arbitrageure die Diskrepanz zwischen den beiden Zinssätzen ausnutzen würden. Entsprechende Tests ergeben, dass beide Zeitreihen I(1) und in der Tat kointegriert sind. Diese möge der Leser selbst nachvollziehen.

Die Schätzung eines Fehlerkorrekturmodells liefert nun folgendes Ergebnis. D(...) ist dabei wieder die EViews-Befehlssymbolik für erste Differenzen und E(-1) ist der verzögerte Fehlerkorrekturterm, der aus den Residuen der Langfristbeziehung, d.h. der Regression der 6- auf die 3-Monatsgeldmarktzinsen, gewonnen wurde.

Dependent Variable: D(Y)		Sample (adjusted): 1990M08 2009M04		
Method: Least Squares		Included observations: 225 after adjustments		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000235	0.003538	-0.066448	0.9471
D(X)	0.998382	0.020120	49.62053	0.0000
E(-1)	-0.101382	0.030066	-3.371935	0.0009
R-squared	0.928136	Mean dependent var		-0.030533
Adjusted R-squared	0.927488	S.D. dependent var		0.194185
S.E. of regression	0.052290	Akaike info criterion		-3.050778
Sum squared resid	0.607003	Schwarz criterion		-3.005230
Log likelihood	346.2125	F-statistic		1433.581
Durbin-Watson stat	1.583538	Prob(F-statistic)		0.000000

Wie wir sehen, ist der geschätzte Koeffizient des Fehlerkorrekturterms negativ und signifikant. Der Wert von -0,1014 bedeutet, dass 10,14 % der im vorhergehenden Monat auftretenen Abweichung der Zinssätze vom langfristigen Gleichgewicht im aktuellen Monat eliminiert wird. Wenn also der 6-Monats-Geldmarktsatz im letzten Monat höher als a priori erwartet über dem 3-Monats-Geldmarktsatz lag, wird der 6-Monats-Geldmarktsatz im Folgemonat um 0,1014 Prozentpunkte sinken, um die langfristige Beziehung zwischen den Zinssätzen wieder herzustellen. Darüber hinaus werden kurzfristige Änderungen des 3-Monats-Satzes schnell durch den 6-Monats-Satz reflektiert, da der Steigungskoeffizient zwischen ihnen bei 0,9984 liegt. Dies ist bei der Effizienz moderner Kapitalmärkte allerdings kein verwunderliches Resultat.

4.7.2.7 Zusammenfassung

Wird ein Regressionsmodell mit Zeitreihendaten geschätzt, sind vor Interpretation der Ergebnisse bzw. Tests auf eine Verletzung der Annahmen des CLRM die folgenden fünf Schritte zu beachten:

1. Alle Variablen sind mittels des Dickey-Fuller-Tests oder des ADF-Tests auf Stationarität zu prüfen.
2. Sind alle Variablen stationär, können wir das Modell schätzen.
3. Sind alle Variablen nicht-stationär, sind die Residuen der Kointegrationsregression mittels des Dickey-Fuller-Tests oder des ADF-Tests (mit kritischen Engle-Granger-Werten und ohne Konstante in der Testgleichung) auf Stationarität zu prüfen. Der Hinweis auf Stationarität belegt Kointegration.
4. Sind die nicht-stationären Variablen kointegriert, kann ein Fehlerkorrekturmodell geschätzt werden.
5. Bei fehlender Kointegration sollte das Modell in ersten Differenzen geschätzt werden, da diese in der Regel stationär sind. Sollte dies nicht der Fall sein, sind höhere Differenzen zu verwenden.

5.

Zusammenfassende Anwendungen aus dem Finanzbereich

Wir haben bisher die Regressionsanalyse ausgehend von der Modelltheorie über die Modellschätzung bis hin zur Analyse von Annahmenverletzungen eingehend behandelt. Die nachfolgenden Beispiele zeigen nun konkrete Anwendungen am Beispiel von Finanzmärkten. Wir beschäftigen uns dabei mit dem Capital Asset Pricing Model und Verfahren zur Beurteilung der Performance von Investmentfonds.

5.1 Capital Asset Pricing Model

Das sog. Capital Asset Pricing Model (CAPM)¹ gehört zu den wichtigsten Modellen im Finanzbereich. Es erklärt die erwartete Risikoprämie eines Wertpapiers j als eine Funktion der erwarteten Risikoprämie eines Portfolios, welches alle öffentlich gehandelten Wertpapiere enthält. Man bezeichnet dieses Portfolio auch als Marktportfolio. Hinter den Risikoprämienv verbirgt sich nichts anderes als die Differenz zwischen der erwarteten Rendite des Wertpapiers $E[R_j]$ und dem Zins einer risikolosen Anlage R_f im ersten Fall bzw. der Rendite des Marktportfolios $E[R_M]$ und R_f im zweiten Fall. Es gilt im Modell

$$E[R_j] - R_f = \beta_j (E[R_M] - R_f). \quad (\text{IV5: 1})$$

Der Ausdruck β_j beschreibt dabei, wie sensibel die Rendite des Wertpapiers j auf Veränderungen der Rendite des Marktportfolios reagiert. Ein Beta von weniger als Eins bedeutet daher, dass die Variation von R_j geringer ist als jene von R_M . Ein Beta von größer als Eins deutet auf ein im Vergleich zum Marktportfolio stärker schwankendes Wertpapier hin.

Damit der Beta-Faktor im Rahmen praktischer Anwendungen eingesetzt werden kann, muss er mittels geeigneter Verfahren geschätzt werden. Dazu müssen wir die Gleichung (IV5: 1) in eine empirisch schätzbare Form bringen (Operationalisierung). Hierbei ergibt sich das Problem, dass das CAPM eine Beziehung bezüglich der erwarteten Renditen darstellt. In der Praxis können wir aber nur die aktuellen oder realisierten Renditen direkt beobachten. Aufgrund von Erwartungsirrtümern entsprechen im Allgemeinen die realisierten Renditen nicht den in der Vergangenheit erwarteten Renditen. Aus diesen Gründen verwenden wir bei der Schätzung des Betas für ein Wertpapier j folgendes ökonometrisches Modell:

$$R_{jt} - R_{ft} = \alpha_j + \beta_j (R_{Mt} - R_{ft}) + \varepsilon_{jt} \quad (\text{IV5: 2})$$

Wir ersetzen also die Erwartungswerte durch die realisierten Werte zu jedem Zeitpunkt und fügen der Beziehung zusätzlich einen Störterm ε hinzu. Dieser sollte einen Erwartungswert von Null aufweisen, falls es bei der Erwartungsbildung nicht

¹ Vgl. Sharpe, W. (1964), Lintner, J. (1965) und Mossin, J. (1966).

zu systematischen Fehlern kommt.² Zusätzlich ist es erforderlich, dass eine Approximation des Marktportfolios gefunden wird, da dieses ebenfalls nicht empirisch beobachtbar ist. Häufig wird ein möglichst breit gestreuter Aktienindex herangezogen. Der risikofreie Zins wird meist mit dem für Dreimonats-Tagesgeld belegt. Für deutsche Aktien bietet sich die Verwendung des DAX (alternativ auch CDAX) und des 3-Monats-Euribors an.³

Wenn wir β_i mit der Regression (IV5: 2) ermitteln, schätzen wir automatisch α_i als Regressionskonstante mit. Ist das CAPM Gültig, sollte das wahre α_i den Wert Null annehmen (vgl. Gleichung (IV5: 1)). Bei positivem α_i würde das Papier mehr als die nach dem CAPM erwartete faire Risikoprämie liefern, bei negativem entsprechend weniger. Man bezeichnet den Schätzwert von α_i als Jensen-Alpha.⁴ Eine Möglichkeit der empirischen Prüfung des CAPM besteht also in einem einfachen t-Test für ein geschätztes α_i . Hinsichtlich dieses und anderer Tests des CAPM⁵ ist jedoch anzumerken, dass diese nicht kritikfrei sind. Es wird nämlich argumentiert, dass das CAPM nur bei akurater Approximation des Marktportfolios vernünftig getestet werden könne. Ein einfacher Aktienindex sei dazu nicht ausreichend.⁶

Dieser Kritik zum Trotz wollen wir im Folgenden die Gleichung (IV5: 2) mit OLS für 4 ausgewählte Unternehmen des DAX schätzen. Dies ist einer der wenigen Fälle, in dem der Schätzung der Regressionskonstanten Aufmerksamkeit geschenkt wird. Um die Schätzung rechtfertigen zu können, müssen wir von einem korrekt spezifizierten Modell ausgehen, damit der Störterm einen Erwartungswert von Null aufweist (vgl. Abschnitt IV 4.2.1).

Wir verwenden zur Schätzung 60 Monatsrenditen von September 2003 bis August 2008. Die nachfolgende Tabelle zeigt für die Renditen der verwendeten Aktien die wichtigsten Parameter der Stichprobenverteilung. Abgesehen von BMW weisen sämtliche Aktienwerte eine positive Durchschnittsrendite (auf Monatsbasis) auf, wobei jene von MAN und Volkswagen mit der größten Volatilität verbunden sind. Wir erkennen außerdem, dass Euribor und Volkswagen rechtsschiefe und alle anderen linksschiefe Verteilungen sind. Im Vergleich zur Normalverteilung (Schiefe 0 und Kurtosis 3) weisen vor allem die Verteilungen von DAX und MAN eine größere Wölbung auf.

	Mittelwert	Standardabw.	Schiefe	Kurtosis
DAX	0,0112	0,0435	-0,9371	5,9285
Euribor	0,0313	0,0108	0,4983	1,7694
BMW	-0,0021	0,0617	-0,4039	3,5913
Daimler	0,0055	0,0813	-0,0581	3,9093
MAN	0,0251	0,0959	-0,6887	6,1258
Volkswagen	0,0298	0,0941	0,2994	2,8597

² Vgl. dazu Elton, E.J., Gruber, M.J., Brown, S.J., Goetzmann, W.N. (2007), Kapitel 15.

³ Streng genommen handelt es sich beim Euribor nicht um eine risikolose Anlagemöglichkeit, da das potentielle Ausfallrisiko nicht besichert ist.

⁴ Vgl. Jensen, M. (1967).

⁵ Vgl. z.B. Fama, E.F., French, K.R. (2004).

⁶ Vgl. Roll, R. (1997).

Die im Vorfeld der Schätzungen durchgeführten Stationaritätstests für die Risikoprämien der einzelnen Wertpapiere und des DAX lieferten keine Hinweise auf Instationarität der Zeitreihen. Dies ist auch nicht verwunderlich, da die Berechnung von Renditen schließlich eine Art Differenzierung der Kurszeitreihe darstellt, die ja gerade eine nicht stationäre Zeitreihe in eine stationäre überführen kann. Wir können daher ohne Gefahr einer Scheinregression die OLS-Schätzungen durchführen. Die folgende Tabelle enthält die Schätzergebnisse, wobei der geschätzte Koeffizient der Konstanten C das Jensen-Alpha darstellt.

Dependent Variable: **BMW-EURIBOR**

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.012664	0.006137	-2.063748	0.0435
DAX-EURIBOR	1.032032	0.121917	8.465034	0.0000
R-squared	0.552665	Durbin-Watson stat		2.350171
Adjusted R-squared	0.544953	F-statistic		71.65680
S.E. of regression	0.043581	Prob(F-statistic)		0.000000

Dependent Variable: **DAIMLER-EURIBOR**

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000576	0.008421	-0.068366	0.9457
DAX-EURIBOR	1.254654	0.167298	7.499534	0.0000
R-squared	0.492310	Durbin-Watson stat		1.928669
Adjusted R-squared	0.483557	F-statistic		56.24302
S.E. of regression	0.059803	Prob(F-statistic)		0.000000

Dependent Variable: **MAN-EURIBOR**

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.025749	0.009215	2.794129	0.0070
DAX-EURIBOR	1.588576	0.183084	8.676779	0.0000
R-squared	0.564847	Durbin-Watson stat		2.463588
Adjusted R-squared	0.557345	F-statistic		75.28650
S.E. of regression	0.065446	Prob(F-statistic)		0.000000

Dependent Variable: **VOLKSWAGEN-EURIBOR**

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011218	0.012461	0.900258	0.3717
DAX-EURIBOR	0.633039	0.247564	2.557071	0.0132
R-squared	0.101313	Durbin-Watson stat		2.088759
Adjusted R-squared	0.085819	F-statistic		6.538611
S.E. of regression	0.088495	Prob(F-statistic)		0.013194

Die Schätzergebnisse weisen keine Probleme von Heteroskedastizität oder Autokorrelation auf,⁷ sodass Newey-West Standardfehler oder andere Korrekturmaßnahmen nicht erforderlich sind.

Bei näherer Betrachtung der Ergebnisse fällt zunächst auf, dass die Betakoeffizienten aller 4 Schätzungen bei einem Signifikanzniveau von 5 % signifikant von Null verschieden sind (BMW, Daimler und MAN sogar bei einem Signifikanzniveau von 1 %). Interessanter ist allerdings, inwieweit sich die Betas signifikant von 1 unterscheiden. Um dies zu prüfen müssen wir die t-Statistiken für die Nullhypothese $H_0: \beta_j = 1$ gegenüber $H_A: \beta_j \neq 1$ berechnen. Wir erhalten diese wie folgt:

	$\hat{\beta}_j - 1$:	$\hat{\sigma}_{\hat{\beta}_j}$	=	t_j
BMW	0,0320		0,1219		0,2625
Daimler	0,2547		0,1673		1,5224
MAN	0,5885		0,1831		3,2141
Volkswagen	-0,3670		0,2476		-1,4822

Bei einem Signifikanzniveau von 5 % und einem Stichprobenumfang von 60 liegt der zur Testentscheidung benötigte kritische t-Wert bei $t_{0,975;58} = 2,00$. H_0 wird nun abgelehnt, wenn der Absolutbetrag der t-Statistik größer als 2,00 ist. Dies ist nur bei MAN der Fall. Wir können also sagen, dass die Renditen von MAN stärker als der Markt schwanken. Für BMW, Daimler und Volkswagen haben wir eine derartige Evidenz nicht.

t-Tests der Nullhypothese $H_0: \alpha_j = 0$ gegenüber $H_A: \alpha_j \neq 0$ der verschiedenen Alphas führen bei einem Signifikanzniveau von 5 % bei BMW und MAN zu einer Ablehnung von H_0 , da die p-Werte kleiner als 0,05 sind. Sie sprechen damit gegen eine Gültigkeit des CAPM. Bei Daimler und Volkswagen hingegen kann H_0 beim gleichen Signifikanzniveau nicht abgelehnt werden (p-Werte größer als 0,05), was nicht gegen das CAPM spricht.

Die Bestimmtheitsmaße der Regressionen liegen bis auf VW bei ungefähr 50 %, d.h. die Hälfte der Renditevariationen können wir nicht durch Veränderungen der Risikoprämie des Marktes erklären. Insbesondere bei Volkswagen trägt das Modell nur wenig zur Erklärung der Renditevariation bei. Für den gegebenen Zeitraum scheint das CAPM hier wesentliche Einflussfaktoren zu vernachlässigen.

5.2 Investmentfondsperformance

Für die Bewertung und Analyse von Kapitalanlagen gibt es zahlreiche Verfahren, die sich durch die Informationsanforderungen an die Daten und die unterstellten Prämissen unterscheiden. Zu den bekanntesten regressionsbasierten Verfahren gehören das unter IV 5.1 behandelte Jensen-Alpha und der Treynor-Mazuy-Ansatz.

⁷ Die p-Werte des White-Tests liegen mit 0,9292 (BMW), 0,4300 (Daimler), 0,5862 (MAN), 0,2249 (Volkswagen) deutlich über einem Signifikanzniveau von 5 %. Auch die Durbin-Watson-Statistik bewegt sich im Bereich um 2, wodurch Freiheit von Autokorrelation erster Ordnung gewährleistet sein dürfte.

Tritt in Gleichung (IV5: 2) an die Stelle der Rendite eines Wertpapiers die Rendite eines Investmentfonds j , so kann für diesen Fonds j das **Jensen-Alpha** α_j geschätzt werden. Diese Schätzung gibt (bei entsprechender Signifikanz) Auskunft über die Differenz zwischen der erzielten Fondsrendite und der Rendite einer passiven Strategie (Investition in das Marktportfolio und das risikolose Wertpapier) mit gleichem systematischen (marktbasierten) Risiko. α_j misst also die risikoadjustierte Performance. Portfoliomanager, die über besondere Selektionsfähigkeiten hinsichtlich der Papiere in ihren Portfolios verfügen, generieren positive Alphas. Selektionsfähigkeit bezeichnet dabei die Eigenschaft, eine überdurchschnittlich renditeträchtige Titelselektion vorzunehmen.

Wir wollen nun auf Basis von Monatsrenditen für einen Zeitraum von Februar 1998 bis Dezember 2004 und drei zufällig ausgewählte Fonds mittels des Jensen-Alpha testen, ob die Fondmanager über Selektionsfähigkeiten verfügen. Als Näherung für das Marktportfolio verwenden wir dabei wieder den DAX und für den risikofreien Zins den 3-Monats-Euribor.

Die im Vorfeld der Schätzungen durchgeführten Stationaritätstests für die Risikoprämien der einzelnen Fonds liefern wieder keine Hinweise auf Instationarität der Zeitreihen. Die OLS-Schätzungen sind im Folgenden dargestellt:

Dependent Variable: FONDS-EURIBOR		ISIN: DE0008471608		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.001947	0.002013	-0.967219	0.3363
DAX-EURIBOR	0.831602	0.025483	32.63410	0.0000
R-squared	0.929318		Durbin-Watson stat	2.435341
Adjusted R-squared	0.928446		F-statistic	1064.984
S.E. of regression	0.018329		Prob(F-statistic)	0.000000

Dependent Variable: FONDS-EURIBOR		ISIN: DE0008485509		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.002557	0.002681	-0.953651	0.3431
DAX-EURIBOR	0.928589	0.033937	27.36234	0.0000
R-squared	0.902374		Durbin-Watson stat	2.274321
Adjusted R-squared	0.901169		F-statistic	748.6976
S.E. of regression	0.024410		Prob(F-statistic)	0.000000

Dependent Variable: FONDS-EURIBOR		ISIN: LU0028514403		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.002388	0.000787	-3.035912	0.0032
DAX-EURIBOR	0.983048	0.009957	98.73399	0.0000
R-squared	0.991759		Durbin-Watson stat	2.042575
Adjusted R-squared	0.991658		F-statistic	9748.401
S.E. of regression	0.007162		Prob(F-statistic)	0.000000



Es wird deutlich, dass nur beim Fonds LU0028514403 die Nullhypothese $H_0: \alpha_j = 0$ gegenüber $H_A: \alpha_j \neq 0$ bei einem Signifikanzniveau von 5 % abgelehnt werden kann. Dieser Fonds weist eine signifikant negative Performance gegenüber dem Benchmarkportfolio DAX auf. Auch bei den beiden anderen Fonds sind die geschätzten Alphas negativ, aber insignifikant. Wir können also für keinen Fonds signifikante Selektionsfähigkeiten nachweisen.⁸ Zudem fällt deutlich auf, dass sich schon durch diesen einfachen Modellansatz ein hoher Anteil der Variabilität der Fondsrenditen (hohe Werte von R^2) erklären lässt.

Die Idee des **Treynor-Mazuy-Ansatzes** ist es, dass ein Portfoliomanager neben Selektions- eventuell auch über Timingfähigkeiten verfügen kann. Verfügt man über Timingfähigkeiten, ist man in der Lage, die Gesamtmarktentwicklung (steigende oder fallende Marktphase) richtig einschätzen zu können und dadurch eine überdurchschnittliche Performance zu erzielen. Bei diesem Ansatz steigt man sein Fondsbeta kontinuierlich mit der Höhe der prognostizierten Marktrisikoprämie. Wird eine positive Risikoprämie des Marktes prognostiziert, strebt der Fondsmanager ein hohes Beta (systematisches Risiko) seines Fonds an. Erwartet er eine negative Prämie, wird der Fonds versuchen ein niedriges Beta zu erzielen. Dies bedeutet, dass in steigenden Marktphasen die Portfoliorendite überproportional zunimmt und in sinkenden Marktphasen unterproportional abnimmt.

Nehmen wir an, ein Fondsmanager besitzt keinerlei Fähigkeiten bei der Wertpapierauswahl, allerdings verfügt er über sehr gute Prognosefähigkeiten bei der Einschätzung des Marktes. Ermitteln wir nun über eine Schätzung von (IV5: 2) das Jensen-Alpha, erhalten wir häufig ein positives Alpha und bescheinigen dem Fondsmanager Selektionsfähigkeiten, die er nicht besitzt. Um sowohl Timing- als auch Selektionsfähigkeiten zu ermitteln, haben Treynor und Mazuy⁹ vorgeschlagen, die Regressionsgleichung zur Schätzung des Jensen-Alpha um einen quadratischen Term zu erweitern:

$$R_{jt} - R_{ft} = \alpha_j + \beta_j(R_{Mt} - R_{ft}) + \gamma_j(R_{Mt} - R_{ft})^2 + \varepsilon_{jt} \quad (\text{IV5: 3})$$

Der quadratische Term erfasst, wie das Beta des Fonds mit Veränderungen der Risikoprämie des Marktes variiert. Es gilt nämlich

$$R_{jt} - R_{ft} = \alpha_j + (R_{Mt} - R_{ft})[\beta_j + \gamma_j(R_{Mt} - R_{ft})] + \varepsilon_{jt},$$

wobei $\gamma_j(R_{Mt} - R_{ft})$ die Veränderung von Beta widerspiegelt. Liegen entsprechende Selektions- bzw. Timingfähigkeiten vor, so lassen sich diese durch die positiven Regressionskoeffizienten α_j und γ_j erfassen. Falls $\gamma_j = 0$ ist, entspricht die quadratische Regressionsfunktion (IV5: 3) genau der einfachen Jensen-Alpha-Regressionsfunktion (IV5: 2) und der Fondsmanager besitzt keine Timingfähigkeiten. Die Schätzung des Treynor-Mazuy-Ansatzes für unsere drei zufällig gewählten Fonds aus dem vorhergehenden Beispiel liefert folgende OLS-Ergebnisse:¹⁰

⁸ Die Vor- und Nachteile des Jensen-Alpha sowie dessen praktische Eignung und Signifikanztests zur Beurteilung von Können oder Glück werden ausführlich von Bodie, Z., Kane, A., Marcus, A. J. (2005) behandelt. Eine Anwendung für Deutschland findet man bei Rottmann, H., Franz, T. (2007).

⁹ Vgl. Treynor, J.L., Mazuy, K.K. (1966).

¹⁰ Eine Anwendung dieses und weiterer Verfahren auf deutsche Fonds findet man bei Rottmann, H., Franz, T. (2008).

Dependent Variable: FONDS-EURIBOR		ISIN: DE0008471608		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004042	0.002254	-1.793527	0.0767
DAX-EURIBOR	0.853554	0.027482	31.05829	0.0000
(DAX-EURIBOR)^2	0.345804	0.177851	1.944349	0.0554
R-squared	0.932508	Durbin-Watson stat		2.292218
Adjusted R-squared	0.930821	F-statistic		552.6613
S.E. of regression	0.018023	Prob(F-statistic)		0.000000

Dependent Variable: FONDS-EURIBOR		ISIN: DE0008485509		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004955	0.003020	-1.640672	0.1048
DAX-EURIBOR	0.953715	0.036825	25.89845	0.0000
(DAX-EURIBOR)^2	0.395807	0.238312	1.660872	0.1007
R-squared	0.905628	Durbin-Watson stat		2.167413
Adjusted R-squared	0.903269	F-statistic		383.8551
S.E. of regression	0.024150	Prob(F-statistic)		0.000000

Dependent Variable: FONDS-EURIBOR		ISIN: LU0028514403		
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.001884	0.000893	-2.108793	0.0381
DAX-EURIBOR	0.977767	0.010894	89.75088	0.0000
(DAX-EURIBOR)^2	-0.083197	0.070502	-1.180077	0.2415
R-squared	0.991900	Durbin-Watson stat		2.088623
Adjusted R-squared	0.991698	F-statistic		4898.520
S.E. of regression	0.007144	Prob(F-statistic)		0.000000

Tests auf Autokorrelation und Heteroskedastizität ergeben, dass die geschätzten Modelle keine offensichtlichen Mängel oder Fehlspezifikationen aufweisen.¹¹ Durch die neue quadratische Variable haben sich die angepassten Bestimmtheitsmaße (wenn auch nur geringfügig) erhöht. Der Treynor-Mazuy-Ansatz erklärt also die Renditevariabilität der Fonds besser als der einfache Jensen-Ansatz.

Bei einem Signifikanzniveau von 5 % können wir lediglich bei Fonds DE0008471608 durch einen t-Test von $H_0: \gamma_j \leq 0$ gegenüber $H_A: \gamma_j > 0$ positive Timingfähigkeiten nachweisen. Der p-Wert der quadratischen Variablen liegt nämlich dafür bei $0,0554/2 = 0,0277$ und damit unter 0,05, was ein Ablehnen von H_0 ermöglicht. Es ist zudem auffällig, dass die Fonds-Alphas alle negativ, jedoch nur bei Fonds LU0028514403 auf dem 5 %-Niveau signifikant sind. Allerdings sind die Alphas bei den beiden Fonds DE0008471608 und DE0008485509 nun deutlich kleiner und zumindest auf einem 10 %-Signifikanzniveau schwach signifikant. In Anbetracht unserer Argumentation aus den Abschnitten IV 4.1.1 und IV 4.2.1 ist bei derartigen Analysen jedoch immer zu bedenken, dass die Alphas ggf. verzerrt geschätzt werden.

¹¹ Die p-Werte des White-Tests liegen mit 0,9025 (DE0008471608), 0,3512 (DE0008485509), 0,3381 (LU0028514403) deutlich über einem Signifikanzniveau von 5 %. Auch die Durbin-Watson-Statistik bewegt sich um 2.

6.

Prognose mit geschätzten Regressionsmodellen

Auf Basis der vorangegangenen Abschnitte IV 1 bis IV 5 sind wir in der Lage Regressionsmodelle aufzustellen, zu schätzen und die Schätzergebnisse zu beurteilen. Häufig ist es für Unternehmen und politische Entscheidungsträger von Interesse nicht nur die Vergangenheit mit der Schätzung eines Modells beschreiben und erklären, sondern auch Prognosen für die Zukunft abgeben zu können. Dies ist Thema des letzten Abschnitts.

6.1 Grundlagen der Prognose

Eine wichtige Anwendung von ökonometrischen Modellen sind Prognosen. Vorhersagen werden ständig in vielen Bereichen der Wirtschaft, wie beispielsweise auf den Finanz-, Rohstoff- und Energiemarkten von Unternehmen (insbesondere in den Bereichen strategische Unternehmensführung und Marktforschung), von Regierungen, Forschungsinstituten und Verbrauchern gemacht. Viele Entscheidungen sind nämlich in die Zukunft gerichtet und hängen deswegen von Einschätzungen der zukünftigen Entwicklungen ab. Um eine ökonometrische Prognose erstellen zu können, schätzen wir Modelle mit vorliegenden Daten aus der Vergangenheit. Mit den geschätzten Modellen generieren wir dann unsere Prognosen.

Grundsätzlich können wir bei der Abgabe von Prognosen für die erklärte Variable zwischen einer Punkt- und einer Intervallprognose unterscheiden. Die *Punktprognose* liefert einen einzigen Prognosewert, während die *Intervallprognose* ein Intervall liefert, in dem die erklärte Variable mit einer gewissen Wahrscheinlichkeit liegt.

Neben dieser Klassifikation ist eine weitere Unterscheidung von Prognosen sinnvoll. Betrachten wir dazu zunächst Abbildung IV 27. Wir bezeichnen das Zeitintervall von 1 bis T, das als Grundlage für die Schätzung eines Regressionsmodells¹ verwendet wurde, als *Schätzperiode*. Man spricht hier auch häufig vom *Stützbereich* des Modells. Nehmen wir nun an, wir wollen mittels des einfachen Modells

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad (\text{IV6: 1a})$$

die Variable Y vorhersagen, so können wir dafür das mit den Daten von 1 bis T geschätzte Modell $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$ verwenden. Werden nun mittels der Ausprägungen der erklärenden Variablen innerhalb des Stützbereichs Prognosen für die erklärte Variable abgeben, sprechen wir von "in-sample"-Prognosen (**IS-Prognosen**). Derartige Prognosen sind uns bereits bekannt, da wir so bisher \hat{Y}_t berechnet und damit die Residuen bestimmt haben.

¹ Auch wenn wir uns im Folgenden hinsichtlich Notation und Argumentation auf die Zeitreihenregression beschränken, ist der Großteil der Aussagen auch für die Querschnittsregression gültig.

Jede Prognose, die sich auf einen Zeitpunkt außerhalb des Stützbereichs bezieht, wird als "out-of-sample"-Prognose (**OS-Prognose** oder eigentliche Vorhersage) bezeichnet. Um die Prognosegüte des Modells zu beurteilen, müssen die tatsächlichen Werte der erklärten Variable mit den Prognosen verglichen werden. Die Differenz $\hat{Y}_{T+1} - Y_{T+1} = e_{T+1}$ wird als **1-Schritt-Prognosefehler** (engl. one-step-ahead forecast error) bezeichnet.² Zukünftige Werte der erklärten Variablen sind aber momentan unbekannt. Deswegen wird die Stichprobe in der Praxis häufig aufgeteilt. Ein Teil der Daten wird dazu verwendet, das interessierende Regressionsmodell zu schätzen (Stützbereich), und der andere Teil wird zur Beurteilung der Prognosegüte genutzt. Man spricht dann von ex post Prognosen im Gegensatz zu ex ante Prognosen. Bei der **ex post Prognose** sind die Ausprägungen der erklärenden Variablen und auch der erklärten Variable bereits bekannt. Die Prognosen können daher mit den tatsächlichen Werten der erklärten Variable verglichen und damit die Prognosegüte des Modells beurteilt werden. Bei **ex ante Prognosen** sind die Ausprägungen der erklären und häufig auch der erklärenden Variablen unbekannt. Für den Wirtschaftswissenschaftler sind gerade ex ante Prognosen interessant, da es gerade die unbekannte erklärte Variable vorherzusagen gilt.

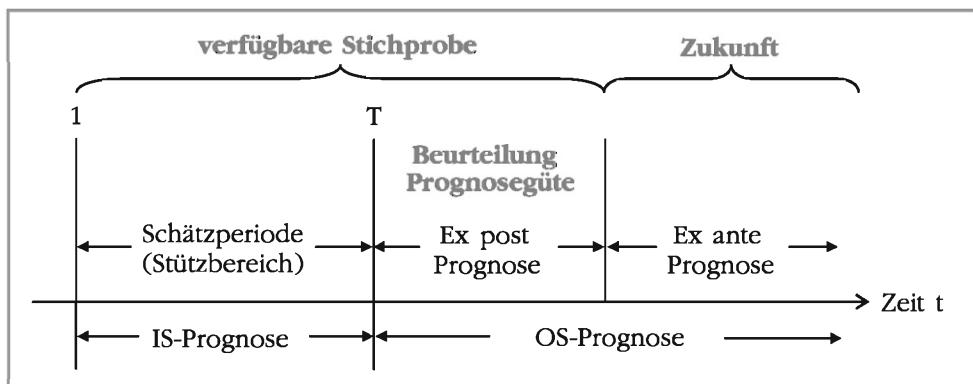


Abbildung IV 27: Prognosearten (Zeitreihenmodelle)

Zusätzlich ist zwischen bedingten und unbedingten Prognosen zu unterscheiden. Bei **bedingten Prognosen** werden bestimmte Ausprägungen für die erklärenden Variablen angenommen (Szenarioanalyse) oder sie werden als bekannt unterstellt. Mit bedingten Prognosen wird daher die Frage "Was wäre, wenn ...?" beantwortet. Eine bedingte Prognose von Y mittels der Schätzung des Modells (IV6: 1a) hätte deshalb die Gestalt

$$\hat{Y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{T+1} . \quad (\text{IV6: 1b})$$

Klassische Fälle von bedingten Prognosen sind IS-Prognosen und ex post Prognosen, da hier immer die tatsächlichen Werte für die erklärenden Variablen angenommen werden können.

² Wir beschränken uns in diesem Lehrbuch zumeist auf die Betrachtung von "one-step-ahead forecasts", d.h. Prognosen der erklärten Variablen für $T + 1$. "multi-step-ahead forecasts", d.h. Prognosen für $T + h$ mit $h > 1$, werden z.B. von Wooldridge, J.M. (2006), Kapitel 18 im Detail behandelt. Sie sind gewöhnlich mit höherer Ungenauigkeit verbunden.

Häufig interessiert uns jedoch nicht die Frage "Was wäre, wenn ...?", sondern wir wollen die bestmögliche Prognose für die erklärte Variable abgeben (**unbedingte Prognose**). Wollen wir mit der Schätzung des Modells (IV6: 1a) eine unbedingte Prognose für Y abgeben, so hätte diese die Gestalt

$$\hat{Y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_{T+1}, \quad (\text{IV6: 1c})$$

da die Ausprägung der erklärenden Variable selbst prognostiziert werden muss, bevor die Prognose für die erklärte Variable abgegeben werden kann. Dies ist in der Regel bei ex ante Prognosen der Fall. In der Praxis werden für die Vorhersage von X häufig autoregressive Modelle verwendet. Nehmen wir z.B. an, wir nutzen ein AR(1)-Modell der Form

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + u_t. \quad (\text{IV6: 2a})$$

Mittels des geschätzten Modells

$$\hat{X}_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_{t-1} \quad (\text{IV6: 2b})$$

und des bekannten Wertes X, wären wir hier in der Lage eine Prognose für die Periode T + 1 abzugeben. Die vorhergesagten X-Werte könnten wir dann in Formel (IV6: 1c) einsetzen und damit Y prognostizieren. Besser und einfacher ist es jedoch, dies in einem Modell zusammenzufassen, in dem Y direkt vom verzögerten X abhängt. Es wird also an Stelle von (IV6: 1a) und (IV6: 2a) direkt das Modell

$$Y_t = \gamma_0 + \gamma_1 X_{t-1} + \eta_t \quad (\text{IV6: 3})$$

geschätzt und für die Prognose von Y verwendet. Dies spart den zusätzlichen Schritt der Vorhersagen von X, um Y für die nächste Periode vorherzusagen (*unbedingte Prognose*).

Im Folgenden wollen wir uns näher mit bedingten und unbedingten Prognosen sowie den Besonderheiten beschäftigen, die in Modellen mit autokorrelierten Störtermen (typisches Problem in Zeitreihenmodellen), deterministischen Trends und integrierten Variablen auftreten. Besonderes Augenmerk legen wir dabei auf die Eigenschaften des Prognosefehlers. Wir werden dabei unter anderem sehen, dass der Prognosefehler folgende Quellen besitzt:

1. Unsicherheit über die wahren Werte der Parameter β_0 und β_1 des zur Prognose verwendeten Modells
2. Unsicherheit über X_t , sofern X_t prognostiziert werden muss
3. Unsicherheit über zukünftige ϵ_t , da zukünftige zufällige Abweichungen nicht bekannt sind
4. Richtige Modellspezifikation: Modelle sind immer Vereinfachungen der Realität. Lassen sie aber wesentliche Aspekte unberücksichtigt, so führt dies zu einer Fehlspezifikation und damit eventuell zu schlechten Prognosen.

Bevor wir uns jedoch diesem Themenkomplex widmen, gilt es noch zu begründen, warum unsere Prognosen häufig auf OLS-Schätzungen basieren. Es gibt nämlich theoretisch eine Vielzahl von Möglichkeiten aus historischem Datenmaterial einen Wert Y_{T+1} zu prognostizieren und es stellt sich dabei natürlich die Frage, welche als die beste angesehen werden kann. Zur Klärung dieser Fragestellung wird

gewöhnlich der *quadrierte Fehler* e_{T+1}^2 herangezogen.³ Er wertet positive und negative Prognosefehler gleich und größere Prognosefehler erhalten ein größeres Gewicht. e_{T+1} ist aufgrund des unbekannten Y_{T+1} aber zum Zeitpunkt T unbekannt und damit eine Zufallsvariable. Jedes mögliche Verfahren zur Prognose von Y_{T+1} kann außerdem nur auf den zum Zeitpunkt T verfügbaren Informationen basieren. Es wird daher die Prognose \hat{Y}_{T+1} gewählt, die den *erwarteten* quadrierten Prognosefehler bei gegebenem heutigen Informationsstand I_T (enthält alle relevanten erklärenden Variablen des Modells, ggf. mit Verzögerung)

$$E(e_{T+1}^2 | I_T) = E[(\hat{Y}_{T+1} - Y_{T+1}) | I_T] \quad (\text{IV6: 4})$$

minimiert. Es lässt sich zeigen, dass der bedingte Erwartungswert von Y_{T+1} gegeben dem heutigen Informationsstand, also $E(Y_{T+1} | I_T)$, die bestmögliche Prognose im Sinne eines möglichst niedrigen erwarteten quadratischen Fehlers darstellt.⁴ Anders ausgedrückt: Wollen wir den erwarteten quadrierten Prognosefehler bei gegebenem heutigen Informationsstand minimieren, so sollten wir als Prognose den bedingten Erwartungswert von Y_{T+1} gegeben dem heutigen Informationsstand wählen. Im Rahmen des klassischen Regressionsmodells, handelt es sich nun bei auf OLS-Schätzungen basierenden Vorhersagen gerade um bedingte Erwartungswerte (vgl. Kapitel IV 2.1). Damit stellen sie die besten bedingten Prognosen dar, da es keine anderen Schätzfunktionen mit niedrigerem erwarteten quadrierten Fehler gibt.

6.2 Bedingte Prognosen

6.2.1 Prognosefehler bei bedingten Prognosen

Nicht immer müssen erklärende Variablen mit Verzögerungen der Form (IV6: 3) vorliegen, um bedingte Prognosen abgeben zu können. Auch in statischen Modellen können wir sie in manchen Fällen als bekannt unterstellen. Dies ist z.B. der Fall, wenn es sich um deterministische Variablen (z.B. saisonale Dummies) oder demografische und ökonomische Variablen handelt, die sich nur langsam und vorhersehbar verändern. Befindet sich etwa ein Halbjahresdummy (0 = erste Jahreshälfte, 1 = zweite Jahreshälfte) als erklärende Variable im Modell, so können wir natürlich ohne Probleme die Ausprägung einer Folgeperiode vorhersagen. Auch eine erklärende Variable "Bevölkerungszahl" können wir relativ "präzise" für das nächste Jahr vorhersagen. In Modellen ohne Verzögerung der Art

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad t = 1, \dots, T \quad (\text{IV6: 4a})$$

können wir daher in diesen Fällen bedingte Prognosen über

$$\hat{Y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{T+1} \quad (\text{IV6: 4b})$$

abgeben. Betrachten wir nun den **Prognosefehler** in diesem Modellkontext etwas näher. Für das einfache Modell (IV6: 4a) ergibt er sich als Differenz aus der Prognose \hat{Y}_{T+1} und dem tatsächlich realisierten Wert Y_{T+1} . Analog zu den Residuen wollen wir den Prognosefehler im Folgenden auch mit e bezeichnen. Es gilt damit

³ Zur Herleitung des OLS-Schätzers verwendeten wir die quadrierten Fehler im Stützbereich.

⁴ Vgl. Wooldridge, J.M. (2002), Kapitel 2.

$$\begin{aligned}
 e_{T+1} &= \hat{Y}_{T+1} - Y_{T+1} \\
 &= (\hat{\beta}_0 + \hat{\beta}_1 X_{T+1}) - (\beta_0 + \beta_1 X_{T+1} + \epsilon_{T+1}) \\
 &= (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)X_{T+1} - \epsilon_{T+1}
 \end{aligned} \tag{IV6: 5}$$

Diese Darstellung des Prognosefehlers macht seine Quellen deutlich. Zum einen geht er aus der stochastischen Natur der OLS-Parameterschätzer hervor. $\hat{\beta}_0$ und $\hat{\beta}_1$ sind Zufallsvariablen, deren Ausprägungen in der Stichprobe mehr oder weniger stark von den wahren Werten β_0 und β_1 der Grundgesamtheit abweichen. Selbst wenn die wahren Regressionsparameter bekannt wären und von einem korrekt spezifizierten Regressionsmodell ausgegangen werden kann, kann jedoch ein Prognosefehler zum anderen auch aufgrund des unbekannten stochastischen Störterms ϵ_{T+1} entstehen.

Aufgrund der Annahmen des CLRM ist der stochastische Störterm ϵ_{T+1} mit dem Erwartungswert Null und der konstanten Varianz σ^2 normalverteilt. Die Normalverteilung gilt damit auch für die OLS-Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$. Da der Prognosefehler nach (IV6: 5) eine lineare Funktion dieser drei normalverteilten Zufallsvariablen ist, ist auch dieser mit dem Erwartungswert⁵

$$E(e_{T+1}) = E(\hat{\beta}_0 - \beta_0) + E(\hat{\beta}_1 - \beta_1)X_{T+1} + E(-\epsilon_{T+1}) = 0 \tag{IV6: 6}$$

und der Varianz⁶

$$\begin{aligned}
 \sigma_f^2 &= E[(e_{T+1})^2] = E[(\hat{\beta}_0 - \beta_0)^2] + E[(\hat{\beta}_1 - \beta_1)^2]X_{T+1}^2 + E[(\epsilon_{T+1})^2] \\
 &\quad + E[(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)]2X_{T+1} \\
 &= \text{Var}(\hat{\beta}_0) + X_{T+1}^2 \text{Var}(\hat{\beta}_1) + \sigma^2 + 2X_{T+1}\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)
 \end{aligned} \tag{IV6: 7}$$

normalverteilt. Setzen wir in (IV6: 7) die Varianzen der OLS-Schätzer (IV2: 23) und (IV2: 24) und ihre Kovarianz

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{X}}{\sum x_t^2} \sigma^2 \tag{IV6: 8}$$

ein,⁷ so erhalten wir nach einigen Umformungen

$$\sigma_f^2 = \sigma^2 \left[1 + \frac{1}{T} + \frac{(X_{T+1} - \bar{X})^2}{\sum x_t^2} \right]. \tag{IV6: 9a}$$

Aus (IV6: 9a) erkennen wir, dass der Prognosefehler auf Änderungen des Stichprobenumfangs T, der Varianz von X und der Differenz zwischen X_{T+1} und \bar{X} reagiert.

⁵ Man beachte dabei, dass $\hat{\beta}_0$ und $\hat{\beta}_1$ unverzerrte Schätzer sind und X_{T+1} bekannt (und damit nicht stochastisch) ist. Aufgrund von (IV6: 6) heißt \hat{Y}_{T+1} im Englischen auch "unbiased linear predictor" von Y_{T+1} .

⁶ Die Kreuzproduktterme, die die OLS-Schätzer und ϵ_{T+1} betreffen, werden bei Erwartungswertbildung Null, da $\hat{\beta}_0 - \beta_0$ und $\hat{\beta}_1 - \beta_1$ von den $\epsilon_1, \dots, \epsilon_T$ abhängen, die wiederum alle von ϵ_{T+1} unabhängig sind.

⁷ Für eine konkrete Herleitung des Ausdrucks (IV6: 8) vgl. z.B. Hill, R.C., Griffiths, W.E., Judge, G.G. (2001) oder Gujarati, D.N. (2003).

Der Prognosefehler fällt umso geringer aus, je höher der Stichprobenumfang T und je höher die Varianz von X ist. Der Prognosefehler ist minimal, wenn X_{T+1} gleich \bar{X} ist, da dann

$$\sigma_f^2 = \sigma^2 [1 + 1 / T] \quad (\text{IV6: 9b})$$

gilt. Die besten Prognosen können somit für Werte X_{T+1} nahe \bar{X} gegeben werden. Dies ist nicht überraschend, da nämlich mit zunehmender Entfernung von X_{T+1} zu \bar{X} der Wert X_{T+1} immer mehr aus dem "Erfahrungsbereich" (Stützbereich) des geschätzten Modells wandert und so die Prognosen an Verlässlichkeit verlieren. Es ist generell gefährlich ein Modell stark über seinen Stützbereich hinaus auszudehnen.

Da σ^2 in praktischen Anwendungen meist nicht bekannt ist, müssen wir seinen Schätzer (IV2: 25) heranziehen, was uns zu folgender geschätzter Prognosefehlervarianz führt:

$$\hat{\sigma}_f^2 = \hat{\sigma}^2 \left[1 + \frac{1}{T} + \frac{(X_{T+1} - \bar{X})^2}{\sum x_t^2} \right] \quad (\text{IV6: 10})$$

Die Wurzel von (IV6: 10) ist der sog. **Standardfehler einer Punktprognose** von Y_{T+1} . Diesen können wir dazu verwenden, den Prognosefehler zu standardisieren:

$$\frac{e_{T+1} - E(e_{T+1})}{\hat{\sigma}_f} = \frac{e_{T+1} - 0}{\hat{\sigma}_f} = \frac{\hat{Y}_{T+1} - Y_{T+1}}{\hat{\sigma}_f}$$

Dieser Quotient folgt einer t-Verteilung mit $T - 2$ Freiheitsgraden, sodass wir aus ihm nach den im Abschnitt III 2 behandelten Vorschriften ein **Prognose- bzw. Konfidenzintervall**

$$\hat{Y}_{T+1} - t_{1-\alpha/2, T-2} \cdot \hat{\sigma}_f \leq Y_{T+1} \leq \hat{Y}_{T+1} + t_{1-\alpha/2, T-2} \cdot \hat{\sigma}_f \quad (\text{IV6: 11})$$

aufstellen können, in das Y_{T+1} mit einer Wahrscheinlichkeit von $1 - \alpha$ fällt. Grafisch lässt sich dieses Konfidenzintervall mittels Abbildung IV 28 veranschaulichen. Für den Fall $X_{T+1} = \bar{X}$ ergibt sich das kleinstmögliche und damit verlässlichste Prognoseintervall, da hier $\hat{\sigma}_f$ sein Minimum annimmt. Je weiter X_{T+1} von \bar{X} entfernt liegt, desto größer wird $\hat{\sigma}_f$, was zu einer Verbreiterung des Intervalls führt.

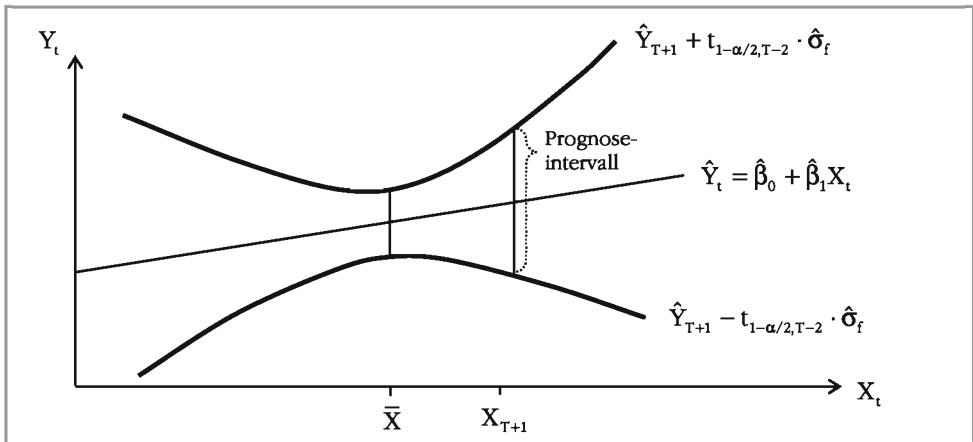


Abbildung IV 28: Prognosekonfidenzintervalle

Es sei erwähnt, dass die hier am Beispiel des einfachen Regressionsmodells beschriebenen Prinzipien auch für das multiple Regressionsmodell Gültigkeit besitzen. Prognosekonfidenzintervalle, die aus einem multiplen Modell entstehen, haben die gleiche Gestalt wie jene in Abbildung IV 28. Im Falle mehrerer erklärender Variablen sind jedoch die mathematischen Herleitungen der Verteilungsparameter des Prognosefehlers komplexer.⁸

6.2.2 Beurteilung der Güte von Prognosen

Nach Abgabe von Prognosen mittels eines geschätzten Regressionsmodells gilt es die Güte dieser Prognosen zu beurteilen. Wir können die Prognosequalität mittels *IS-Prognosen* und/oder *OS-Prognosen* beurteilen. In Regressionsmodellen handelt es sich bei den "in sample"-Kriterien z.B. um R^2 , \bar{R}^2 oder AIC. Sind das Ziel jedoch echte Vorhersagen, so ist es aber besser, "out of sample"-Kriterien zur Messung der Vorhersagequalität zu verwenden. Denn ein Modell kann einen guten Fit im Stützbereich haben, aber trotzdem ziemlich schlechte Vorhersagefähigkeiten aufweisen. Um OS-Prognosefehler zu erhalten, verwenden wir in der Praxis häufig ex post Prognosen.

Um die Nähe zwischen prognostizierten und tatsächlich realisierten Werten zu quantifizieren, verwendet man in der Praxis häufig verschiedene **Prognosegütemeße**, von denen wir im Folgenden den mittleren, den mittleren absoluten und den mittleren quadratischen Fehler sowie den Theilschen Ungleichheitskoeffizienten kurz vorstellen wollen.⁹

Das einfachste Maß der Treffgenauigkeit von Prognosen ist der **mittlere Fehler** (ME, engl. mean error). Er ist nichts anderes als das arithmetische Mittel der auftretenden Prognosefehler, d.h.

$$ME = \frac{1}{v} \sum_{t=1}^v (Y_t^p - Y_t^r), \quad (\text{IV6: 12})$$

wobei v die Anzahl der Vergleichsperioden, Y_t^p ein prognostizierter Wert von Y_t und Y_t^r ein realisierter Wert von Y_t ist.¹⁰ Ein negativer (positiver) Wert von ME entspricht also einer Unter-(Über-)schätzung der tatsächlichen Realisationen. Gute Prognosen sollten einen mittleren Fehler von Null aufweisen. Werte, die davon abweichen, deuten auf eine Verzerrung (engl. bias) der Prognosen hin, d.h. auf eine systematische Über- oder Unterschätzung.

Der Nachteil von ME liegt darin, dass sich große positive und negative Werte gegenseitig aufheben können. Gerade daher kommt die durchschnittliche Qualität

⁸ Vgl. z.B. Greene, W.H. (2008), S. 99 ff. Als umfassendes Lehrbuch zum Thema Prognosen empfiehlt sich zudem Diebold, F.X. (2004).

⁹ Einen ausgezeichneten Überblick über weitere Verfahren zur Beurteilung der Güte von Prognosen (z.B. grafische Fehlerdarstellungen, Wendepunktanalysen und ökonometrische Methoden zur Beurteilung der Unverzerrtheit und Effizienz von Prognosen) bietet z.B. Döpke, J. (2004).

¹⁰ Wir weichen hier bewusst von der bisherigen Notation ab, da wir so im Folgenden die Mittelwerte und Varianzen der prognostizierten und der tatsächlich realisierten Reihe besser unterscheiden können.

von Prognosen im **mittleren absoluten Fehler** (MAE, engl. mean absolute error) besser zum Ausdruck:

$$\text{MAE} = \frac{1}{v} \sum_{t=1}^v |Y_t^p - Y_t^r| \quad (\text{IV6: 13})$$

Die Betrachtung des mittleren Fehlers wird in der Praxis meist um die der **Fehlervarianz** (EV, engl. error variance) ergänzt. Diese ist definiert als

$$\text{EV} = \frac{1}{v} \sum_{t=1}^v [(Y_t^p - Y_t^r) - \text{ME}]^2. \quad (\text{IV6: 14})$$

Die Fehlervarianz gibt Auskunft über die Volatilität der Prognosefehler. Wünschenswert ist ein niedriger Wert der Fehlervarianz, da ein solcher bedeutet, dass die Prognosefehler nicht weit von ihrem Durchschnitt abweichen, d.h. es zu keinen krassen Ausreißern bei der Vorhersagegenauigkeit kommt.

Die Vorzeichen der Prognosefehler können auch auf eine andere Weise als in Formel (IV6: 13) eliminiert werden. Der **mittlere quadratische Fehler** (MSE, engl. mean squared error)

$$\text{MSE} = \frac{1}{v} \sum_{t=1}^v (Y_t^p - Y_t^r)^2 \quad (\text{IV6: 15a})$$

bzw. die **Wurzel aus dem mittleren quadratischen Fehler** (RMSE, engl. root mean squared error)

$$\text{RMSE} = \sqrt{\frac{1}{v} \sum_{t=1}^v (Y_t^p - Y_t^r)^2} \quad (\text{IV6: 15b})$$

sind dazu die gebräuchlichsten Maße. Diese beiden Kennzahlen haben im Gegensatz zum mittleren absoluten Fehler die Eigenschaft, dass große Fehler stärker gewichtet werden als kleine. Zur besseren Interpretation wird der mittlere quadratische Fehler meist in seine Quellen zerlegt. Es lässt sich zeigen, dass

$$\text{MSE} = (\bar{Y}^p - \bar{Y}^r)^2 + (\sigma_p - \sigma_r)^2 + 2(1 - \rho)\sigma_p\sigma_r \quad (\text{IV6: 16})$$

gilt, wobei \bar{Y}^p , \bar{Y}^r , σ_p und σ_r die Mittelwerte und Standardabweichungen der Reihen Y_t^p und Y_t^r sind und ρ ihr Korrelationskoeffizient ist. Man erkennt, dass der MSE mit dem Ausmaß der Verzerrung (erster Term in der Gleichung), Unterschieden der Variabilität (zweiter Term) und geringer werdender Korrelation zwischen prognostizierter und realisierter Variablen zunimmt.

Die Beurteilung von Prognosen anhand der bisher beschriebenen Maße stößt auf ein prinzipielles Hindernis: Es fehlt an einem Vergleichsmaßstab, der es erlaubt zu entscheiden, ob eine Prognose hinreichend gut oder verbessерungsfähig ist. So kann z.B. nicht entschieden werden, ob ein mittlerer absoluter Fehler von einem Prozentpunkt eine "gute" oder eine "schlechte" Prognoseleistung ist. Eine Möglichkeit, diesem Problem zu begegnen, ist der Vergleich der Prognose mit einer sog. "*naiven*" Prognose. In der Praxis verwendet man häufig das Random Walk Modell

für die Abgabe einer "naiven" Prognose. Der heute realisierte Wert einer Variablen stellt danach die "naive" Prognose für morgen dar.¹¹

Liegt der RMSE der "naiven" Prognose, $\text{RMSE}_{\text{naiv}}$, sowie derjenige der zu beurteilenden Prognose vor, so kann damit der sog. **Theilsche Ungleichheitskoeffizient** (TUK) berechnet werden. Für diesen gilt

$$\text{TUK} = \frac{\text{RMSE}}{\text{RMSE}_{\text{naiv}}} .12 \quad (\text{IV6: 17})$$

Diese Kennzahl lässt sich recht einfach interpretieren: Ist $\text{TUK} = 1$, ist die Prognosegüte unseres Modells unzureichend, da beide RMSE gleich groß sind. Wir hätten also ebenso auf die Modellschätzung verzichten und lediglich eine "naive" Prognose abgeben können. Ist $\text{TUK} < 1$ ($\text{TUK} > 1$), so schneidet das geschätzte Modell bei der Prognose besser (schlechter) ab als der "naive" Vergleichsmaßstab.

Zusammenfassend können wir also sagen, dass die Maße (IV6: 12) - (IV6: 15b) vornehmlich zum Vergleich verschiedenartig spezifizierter Prognosemodelle dienen (vgl. dazu Beispiel IV 6-1). Liegt nur ein einziges Modell vor, dessen Prognosegüte beurteilt werden soll, wird stattdessen auf (IV6: 17) zurückgegriffen.

Beispiel IV 6-1

Zinsprognose



Mit Hilfe von Jahresdaten von 1960 bis 2008 aus der Online-Datenbank der Deutschen Bundesbank wollen wir im Folgenden die Prognosegüte von verschiedenen Modellen zur Erklärung des Dreimonatszinses Y auf dem deutschen Geldmarkt untersuchen. Als exogene erklärende Variablen dienen dabei die deutsche Inflationsrate X_1 sowie das BIP-Wachstum X_2 . Konkret schätzen wir die folgenden Modelle:

- Modell 0: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$
- Modell 1: $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$
- Modell 2: $Y_t = \beta_0 + \beta_1 X_{t-1} + \varepsilon_t$
- Modell 3: $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{2t} + \varepsilon_t$
- Modell 4: $Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{2t-1} + \varepsilon_t$

Das Modell 0 haben wir hinzugefügt, um zu beurteilen, ob mit dem um eine Periode verzögerten Zins als erklärender Variable eine bessere Prognosegüte erreicht werden kann als mit den exogenen Variablen X_1 und X_2 . Wir schätzen alle Modelle für den Zeitraum 1960 bis 2000 (Stützbereich) und bilden damit bedingte Prognosen für die Jahre 2001 bis 2008. In EViews müssen wir dabei beachten, dass bei Schätzung und Prognose stets genau das "sample" angegeben werden muss, das Schätzungs- bzw. Prognosegrundlage sein soll. Die Prognose selbst können wir über den Button "Forecast" im Schätzergebnisfenster erzeugen. Dabei ist wiederum zu beachten, dass die von uns behandelten bedingten 1-Schritt-Prognosen in EViews als "static forecasts" geführt werden.

¹¹ Alternativ ist als "naive" Prognose z.B. auch der Mittelwert während des Stützzeitraums oder ein Fortschreiben mit der durchschnittlichen Wachstumsrate während des Stützzeitraums denkbar.

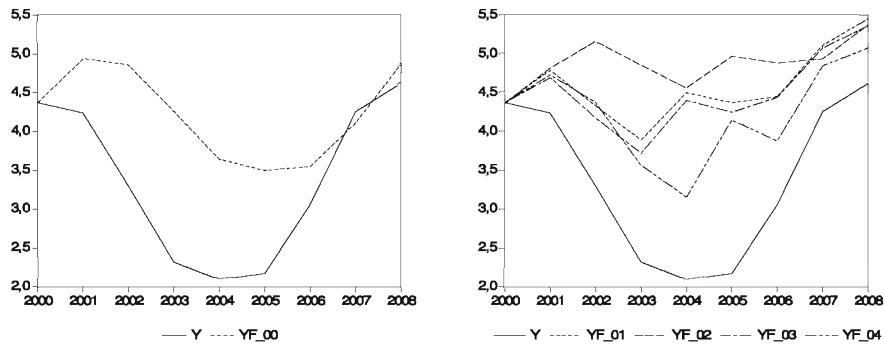
¹² Es handelt sich hierbei nicht um die ursprünglich von Theil, H. (1966) vorgeschlagene, jedoch eine in der Praxis gewöhnlich bevorzugte Version des TUK. Es ist außerdem anzumerken, dass weitere TUK-Varianten existieren und in den gängigen Softwarepaketen unterschiedliche implementiert sind. Auch EViews nutzt eine modifizierte Variante von (IV6: 17). Es empfiehlt sich daher stets ein Blick in die zugehörigen Softwarehandbücher.

Die Schätzergebnisse unserer Modelle zeigen für den Stützbereich folgendes Bild, wobei Standardfehler in Klammern angegeben und eine Signifikanz auf einem Niveau von 5 % bzw. 1 % wie in Fachpublikationen üblich mit * bzw. ** gekennzeichnet sind:

	Modell 0	Modell 1	Modell 2	Modell 3	Modell 4
C	2,1760*	2,8506**	3,8681**	2,6803**	2,4653**
	(0,8158)	(0,5333)	(0,7062)	(0,7189)	(0,8787)
Y(-1)	0,6325**				
	(0,1264)				
X1		1,0005**		1,0098**	
		(0,1501)		(0,1540)	
X1(-1)			0,6678**		0,7447**
			(0,1967)		(0,1877)
X2				0,0458	
				(0,1280)	
X2(-1)					0,3763*
					(0,1543)
R ²	0,3971	0,5325	0,2327	0,5341	0,3390
\bar{R}^2	0,3812	0,5205	0,2125	0,5096	0,3033
AIC	4,2299	3,9513	4,4710	3,9967	4,3719

Eine erste Betrachtung dieser Ergebnisse zeigt, dass Modell 1 den besten Stichprobenfit (größtes angepasstes Bestimmtheitsmaß bzw. kleinster AIC-Wert) aufweist. Die Einbeziehung weiterer unverzögterer und auch verzögter exogener Variablen (Modelle 2 bis 4) und auch ein rein autoregressives Modell für den Zins (Modell 0) liefern keinen besseren Stichprobenfit.

Beurteilen wir nun die Prognosegüte der von uns geschätzten Modelle. In EViews können wir uns dazu automatisch die prognostizierten Zeitreihen ausgeben sowie eine Reihe von Prognosegütemaßen berechnen lassen. Die Gegenüberstellung der Prognosen und der tatsächlichen Beobachtungen von 2001 bis 2008 sowie der RMSE zeigen folgendes Bild:



	Modell 0	Modell 1	Modell 2	Modell 3	Modell 4
RMSE	1,1827	1,4938	1,8891	1,4022	1,0745

Sowohl die Grafiken als auch die Fehlermaße zeigen zunächst, dass sich die Prognosegüte der Modelle mit kontemporären erklärenden Variablen (Modell 1 und 3) kaum von einander unterscheidet. Eine Verzögerung der erklärenden Variable in Modell 1, die zu Modell 2 führt, bringt eine Verschlechterung der Prognosegüte mit sich, wohingegen ein analoges Vorgehen bei Modell 3, das zu Modell 4 führt, die Prognosegüte verbessert. Modell 4 liefert gemessen am RMSE sogar bessere Prognosen als das einfache AR(1)-Modell (Modell 0). Interessant ist es auch zu sehen, dass das Modell 1, welches im Stützbereich den besten "fit" aufwies, nun als das zweitschlechteste Prognosemodell erscheint.

Die Prognosen aus allen Modellen liegen besonders zwischen 2003 und 2006 deutlich von den wahren Werten entfernt. Eine mögliche Erklärung dieser starken Verzerrungen liefern die p-Werte des Breusch-Godfrey LM-Tests auf Autokorrelation erster Ordnung für die geschätzten Modelle (Modell 0: 0,0260; Modell 1: 0,0009; Modell 2: 0,0007, Modell 3: 0,0003, Modell 4: 0,0007). Sie geben einen deutlichen Hinweis auf Autokorrelation. Eine Berücksichtigung dieser Autokorrelation bei der Prognose könnte daher die Prognosegüte verbessern (vgl. Abschnitt IV 6.2.3), d.h. der starken systematischen Verzerrung begegnen.

6.2.3 Prognose bei Vorliegen von Autokorrelation

In den vorhergehenden Abschnitten erstellten wir Prognosen für lineare Modelle mit Störtermen, die die Annahmen des klassischen Regressionsmodells erfüllten. Sind die stochastischen Störterme eines Modells autokorreliert, so führt die Berücksichtigung dieser Effekte oft zu besseren Prognosen. Betrachten wir dazu das einfache Regressionsmodell

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad \text{mit} \quad \epsilon_t = \rho \epsilon_{t-1} + u_t, \quad (\text{IV6: 18})$$

in dem $|\rho| < 1$ und $u_t \sim N(0, \sigma_u^2)$ gilt. Unter der Annahme, dass β_0 , β_1 und ρ bekannt sind, kann prinzipiell die Prognose

$$\hat{Y}_{T+1} = \beta_0 + \beta_1 X_{T+1} + \hat{\epsilon}_{T+1} \quad (\text{IV6: 19})$$

abgegeben werden. Da nun $\epsilon_t = \rho \epsilon_{t-1} + u_t$ gilt, ist $E[\epsilon_{T+1} | \epsilon_T] = \rho \epsilon_T$ und wir erhalten deshalb

$$\hat{Y}_{T+1} = \beta_0 + \beta_1 X_{T+1} + \rho \epsilon_T. \quad (\text{IV6: 20})$$

Der Anteil ρ des Störterms wird also in die nächste Periode übertragen. Aufgrund $Y_{T+1} = \beta_0 + \beta_1 X_{T+1} + \epsilon_{T+1}$ und (IV6: 20) ergibt sich der Prognosefehler e_{T+1} damit als

$$\begin{aligned} e_{T+1} &= Y_{T+1} - \hat{Y}_{T+1} = (\beta_0 + \beta_1 X_{T+1} + \epsilon_{T+1}) - (\beta_0 + \beta_1 X_{T+1} + \rho \epsilon_T) \\ &= \epsilon_{T+1} - \rho \epsilon_T = u_{T+1}. \end{aligned} \quad (\text{IV6: 21})$$

Der Prognosefehler ist damit mit einem Erwartungswert von Null und der Varianz

$$\begin{aligned} \sigma_e^2 &= E[(\epsilon_{T+1} - \rho \epsilon_T)^2] = E(\epsilon_{T+1}^2) - 2\rho E(\epsilon_T \epsilon_{T+1}) + \rho^2 E(\epsilon_T^2) \\ &= E(\epsilon_{T+1}^2) - 2\rho^2 E(\epsilon_T^2) + \rho^2 E(\epsilon_T^2) \\ &= (1 - \rho^2) \sigma_e^2 \end{aligned} \quad (\text{IV6: 22})$$

normalverteilt.¹³ Wie wir daran erkennen, kann durch die Berücksichtigung der Autokorrelation die Varianz des Prognosefehlers um den Faktor $1 - \rho^2$ verringert werden.

¹³ Der Übergang von Zeile 1 zu Zeile 2 in (IV6: 22) erfolgt, da $\epsilon_{T+1} = \rho \epsilon_T + u_{T+1}$ gilt und ϵ_T und u_{T+1} unkorreliert sind. Zeile 3 folgt aufgrund von $E(\epsilon_{T+1}^2) = E(\epsilon_T^2) = \sigma_e^2$.

In der Praxis sind natürlich β_0 , β_1 und ρ unbekannt, sodass unsere Prognose \hat{Y}_{T+1} mittels ihrer Schätzungen generiert werden muss. Setzen wir die entsprechenden Schätzer in (IV6: 20) ein und verwenden wir statt des enthaltenen Störterms der Periode T das entsprechende Residuum, so erhalten wir

$$\hat{Y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{T+1} + \hat{\beta} e_T. \quad (\text{IV6: 23})$$

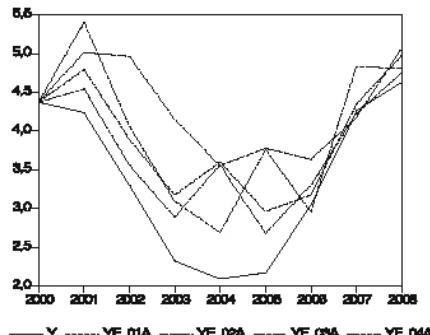
Es lässt sich zeigen, dass der Erwartungswert des Prognosefehlers, der aus der Verwendung von (IV6: 23) hervorgeht, mit zunehmenden Stichprobenumfang gegen Null geht. Die Vorhersagevarianzen werden jedoch aufgrund der zusätzlichen Ungenauigkeit, die durch die Schätzung der Parameter entsteht, größer. Falls β_0 und β_1 geschätzt werden, ρ jedoch bekannt ist, gilt (IV6: 23) als beste lineare unverzerrte Vorhersage. Falls ρ zusätzlich geschätzt werden muss, beinhaltet e_T zwar auch Informationen über e_{T+1} , aber durch die Schätzung von ρ ergibt sich eine zusätzliche Ungenauigkeit. Deshalb ist insbesondere in kleinen Stichproben nicht klar, ob (IV6: 23) dem einfachen Vorhersagemodell $\hat{Y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{T+1}$ überlegen ist.

Beispiel IV 6-2

Zinsprognose und Autokorrelation I



In Beispiel IV 6-1 hatten wir bereits den Verdacht geäußert, dass die schlechte Prognosegüte der von uns geschätzten Modelle auf nicht berücksichtigte Autokorrelation zurückzuführen sein könnte. Schätzen wir unsere Modelle 1 bis 4 unter Verwendung der in Abschnitt IV 4.5.4.1 behandelten FGLS-Methode (p -Werte des Breusch-Godfrey Autokorrelationstests sind nun nicht mehr auffällig) und prognostizieren mit den neuen Schätzergebnissen, so erhalten wir folgende Ergebnisse:



	Modell 1	Modell 2	Modell 3	Modell 4
RMSE	0,7440	1,2225	0,6093	0,8492

Wir erkennen deutlich, dass wir durch Berücksichtigung der Autokorrelation in den Modellschätzungen bessere Prognosen erzielen können. Unter unseren Modellen ist nun Modell 3 zu bevorzugen, da dieses den geringsten RMSE aufweist.

Neben der gerade beschriebenen Vorgehensweise zur Berücksichtigung von Autokorrelation bei Prognosemodellen ist in der Praxis vor allem die folgende beliebt:

In Abschnitt IV 4.5.1 haben wir gesehen, dass wir durch Modellierung der dynamischen Beziehungen zwischen Y und X (Aufnehmen von verzögerten Y und X in

die Modellgleichung) im Falle von Autokorrelation ein Modell ohne autokorrelierten Störterm erhalten können. Eine solche Vorgehensweise bietet sich natürlich auch im Prognosekontext an, um bestehender Autokorrelation zu begegnen (vgl. Beispiel IV 6-3). Der Vorteil dieser alternativen Methodik liegt darin, dass ein derartig spezifiziertes dynamisches Modell weniger restriktiv ist, als ein Modell mit autokorreliertem Störterm.¹⁴

Beispiel IV 6-3

Zinsprognose und Autokorrelation II



Zur Veranschaulichung dynamischer Modellierungen zur Begegnung von Autokorrelation schätzen wir für unser Beispiel IV 6-1 die folgenden drei Spezifikationen. Wir beginnen dabei mit der allgemeinsten Spezifikation 5 und schließen schrittweise nicht signifikante Variablen aus, bis wir schließlich bei Spezifikation 7 angelangt sind:

$$\text{Modell 5: } Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_{1t} + \beta_3 X_{1t-1} + \beta_4 X_{2t} + \beta_5 X_{2t-1} + \varepsilon_t$$

$$\text{Modell 6: } Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_{1t} + \beta_3 X_{1t-1} + \beta_4 X_{2t} + \varepsilon_t$$

$$\text{Modell 7: } Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_{1t} + \beta_3 X_{1t-1} + \varepsilon_t$$

Die Modellschätzungen liefern zunächst folgende Ergebnisse, für die die p-Werte des Breusch-Godfrey Autokorrelationstests nun nicht mehr auffällig sind:

	Modell 5	Modell 6	Modell 7
C	0,8719 (0,9979)	0,9124 (0,9663)	1,7066** (0,6113)
Y(-1)	0,5685** (0,1576)	0,5728** (0,1541)	0,4988** (0,1376)
X1	1,1643** (0,2657)	1,1869** (0,2403)	1,2810** (0,2237)
X1(-1)	-0,7881** (0,2853)	-0,8131** (0,2567)	-0,8741** (0,2506)
X2	0,1449 (0,1535)	0,1541 (0,1454)	
X2(-1)	0,0279 (0,1307)		
R ²	0,6953	0,6949	0,6851
\bar{R}^2	0,6505	0,6601	0,6589
AIC	3,7473	3,6987	3,6803

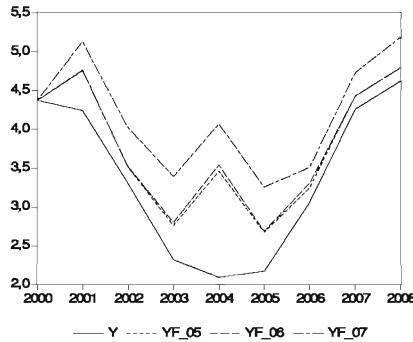
Standardfehler in Klammern, * p < 0,05, ** p < 0,01

Im Vergleich zu den im Beispiel IV 6-1 geschätzten Modellen fällt zunächst auf, dass wir in den Modellen 5 bis 7 einen deutlich besseren Stichprobenfit erzielen. Legen wir das in der Praxis am häufigsten verwendete AIC zugrunde, so können wir sagen, dass Modell 7 die Stichprobe am besten beschreibt.

Betrachten wir nun die Prognosegüte unserer drei neuen Modelle: Wieder sehen wir, dass ein Modell mit gutem Stichprobenfit (Modell 7) eine relativ schlechte Prognosegüte aufweist. Im Vergleich zu den Modellen, die Autokorrelation unberücksichtigt lassen, erkennen

¹⁴ Vgl. dazu z.B. Johnston, J., Dinardo, J. (1997) sowie Abschnitt IV 4.5.1.

wir, dass unsere dynamische Modellierung zu einer Verbesserung der Prognosegüte beigetragen hat. Gegenüber der Korrektur mit der AR(1)-Methode lässt sich ebenso eine Verbesserung feststellen. Im Vergleich zu Beispiel IV 6-2 kann nun nach RMSE Modell 5 und nicht länger das angepasste Modell 3 als bestes Prognosemodell angesehen werden.



	Modell 5	Modell 6	Modell 7
RMSE	0,5852	0,6159	1,0182

6.2.4 Trendprognosen

Wie wir bereits in Abschnitt IV 4.7.2.3 gesehen haben, kann als erklärende Variable bei Zeitreihenregressionen auch ein linearer deterministischer Trend vorkommen. Es ist daher auch relevant, näher zu betrachten, wie mit derartig gestalteten Modellen bei Prognosen vorzugehen ist. Betrachten wir dazu folgendes einfache Modell mit linearem deterministischen Trend:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad (\text{IV6: 24a})$$

Eine Prognose mit der Schätzung eines derartigen Modells ist denkbar einfach, da die Variable t bekannt ist. Es gilt

$$\hat{Y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 (T + 1). \quad (\text{IV6: 24b})$$

Haben wir also unser Modell (IV6: 24a) z.B. für einen Zeitraum von $T = 40$ Jahren, d.h. $t = 1, 2, \dots, 40$, geschätzt, so ist unsere Prognose \hat{Y}_{41} gleich $\hat{\beta}_0 + \hat{\beta}_1 \cdot 41$.

Oft verwendet man in der **Praxis** jedoch *logarithmierte Werte von Y* , da ökonomische Variablen eher mit einer konstanten Rate wachsen, als einen linearen Trend aufweisen. Wir schätzen also häufig Modelle der Form

$$\ln Y_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad (\text{IV6: 25a})$$

Durch Einsetzen von $t = T + 1$ in die geschätzte Modellgleichung können wir dann jedoch nur $\ln Y_{T+1}$ prognostizieren. Wir interessieren uns aber für die Prognose von Y_{T+1} . Um diese zu erhalten dürfen wir nicht schlicht $\exp[\hat{\beta}_0 + \hat{\beta}_1(T + 1)] \cdot \exp(\varepsilon_{T+1})$ mit $\exp[\hat{\beta}_0 + \hat{\beta}_1(T + 1)]$ schätzen, da wir bei einer derartigen Vorgehensweise Y_{T+1} systematisch unterschätzen. Obwohl nämlich der Erwartungswert von ε_t Null ist, ist der Erwartungswert von $\exp(\varepsilon_t)$ aufgrund der Jensenschen Ungleichung (vgl. Ab-

schnitt II 2.5) nicht Eins, sondern größer als Eins. Stattdessen empfiehlt sich folgendes Vorgehen:¹⁵ Nachdem in einem ersten Schritt (IV6: 25a) geschätzt wurde, wird in einem zweiten Schritt Y_t (ohne Konstante) auf $\exp[\hat{\beta}_0 + \hat{\beta}_1 t]$ regressiert. Der geschätzte Steigungskoeffizient $\hat{\gamma}$ dieser Regression wird dann wie folgt bei der Prognose von Y_{T+1} eingesetzt:¹⁶

$$\hat{Y}_{T+1} = \hat{\gamma} \cdot \exp[\hat{\beta}_0 + \hat{\beta}_1(T + 1)] \quad (\text{IV6: 25b})$$

Beispiel IV 6-4

Deterministische Trendmodelle



In diesem Beispiel wollen wir die monatliche Zeitreihe ausstehender Konsumentenkredite in den USA (Quelle: Board of Governors des Federal Reserve Systems) von Januar 1943 bis Oktober 2009 (Werte in Milliarden Dollar) näher untersuchen. Konkret wollen wir den Wert für Oktober 2009 prognostizieren und schätzen dazu die deterministischen Trendmodelle (IV6: 24a) und (IV6: 25a) für die Stützperiode von Januar 1943 bis September 2009 (d.h. $T = 801$). Wir erhalten die folgenden Schätzergebnisse:

	Hauptregression		Hilfsregression
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$
Nicht-Log-Modell	-530,5627** (25,5921)	2,8295** (0,0553)	
Log-Modell	2,3287** (0,0163)	0,0074** (0,0000)	1,000** (0,000)

Standardfehler in Klammern, * $p < 0,05$, ** $p < 0,01$

Wir können nun damit die ausstehenden Konsumentenkredite (KK) im Oktober 2009 prognostizieren:

$$\text{Nicht-Log-Modell: } \hat{KK}_{10/2009} = -530,5627 + 2,8295 \cdot 802 = 1.738,70$$

$$\text{Log-Modell: } \hat{KK}_{10/2009} = 1 \cdot \exp(2,3287 + 0,0074 \cdot 802) = 3.879,65$$

Ein Vergleich mit den tatsächlichen ausstehenden Konsumentenkrediten im Oktober 2009, also 2.482,90 Milliarden Dollar, zeigt, dass keines der beiden Modelle brauchbare Prognosen liefert. Eine Begründung für diese schlechten Prognoseergebnisse liefern die nachfolgenden Ausführungen.

In vielen Fällen genügt es nicht, allein die Trendkomponente zu modellieren. Häufig weisen Zeitreihen neben einem Trend auch noch saisonale oder konjunkturelle Muster auf. So können wir beispielsweise ein lineares Trendmodell mit einem AR(1)-Prozess und Saisondummies kombinieren. In dieser Einführung wollen wir aber nicht weiter darauf eingehen und verweisen für eine leicht verständliche Darstellung auf Diebold, F.X. (2004).

Deterministische Trendmodelle sollten bei ökonomischen Variablen außerdem nur mit großer Zurückhaltung angewendet werden, insbesondere wenn weit in die Zu-

¹⁵ Vgl. Wooldridge, J.M. (2006), Kapitel 18.

¹⁶ Eine andere Vorgehensweise besteht darin, den Korrekturfaktor γ mit $\exp(\hat{\sigma}^2 / 2)$ zu schätzen, wobei $\hat{\sigma}^2$ die geschätzte Varianz von ε in (IV6: 25a) darstellt. Dies erfordert aber die Annahme der Normalverteilung des Störterms.

kunft vorausgesagt wird.¹⁷ Es bestehen hierbei nämlich generell zwei Gefahren: Erstens kann eine schlechte Vorhersage resultieren, falls es am Ende des Stützbereichs oder im Prognosebereich zu einer *Veränderung des Trendverhaltens* kommt. Zweitens folgen die meisten *ökonomischen Variablen eher stochastischen als deterministischen Trends*. Bei integrierten Variablen mit Drift führt die Verwendung eines deterministischen Trendmodells zu einer starken Erhöhung der Varianz des Vorhersagefehlers (vgl. nachfolgende Ausführungen).

Viele ökonomische Variablen folgen eher stochastischen als deterministischen Trends. Für Variablen mit Unit Roots eignet sich die bisher beschriebene Vorgehensweise nicht. Betrachten wir dazu einen einfachen Random Walk mit Drift δ (vgl. Abschnitt IV 4.7.2.2). Für diesen gilt h Perioden in der Zukunft

$$Y_{T+h} = Y_T + \delta \cdot h + \epsilon_{T+1} + \dots + \epsilon_{T+h}, \quad (\text{IV6: 26})$$

wobei jedes ϵ_{T+h} einen bedingten Erwartungswert von Null und eine konstante Varianz σ^2 besitzt.¹⁸ Der bedingte Erwartungswert von Y_{T+h} gegeben den Informationen der Periode T ist daher $Y_T + \delta \cdot h$ und die Varianz des Prognosefehlers $\sigma^2 \cdot h$. Die Varianz des Vorhersagefehlers bei stochastischen Trends wächst somit proportional zum Vorhersagehorizont. Gerade daher ist es sehr schwierig, Random Walks mit (oder ohne) Drift weit in die Zukunft zu prognostizieren.

Verwenden wir zur Prognose eines Random Walks mit Drift ein deterministisches Trendmodell, so gilt

$$Y_{T+h} = Y_0 + \delta(T + h) + \underbrace{\epsilon_1 + \epsilon_2 + \dots + \epsilon_{T+h}}_{u_{T+h}}, \quad (\text{IV6: 27})$$

wobei Y_0 die Rolle des Achsenabschnitts β_0 im linearen Trendmodell einnimmt. Der Prognosefehler besitzt noch immer einen bedingten Erwartungswert von Null, weist nun jedoch eine Varianz von $\sigma^2 \cdot (T + h)$ auf. Nutzen wir also ein deterministisches Trendmodell $\beta_0 + \beta_1(T + h)$ zur Prognose von Y_{T+h} , liegt die Varianz des Prognosefehlers bei $\sigma^2 \cdot (T + h)$, wohingegen sie mit dem Modell $Y_T + \delta \cdot h$ nur bei $\sigma^2 \cdot h$ liegt. Die Varianz des Prognosefehlers kann also bei einem deterministischen Trendmodell viel größer sein als beim Modell $Y_T + \delta \cdot h$, insbesondere wenn T groß ist. Deswegen sollten wir keine deterministischen Trendmodelle nutzen, um integrierte Variablen zu prognostizieren.

6.3 Unbedingte Prognosen

Im Rahmen bedingter Prognosen sind wir bisher davon ausgegangen, dass die Werte der erklärenden Variablen für die Prognoseperioden bekannt sind. Vorallem bei ex ante Prognosen ist aber eine solche Annahme äußerst unrealistisch. In einem solchen Fall müssen zunächst die erklärenden Variablen prognostiziert werden, bevor eine Prognose für die erklärte Variable abgegeben werden kann. Eine typische Prognose für die erklärte Variable eines Modells

¹⁷ Hier ist in den Prognoseformeln (IV6: 24b) bzw. (IV6: 25b) der Term $T + 1$ durch $T + h$ zu ersetzen, wobei h bestimmt, wie weit (wie viele Perioden) in die Zukunft prognostiziert werden soll.

¹⁸ Der Driftterm kann mit dem arithmetischen Mittel der ersten Differenzen von Y_t konsistent geschätzt werden.

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (\text{IV6: 28a})$$

hätte also hier die Form

$$\hat{Y}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_{t+1}. \quad (\text{IV6: 28b})$$

Feldstein, M. (1971) zeigt an einem einfachen Beispiel formal auf, dass bei einem korrekt spezifizierten Modell auch bei unbedingten Prognosen der *Erwartungswert* des Prognosefehlers bei Null liegt, jedoch die *Varianz* des Prognosefehlers im Vergleich zu bedingten Prognosen *höher* ist, wenn die Prognose der erklärenden Variable nicht exakt ist. Dies ist aber nicht verwunderlich, da schließlich durch die Schätzung der erklärenden Variable eine weitere Unsicherheits- und damit Fehlerquelle auftritt. Er zeigt außerdem, dass \hat{Y}_{t+1} *nicht mehr normalverteilt* ist. Dies hat zur Folge, dass wir keine Prognosekoeffizientenintervalle mehr ableiten können.¹⁹

Beispiel IV 6-5

Unbedingte Zinsprognose



Zur Veranschaulichung der Güte unbedingter Prognosen schätzen wir im Folgenden auf Basis des Datenmaterials aus Beispiel IV 6-1 zunächst ein einfaches AR(1)-Modell für die Inflationsrate für den Stützbereich von 1960 bis 2000 und prognostizieren damit die Inflationsrate für die Jahre 2001 bis 2008. Diese Prognosewerte verwenden wir dann, um unbedingte Prognosen für das Zinsniveau abzugeben. Konkret schätzen wir also für den Zeitraum 1960 bis 2000 die Modelle

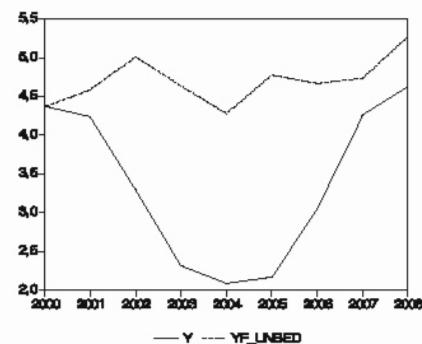
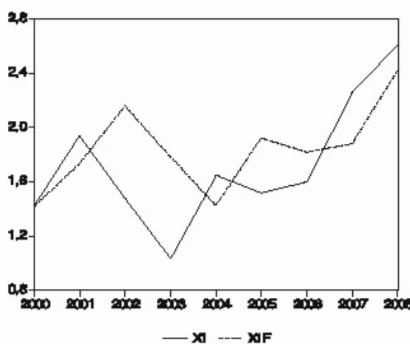
$$\text{Inflationsmodell: } X_{1t} = \alpha_0 + \alpha_1 X_{1t-1} + u_t$$

$$\text{Modell 1: } Y_t = \beta_0 + \beta_1 X_{1t} + \varepsilon_t$$

und verwenden dann die Prognosen aus dem Inflationsmodell in der Schätzung von Modell 1 zur Prognose des Zinsniveaus, d.h. für die erste Prognose außerhalb des Stützbereichs gilt

$$\hat{Y}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_{1t+1} + \varepsilon_t.$$

Betrachten wir nun sowohl die Güte des Prognosemodells für die Inflationsrate als auch die Güte der unbedingten Prognosen. Wir erhalten folgende Ergebnisse:



	Inflationsmodell	Modell 1 (unbedingt)
RMSE	0,4317	1,7031

¹⁹ Es lässt sich aber eine grobe Schätzung des Konfidenzintervalls generieren, wie z.B. Pindyck, R.S., Rubinfeld, D.L. (1998), Kapitel 8, zeigen.

Wir erkennen zunächst, dass das autoregressive Modell brauchbare Prognoseeigenschaften aufweist. Vergleichen wir die unbedingte Prognose mit der bedingten aus dem Beispiel IV 6-1 ($\text{RMSE}_{\text{Modell 1}} = 1,4938$), so hat sich wie erwartet die Prognosegüte durch die Voraussage der Inflationsrate verschlechtert. Modell 2, welches an Stelle der Schätzung eines AR(1)-Modells für die Inflationsrate direkt die um eine Periode verzögerte Inflationsrate als erklärende Variable verwendet, schneidet nur leicht schlechter ab, als das unbedingte Prognosemodell.

6.4 Zusammenfassung

Fassen wir zusammen, so hat uns Abschnitt IV 6 gezeigt, dass selbst Regressionsmodelle mit einem guten "fit" und signifikanten Parametern ungenaue Prognosen liefern können. Auch Autokorrelation des Störterms kann die Prognosegüte entscheidend negativ beeinflussen und sollte daher bei der Prognosebildung "berücksichtigt" werden (z.B. mittels FGLS oder dynamischer Modellierung durch Einbeziehung verzögerter Y und X). Ist eine Prognose der erklärenden Variablen erforderlich, so kann eine Prognose sogar noch unpräziser werden. Darüber hinaus ist zu bemerken, dass es durchaus möglich ist, dass ein Modell, welches gute bedingte Prognosen liefert, sich bei unbedingten Prognosen als mangelhaft erweist. Man sollte daher ein Modell mit hohen Prognosefehlern nicht generell ablehnen, wenn dies hauptsächlich auf die Prognose der erklärenden Variablen zurückzuführen ist. Deshalb sind *geeignete Modelle zur Untersuchung kausaler Zusammenhänge nicht notwendigerweise gute Prognosemodelle und umgekehrt*. Offensichtlich ist dies bei reinen Zeitreihenmodellen (sog. ARMA-Modelle), die die Entwicklung einer Variablen nur aus ihrer eigenen Vergangenheit heraus und evtl. weiteren deterministischen Variablen erklären.²⁰

²⁰ Für einen tiefergehenden Einblick in diese Modellklasse vgl. z.B. Diebold, F.X. (2004) oder Tsay, R.S. (2005).

7.

Aufgaben

OLS-Methodik

Aufgabe O1

Welche Transformationen sind erforderlich um die folgenden Funktionen mittels OLS schätzen zu können? Welche Funktionen sind nicht mit OLS schätzbar?

- | | |
|------------------------------------------------------|----------------------------------------------------------|
| a) $Y_i = \beta_0 + \beta_1 X_i^2 + \varepsilon_i$ | d) $Y_i = e^{\beta_0 + \beta_1 X_i} + \varepsilon_i$ |
| b) $Y_i = \beta_0 X_i^{\beta_1} + \varepsilon_i$ | e) $\ln Y_i = \beta_0 + \beta_1 \ln X_i + \varepsilon_i$ |
| c) $Y_i = e^{\beta_0 + \beta_1 X_i + \varepsilon_i}$ | f) $Y_i = \beta_0 X_i^{\beta_1} e^{\varepsilon_i}$ |

Aufgabe O2

Gegeben seien die folgenden 5 Beobachtungen von $Y_i = \{5, 2, 3, 4, 1\}$ und $X_i = \{3, 2, 1, 3, 2\}$. Schätzen Sie damit die Parameter der folgenden Regressionsmodelle sowohl "von Hand" als auch mit der Hilfe von MS Excel:

- a) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
- b) $Y_i = \beta_0 + \beta_1 \ln X_i + \varepsilon_i$
- c) $Y_i = \beta_0 + \beta_1 X_i^2 + \varepsilon_i$

Beurteilen Sie außerdem, welches Modell den größten Anteil der Variabilität von Y erklären kann!



Aufgabe O3

Nehmen Sie an, Ihnen ist bekannt, dass in einem einfachen linearen Regressionsmodell $\beta_0 = 0$ und $\beta_1 \neq 0$ gilt. Leiten Sie für diesen Fall den OLS-Schätzer für β_1 her! Führen Sie die gleiche Herleitung für den Schätzer von β_0 durch, wenn $\beta_1 = 0$ und $\beta_0 \neq 0$ gilt!

Aufgabe O4

Nehmen Sie an, Ihnen liegen aus einer einfachen Regression folgende Werte vor:

$$\begin{aligned}\sum (X_i - \bar{X})(Y_i - \bar{Y}) &= 655 & \sum (X_i - \bar{X})^2 &= 1.305 & n &= 8 \\ \sum X_i &= 475 & \sum Y_i &= 405\end{aligned}$$

Zeigen Sie mittels dieser Zahlen, dass die sich hier ergebende Regressionsgerade durch den Punkt (\bar{X}, \bar{Y}) verläuft!

Aufgabe O5

Nehmen Sie an, Sie wollen ein Regressionsmodell $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ schätzen. Welche Auswirkung hat

- bei unveränderten Y-Werten die Multiplikation aller X-Werte mit 10
 - bei unveränderten X-Werten die Multiplikation aller Y-Werte mit 10
- auf die Schätzungen $\hat{\beta}_0$ und $\hat{\beta}_1$ und die Schätzung für die Varianz des Störterms?

Aufgabe O6

Nehmen Sie an, Ihnen liegen aus einer einfachen Regression $n = 50$, $\bar{X} = 70,55$, die geschätzte Varianz des Störterms 2,05 und die geschätzte Varianz von $\hat{\beta}_1$ mit 0,001 vor. Wie hoch sind (a) die Residuenquadrate summe, (b) die Summe $\sum(X_i - \bar{X})^2$ und Varianz von X sowie (c) die Summe der quadrierten X_i ?

Aufgabe O7

Gegeben seien die Y-Werte 5, 2, 3, 2, -2 und die korrespondierenden X-Werte 3, 2, 1, -1, 0. Die OLS-Schätzung von $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ liefert $\hat{\beta}_0 = \hat{\beta}_1 = 1$. Schätzen Sie die Varianz des Störterms und von $\hat{\beta}_1$!

Aufgabe O8

Bestimmen Sie anhand der im Folgenden gegebenen Werte die jeweils gesuchte Größe!

- Gegeben: $\sum(Y_i - \bar{Y})^2 = 631,63$, $\sum e_i^2 = 182,85$
Gesucht: R^2
- Gegeben: $n = 20$, $\sum Y_i^2 = 5.930,94$, $\sum(Y_i - \bar{Y})^2 = 788,52$, $\bar{Y} = 16,035$,
ESS = 666,72
Gesucht: R^2
- Gegeben: $R^2 = 0,7911$, TSS = 552,36, $n = 20$, K = 1
Gesucht: $\hat{\sigma}^2$

Aufgabe O9

Nehmen Sie an, eine Regression liefert die folgenden Ergebnisse (Standardfehler in Klammern):

$$\hat{Y}_i = 5,83 + 0,869X_i \quad R^2 = 0,756 \\ (1,23) (0,117)$$

Wie würden die Regressionsergebnisse aussehen, wenn

- alle Werte von X vor der Schätzung durch 10 dividiert werden?
- alle Werte von Y vor der Schätzung durch 10 dividiert werden?
- alle Werte von X und Y vor der Schätzung durch 10 Dividiert werden?

Berücksichtigen Sie bei Ihren Antworten auch die Erkenntnisse aus Aufgabe O5!

**Aufgabe O10**

Von einer ökonomischen Variable Y_t für die Jahresdaten von 1950 bis 1997 vorliegen, wird vermutet, dass sie aufgrund technologischen Fortschritts in einer der drei folgenden funktionalen Formen positiv vom Zeitverlauf t (1, ..., 48) abhängt:

- $Y_t = \beta_0 + \beta_1 t + \varepsilon_t$
- $Y_t = \beta_0 + \beta_1 \ln(t) + \varepsilon_t$
- $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$

Schätzen Sie jede der drei Gleichungen mittels des Zahlenmaterials im Online-Service und beurteilen Sie anhand der grafischen Darstellung der geschätzten Funktion und der Residuen sowie mittels eines geeigneten Maßes, welche Gleichung zu präferieren ist!

Aufgabe O11

Es soll das Modell $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ mittels der Daten der nachfolgenden Tabelle "von Hand" geschätzt werden:

Y_i	X_{1i}	X_{2i}
1	0	1
2	1	-2
3	2	1
-1	-2	0
0	1	-1
-1	-2	-1
2	0	1
1	-1	1
2	1	0

- Schätzen Sie die Modellparameter β_0 , β_1 und β_2 !
- Schätzen Sie σ^2 und den Standardfehler von $\hat{\beta}_1$!
- Wie hoch sind RSS, TSS, ESS und R^2 ?

Aufgabe O12

Aus dem EViews-Output einer Regression mit zwei erklärenden Variablen können Sie die Werte "Included observations" = 40, "S.E. of regression" = 5,146058, "Sum squared resid" = 979,8306, "Mean dependent var" = 16,97628 und "S.D. dependent var" = 13,45222 entnehmen. Wie hoch sind die Varianz von Y , TSS und R^2 ?

Hypothesentests, Modellspezifikation und Multikollinearität

Aufgabe I1

- Ist eine OLS-Schätzung im einfachen Regressionsmodell bei $n = 20$ ($n = 100$) signifikant von Null verschieden, wenn ihr t-Wert bei 2 liegt und $\alpha = 0,05$ ist?
- Wie lauten jeweils die Testentscheidungen auf Basis der p-Werte 0,005 und 0,0108 bei einem zweiseitigen Signifikanztest ($\alpha = 0,05$ und $\alpha = 0,01$)?

Aufgabe I2

Hanson, G.H. und Spilimbergo, A. (1999) untersuchten mittels Monatsdaten von Januar 1968 bis August 1996 den Effekt des mexikanischen Lohnniveaus und von Grenzkontrollen auf die Anzahl aufgegriffener illegaler Einwanderer an der Grenze zwischen den USA und Mexiko.

- Sind A_t die Anzahl illegaler Einwanderer, die im Monat t von der US-Grenzkontrolle festgenommen werden, und E_t die Personenstunden, die im Monat t zur Grenzkontrolle aufgewendet werden, so wurde folgendes Schätzergebnis erzielt, wobei die Schätzung der Konstanten im Papier nicht aufgeführt wurde:

$$\ln\left(\frac{A_t}{A_{t-1}}\right) = \hat{\beta}_0 + 0,510 \cdot \ln\left(\frac{E_t}{E_{t-1}}\right) \quad \hat{\sigma}_{\hat{\beta}_1} = 0,126$$

Interpretieren Sie diese Gleichung und testen Sie die Hypothese $H_0: \beta_1 = 0$ gegenüber $H_A: \beta_1 \neq 0$ bei einer Irrtumswahrscheinlichkeit von 5 %! Konstruieren Sie außerdem ein 95 %-Konfidenzintervall für β_1 und interpretieren Sie dieses!

- Bezeichnet man mit L_t das reale Stundenlohniveau im mexikanischen Manufaktursektor, so ergab die Studie zudem folgendes Ergebnis:

$$\ln\left(\frac{A_t}{A_{t-1}}\right) = \hat{\beta}_0 - 0,550 \cdot \ln\left(\frac{L_t}{L_{t-1}}\right) \quad \hat{\sigma}_{\hat{\beta}_1} = 0,169$$

Interpretieren Sie diese Gleichung und testen Sie bei einer Irrtumswahrscheinlichkeit von 5 % die Hypothese $H_0: \beta_1 \geq 0$ gegenüber $H_A: \beta_1 < 0$!

Aufgabe I3

Eine von Hill, R.C., Griffiths, W.E. und Judge, G.G. (2001) durchgeföhrte Querschnittsregression $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ mittels derer versucht wird, das durchschnittliche Einkommen Y (in Tsd. Dollar) volljähriger Männer in den 50 amerikanischen Staaten und dem District Columbia durch den prozentualen Anteil volljähriger Männer (im jew. Staat) mit Hochschulabschluss X zu erklären, lieferte Folgendes:

Dependent Variable: Y

Sample: 1 51

Method: Least Squares

Included Observations: 51

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	(a)	2,173807	1,257	(c)
X	0,180141	(b)	5,754	0,0001

- Bestimmen Sie die fehlenden Werte (a) bis (c) in obiger Übersicht!

- b) Konstruieren Sie ein 99 %-Konfidenzintervall für den Steigungsparameter!
- c) Testen Sie die Hypothese, dass der Steigungsparameter 0,2 ist, gegenüber der Hypothese, dass er von 0,2 verschieden ist ($\alpha = 0,05$)! Interpretieren Sie Ihr Ergebnis im Hinblick auf seine ökonomische Bedeutung!

Aufgabe I4

Caulkins, J.P. und Padman, R. (1993) untersuchten u.a. die Entwicklung des Kokainpreises Y (Dollar je Gramm) in Kalifornien von 1984 bis 1991 mittels des einfachen Modells $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$, in dem X_1 die Grammmzahl je Verkauf, X_2 die Qualität des Kokains (ausgedrückt als prozentuale Reinheit) und eine Zeitvariable X_3 mit 1984 = 1, ..., 1991 = 8 darstellen.



- a) Nutzen Sie die Monatsdaten aus dem Online-Service (Auszug aus dem der Studie zugrunde liegenden Datensatz) zur Schätzung des obigen Modells! Haben alle Koeffizienten das erwartete Vorzeichen?
- b) Wieviel Prozent der Variation des Kokainpreises wird durch die drei gewählten Variablen erklärt?
- c) In Anbetracht der Tatsache, dass bei größerer Verkaufshäufigkeit das Risiko steigt, von der Polizei gefasst zu werden, könnte man behaupten, dass die Verkäufer bei größeren Mengen niedrigere Preise verlangen. Bilden Sie die Hypothesen, mit denen sich eine derartige Behauptung testen lässt, und führen Sie den Test durch!
- d) Testen Sie die Hypothese, dass die Kokainqualität keinen Einfluss auf den Preis hat, gegenüber jener, dass es eine Qualitätsprämie auf dem "Markt" gibt!
- e) Wie hoch ist die durchschnittliche jährliche Veränderung des Kokainpreises? Begründen Sie die Richtung der Veränderung!

Aufgabe I5



Schätzen Sie mittels des fiktiven Zahlenmaterials im Online-Service die folgende Produktionsfunktion $Y_t = \beta_0 X_{1t}^{\beta_1} X_{2t}^{\beta_2} e^{\varepsilon_t}$, in der Y das Bruttonsozialprodukt, X_1 den Arbeits- und X_2 den Kapitaleinsatz darstellen und testen Sie die folgenden Hypothesen bei einem Signifikanzniveau von 1 %:

- a) $H_0: \beta_1 = 0$ gegenüber $H_A: \beta_1 \neq 0$
- b) $H_0: \beta_1 = 0, \beta_2 = 0$ gegenüber $H_A: \beta_1 \neq 0$ und/oder $\beta_2 \neq 0$
- c) $H_0: \text{"Konstante Skalenerträge"}$ gegenüber $H_A: \text{"Keine konstanten Skalenerträge"}$, d.h. $H_0: \beta_1 + \beta_2 = 1$ gegenüber $H_A: \beta_1 + \beta_2 \neq 1$

Untersuchen Sie auch den Einfluss von Multikollinearität in diesem Modell und beurteilen Sie das geschätzte Modell im Hinblick auf Ihre Ergebnisse!

Aufgabe I6

In ein Modell $Y_t = \beta_0 + \beta_1 X_{1t} + \varepsilon_t$ wurde eine überflüssige Variable X_2 integriert, so dass ein neues Modell $Y_t = \beta_0 + \beta_1' X_{1t} + \beta_2' X_{2t} + \varepsilon_t'$ mit $\varepsilon_t' = \varepsilon_t - \beta_2' X_{2t}$ resultiert (vgl. dazu Abschnitt IV 4.1.2).

Zeigen Sie formal, dass dies keine Verzerrung des OLS-Schätzers des Steigungsparameters von X_1 zur Folge hat, d.h. dass $E(\hat{\beta}_1') = \beta_1$ gilt! Gehen Sie dabei von (IV1: 18) aus und beachten Sie, dass $\sum y_i x_{1i} = \sum Y_i x_{1i}$ gilt.

Aufgabe I7

Die Schätzung eines Modells $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ mit 20 Beobachtungen er gab neben $\hat{\sigma}^2 = 2,5193$ und $R^2 = 0,9466$ folgende Parameterschätzungen und ge schätzte Varianz-Kovarianz-Matrix:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0,9659 \\ 0,6991 \\ 1,7769 \end{bmatrix} \quad \widehat{\text{Cov}} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0,2181 & 0,0192 & -0,0503 \\ 0,0192 & 0,0485 & -0,0312 \\ -0,0503 & -0,0312 & 0,0371 \end{bmatrix}$$

- Bestimmen Sie ein Konfidenzintervall für β_1 und testen Sie die Nullhypothese $H_0: \beta_1 = 1$ gegenüber $H_A: \beta_1 \neq 1$! Gehen Sie dabei jeweils von einer Irrtumswahr scheinlichkeit von 5 % aus!
- Testen Sie mittels eines F-Tests die Hypothese $H_0: \beta_1 = \beta_2 = 0$ gegenüber $H_A: \text{"mindestens ein } \beta \text{ ist von Null verschieden!"}$

Aufgabe I8

Nehmen Sie an, die Schätzung eines Modells $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ mittels 63 Beobachtungen liefere folgende Ergebnisse:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} \quad \widehat{\text{Cov}} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

Testen Sie mittels dieser Daten und geeigneter t-Test-Variationen die folgenden Nullhypotesen bei $\alpha = 0,05$:

- $\beta_1 = 0$
- $\beta_0 + \beta_1 = 4$
- $\beta_1 + 2\beta_2 = 2$

Aufgabe I9

Schätzen Sie mittels der fiktiven Daten im Online-Service die folgende einfache Kostenfunktion $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i$ für 28 Unternehmen einer Branche, in der Y_i die Gesamtkosten der i -ten Unternehmung und X_i die gefertigten Stückzahlen beschreiben.

- Testen Sie die Hypothese, dass eine lineare Kostenfunktion für die Beschreibung der Daten ausreichend ist!
- Testen Sie, ob eine quadratische Kostenfunktion die Daten ausreichend beschreibt!
- Schätzen Sie die logarithmische Kostenfunktion $\ln Y_i = \beta_0 + \beta_1 \ln X_i + \varepsilon_i$! Führen Sie für diese und die Funktion $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i$ RESET-Tests (Einbeziehung von einer und zwei Proxyvariablen \hat{Y}^2 bzw. \hat{Y}^3) durch und beurteilen Sie, ob die logarithmische Funktion der polynomiauen vorzuziehen ist! Ist eine logarithmische Kostenfunktion unter ökonomischen Gesichtspunkten sinnvoll?
- Begründen Sie, warum im Modell, welches beim RESET-Test geschätzt wird, nicht auch \hat{Y}_i als Proxyvariable einbezogen wird?

Dummy-Variablen

Aufgabe D1

Ein jüngst gegründeter Ökonometrielehrstuhl einer Universität interessiert sich dafür, ob es den Studenten beim Berufseinstieg einen finanziellen Vorteil bringt, wenn sie das Fach Ökonometrie belegt haben. Mit Datenmaterial zu den letzten 50 Absolventen (vgl. Online-Service) soll daher ein Modell $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_{1i} + \epsilon_i$ geschätzt werden, wobei Y für das Einstiegsgehalt in US-Dollar, X für den Notendurchschnitt auf einer amerikanischen 4,0-Skala (d.h. je höher der Notenwert, desto besser) und D_1 für eine Dummy-Variable steht, die den Wert 1 annimmt, wenn der Student i Ökonometrie belegt hatte, und andernfalls 0 ist.

- Nutzen Sie das Datenmaterial des Online-Services zur Schätzung dieses Modells! Interpretieren Sie die ökonomische Bedeutung Ihrer Schätzergebnisse!
- Nehmen Sie an, es liegt eine Dummy-Variable D_2 vor, die den Wert 1 annimmt, wenn der betrachtete Student weiblich ist, und andernfalls 0 ist. Wie würden Sie das gegebene Modell modifizieren, wenn Sie untersuchen möchten, ob Frauen ein niedrigeres Einstiegsgehalt als Männer aufweisen? Schätzen Sie Ihr neues Modell und interpretieren Sie die Ergebnisse!
- Wie ist das Ausgangsmodell zu modifizieren, wenn untersucht werden soll, ob die Belegung des Fachs Ökonometrie für Männer und Frauen den gleichen finanziellen Vorteil bringt? Schätzen und interpretieren Sie Ihr neues Modell!



Aufgabe D2

Ein Modell zur Erklärung der Preise Y von Einfamilienhäusern in einer ländlichen Region im Zeitraum von 2000 bis 2004 liefere folgende Schätzergebnisse:

$$\hat{Y}_i = 21.456 + 1.900X_{1i} - 1.000X_{2i} - 4.393D_{1i} - 10.435D_{2i} - 13.174D_{3i} - 19.041D_{4i}$$

Alle Parameter sind signifikant. X_1 und X_2 beschreiben die Quadratmeterzahl und das Alter der Gebäude. D_1 bis D_4 sind Jahres-Dummies (Verkaufsjahr) für die Jahre 2001 bis 2004, wobei das Jahr 2000 als Referenzkategorie anzusehen ist.

- Interpretieren Sie die geschätzten Koeffizienten des Modells!
- Was passiert bei der Schätzung des Modells, wenn auch für das Jahr 2000 eine Dummy-Variable ins Modell integriert wird?



Aufgabe D3

Nehmen Sie an, aus einer Datenerhebung in zwei Wohngebieten, von denen eines in der Nähe einer Universität und das andere etwas weiter von dieser entfernt liegt, resultieren 1.000 Beobachtungen für die Variablen Y: Preis der Immobilie in Euro, X_1 : Wohnfläche in m^2 , X_2 : Alter der Immobilie in Jahren und die Dummyvariablen D_1 : Universitätsnähe (1 = hoch, 0 gering), D_2 : Pool (1 = ja, 0 = nein), D_3 : Kamin (1 = ja, 0 = nein).

- Schätzen Sie mittels der fiktiven Daten im Online-Service das Modell

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 X_{1i} + \beta_3 X_{1i} D_{1i} + \beta_4 X_{2i} + \beta_5 D_{2i} + \beta_6 D_{3i} + \epsilon_i$$

und interpretieren Sie die Ergebnisse!

- b) In empirischen Studien wird häufig der logarithmierte Immobilienpreis als erklärende Variable verwendet, d.h. linksseitige Semi-Log-Formen bei der Modellbildung herangezogen. Schätzen Sie das Modell aus a) mit einer derartigen Modifikation und interpretieren Sie die Schätzungen der Parameter von X_1 und X_2 !
- c) Berechnen Sie die durchschnittliche prozentuale Immobilienpreisänderung, die durch die Präsenz eines Pools entsteht!
- d) Berechnen Sie die durchschnittliche prozentuale Immobilienpreisänderung von Häusern mit einer Fläche von 2.500 m^2 , die durch Universitätsnähe entsteht!

Heteroskedastizität und Autokorrelation

Aufgabe HA1

Betrachten Sie ein Modell $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, in dem Y_i für die Lebensmittelausgaben und X_i für das Einkommen des i-ten Haushalts stehen.

- Schätzen Sie dieses Modell mittels des fiktiven Datenmaterials aus dem Online-Service und testen Sie mit Hilfe des White-Tests auf Heteroskedastizität!
- Schätzen Sie das Modell unter den Annahmen, dass $\text{Var}(\varepsilon_i)$ gleich

- $\sigma^2 \ln X_i$
- $\sigma^2 X_i$
- $\sigma^2 X_i^2$
- $\sigma^2 \sqrt{X_i}$

ist, und beurteilen Sie die Sensitivität der Parameterschätzungen und der OLS-Standardfehler auf Änderungen der Heteroskedastizitätsspezifikation! Nutzen Sie außerdem für jede Spezifikation den White-Test um zu prüfen, ob die Heteroskedastizität eliminiert werden konnte!



Aufgabe HA2

Dougerty, C. (1992) spezifiziert zur Erklärung der Bildungsausgaben pro Kopf Y durch das BIP pro Kopf X das einfache Regressionsmodell $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ und schätzt es mit Daten zu 34 Ländern aus dem Jahr 1980 (vgl. Online-Service).



- Warum ist hier zu befürchten, dass die Störtermvarianz heteroskedastisch ist?
- Schätzen Sie das Modell mit OLS und untersuchen Sie eine grafische Darstellung der quadrierten Residuen auf Anzeichen für Heteroskedastizität?
- Testen Sie mittels des Breusch-Pagan LM-Tests auf Heteroskedastizität!
- Konstruieren Sie sowohl mit den Standardfehlern aus b) als auch mit den von White vorgeschlagenen heteroskedastizitätsrobusten Standardfehlern ein 95 %-Konfidenzintervall für β_1 ! Beurteilen Sie das Konfidenzintervall, welches der Heteroskedastizität nicht Rechnung trägt!
- Schätzen Sie das Modell unter der Annahme $\text{Var}(\varepsilon_i) = \sigma^2 X_i$ und konstruieren Sie erneut ein 95 %-Konfidenzintervall für β_1 ! Was können Sie über die Intervallbreite im Vergleich zu denen aus d) aussagen?

Aufgabe HA3



Bei einer Erhebung in 38 Haushalten einer Kleinstadt konnten die Lebensmittelausgaben Y (in Tausend Euro), das Haushalteinkommen X_1 (in Tausend Euro) und die Anzahl der im Haushalt lebenden Personen X_2 festgestellt werden.

- Schätzen Sie das Modell $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ mit den fiktiven Zahlen im Online-Service und interpretieren Sie Ihre Ergebnisse!
- Gibt eine grafische Analyse der quadrierten Residuen Hinweise auf Heteroskedastizität? Untermauern Sie Ihre Ergebnisse durch den White-Test und den Breusch-Pagan LM-Test!
- Bestimmen Sie die White-Standardfehler für die Modellschätzung! Kam es zu einer Über- oder Unterschätzung der Präzision der OLS-Schätzer?
- In Abhängigkeit welcher Variablen würden Sie $\text{Var}(\varepsilon_i)$ im Rahmen einer WLS-Schätzung des vorliegenden Modells modellieren? Warum ist es häufig besser die Verwendung von White-Standardfehlern der WLS-Schätzung vorzuziehen?

Aufgabe HA4

Betrachten Sie das Modell $\ln Y_t = \beta_0 + \beta_1 \ln X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t$, welches das logarithmierte Investitionsaufkommen $\ln Y$ Deutschlands durch das logarithmierte BIP $\ln X_1$ und das Zinsniveau X_2 in Prozent (jeweils real) zwischen 1955 und 2009 erklären soll.

- Schätzen Sie das Modell mit den im Online-Service zu findenden Daten der Deutschen Bundesbank und interpretieren Sie die Ergebnisse!
- Gibt eine grafische Darstellung der Residuen Hinweise auf Autokorrelation?
- Testen Sie mit dem Durbin-Watson-Test auf positive Autokorrelation erster Ordnung ($\alpha = 0,01$)!
- Beschreiben Sie allgemein, welche Konsequenzen Autokorrelation für das gegebene Modell nach sich zieht und geben Sie Möglichkeiten an, Autokorrelation erster und höherer Ordnung Rechnung zu tragen!

Aufgabe HA5

Das sog. Erfahrungskurvenkonzept besagt, dass Personal im Zeitverlauf Erfahrung sammelt und damit bei der Arbeitserfüllung effizienter wird. Unmittelbare Folge daraus ist, dass immer weniger Arbeitsstunden und damit weniger Kosten erforderlich werden, um ein Produkt herzustellen. Zur empirischen Überprüfung dieses Konzepts wird in der Praxis meist das einfache Modell $\ln Y_t = \beta_0 + \beta_1 \ln X_{t-1} + \varepsilon_t$ verwendet (vgl. z.B. Berndt, E. (1991)), in dem Y_t die Stückkosten in t und X_{t-1} die kumulative Produktionsmenge bis t - 1 darstellt.

- Schätzen Sie dieses Modell mit dem fiktiven Datenmaterial im Online-Service und interpretieren Sie die Ergebnisse!
- Gibt die Durbin-Watson-Statisik Hinweise auf Autokorrelation erster Ordnung? Zu welchem Ergebnis kommt der Breusch-Godfrey LM-Tests beim Test auf Autokorrelation erster Ordnung? Verwenden Sie jeweils $\alpha = 0,05$!
- Führen Sie eine GLS-Schätzung durch, welche Autokorrelation erster Ordnung Rechnung trägt, und interpretieren Sie die Ergebnisse! Wiederholen Sie Ihre Be- trachtung aus b) für die Ergebnisse der GLS-Schätzung!
- Vergleichen Sie die 95 %-Konfidenzintervalle, die
 - aus den Standard-OLS-Ergebnissen,
 - unter Verwendung von Newey-West-Standardfehlern und
 - einer GLS-Schätzung zur Berücksichtigung eines AR(1)-Störterms für β_1 resultieren!
- Nehmen Sie die vorliegenden Zeitreihen aus dem Online-Service näher in Augenschein und beurteilen Sie ihre bisherigen Schätzergebnisse im Hinblick auf den vorliegenden Stichprobenumfang!
- Bei kritischer Betrachtung der hier verwendeten Regressionsfunktion ist es schwer vorstellbar, dass diese für eine Erklärung von Stückkosten wirklich geeignet ist. Vielmehr ist es wahrscheinlich, dass eine Fehlspezifikation vorliegt. Beurteilen Sie Ihre bisherigen Schätzergebnisse vor dem Hintergrund dieses Arguments und geben Sie beispielhaft an, wie sich das vorliegende Modell verbessern ließe!



Instrumentenvariablenabschätzung

Aufgabe IV1

Mittels der Ergebnisse einer Befragung von 50 zufällig ausgewählten Arbeitnehmern soll ein Modell $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ geschätzt werden, das die Ersparnisse Y in einem bestimmten Jahr durch das Einkommen X in diesem Jahr erklären soll.



- Schätzen Sie dieses Modell mittels der fiktiven Daten des Online-Services mit OLS und interpretieren Sie die Ergebnisse?
- Die Variable Z beschreibt das durchschnittliche Einkommen einer jeden befragten Person über die letzten 10 Jahre (ohne das aktuelle Jahr) und kann im Zuge einer IV-Schätzung als Instrument herangezogen werden. Nehmen Sie an, die Variable X wurde bei der Stichprobenerhebung z.T. fehlerhaft erfasst! Wie lauten die Ergebnisse einer IV-Schätzung unter Verwendung des Instruments Z? Interpretieren Sie die Ergebnisse!
- Beurteilen Sie mittels des Hausman-Tests, ob die erklärende Variable und der Störterm des betrachteten Modells korreliert sind und damit die OLS-Schätzung unter a) verzerrt war!

Aufgabe IV2

Ein nützliches Hilfsmittel in der Ökonometrie ist die Monte Carlo Simulation. Bei ihr werden aus einem bekannten Modell künstliche Daten erzeugt, die es erlauben, alternative Schätzverfahren unter verschiedenen Bedingungen zu vergleichen.



Spezifizieren wir ein einfaches Regressionsmodell mit $\beta_0 = 3$ und $\beta_1 = 1$, so gilt für seine systematische Komponente $E(Y_i | X_i) = 3 + 1 \cdot X_i$. Nehmen wir nun an, X ist mit einem Erwartungswert von Null und einer Varianz von 2 normalverteilt, der Störterm ε ist standardnormalverteilt und die Kovarianz zwischen X und ε sei 0,9. Wir können auf Basis dieser Annahmen zunächst zufällige Werte für X und ε erzeugen und über $Y_i = 3 + 1 \cdot X_i + \varepsilon_i$ Stichprobenbeobachtungen von Y erzeugen. Im Online-Service finden Sie 500 Beobachtungen zu den Variablen X und Y, die auf diese Weise künstlich generiert wurden.

- Stellen Sie die 500 Beobachtungen inkl. der Regressionsfunktion der Grundgesamtheit grafisch dar! Was stellen Sie fest?
- Schätzen Sie das Modell $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ bei den Stichprobenumfängen $T = 10$, $T = 20$, $T = 100$ und $T = 500$ Beobachtungen (d.h. den ersten 10, 20, ... Beobachtungen) mittels OLS! Was stellen Sie im Hinblick auf die Veränderung der Parameterschätzungen fest?
- Die im Online-Service zu findenden Variablen Z_1 und Z_2 wurden unter den Bedingungen, dass sie jeweils standardnormalverteilt, mit X korreliert (konkret $Cov(X, Z_1) = 0,9$ und $Cov(X, Z_2) = 0,4$) und mit ε unkorreliert sind, ebenfalls zufällig erzeugt. Führen Sie eine IV-Schätzung des Modells $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ unter alleiniger Nutzung von Z_1 , alleiniger Nutzung von Z_2 und gemeinsamer Nutzung von Z_1 und Z_2 als Instrumente für $T = 10$, $T = 20$, $T = 100$ und $T = 500$ durch! Was können Sie nun zu den Veränderungen der Parameterschätzungen aussagen? Welche Instrumentenkonstellation ist zu bevorzugen?

Aufgabe IV3

Ein Beratungsunternehmen wurde damit betraut, die Effizienz von 75 Brauereien in einem bestimmten Jahr zu untersuchen. Dazu schätzt es ein Modell $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \epsilon_t$, in dem Y für einen Bieroutputindex steht, der sowohl Produktionsmenge als auch -qualität berücksichtigt. Y wird durch die Berufserfahrung X_1 des Braumeisters (in Jahren), einen Index für Kapitaleinsatz X_2 und einen Index für Arbeitseinsatz X_3 erklärt.

- Schätzen Sie das Modell mittels des fiktiven Datenmaterials im Online-Service und beurteilen Sie die Ergebnisse! Nutzen Sie die Ergebnisse für eine Prognose des Outputs Y einer neu gegründeten Brauerei, die den Stichprobendurchschnitt des Kapital- und Arbeitseinsatzes aufweist sowie von einem Braumeister geführt wird, der 10 Jahre Berufserfahrung besitzt!
- Das Beratungsunternehmen befürchtet, dass die Variable X_1 mit dem Störterm des Modells korreliert ist und führt daher einen Hausman-Test mit der Variable Z (Alter des Braumeisters) als Instrument für X_1 durch. Die übrigen Variablen werden als exogen angenommen. Zu welchem Resultat kommt der Test?
- Führen Sie eine IV-Schätzung mit dem Instrument Z durch und geben Sie unter Verwendung der neuen Schätzergebnisse eine korrigierte Version der Prognose aus a) an!

Aufgabe IV4

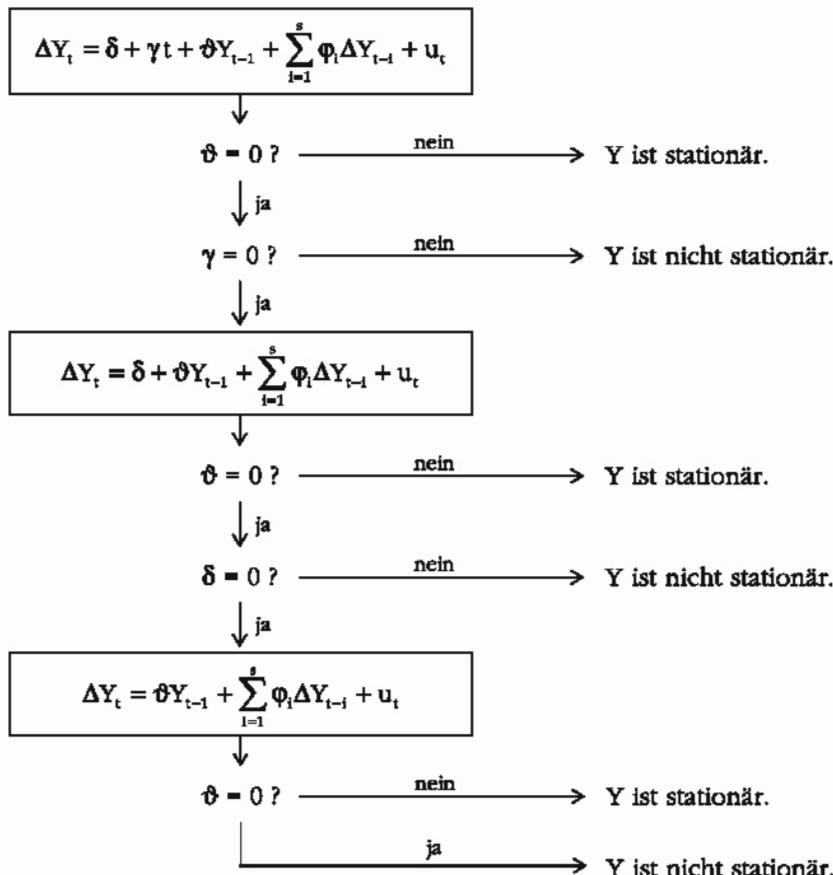
Im Online-Service finden Sie Daten zum realen Pro-Kopf-Konsum Y und zum realen Pro-Kopf-Einkommen X der USA vom 1. Quartal 1947 bis zum 4. Quartal 1980. Zur Erklärung des Pro-Kopf-Konsums durch das Pro-Kopf-Einkommen in der aktuellen Periode und in weiter zurückliegenden Perioden ließe sich ein distributives Lag-Modell der Form $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \dots + \epsilon_t$ spezifizieren. Aufgrund der Probleme dieser Modellklasse empfiehlt sich jedoch die Schätzung eines durch Koyck-Transformation resultierenden ARDL-Modells $Y_t = \gamma_0 + \beta_0 X_t + \gamma_1 Y_{t-1} + u_t$.

- Schätzen Sie das genannte ARDL-Modell mit OLS und testen Sie mit einem geeigneten Testverfahren auf Autokorrelation erster Ordnung! Begründen Sie die Wahl des Testverfahrens und beschreiben Sie die Folgen des Testergebnisses für die OLS-Schätzung!
- Führen Sie eine IV-Schätzung des ARDL-Modells mit dem um eine Periode verzögerten Pro-Kopf-Einkommen als Instrument für Y_{t-1} durch! Wiederholen Sie die IV-Schätzung und verwenden Sie nun als Instrumente für Y_{t-1} sowohl das um eine als auch das um zwei Perioden verzögerte Pro-Kopf-Einkommen! X_t instrumentiert sich in beiden Fällen selbst. Vergleichen Sie die Ergebnisse!
- Begründen Sie, warum die unter b) vorgenommene Instrumentierung problematisch für die Konsistenz der IV-Schätzer sein könnte!

Regression mit Zeitreihendaten

Aufgabe ZR1

Nach Frenkel, Funke und Koske (2003) empfiehlt sich zur Begrenzung des Risikos einer falschen Spezifikation der Testgleichung beim Dickey-Fuller Stationaritätstest ein sequenzielles Vorgehen, bei dem zunächst mit der allgemeinsten Spezifikation begonnen und anschließend auf Basis der Testergebnisse eine sukzessive Modelleinschränkung erfolgt. Diese Methodik ist durch folgendes Schema beschrieben:



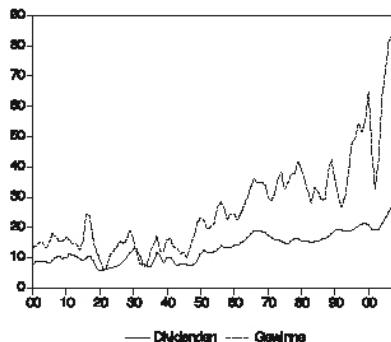
Im Online-Service finden Sie einen Datensatz mit wesentlichen Kenngrößen zur wirtschaftlichen Entwicklung der USA (Quellen: US Department of Commerce, Bureau of Economic Analysis sowie Homepage von R. Shiller). Zu diesen zählen die realen Konsumausgaben (Y_1) und das reale verfügbare Einkommen (Y_2) von Privatpersonen für die Jahre 1950 bis 2008 sowie die realen Dividenden (Y_3) und der reale Gewinn (Y_4) der Unternehmen des S&P 500 für die Jahre 1900 bis 2008.

Nutzen Sie das beschriebene sequenzielle Vorgehen um die Reihen auf Stationarität zu testen! Geben Sie außerdem den Integrationsgrad der einzelnen Reihen an!

Aufgabe ZR2

Betrachten Sie erneut den Datensatz aus Aufgabe ZR1! Es soll damit ein einfaches Regressionsmodell $Y_3 = \beta_0 + \beta_1 Y_4 + \varepsilon_i$ geschätzt werden, welches die Dividenden (Y_3) durch die Gewinne (Y_4) erklärt.

- a) Schätzen Sie das Modell! Kann es dabei zu einer typischen Scheinregression kommen? Begründen Sie Ihre Aussagen durch Rückgriff auf Ihre Ergebnisse bei Aufgabe ZR1 und die nachfolgende Grafik!



- b) Sind Dividenden und Gewinne bei Zulassen einer Irrtumswahrscheinlichkeit von 10 % kointegriert? Wie ist das geschätzte Modell aus a) zu interpretieren, wenn Sie beim Test eine Kointegrationsbeziehung feststellen?
- c) Schätzen Sie (sofern möglich) zur Beurteilung der Kurz- und Langfristbeziehung zwischen Dividenden und Gewinnen ein Fehlerkorrekturmödell! Interpretieren Sie Ihre Ergebnisse!

V LÖSUNGEN

1.

Kapitel I – Deskriptive Statistik

Aufgabe G1

- | | |
|-------------------|-------------------|
| a) Bewegungsmasse | d) Bestandsmasse |
| b) Bewegungsmasse | e) Bewegungsmasse |
| c) Bestandsmasse | |

Aufgabe G3

- | | |
|--------------------|-----------------|
| a) Verhältnisskala | d) Nominalskala |
| b) Intervallskala | e) Ordinalskala |
| c) Ordinalskala | f) Ordinalskala |

Aufgabe EH1

$$\bar{x}_{\text{geo}} = \sqrt[2]{\left(1 + \frac{160 - 100}{100}\right) \cdot \left(1 + \frac{80 - 160}{160}\right)} = \sqrt[2]{1,6 \cdot 0,5} = 0,8944$$

$$\rightarrow \text{Wachstumsrate} = 0,8944 - 1 = -0,1056 = -10,56 \%$$

Aufgabe EH3

$$\text{Arbeitslosenquote}_{\text{gesamt}} = \frac{7,806 \cdot 0,174 + 32,505 \cdot 0,078}{7,806 + 32,505} = 0,0966 = 9,66 \%$$

Aufgabe EH5

Verteilung A:

$$\bar{x}_A = \frac{1}{10} \cdot (14 \cdot 10) = 14, \quad s_A^2 = \frac{1}{10} \cdot (14^2 \cdot 10) - 14^2 = 0 \rightarrow s_A = 0, \quad SP_A = 14 - 14 = 0$$

Verteilung B:

$$\bar{x}_B = \frac{1}{10} \cdot (12 \cdot 1 + 13 \cdot 2 + 14 \cdot 4 + 15 \cdot 2 + 16 \cdot 1) = 14,0$$

$$s_B^2 = \frac{1}{10} \cdot (12^2 \cdot 1 + 13^2 \cdot 2 + 14^2 \cdot 4 + 15^2 \cdot 2 + 16^2 \cdot 1) - 14,0^2 = 1,20 \rightarrow s_B = 1,01$$

$$SP_B = 16 - 12 = 4$$

Verteilung C:

$$\bar{x}_C = \frac{1}{10} \cdot (9 \cdot 2 + 12 \cdot 2 + 13 \cdot 1 + 15 \cdot 1 + 16 \cdot 1 + 17 \cdot 1 + 18 \cdot 1 + 19 \cdot 1) = 14$$

$$s_c^2 = \frac{1}{10} \cdot (9^2 \cdot 2 + 12^2 \cdot 2 + 13^2 \cdot 1 + 15^2 \cdot 1 + 16^2 \cdot 1 + 17^2 \cdot 1 + 18^2 \cdot 1 + 19^2 \cdot 1) - 14^2 = 11,40$$

$$\rightarrow s_c = 3,38$$

$$SP_c = 19 - 9 = 10$$

Da Verteilung A nur die Beobachtung $a_1 = 14$ mit Häufigkeit $h(a_1) = 10$ besitzt, ist sie naturgemäß symmetrisch. Für Verteilung B treten die Merkmalsausprägungen, die sich jeweils nur um 1 unterscheiden, mit den Häufigkeiten 1, 2, 4, 2, 1 auf, die ein klar symmetrisches Muster zeigen. Verteilung C weist eine leicht rechtsschiefe (bzw. linkssteile) Tendenz auf, da die Häufigkeiten im linken Bereich der Verteilung größer sind als im rechten.

Aufgabe EH7

$$\bar{x}_M = \frac{1}{7} \cdot (0,082 + 0,091 + 0,079 + 0,070 + 0,084 + 0,092 + 0,089) = 0,0839$$

$$s_M = \sqrt{\frac{1}{7} \cdot (0,082^2 + 0,091^2 + 0,079^2 + 0,070^2 + 0,084^2 + 0,092^2 + 0,089^2) - 0,0839^2} = 0,0067$$

$$VC_M = \frac{0,0067}{0,0839} = 0,0799$$

$$\bar{x}_F = \frac{1}{7} \cdot (0,052 + 0,049 + 0,032 + 0,041 + 0,052 + 0,061 + 0,059) = 0,0494$$

$$s_F = \sqrt{\frac{1}{7} \cdot (0,052^2 + 0,049^2 + 0,032^2 + 0,041^2 + 0,052^2 + 0,061^2 + 0,059^2) - 0,0494^2} = 0,0095$$

$$VC_F = \frac{0,0095}{0,0494} = 0,1923$$

Anhand der Variationskoeffizienten kann ausgesagt werden, dass die Alkoholkonzentration bei den Frauen stärker streut als bei den Männern.

Aufgabe EH9

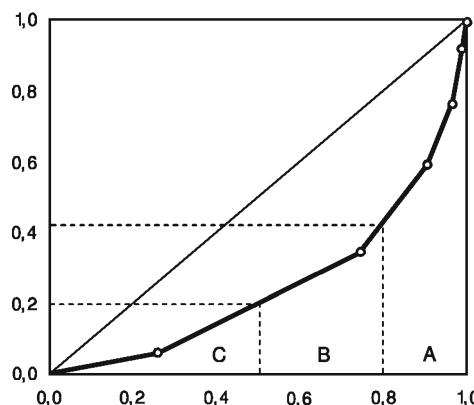
Mittels der folgenden Tabelle lassen sich der gesuchte Mittelwert 37,50, Varianz 468,75, Standardabweichung 21,65 und der Variationskoeffizient 0,58 bestimmen:

j	a_j^m	h_j	f_j	$f_j \cdot a_j^m$	$f_j \cdot (a_j^m)^2$
1	10	20	0,20	2,00	20,00
2	20	10	0,10	2,00	40,00
3	30	30	0,30	9,00	270,00
4	40	5	0,05	2,00	80,00
5	50	5	0,05	2,50	125,00
6	60	10	0,10	6,00	360,00
7	70	20	0,20	14,00	980,00
Σ	-	100	1,00	37,50	1.875,00

Aufgabe EH11

j	f _j	p _j	u _j	v _j
1	0,26	0,06	0,26	0,06
2	0,49	0,29	0,75	0,35
3	0,17	0,25	0,91	0,59
4	0,06	0,18	0,97	0,77
5	0,02	0,16	0,99	0,93
6	0,01	0,07	1,00	1,00

Mit diesen Werten können wir die auf der Folgeseite abgebildete Lorenz-Kurve skizzieren. Der Grafik ist zu entnehmen, dass über 55 % der Gesamtbeschaffungskosten auf diejenigen 20 % der Rechnungen mit den größten Beträgen entfallen. A-Güter machen also 55 % der Gesamtbeschaffungskosten aus. Aufgrund dieses hohen Kostenanteils ist ihnen bei der Planung zukünftiger Beschaffungsmaßnahmen besondere Aufmerksamkeit zu schenken.

**Aufgabe MH1**

Die drei gesuchten bedingten Verteilungen (Zeilen) haben folgendes Aussehen:

Y Bedingung	nie	gelegentlich	regelmäßig
BWL	$\frac{30}{270} = 0,11$	$\frac{40}{270} = 0,15$	$\frac{200}{270} = 0,74$
VWL	$\frac{20}{90} = 0,22$	$\frac{50}{90} = 0,56$	$\frac{20}{90} = 0,22$
Mathematik	$\frac{80}{140} = 0,57$	$\frac{40}{140} = 0,29$	$\frac{20}{140} = 0,14$

Die bedingten Verteilungen geben hier an, welcher Anteil unter den Studenten der einzelnen Studiengänge nie, gelegentlich und häufig Sport treibt. So gibt uns etwa der Wert 0,11 an, dass sich 11 % der BWL-Studenten nie sportlich betätigen.

Aufgabe MH3

i	h_i	a_i^m	$h_i \cdot a_i^m$	$h_i \cdot (a_i^m)^2$
1	127	20	2.540	50.800
2	271	60	16.260	975.600
3	102	100	10.200	1.020.000
Σ	500		29.000	2.046.400

j	h_j	b_j^m	$h_j \cdot b_j^m$	$h_j \cdot (b_j^m)^2$
1	163	150	24.450	3.667.500
2	217	450	97.650	43.942.500
3	48	750	36.000	27.000.000
4	72	1.050	75.600	79.380.000
Σ	500		233.700	153.990.000

Daraus ergibt sich:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^k h_i \cdot a_i^m = \frac{1}{500} \cdot 29.000 = 58 \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^r h_j \cdot b_j^m = \frac{1}{500} \cdot 233.700 = 467,40$$

$$s_x = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^k (a_i^m)^2 \cdot h_i - \bar{x}^2} = \sqrt{\frac{1}{500} \cdot 2.046.400 - 58^2} = 27,00$$

$$s_y = \sqrt{\frac{1}{n} \cdot \sum_{j=1}^r (b_j^m)^2 \cdot h_j - \bar{y}^2} = \sqrt{\frac{1}{500} \cdot 153.990.000 - 467,40^2} = 299,19$$

i,j	h_{ij}	a_i^m	b_j^m	$h_{ij} \cdot a_i^m \cdot b_j^m$
1,1	84	20	150	252.000
1,2	43	20	450	387.000
1,3	0	20	750	0
1,4	0	20	1.050	0
2,1	76	60	150	684.000
2,2	160	60	450	4.320.000
2,3	23	60	750	1.035.000
2,4	12	60	1.050	756.000
3,1	3	100	150	45.000
3,2	14	100	450	630.000
3,3	25	100	750	1.875.000
3,4	60	100	1.050	6.300.000
Σ	500			16.284.000

Daraus ergibt sich:

$$\text{Cov}_{xy} = \frac{1}{n} \cdot \sum_{i=1}^k \sum_{j=1}^r h_{ij} \cdot a_i^m \cdot b_j^m - \bar{x} \cdot \bar{y} = \frac{1}{500} \cdot 16.284.000 - 58 \cdot 467,40 = 5.458,80$$

$$r_{xy} = \frac{\text{COV}_{xy}}{s_x \cdot s_y} = \frac{5.458,80}{27,00 \cdot 299,19} = 0,68$$

Aufgabe MH5

- a) dichotome Merkmale
 b) Bei stat. Unabhängigkeit, sind folgende absolute Häufigkeiten zu erwarten:

	Erwerbspersonen	Nichterwerbspersonen	Σ
Männlich	19.578,8	20.751,2	40.330
Weiblich	20.467,2	21.692,8	42.160
Σ	40.046	42.444	82.490

Hiermit folgt:

$$\chi^2 = \frac{(22.232 - 19.578,8)^2}{19.578,8} + \frac{(18.098 - 20.751,2)^2}{20.751,2} + \frac{(17.814 - 20.467,2)^2}{20.467,2} + \frac{(24.346 - 21.692,8)^2}{21.692,8} = 1.367,22$$

$$K_{XY} = \sqrt{\frac{1.367,22}{82.490+1.367,22}} = 0,13 \rightarrow K_{XY}^{\text{norm}} = \frac{0,13}{\sqrt{\frac{\min(2;2)-1}{\min(2;2)}}} = \frac{0,13}{\sqrt{\frac{1}{2}}} = 0,18$$

c)

	Erwerbspersonen	Nichterwerbspersonen	Σ
Männlich	40.330	0	40.330
Weiblich	0	42.160	42.160
Σ	40.330	42.160	82.490

$$\text{Damit ist } \chi^2 = 82.490 \text{ und } K_{XY} = \frac{1}{\sqrt{2}} \rightarrow K_{XY}^{\text{norm}} = \frac{\sqrt{\frac{1}{2}}}{\sqrt{\frac{1}{2}}} = 1.$$

Es liegt also maximale Abhängigkeit vor.

Aufgabe P1

Kommt es zwischen Basis- und Berichtsperiode zu keiner Veränderung der Verbrauchsmengen sind die Preisindizes nach Laspeyres und Paasche identisch.

Aufgabe P3

$$a) p_{A,B}^L = \frac{\sum_{i=1}^n p_B^{(i)} \cdot q_A^{(i)}}{\sum_{i=1}^n p_A^{(i)} \cdot q_A^{(i)}} = \frac{1,15 \cdot 400 + 1,10 \cdot 700 + 1,01 \cdot 350}{0,96 \cdot 400 + 0,90 \cdot 700 + 0,87 \cdot 350} = \frac{1.583,50}{1.318,50} = 1,2010$$

$$p_{A,B}^P = \frac{\sum_{i=1}^n p_B^{(i)} \cdot q_B^{(i)}}{\sum_{i=1}^n p_A^{(i)} \cdot q_B^{(i)}} = \frac{1,15 \cdot 700 + 1,10 \cdot 600 + 1,01 \cdot 250}{0,96 \cdot 700 + 0,90 \cdot 600 + 0,87 \cdot 250} = \frac{1.717,50}{1.429,50} = 1,2015$$

$$b) \quad u_{A,B} = \frac{\sum_{i=1}^n p_B^{(i)} \cdot q_A^{(i)}}{\sum_{i=1}^n p_A^{(i)} \cdot q_A^{(i)}} = \frac{1.717,50}{1.318,50} = 1,3026$$

Aufgabe P5

Ein Vergleich ist nur möglich, wenn beide Indexreihen dieselbe Basis besitzen.
Eine Umbasierung von Index B auf das neue Basisjahr 2003 ist daher sinnvoll:

	A	B
2003	100	$\frac{I_B^{2003}}{I_B^{2003}} \cdot 100 = \frac{140}{140} \cdot 100 = 100,00$
2004	104	$\frac{I_B^{2004}}{I_B^{2003}} \cdot 100 = \frac{150}{140} \cdot 100 = 107,14$
2005	115	$\frac{I_B^{2005}}{I_B^{2003}} \cdot 100 = \frac{155}{140} \cdot 100 = 110,71$
2006	120	$\frac{I_B^{2006}}{I_B^{2003}} \cdot 100 = \frac{157}{140} \cdot 100 = 112,14$

2.

Kapitel II – Wahrscheinlichkeitsrechnung

Aufgabe W1

Die Ortskennbuchstaben sind als fix zu betrachten und bei der Berechnung der Möglichkeiten mit dem Faktor 1 anzusetzen. Was die Buchstabengruppen anbelangt, verbleiben nach Zeugenaussage nur 3 mögliche Ausprägungen. Zu den Zahlenkombinationen ist folgendes auszusagen:

	Stelle 1	Stelle 2	Stelle 3
Variante A	1	Ziffern von 0 bis 9	2
Variante B	1	2	Ziffern von 0 bis 9

Hierbei ist aber zu beachten, dass die Kombination 122 nicht doppelt erfasst werden darf. Letztendlich sind es also lediglich 19 Kombinationen. Für die Anzahl der Kfz-Kennzeichen gilt daher $1 \cdot 3 \cdot 19 = 57$.

Gesuchte Wahrscheinlichkeiten:

$$a) P = \frac{1 \cdot 1 \cdot 19}{57} = \frac{1}{3}$$

$$b) P = \frac{1 \cdot 3 \cdot 10}{57} = \frac{30}{57} = \frac{10}{19}$$

$$c) P = \frac{1 \cdot 3 \cdot 1}{57} = \frac{3}{57} = \frac{1}{19}$$

$$d) P = \frac{1 \cdot 3 \cdot 10}{57} = \frac{30}{57} = \frac{10}{19}$$

$$e) P = \frac{1 \cdot 3 \cdot 2}{57} = \frac{6}{57} = \frac{2}{19}$$

$$f) P = \frac{1 \cdot 3 \cdot 9}{57} = \frac{27}{57} = \frac{9}{19}$$

Aufgabe W3

$$P(A) = \frac{18}{37}, \quad P(B) = \frac{19}{37}, \quad P(C) = \frac{20}{37}$$

Aufgabe W5

Da kein Bewerber zweimal "gezogen" werden kann (Ziehen ohne Zurücklegen) und die Reihenfolge bei der Auswahl keine Rolle spielt, ergibt sich die Anzahl der möglichen Bewerber- bzw. Stellenbesetzungskombinationen wie folgt:

$$\binom{4}{2} = \frac{4 \cdot 3}{2 \cdot 1} = 6$$

Die beiden besten Bewerber stellen nun nur eine dieser 6 Kombinationen dar. Die gesuchte Wahrscheinlichkeit liegt daher bei $P = 1/6$.

Aufgabe W7

a) $P = 2/5$

b) Ziehen mit Zurücklegen: Ziehen ohne Zurücklegen:

$$P = \frac{3 \cdot 3 \cdot 3}{5 \cdot 5 \cdot 5} = 0,216$$

$$P = \frac{\binom{3}{3}}{\binom{5}{3}} = \frac{1}{\frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1}} = \frac{3 \cdot 2 \cdot 1}{5 \cdot 4 \cdot 3} = 0,10$$

Aufgabe W9

a) $P(A_1) = \frac{10}{36}$, $P(A_2) = \frac{6}{36}$, $P(A_1 \cap A_2) = \frac{4}{36}$, $P(A_1 | A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{\frac{4}{36}}{\frac{6}{36}} = \frac{4}{6}$

b) Nach dem Satz von Bayes ergibt sich die gesuchte Wahrscheinlichkeit als

$$P(A_2 | A_1) = \frac{P(A_1 | A_2) \cdot P(A_2)}{P(A_1)} = \frac{\frac{4}{6} \cdot \frac{6}{36}}{\frac{10}{36}} = \frac{2}{5}.$$

Veranschaulichung:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- ... Augensumme ≥ 9
- erster Würfel zeigt 6
relevant für gesuchte Wahrscheinlichkeit

Aufgabe W11

Es sei A: "Erster Student findet die Lösung", B: "Zweiter Student findet die Lösung" und C: "Wenigstens einer der beiden Studenten findet die Lösung". Für die gesuchte Wahrscheinlichkeit gilt dann

$$P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,70 + 0,70 - 0,70 \cdot 0,70 = 0,91$$

oder alternativ

$$P(C) = P(A \cup B) = 1 - P(\bar{A} \cap \bar{B}) = 1 - 0,30 \cdot 0,30 = 0,91.$$

Aufgabe W13

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y) = 0,40 + 0,70 - 0,20 = 0,90$$

Aufgabe W15

a) $P(A) = \frac{7 \cdot 1 \cdot 1}{7 \cdot 7 \cdot 7} = \frac{7}{343} = 0,0204$

c) $P(C) = \frac{1 \cdot 1 \cdot 1}{7 \cdot 7 \cdot 7} = \frac{1}{343} = 0,0029$

b) $P(B) = \frac{7 \cdot 6 \cdot 5}{7 \cdot 7 \cdot 7} = \frac{210}{343} = 0,6122$

d) $P(D) = \frac{7 \cdot 1 \cdot 6 \cdot 3}{7 \cdot 7 \cdot 7} = \frac{126}{343} = 0,3673$

Zur Erklärung von P(D):

- Erster Student hat an einem beliebigen Tag Geburtstag: 7 Möglichkeiten
- Zweiter Student hat genau am gleichen Tag Geburtstag: 1 Möglichkeit
- Dritter Student hat an einem anderen Tag Geburtstag: 6 Möglichkeiten

Daraus ergeben sich insgesamt $7 \cdot 1 \cdot 6 = 42$ Möglichkeiten. Diese beschreiben allerdings nur eine der drei möglichen Anordnungen von GLEICH, GLEICH, ANDERS. Daher ist der Wert 42 noch mit 3 zu multiplizieren, um die Gesamtmöglichkeiten zur Berechnung der Wahrscheinlichkeit zu erhalten.

Aufgabe W17

Es sei A_i : "Fehler bei Arbeitsgang i" mit $i = 1, 2$. Da die beiden Ereignisse voneinander unabhängig sind, gilt

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1) \cdot P(A_2),$$

$$P(A_2) = \frac{P(A_1 \cup A_2) - P(A_1)}{1 - P(A_1)} = \frac{0,06 - 0,04}{1 - 0,04} = 0,0208.$$

Aufgabe W19

a) 25 Spieler gesamt – 3 Torhüter – 3 Verletzte – 1 Gesperrter = 18 Verbleibende
18 Verbleibende – 5 Stürmer – 7 Mittelfeldspieler = 6 Verteidiger

b) Mögliche Aufstellungen:

$$\underbrace{5 \cdot 4}_{\text{Sturm}} \cdot \underbrace{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}_{\text{Mittelfeld}} \cdot \underbrace{6 \cdot 5 \cdot 4}_{\text{Abwehr}} \cdot \underbrace{3}_{\text{Tor}} = 18.144.000$$

Aufgabe W21

$$P = \frac{\binom{60}{10}}{\binom{100}{10}} = 0,0044$$

Aufgabe W23

a) $P(A) = 0,50^4 = 0,0625$

b) $P(B) = \binom{4}{2} \cdot 0,50^4 = 6 \cdot 0,0625 = 0,375$

c) $P(C) = 0,50^4 = 0,0625$

Hier wird das Urnenmodell des Ziehens mit Zurücklegen zugrunde gelegt.

Aufgabe W25

a) $P(A) = \frac{1}{10}$, $P(B) = \frac{1}{10 \cdot 9 \cdot 8} = 0,001389$, $P(C) = \frac{1}{\binom{10}{3}} = \frac{1}{120} = 0,0083$

b) Unter Berücksichtigung des Wetttipps:

$$P(A) = 1, \quad P(B) = \frac{1}{1 \cdot 9 \cdot 8} = 0,01389, \quad P(C) = \frac{1}{\binom{9}{2}} = \frac{1}{36} = 0,0278$$

Aufgabe Z1

Realisation ZV x_1	Elementar- ereignis ω	Wahrscheinlich- keit $P(\omega)$	Wahrscheinlich- keitsfunktion $f(y_i)$	Verteilungs- funktion $F(y_i)$
$x_1 = 0$	$\omega_1 = ZZZ$	$P(\omega_1) = 0,125$	0,125	0,125
	$\omega_2 = ZZK$	$P(\omega_2) = 0,125$		
$x_2 = 1$	$\omega_3 = ZKZ$	$P(\omega_3) = 0,125$	0,375	0,500
	$\omega_4 = KZZ$	$P(\omega_4) = 0,125$		
$x_3 = 2$	$\omega_5 = ZKK$	$P(\omega_5) = 0,125$		
	$\omega_6 = KZK$	$P(\omega_6) = 0,125$	0,375	0,875
$x_4 = 3$	$\omega_7 = KKZ$	$P(\omega_7) = 0,125$		
	$\omega_8 = KKK$	$P(\omega_8) = 0,125$	0,125	1,000

Aufgabe Z3

$$E(X) = 1 \cdot 0,2 + 2 \cdot 0,8 = 1,8$$

Aufgabe Z5

Die Werte $f(x_i)$ zeigen deutlich, dass es sich hier um eine symmetrische Verteilung handelt. Der Erwartungswert muss also $E(X) = 5$ sein.

Aufgabe Z7

Kursänderung	Gewinn	Wahrscheinlichkeit	$g_i \cdot f(g_i)$	$[g_i - E(G)]^2 \cdot f(g_i)$
	g_i	$f(g_i)$		
-0,30	-300	0,1	-30	14.440
-0,10	-100	0,1	-10	3.240
0,00	0	0,4	0	2.560
+0,20	+200	0,2	+40	2.880
+0,40	+400	0,2	+80	20.480
-	-	1,0	$E(G) = +80$	$Var(G) = 43.600$

Aus der Spekulation ist also ein Gewinn von 80 Euro bei einer Varianz von 43.600 Euro² bzw. Standardabweichung von 208,81 Euro zu erwarten.

Aufgabe Z9

a) 1. $f(x) > 0$ für $0 \leq x \leq 1$

2. $\int_{x_0}^1 f(x) dx = \int_0^1 2x dx = [x^2]_0^1 = 1$

b) $F(x) = \begin{cases} 0 & \text{für } x < 0 \\ x^2 & \text{für } 0 \leq x \leq 1 \\ 1 & \text{für } x > 1 \end{cases}$

c) $P(0,3 \leq X \leq 0,7) = \int_{0,3}^{0,7} 2x dx = [x^2]_{0,3}^{0,7} = 0,49 - 0,09 = 0,4$

$$P(X > 0,8) = \int_{0,8}^1 2x dx = [x^2]_{0,8}^1 = 1 - 0,64 = 0,36$$

oder: $P(X > 0,8) = 1 - P(X \leq 0,8) = 1 - \int_0^{0,8} 2x dx = 1 - [x^2]_0^{0,8} = 1 - 0,64 = 0,36$

Aufgabe Z11

a) Allgemein gilt $P(\mu_x - k \cdot \sigma_x < X < \mu_x + k \cdot \sigma_x) \geq 1 - \frac{1}{k^2}$.

Hier ist $\mu_x - k \cdot \sigma_x = 4 \rightarrow 8 - 2k = 4 \rightarrow k = 2$ bzw. $\mu_x + k \cdot \sigma_x = 12 \rightarrow 8 + 2k = 12 \rightarrow k = 2$, sodass sich für die gesuchte Wahrscheinlichkeit $P(4 < X < 12) \geq 3/4$ ergibt.

b) Allgemein gilt $P(|X - \mu_x| \geq k \cdot \sigma_x) \leq \frac{1}{k^2}$.

Es ist $P(X \leq 2 \cup X \geq 14) = P(X - 8 \leq 2 - 8 \cup X - 8 \geq 14 - 8) = P(|X - 8| \geq 6)$, woraus $k = 6 / \sigma_x = 6 / 2 = 3$ und damit $P(X \leq 2 \cup X \geq 14) \leq 1/9$ folgt.

Aufgabe Z13

Es sei X_i : "Augenzahl beim i-ten Wurf". Damit gilt $X = X_1 + X_2 + \dots + X_n$ sowie $Y = X_1 \cdot X_2 \cdot \dots \cdot X_n$.

Da $E(X_i) = \frac{1}{6} \cdot (1+2+3+4+5+6) = 3,5$ gilt, kann $E(X)$ wie folgt bestimmt werden:

$$E(X) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) = n \cdot 3,5$$

Da die einzelnen X_i unabhängig sind, darf auch folgendes gerechnet werden:

$$E(Y) = E(X_1 \cdot X_2 \cdot \dots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \dots \cdot E(X_n) = 3,5^n$$

Aufgabe Z15

i	A			B		
	r_i^A	$r_i^A - E(R^A)$	$[r_i^A - E(R^A)]^2$	r_i^B	$r_i^B - E(R^B)$	$[r_i^B - E(R^B)]^2$
1	0,385	0,245	0,060	-0,225	-0,245	0,060
2	0,140	0	0	0,020	0	0
3	-0,105	-0,245	0,060	0,265	0,245	0,060

$E(R)$	$1/3 \cdot (0,385 + 0,140 - 0,105) = 0,14$	$1/3 \cdot (-0,225 + 0,020 + 0,265) = 0,02$
$Var(R)$	$1/3 \cdot (0,060 + 0 + 0,060) = 0,04$	$1/3 \cdot (0,060 + 0 + 0,060) = 0,04$
$\sigma(R)$	0,20	0,20
i	$[r_i^A - E(R^A)] \cdot [r_i^B - E(R^B)]$	
1	-0,060	
2	0	
3	-0,060	
$Cov(R^A, R^B)$	$1/3 \cdot (-0,060 + 0 - 0,060) = -0,04$	
$\rho(R^A, R^B)$	$-0,04 / (0,20 \cdot 0,20) = -1$	

Der Korrelationskoeffizient von -1 bedeutet, dass wenn die Rendite einer der beiden Aktien steigt (fällt), die Rendite der anderen Aktie in gleichem Ausmaß fällt (steigt). Ein Portfolio, das 50-50 in A und B investiert, liefert die erwartete Rendite

$$E(R^P) = 0,5 \cdot E(R^A) + 0,5 \cdot E(R^B) = 0,5 \cdot 0,14 + 0,5 \cdot 0,02 = 0,08.$$

Da die Korrelation bei -1 liegt und die Varianzen der beiden Aktien gleich sind, besteht für den Anleger bei Investition in ein solches Portfolio keinerlei Risiko. Jede Renditeänderung eines Papiers wird durch die gegenläufige Entwicklung des anderen Papiers sofort wieder ausgeglichen. Es gilt daher für die Varianz der Rendite des Portfolios

$$Var(R^P) = 0,5^2 \cdot 0,04 + 0,5^2 \cdot 0,04 + 2 \cdot 0,5 \cdot 0,5 \cdot 0,2 \cdot 0,2 \cdot (-1) = 0.$$

Aufgabe Z17

x	y = 1	y = 2	y = 3	$f_x(x)$
1	0,45	0,20	0,05	0,70
2	0,15	0,10	0,05	0,30
$f_y(y)$	0,60	0,30	0,10	1,00

a) $f(1; 1) = 0,45$

Der Doktorand schätzt die Wahrscheinlichkeit, dass er sowohl für formale Korrektheit als auch für praktische Verwendbarkeit in Klasse 1 fällt, recht hoch ein.

b) $f_y(1) = 0,60$

Mit dieser Wahrscheinlichkeit geht der Doktorand davon aus, für praktische Verwendbarkeit (ungeachtet formaler Korrektheit) mit 1 beurteilt zu werden.

c) $f(X = 2 | Y = 3) = \frac{f(2, 3)}{f_y(3)} = \frac{0,05}{0,10} = 0,50$

Hat der Doktorand bereits für praktische Verwendbarkeit eine 3 erhalten, so stehen seiner Ansicht nach die Chancen für eine 2 bei formaler Korrektheit bei 50 %.

$$d) E(X) = \sum_i x_i \cdot f_X(x_i) = 1 \cdot 0,7 + 2 \cdot 0,3 = 1,3$$

$$\text{Var}(X) = \sum_i x_i^2 \cdot f_X(x_i) - [E(X)]^2 = 1^2 \cdot 0,7 + 2^2 \cdot 0,3 - 1,3^2 = 0,21$$

$$E(Y) = \sum_j y_j \cdot f_Y(y_j) = 1,5$$

$$\text{Var}(Y) = \sum_j y_j^2 \cdot f_Y(y_j) - [E(Y)]^2 = 2,7 - 1,5^2 = 0,45$$

Wie wir an den Erwartungswerten erkennen, vermutet der Doktorand eine sehr gute Bewertung seiner Arbeit bei beiden Beurteilungskriterien. So erwartet er z.B. in formaler Korrektheit eine Bewertung von 1,3 bei einer Standardabweichung von $\sigma_x = 0,46$. Da die Standardabweichung bei praktischer Verwertbarkeit mit $\sigma_y = 0,67$ deutlich höher ist, erkennen wir, dass der Doktorand das "Risiko" einer Abweichung von der erwarteten Bewertung hier als größer einschätzt.

e)

x_i	1	2
$f(x_i 1)$	$0,45/0,60 = 0,75$	$0,15/0,60 = 0,25$

$$E(X | 1) = \sum_i x_i \cdot f(x_i | 1) = 1 \cdot 0,75 + 2 \cdot 0,25 = 1,25$$

$$\text{Var}(X | 1) = \sum_i x_i^2 \cdot f(x_i | 1) - [E(X | 1)]^2 = 1^2 \cdot 0,75 + 2^2 \cdot 0,25 - 1,25^2 = 0,1875$$

Betrachten wir den Erwartungswert, so können wir sagen, dass wenn bereits bei praktischer Verwendbarkeit eine 1 erreicht wurde, für formale Korrektheit eine 1,25 zu erwarten ist.

$$f) E(XY) = \sum_i \sum_j x_i \cdot y_j \cdot f(x_i, y_j) = 2$$

$$\rho_{XY} = \frac{E(XY) - E(X) \cdot E(Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{2 - 1,3 \cdot 1,5}{\sqrt{0,21 \cdot 0,45}} = 0,16$$

Die Korrelation zwischen beiden Beurteilungskriterien ist zwar positiv, doch als relativ gering anzusehen. Eine bessere (schlechtere) Beurteilung bei X bedeutet also nur geringfügig auch eine bessere (schlechtere) bei Y und umgekehrt.

- g) X und Y sind stochastisch abhängig, da z.B. $[f_X(1) \cdot f_Y(1) = 0,70 \cdot 0,60 = 0,42] \neq [f(1,1) = 0,45]$ gilt, was für alle i, j nachweisbar ist. Außerdem ist $\rho_{XY} \neq 0$.

Aufgabe V1

X: "Anzahl verstorbener Personen" ist $B(20; 0,1)$ -verteilt.

$$a) P(20 - X \geq 18) = P(X \leq 2) = F_B(2) = f_B(0) + f_B(1) + f_B(2)$$

$$= \binom{20}{0} \cdot 0,1^0 \cdot 0,9^{20} + \binom{20}{1} \cdot 0,1^1 \cdot 0,9^{19} + \binom{20}{2} \cdot 0,1^2 \cdot 0,9^{18}$$

$$= 0,1216 + 0,2702 + 0,2852 = 0,6769$$

$$\begin{aligned} b) \quad P(20 - X \geq 1) &= P(X \leq 19) = 1 - P(X = 20) = 1 - f_B(20) \\ &= 1 - \binom{20}{20} \cdot 0,1^{20} \cdot 0,9^0 = 1 - 0 = 1 \end{aligned}$$

Aufgabe V3

Da $n \leq 0,1N$ gilt, kann die hypergeometrisch verteilte Zufallsvariable X : "Anzahl B-Wähler" durch eine $B(10; 0,45)$ -Verteilung approximiert werden. Eine Näherung durch die Normalverteilung ist nicht möglich, da $np(1-p) \geq 9$ nicht gilt. Die Verteilungstabellen liefern daher $P(X > 5) = 1 - P(X \leq 5) \approx 1 - F_B(5) = 1 - 0,7384 = 0,2616$.

Aufgabe V5

X: "Anzahl der Störungen pro Tag" ist binomialverteilt, jedoch aufgrund von $n > 50$ und $p < 0,1$ approximativ $Po(10.000 \cdot 0,0005 = 5)$ -verteilt.

$$a) \quad P(X = 5) \approx f_{Po}(5) = \frac{5^5}{5!} \cdot e^{-5} = 0,1755$$

Die Verteilungstabellen liefern folgende Ergebnisse:

$$b) \quad P(X \geq 10) = 1 - P(X \leq 9) \approx 1 - F_{Po}(9) = 1 - 0,9682 = 0,0318$$

$$c) \quad P(X \leq 20) \approx F_{Po}(20) = 1$$

Aufgabe V7

X: "Anzahl ablehnender Bürger" ist $H(10.000; 500; 50)$ verteilt. Es empfiehlt sich jedoch eine Approximation durch die $Po(50 \cdot 500/10.000 = 2,5)$ -Verteilung, da $n \geq 50$, $M/N \leq 0,1$ und $n \leq 0,1N$ erfüllt ist. Durch Nachschlagen erhalten wir damit

$$P(7 \leq X \leq 10) \approx F_{Po}(10) - F_{Po}(6) = 0,9999 - 0,9858 = 0,0141.$$

Aufgabe V9

Die Wahrscheinlichkeit, dass die Abfertigung x Minuten länger dauert, nachdem man bereits w Minuten gewartet hat, beträgt nach der Definition der bedingten Wahrscheinlichkeit

$$\begin{aligned} P(X \leq w+x | X > w) &= \frac{P[(X \leq w+x) \cap (X > w)]}{P(X > w)} = \frac{P(w < X \leq w+x)}{1 - P(X \leq w)} \\ &= \frac{F_{Ex}(w+x) - F_{Ex}(w)}{1 - F_{Ex}(w)} = \frac{1 - e^{-\lambda \cdot (w+x)} - (1 - e^{-\lambda \cdot w})}{1 - (1 - e^{-\lambda \cdot w})} \\ &= \frac{e^{-\lambda \cdot w} - e^{-\lambda \cdot (w+x)}}{e^{-\lambda \cdot w}} = 1 - e^{-\lambda \cdot x} = P(X \leq x). \end{aligned}$$

Die Wartezeit ist somit völlig unabhängig davon, wie viel Wartezeit w schon verstrichen ist. Für die Werte der Dichtefunktion gilt daher

$$f_{Ex}(w+x | X > w) = \frac{f_{Ex}(w+x)}{1 - F_{Ex}(w)} = \frac{\lambda \cdot e^{-\lambda \cdot (w+x)}}{e^{-\lambda \cdot w}} = \lambda \cdot e^{-\lambda \cdot x} = f_{Ex}(x).$$

Aufgabe V11

$$a) P(X \geq 1.200) = 1 - \Phi\left(\frac{1.200 - 900}{100}\right) = 1 - \Phi(3) = 1 - 0,9987 = 0,0013$$

$$b) P(X \leq 650) = \Phi\left(\frac{650 - 900}{100}\right) = \Phi(-2,5) = 1 - \Phi(2,5) = 1 - 0,9938 = 0,0062$$

$$c) P(750 \leq X \leq 1.050) = \Phi(1,5) - \Phi(-1,5) = 2 \cdot \Phi(1,5) - 1 = 2 \cdot 0,9332 - 1 = 0,8664$$

Aufgabe V13

$$P(\mu - c \leq X \leq \mu + c) = 2 \cdot \Phi\left(\frac{c}{\sigma}\right) - 1 \rightarrow 0,98 = 2 \cdot \Phi\left(\frac{1}{\sigma}\right) - 1 \leftrightarrow \Phi\left(\frac{1}{\sigma}\right) = 0,99$$

Nachschlagen in der Verteilungstabelle liefert $z = 2,33$.

$$\rightarrow \frac{1}{\sigma} = 2,33 \leftrightarrow \sigma = 0,4292$$

Aufgabe V15

Die Wartezeit ist stetig $G(0; 10)$ -verteilt mit folgender Dichtefunktion:

$$f_G(x) = \begin{cases} 0,1 & 0 \leq x \leq 10 \\ 0 & \text{sonst} \end{cases}$$

Für die gesuchte Wahrscheinlichkeit gilt daher

$$P(X > 5) = P(5 < X \leq 10) = f_G(x) \cdot (10 - 5) = 0,1 \cdot 5 = 0,5.$$

Aufgabe V17

W : "Anteil fehlerhafter Belege in der Stichprobe", X : "Anzahl fehlerhafter Belege in der Stichprobe"

Es besteht hier die Möglichkeit die vorliegende Hypergeometrische Verteilung durch eine Poissonverteilung mit $\lambda = np = 3.000 \cdot 0,005 = 15$ zu approximieren, da $n \geq 50$, $M/N \leq 0,1$ und $n \leq 0,1N$ erfüllt ist. Es gilt daher

$$P(W > 0,01) = P(X > 30) = 1 - P(X \leq 30) \approx 1 - F_{po}(30) = 1 - 0,9998 = 0,0002.$$

Da $np(1-p) \geq 9$ und $n \leq 0,1N$, kann alternativ auch unter Vernachlässigung des Korrekturfaktors $(N-n)/(N-1)$ durch eine Normalverteilung mit $\mu = np = 15$ und $\sigma^2 = np(1-p) = 14,93$ approximiert werden. Es gilt dann

$$P(W > 0,01) = P(X > 30) = 1 - P(X \leq 30) \approx 1 - F_N(30 + 0,5)$$

$$= 1 - \Phi\left(\frac{30,5 - 15}{\sqrt{14,93}}\right) = 1 - \Phi(4,01) = 0,0000$$

3.

Kapitel III – Induktive Statistik

Aufgabe PS1

Beim Ziehen mit Zurücklegen sind die X_i ($i = 1, 2, \dots, n$) unabhängig, sodass

$$\begin{aligned} E(\hat{\sigma}^2) &= E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2\right] \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n [(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2]\right] \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) + n(\bar{X} - \mu)^2\right] \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2\right] \quad \text{da } \sum_{i=1}^n (X_i - \mu) = n\bar{X} - n\mu \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n E[(X_i - \mu)^2] - nE[(\bar{X} - \mu)^2] \right]. \end{aligned}$$

Da gerade $E[(X_i - \mu)^2] = \sigma^2$ und $E[(\bar{X} - \mu)^2] = \sigma_{\bar{X}}^2 = \sigma^2 / n$, ergibt sich

$$E(\hat{\sigma}^2) = \frac{1}{n-1} \left[n\sigma^2 - n \frac{\sigma^2}{n} \right] = \frac{1}{n-1} \sigma^2(n-1) = \sigma^2.$$

Aufgabe PS3

\hat{p} : "Anteil der optimistischen Unternehmer in der Stichprobe"

Da $np(1-p) \geq 9$ und $n \leq 0,1N$ erfüllt sind, kann die vorliegende Hypergeometrische Verteilung unter Vernachlässigung des Endlichkeitskorrekturfaktors durch folgende Normalverteilung approximiert werden:

$$N\left(p, \frac{p(1-p)}{n}\right) \text{ wobei } p = \frac{M}{N} \quad \rightarrow \quad N\left(0,4; \frac{0,4 \cdot 0,6}{200}\right) = N(0,4; 0,0012)$$

Vernachlässigt man die Stetigkeitskorrektur, erhält man

$$P(\hat{p} \leq b) \approx F_N(b) = \Phi\left(\frac{b - 0,4}{\sqrt{0,0012}}\right) = 0,8.$$

Die tabellierte Standardnormalverteilung liefert $z_{0,8} = 0,84$, sodass

$$\frac{b - 0,4}{\sqrt{0,0012}} = 0,84 \quad \rightarrow \quad b = 0,4291.$$

Aufgabe PS5

a) $\mu = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{3} \cdot (2 + 4 + 6) = 4 \text{ g} \quad E(\bar{X}) = \mu = 4 \text{ g}$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \mu^2 = \frac{56}{3} - 4^2 = 2,67 \text{ g}^2 \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} = \frac{2,67}{2} = 1,33 \text{ g}^2$$

Der Vollständigkeit halber sei erwähnt, dass sich die Werte $E(\bar{X})$ und $\text{Var}(\bar{X})$ auch aus der Verteilung in Teilaufgabe b) über folgende bekannten Formeln berechnen ließen:

$$E(\bar{X}) = \sum \bar{x} \cdot f(\bar{x}), \quad \text{Var}(\bar{X}) = \sum \bar{x}^2 \cdot f(\bar{x}) - [E(\bar{X})]^2$$

b)

Stichprobe	Stichprobeneinzelwerte	Sichprobenmittel
j	x_{j1}	\bar{x}_j
1	2	2
2	2	3
3	2	4
4	4	3
5	4	4
6	4	5
7	6	4
8	6	5
9	6	6

Es ergibt sich damit die Wahrscheinlichkeitsverteilung:

\bar{x}	2	3	4	5	6
$f(\bar{x})$	1/9	2/9	3/9	2/9	1/9

Aufgabe PS7

\hat{p} : "Anteil Koalitionswähler in Stichprobe"

a) Da $np(1-p) \geq 9$ und $n \leq 0,1N$ erfüllt sind, gilt

$$\hat{p} \approx N(E(\hat{p}), \text{Var}(\hat{p})) = N\left(p, \frac{p(1-p)}{n}\right) = N(0,52; 0,0005).$$

Für die gesuchte Wahrscheinlichkeit gilt daher

$$P(\hat{p} < 0,5) \approx F_N(0,5) = \Phi\left(\frac{0,5 - 0,52}{\sqrt{0,0005}}\right) = \Phi(-0,89) = 1 - \Phi(0,89) = 0,1867.$$

$$\text{b) } P(\hat{p} < 0,5) = 0,0367 \rightarrow \Phi\left(\frac{0,5 - 0,52}{\sqrt{\frac{0,52 \cdot 0,48}{n}}}\right) = 0,0367$$

Nachschlagen in der Verteilungstabelle liefert $z_{0,0367} = -1,79$.

$$\begin{aligned} \frac{0,5 - 0,52}{\sqrt{\frac{0,52 \cdot 0,48}{n}}} &= -1,79 \rightarrow \sqrt{\frac{0,52 \cdot 0,48}{n}} = 0,0112 \\ &\rightarrow \frac{0,52 \cdot 0,48}{n} = 0,0001 \rightarrow n = 2.496 \end{aligned}$$

Aufgabe PS9

Gegeben sei eine konkrete Stichprobe x_1, \dots, x_n vom Umfang n aus einer normalverteilten Grundgesamtheit $X \sim N(\mu, \sigma^2)$ mit der Dichte

$$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

Die Likelihood-Funktion $L(\mu, \sigma^2) \equiv L(\mu, \sigma^2 | x_1, \dots, x_n)$ der konkreten Stichprobe ist

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n f_x(x_i | \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \\ &= \prod_{i=1}^n \left(\frac{1}{\sigma^2 2\pi} \right)^{\frac{1}{2}} \cdot e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = \left(\frac{1}{\sigma^2 2\pi} \right)^{\frac{n}{2}} \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2}. \end{aligned}$$

Wegen

$$\ln \left(\frac{1}{\sigma^2 2\pi} \right)^{\frac{n}{2}} = \frac{n}{2} \ln \left(\frac{1}{\sigma^2 2\pi} \right) = -\frac{n}{2} \ln (\sigma^2 2\pi) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi$$

besitzt die Log-Likelihood-Funktion die Form

$$\ln L(\mu, \sigma^2) = \ln \left(\frac{1}{\sigma^2 2\pi} \right)^{\frac{n}{2}} + \ln e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2} = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Die *partiellen Ableitungen* erster Ordnung der Funktion bezüglich μ und σ^2 sind

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) \quad \text{und} \quad \frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2.$$

Notwendige Bedingung für ein Maximum der Log-Likelihood-Funktion ist, dass beide partiellen Ableitungen gleich Null sind. Es muss das Gleichungssystem

$$\begin{aligned} \frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \hat{\mu}) &= 0 \\ -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\mu})^2 &= 0 \end{aligned}$$

gelöst werden. Für die erste Gleichung gilt

$$\frac{2}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu}) = 0 \leftrightarrow \sum_{i=1}^n (x_i - \hat{\mu}) = \sum_{i=1}^n x_i - n\hat{\mu} = 0.$$

Dies ist für $\hat{\mu} = \bar{x}$ erfüllt. Einsetzen von \bar{x} in die zweite Gleichung und Auflösen nach $\hat{\sigma}^2$ liefert

$$\frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n}{2\hat{\sigma}^2} \leftrightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2.$$

Auf die Prüfung der hinreichenden Bedingung für ein Maximum soll verzichtet werden. Das arithmetische Mittel \bar{x} ist die ML-Schätzung für μ und die mittlere quadratische Abweichung bezüglich \bar{x} , d.h. s^2 , die ML-Schätzung für σ^2 . Die zugehörigen ML-Schätzer sind also \bar{X} und S^2 .

Aufgabe IS1

a) Intervallschätzung für μ bei normalverteilter GG und bekannter Varianz σ^2 :

Benötigte Werte:

$$z = z_{\frac{1-0,05}{2}} = z_{0,975} = 1,96$$

$$\bar{x} = \frac{1}{10} (20,1 + \dots + 19,9) = 20,31$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0,5}{\sqrt{10}} = 0,09$$

Intervall:

$$\bar{x} - z\sigma_{\bar{x}} \leq \mu \leq \bar{x} + z\sigma_{\bar{x}} \rightarrow 20,31 - 1,96 \cdot 0,09 \leq \mu \leq 20,31 + 1,96 \cdot 0,09 \\ 20,13 \leq \mu \leq 20,49$$

b) Intervallschätzung für μ bei normalverteilter GG und unbekannter Varianz σ^2 :

Neu benötigte Werte:

$$t = t_{\frac{1-0,05}{2}, 10-1} = t_{0,975; 9} = 2,26$$

$$\hat{\sigma} = \sqrt{\frac{1}{10-1} ((20,1 - 20,31)^2 + \dots + (19,9 - 20,31)^2)} = \sqrt{0,11} = 0,33$$

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{0,33}{\sqrt{10}} = 0,10$$

Intervall:

$$\bar{x} - t\hat{\sigma}_{\bar{x}} \leq \mu \leq \bar{x} + t\hat{\sigma}_{\bar{x}} \rightarrow 20,31 - 2,26 \cdot 0,10 \leq \mu \leq 20,31 + 2,26 \cdot 0,10 \\ 20,08 \leq \mu \leq 20,54$$

Anmerkung: Wie hier sehr schön zu erkennen ist, führt die Intervallschätzung bei unbekannter Varianz zu breiteren Intervallen.

Aufgabe IS3

a) Intervallschätzung für μ bei beliebig verteilter GG und unbekannter Varianz σ^2 :

$n > 30 \rightarrow$ Anwendbarkeit des zentralen Grenzwertsatzes

$$z = z_{\frac{1-0,05}{2}} = z_{0,975} = 1,96$$

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{25}{\sqrt{60}} = 3,23$$

Intervall:

$$\bar{x} - z\hat{\sigma}_{\bar{x}} \leq \mu \leq \bar{x} + z\hat{\sigma}_{\bar{x}} \quad \rightarrow \quad 550 - 1,96 \cdot 3,23 \leq \mu \leq 550 + 1,96 \cdot 3,23 \\ 543,67 \leq \mu \leq 556,33$$

Der tatsächliche Mittelwert μ der Grundgesamtheit liegt also mit 95 % Wahrscheinlichkeit zwischen 543,67 km und 556,33 km.

- b) Da die Varianz σ^2 der Grundgesamtheit nicht bekannt ist, ist die Länge L des Konfidenzintervalls zufallsabhängig. Es kann somit kein exakter Mindeststichprobenumfang n berechnet, jedoch unter Nutzung von $\hat{\sigma}^2$ abgeschätzt werden.

Aufgabe IS5

Intervallschätzung für den Anteilswert p:

$$\hat{p} = \frac{x}{n} = \frac{50}{200} = 0,25 \quad n\hat{p}(1-\hat{p}) \geq 9? --- 200 \cdot 0,25 \cdot 0,75 = 37,5$$

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0,25 \cdot 0,75}{200}} = 0,0306 \quad z = z_{\frac{1-0,0455}{2}} = z_{0,9773} = 2$$

Intervall:

$$\hat{p} - z\hat{\sigma}_{\hat{p}} \leq p \leq \hat{p} + z\hat{\sigma}_{\hat{p}} \quad \rightarrow \quad 0,25 - 2 \cdot 0,0306 \leq p \leq 0,25 + 2 \cdot 0,0306 \\ 0,1888 \leq p \leq 0,3112$$

Aufgabe IS7

a) Intervallschätzung für den Anteilswert p:

$$\hat{p} = \frac{80}{1.000} = 0,08 \quad n\hat{p}(1-\hat{p}) \geq 9? --- 1.000 \cdot 0,08 \cdot 0,92 = 73,6$$

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0,08 \cdot 0,92}{1.000}} = 0,0086 \quad z = z_{\frac{1-0,01}{2}} = z_{0,995} = 2,58$$

Intervall:

$$\hat{p} - z\hat{\sigma}_{\hat{p}} \leq p \leq \hat{p} + z\hat{\sigma}_{\hat{p}} \quad \rightarrow \quad 0,08 - 2,58 \cdot 0,0086 \leq p \leq 0,08 + 2,58 \cdot 0,0086 \\ 0,0578 \leq p \leq 0,1022$$

Mit 99 %iger Wahrscheinlichkeit kann man also davon ausgehen, dass der tatsächliche Anteil der Wähler der Partei A zwischen 5,78% und 10,22% liegt.

- b) Wenn $p \leq 0,1$ ist, gilt sicher $\sqrt{p(1-p)} \leq 0,3$. Damit die Länge L kleiner oder gleich 1 % bzw. 0,01 ist, muss n größer oder gleich folgendem Wert sein:

$$n = \left(\frac{2z\sqrt{p(1-p)}}{L} \right)^2 = \left(\frac{2 \cdot 2,576 \cdot 0,3}{0,01} \right)^2 = 23.888,79 \rightarrow n = 23.889$$

Aufgabe T1

a) Einstichprobentest für den Anteilswert:

$$(1) H_0: p \leq 0,05 \quad H_A: p > 0,05 \quad \alpha = 0,02$$

(2) Konkrete Ausprägung der Teststatistik V:

$$v = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{105}{1.500} - 0,05}{\sqrt{\frac{0,05 \cdot 0,95}{1.500}}} = \frac{0,07 - 0,05}{0,0056} = 3,57$$

Verteilung von V (falls H_0 richtig):

$$np_0(1-p_0) \geq 9 ? \text{ hier: } 1.500 \cdot 0,05 \cdot 0,95 = 71,25 \rightarrow V \approx N(0;1)$$

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > z_{1-\alpha}$, wobei $z_{1-\alpha} = z_{0,98} = 2,05$.

(4) Entscheidung:

$$(v = 3,57) > (z_{1-\alpha} = 2,05) \rightarrow H_0 \text{ kann abgelehnt werden.}$$

b) Bestätigung von H_A , wenn

$$\frac{\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}{z_{1-\alpha}} > z_{1-\alpha} \rightarrow \frac{\frac{\hat{p} - 0,05}{0,0056}}{2,05} > 2,05 \leftrightarrow \hat{p} > 0,0615.$$

Aufgabe T3

Approximativer Zweistichproben-Gaußtest:

$$(1) H_0: \mu_1 \leq \mu_2 \quad H_A: \mu_1 > \mu_2 \quad \alpha = 0,01$$

(2) Konkrete Ausprägung der Teststatistik V:

$$\bar{x}_2 = \frac{5 \cdot 2 + \dots + 34 \cdot 1}{2 + \dots + 1} = 21,55 \quad \hat{\sigma}_2^2 = \frac{1}{40-1} [2 \cdot (5-21,55)^2 + \dots + 1 \cdot (34-21,55)^2] = 40,87$$

$$v = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}{n_1 + n_2}}} = \frac{25 - 21,55}{\sqrt{\frac{7}{100} + \frac{40,87}{40}}} = \frac{3,45}{1,04} = 3,32$$

Verteilung von V (falls H_0 richtig):

GG beliebig verteilt, Varianzen σ_1^2 und σ_2^2 unbekannt, $n_1, n_2 > 30 \rightarrow V \approx N(0;1)$

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > z_{1-\alpha}$, wobei $z_{1-\alpha} = z_{0,99} = 2,33$.

(4) Entscheidung:

$$(v = 3,32) > (z_{1-\alpha} = 2,33) \rightarrow H_0 \text{ kann abgelehnt werden.}$$

Damit ist statistisch bestätigt, dass das Verkehrsaufkommen vor Bau der Umgehungsstraße größer war. Die neue Verkehrsführung hat also ihren Zweck erfüllt und die Kreuzung entlastet.

Aufgabe T5

Differenzentest (approximativer Gaußtest):

$$(1) H_0: \delta = 0 \quad H_A: \delta \neq 0 \quad \alpha = 0,01$$

(2) Konkrete Ausprägung der Teststatistik V:

$$v = \frac{\bar{d}}{\hat{\sigma}_D} \sqrt{n} = \frac{20}{8} \sqrt{50} = 17,68$$

Verteilung von V (falls H_0 richtig):

Standardnormalverteilung, da GG beliebig verteilt und $n > 30$.

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $|v| > z_{1-\alpha/2}$, wobei $z_{1-\alpha/2} = z_{0,995} = 2,58$.

(4) Entscheidung:

($|v| = 17,68 > (z_{1-\alpha/2} = 2,58)$ → H_0 kann deutlich abgelehnt werden.

Die mittleren Pulsfrequenzen in der Grundgesamtheit unterscheiden sich also signifikant.

Der Vorteil einer verbundenen Stichprobe ist hier, dass individuelle Unterschiede der Merkmalsträger (natürliche unterschiedliche Pulsfrequenzen der einzelnen Männer) eliminiert werden.

Aufgabe T7

a) Korrelationstest:

$$(1) H_0: \rho \leq 0 \quad H_A: \rho > 0 \quad \alpha = 0,05$$

(2) Konkrete Ausprägung der Teststatistik V:

$$v = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,08 \cdot \sqrt{98}}{\sqrt{1-0,08^2}} = 0,79$$

Verteilung von V (falls H_0 richtig):

t-Verteilung mit $n - 2$ Freiheitsgraden

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > t_{1-\alpha, n-2}$, wobei $t_{1-\alpha, n-2} = t_{0,95, 98} = 1,66$.

(4) Entscheidung:

($v = 0,79 < (t_{1-\alpha, n-2} = 1,66)$ → H_0 kann nicht abgelehnt werden.

H_A kann somit nicht angenommen und $\rho > 0$ nicht bewiesen werden.

b) Damit der Korrelationskoeffizient r signifikant ist, muss ein großer Stichprobenumfang n vorliegen. Gerade daher ist das Fraktil der t-Verteilung ungefähr gleich dem entsprechenden Fraktil der Standardnormalverteilung. H_0 wird also abgelehnt, wenn

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} > z_{0,95} \rightarrow \frac{0,08\sqrt{98}}{\sqrt{1-0,08^2}} > 1,64 \rightarrow n-2 > 417,56 \rightarrow n > 419,56 \approx 420.$$

Aufgabe T9

Chi-Quadrat-Anpassungstest:

$$(1) H_0: X \text{ ist wie } X' \text{ verteilt.} \quad \alpha = 0,05$$

$H_A: X \text{ ist nicht wie } X' \text{ verteilt.}$

(2) Konkrete Ausprägung der Teststatistik V:

Klasse A _j	0	1	2	3	4	Σ
h_j	11	28	33	18	10	100
p_j	0,15	0,20	0,30	0,20	0,15	1,00
h_j^e	15	20	30	20	15	100
$(h_j - h_j^e)^2$	16	64	9	4	25	
$(h_j - h_j^e)^2 / h_j^e$	1,067	3,200	0,300	0,200	1,667	6,43

Verteilung von V (falls H_0 richtig):

Chi-Quadrat-Verteilung mit $k - 1$ Freiheitsgraden, da $h_j^e \geq 5$ für alle j und keine Schätzung von Funktionalparametern erforderlich.

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > \chi^2_{1-\alpha, k-1}$, wobei $\chi^2_{1-\alpha, k-1} = \chi^2_{0,95, 4} = 9,49$.

(4) Entscheidung:

$(v = 6,43) < (\chi^2_{1-\alpha, k-1} = 9,49) \rightarrow H_0$ kann nicht abgelehnt werden.

Es lässt sich also nicht nachweisen, dass der neue Werkstoff die Verteilung der Anzahl der Störfälle verändert hat.

Aufgabe T11

Chi-Quadrat-Homogenitätstest:

(1) H_0 : Die Qualifikation der beiden Gruppen stimmt überein. $\alpha = 0,05$

H_A : Die Qualifikation der beiden Gruppen stimmt nicht überein.

(2) Konkrete Ausprägung der Teststatistik V:

Note	Männer	Frauen	Σ
1	15 13,33	5 6,67	20
2	35 33,33	15 16,67	50
3	50 46,67	20 23,33	70
4	60 46,67	10 23,33	70
5	40 60	50 30	90
Σ	200	100	300

$$v = \sum_{i=1}^5 \sum_{j=1}^2 \frac{(h_{ij} - h_{ij}^e)^2}{h_{ij}^e} = \frac{(15 - 13,33)^2}{13,33} + \frac{(5 - 6,67)^2}{6,67} + \dots + \frac{(50 - 30)^2}{30} = 33,01$$

Verteilung von V (falls H_0 richtig):

Chi-Quadrat-Verteilung mit $(k - 1)(m - 1)$ Freiheitsgraden, da $h_{ij}^e \geq 5$ für alle i und j .

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > \chi^2_{1-\alpha, (k-1)(m-1)}$,
wobei $\chi^2_{1-\alpha, (k-1)(m-1)} = \chi^2_{0,95; 4} = 9,49$.

(4) Entscheidung:

$(v = 33,01) > (\chi^2_{1-\alpha, (k-1)(m-1)} = 9,49) \rightarrow H_0$ kann abgelehnt werden.

Man kann also davon ausgehen, dass die Qualifikation der beiden Gruppen nicht übereinstimmt, d.h. nicht homogen ist.

Aufgabe T13

Chi-Quadrat-Unabhängigkeitstest:

(1) H_0 : Berufsausbildung und Prüfungsergebnis sind unabhängig. $\alpha = 0,05$

H_A : Berufsausbildung und Prüfungsergebnis sind abhängig.

(2) Konkrete Ausprägung der Teststatistik V:

	Ausbildung	keine Ausbildung	Σ
geeignet	20	10	30
nicht geeignet	40	30	70
Σ	60	40	100

$$v = \frac{n(h_{11}h_{22} - h_{12}h_{21})^2}{h_1h_2h_{12}h_{21}} = \frac{100 \cdot (30 \cdot 20 - 40 \cdot 10)^2}{70 \cdot 30 \cdot 60 \cdot 40} = 0,79$$

Verteilung von V (falls H_0 richtig):

Chi-Quadrat-Verteilung mit $(k - 1)(m - 1)$ Freiheitsgraden, da $h_{ij}^e \geq 5$ für alle i und j ($h_{11}^e = 42$, $h_{22}^e = 12$, $h_{12}^e = 28$, $h_{21}^e = 18$).

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > \chi^2_{1-\alpha, (k-1)(m-1)}$,
wobei $\chi^2_{1-\alpha, (k-1)(m-1)} = \chi^2_{0,95; 1} = 3,84$.

(4) Entscheidung:

$(v = 0,79) < (\chi^2_{1-\alpha, (k-1)(m-1)} = 3,84) \rightarrow H_0$ kann nicht abgelehnt werden.

Man kann also nicht statistisch nachweisen, dass die Berufsausbildung und die Prüfungsergebnisse voneinander abhängen.

Aufgabe T15

Einfache Varianzanalyse:

(1) $H_0: \mu_1 = \mu_2 = \mu_3$ $\alpha = 0,05$

H_A : Mindestens zwei der μ_i sind verschieden.

(2) Konkrete Ausprägung der Teststatistik V:

Aus der nachfolgenden Tabelle lassen sich die Werte q_1 und q_2 und damit die Realisation v berechnen:

Feld (i) Weizen (j)	1	2	3	4	5	6	7	Stichproben-summe	Stichproben-mittel
W_1	12	13	10	9	11			55	11
W_2	9	11	10	11	13	12		66	11
W_3	10	9	11	13	10	9	8	70	10
Stichproben-gesamtsumme									
Stichproben-gesamtmittel									

$$q_1 = \sum_{j=1}^3 n_j \bar{x}_j^2 - n \bar{x}_{Ges}^2 = (5 \cdot 11^2 + 6 \cdot 11^2 + 7 \cdot 10^2) - (5+6+7) \cdot 10,61^2 \\ = 2.031 - 2.026,30 = 4,70$$

$$q_2 = \sum_{j=1}^3 \sum_{i=1}^{n_j} x_{ji}^2 - \sum_{j=1}^3 n_j \bar{x}_j^2 = (12^2 + 13^2 + 10^2 + \dots + 8^2) - 2.031 = 2.067 - 2.031 = 36 \\ v = \frac{(n-r)q_1}{(r-1)q_2} = \frac{((5+6+7)-3) \cdot 4,70}{(3-1) \cdot 36} = \frac{70,5}{72} = 0,98$$

Verteilung von V (falls H_0 richtig):

F-Verteilung mit $r-1$ und $n-r$ Freiheitsgraden, da GG normalverteilt.

(3) Entscheidungsregel:

H_0 ablehnen gegenüber H_A , wenn $v > f_{1-\alpha, r-1, n-r}$,
wobei $f_{1-\alpha, r-1, n-r} = f_{0,95; 2; 15} = 3,68$.

(4) Entscheidung:

$(v = 0,98) < (f_{1-\alpha, r-1, n-r} = 3,68) \rightarrow H_0$ kann nicht abgelehnt werden.

Es lässt sich also nicht statistisch nachweisen, dass die drei Weizensorten im Mittel unterschiedliche Erträge liefern.

4. Kapitel IV – Ökonometrie

Aufgabe O1

- a) $Y_i = \beta_0 + \beta_1 Z_i + \epsilon_i$ $Z_i = X_i^2$
 c) $Z_i = \beta_0 + \beta_1 X_i + \epsilon_i$ $Z_i = \ln Y_i$
 e) $Z_{1i} = \beta_0 + \beta_1 Z_{2i} + \epsilon_i$ $Z_{1i} = \ln Y_i, Z_{2i} = \ln X_i$
 f) $Z_{1i} = \alpha + \beta_1 Z_{2i} + \epsilon_i$ $Z_{1i} = \ln Y_i, Z_{2i} = \ln X_i, \alpha = \ln \beta_0$

Nicht schätzbar: b), d)

Aufgabe O3

$$\beta_0 = 0 \rightarrow \sum e_i^2 = \sum (Y_i - \hat{\beta}_1 X_i)^2 \rightarrow \frac{d \sum e_i^2}{d \hat{\beta}_1} = -2 \sum (Y_i - \hat{\beta}_1 X_i) X_i = 0$$

$$\rightarrow \sum (Y_i X_i - \hat{\beta}_1 X_i^2) = 0 \leftrightarrow \sum Y_i X_i - \hat{\beta}_1 \sum X_i^2 = 0 \leftrightarrow \hat{\beta}_1 = \frac{\sum Y_i X_i}{\sum X_i^2}$$

$$\beta_1 = 0 \rightarrow \sum e_i^2 = \sum (Y_i - \hat{\beta}_0)^2 \rightarrow \frac{d \sum e_i^2}{d \hat{\beta}_0} = -2 \sum (Y_i - \hat{\beta}_0) = 0$$

$$\rightarrow n \bar{Y} - n \hat{\beta}_0 = 0 \leftrightarrow \hat{\beta}_0 = \bar{Y}$$

Aufgabe O5

- a) Definieren wir das neue Modell $\tilde{Y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_i + \tilde{\epsilon}_i$ mit $\tilde{X}_i = 10X_i$, so wird die Schätzung des Steigungsparameters 10-mal kleiner und die der Konstanten bleibt unverändert. Es gilt nämlich

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X}) Y_i}{\sum (X_i - \bar{X})^2} = \frac{\sum (10X_i - 10\bar{X}) Y_i}{\sum (10X_i - 10\bar{X})^2} = \frac{10 \sum (X_i - \bar{X}) Y_i}{100 \sum (X_i - \bar{X})^2} = \frac{\hat{\beta}_1}{10},$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \bar{Y} - \frac{\hat{\beta}_1}{10} \cdot 10 \bar{X} = \hat{\beta}_0.$$

Alternative Begründung: Da Y sich nicht ändern darf, muss $\tilde{\beta}_0 = \beta_0$ und auch $\tilde{\beta}_1 = \frac{1}{10} \beta_1$ gelten, da $\frac{1}{10} \beta_1 \cdot 10X_i = \beta_1 X_i$.

Da $e_i = Y_i - \tilde{\beta}_0 - \tilde{\beta}_1 \tilde{X}_i = Y_i - \beta_0 - \beta_1 X_i$ unverändert ist, bleibt auch die Schätzung der Varianz des Störterms unverändert.

- b) Definieren wir das neue Modell $\tilde{Y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_i + \tilde{\epsilon}_i$ mit $\tilde{Y}_i = 10Y_i$, so wird sowohl die Schätzung des Steigungsparameters als auch die der Konstanten 10-mal größer. Es gilt nämlich

$$\hat{\beta}_1 = \frac{\sum x_i (\tilde{Y}_i - \bar{\tilde{Y}})}{\sum x_i^2} = \frac{\sum x_i (10Y_i - 10\bar{Y})}{\sum x_i^2} = \frac{10 \sum x_i (Y_i - \bar{Y})}{\sum x_i^2} = 10 \hat{\beta}_1,$$

$$\hat{\beta}_0^* = \bar{Y}^* - \hat{\beta}_1^* \bar{X} = 10\bar{Y} - 10\hat{\beta}_1 \bar{X} = 10\hat{\beta}_0.$$

Alternative Begründung: Multiplikation beider Seiten von $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ mit 10 liefert $\underbrace{10Y_i}_{Y'_i} = \underbrace{10\beta_0}_{\hat{\beta}_0^*} + \underbrace{10\beta_1 X_i}_{\hat{\beta}_1^* X_i} + \underbrace{10\epsilon_i}_{\epsilon'_i}$ und somit $\hat{\beta}_0^* = 10\beta_0$ und $\hat{\beta}_1^* = 10\beta_1$.

Die Schätzung für die Varianz des Störterms wird 100-mal größer, da

$$\epsilon_i'^2 = (Y_i^* - \hat{\beta}_0^* - \hat{\beta}_1^* X_i)^2 = (10Y_i - 10\hat{\beta}_0 - 10\hat{\beta}_1 X_i)^2 = 100(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = 100\epsilon_i^2.$$

Aufgabe O7

Die Residuen ergeben sich zu 1, -1, 1, 2 und -3, was folgendes Ergebnis liefert:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{16}{3}; \text{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2} = \frac{\frac{16}{3}}{10} = 0,5333$$

Aufgabe O9

a) $\hat{Y}_i^* = 5,83 + 8,69 X_i^*$ mit $X_i^* = X_i / 10$
 $(1,23) \quad (1,17)$

b) $\hat{Y}_i^* = 0,583 + 0,0869 X_i$ mit $\hat{Y}_i^* = \hat{Y}_i / 10$
 $(0,123) \quad (0,0117)$

c) $\hat{Y}_i^* = 0,583 + 0,869 X_i^*$ mit $\hat{Y}_i^* = \hat{Y}_i / 10$ und $X_i^* = X_i / 10$
 $(0,123) \quad (0,117)$

Die Werte von R^2 bleiben in allen drei Fällen unverändert.

Aufgabe O11

a) $\bar{Y} = 1, \bar{X}_1 = 0, \bar{X}_2 = 0$

$$\sum y_i x_{1i} = 13, \sum x_{1i}^2 = 16, \sum y_i x_{2i} = 4, \sum x_{2i}^2 = 10, \sum x_{1i} x_{2i} = 0$$

$$\hat{\beta}_1 = \frac{(\sum y_i x_{1i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2} = 0,8125$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{1i}^2) - (\sum y_i x_{1i})(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2} = 0,4$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 = 1$$

b) $e = [-0,4; 0,9875; -0,025; -0,375; -1,4125; 0,025; 0,6; 0,4125; 0,1875]$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-K-1} = \frac{3,8375}{9-2-1} = 0,6396$$

$$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{\hat{\sigma}^2}{\sum x_{1i}^2(1-r_{12}^2)}} = 0,1999$$

$$\text{Dabei ist } r_{12}^2 = \frac{\sum x_{1i}x_{2i}}{\sqrt{\sum x_{1i}^2 \sum x_{2i}^2}} = \frac{0}{\sqrt{16 \cdot 10}} = 0.$$

c) $\text{RSS} = \sum e_i^2 = 3,8375 \quad \text{TSS} = \sum (Y_i - \bar{Y})^2 = 16$
 $\text{ESS} = \text{TSS} - \text{RSS} = 12,1625 \quad R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{12,1625}{16} = 0,7602$

Aufgabe I1

- a) Für $n = 20$ gilt $|t| < t_{1-\alpha/2, n-K-1}$ bzw. $2 < (t_{0,975, 18} = 2,10)$, d.h. keine Signifikanz. Für $n = 100$ erhalten wir $2 > (z_{0,975} = 1,96)$, d.h. Signifikanz.
b) Für $p = 0,005$ wird H_0 sowohl bei einer Irrtumswahrscheinlichkeit von 5 % als auch bei 1 % abgelehnt. Für $p = 0,0108$ wird H_0 nur bei einer Irrtumswahrscheinlichkeit von 5 %, jedoch nicht bei 1 % abgelehnt.

Aufgabe I3

- a) (a) $\hat{\beta}_0 = t_0 \cdot \hat{\sigma}_{\hat{\beta}_0} = 1,257 \cdot 2,1738 = 2,732$
(b) $\hat{\sigma}_{\hat{\beta}_1} = \hat{\beta}_1 / t_1 = 0,1801 / 5,754 = 0,0313$
(c) $p\text{-Wert} = 2 \cdot (1 - P(t < 1,257)) = 2 \cdot (1 - 0,8926) = 0,2148$
- b) $\hat{\beta}_1 \pm t_{1-\alpha/2, n-K-1} \cdot \hat{\sigma}_{\hat{\beta}_1} = 0,1801 \pm 2,58 \cdot 0,0313 \rightarrow 0,099 \leq \beta_1 \leq 0,261$
- c) Da der Absolutbetrag von $t_1 = (0,1801 - 0,2) / 0,0313 = -0,636$ kleiner ist als der kritische Wert eines zweiseitigen Tests mit 5 % Irrtumswahrscheinlichkeit und 49 Freiheitsgraden, also 2,01, kann H_0 nicht abgelehnt werden. Man hätte hier auf Grund des zentralen Grenzwertsatzes auch das 97,5 %-Fraktil der Standardnormalverteilung (1,96) als zweiseitige Testschranke verwenden können.

Die Nullhypothese unterstellt, dass eine Erhöhung des Anteils volljähriger männlicher Hochschulabsolventen um einen Prozentpunkt zu einer Erhöhung des durchschnittlichen Einkommens um 200 Dollar führt. Eine Nicht-Ablehnung ermöglicht aber bekannterweise keinen Beweis.

Aufgabe I5

Zum Test der gegebenen Hypothesen muss das Modell zunächst geschätzt werden. Dazu ist es erforderlich die Modellgleichung zu logarithmieren, da andernfalls OLS nicht angewendet werden kann. Wir sind in der Lage die Steigungsparameter als partielle Elastizitäten aufzufassen. Die Ergebnisse der Tests für a) bis c) sind in der nachfolgenden Tabelle dargestellt. In allen Fällen kann H_0 bei einem Signifikanzniveau von 1 % abgelehnt werden. [Hinweis: Die Werte von a) und b) erhalten wir direkt aus dem Regressionsoutput. Der Test c) ist in EViews nach Eingabe von c(1) + c(2) = 1 bei View / Coefficient Tests / Wald Coefficient Restrictions zu finden.]

	H_0	t-Wert	F-Wert	p-Wert
a)	$\beta_1 = 0$	3,05	-	0,0045
b)	$\beta_1 = \beta_2 = 0$	-	2,810,68	0,0000
c)	$\beta_1 + \beta_2 = 1$	-	276,51	0,0000

Der Korrelationskoeffizient zwischen X_1 und X_2 liegt bei 0,9151. Eine Regression von X_1 auf X_2 oder umgekehrt würde daher ein Bestimmtheitsmaß von $0,9151^2 = 0,8374$ liefern. Dies deutet evtl. auf einen gewissen Grad an Multikollinearität hin.

Werfen wir jedoch einen Blick auf das Regressionsergebnis mit seinen relativ hohen t-Werten und einem sehr hohen Bestimmtheitsmaß, so zeigt sich ein anderes Bild:

$$\ln(Y_t) = 0,0015 + 0,1391 \ln(X_{1t}) + 0,6484 \ln(X_{2t}) \quad R^2 = 0,9943 \\ 0,8034 \quad 3,0525 \quad 27,4196$$

Das klassische Symptom von Multikollinearität sind sehr niedrige t-Werte bei einem hohen Bestimmtheitsmaß. Multikollinearität führt also i.d.R. zwar zu einem Modell mit guter Erklärung der Variabilität von Y, jedoch sind wir nicht in der Lage die Einflüsse der einzelnen erklärenden Variablen mit einem vernünftigen Grad an Präzision zu schätzen. Hier liegt dieses Symptom allerdings nicht vor. Unsere Steigungsparameter sind hoch signifikant und auch das R^2 ist sehr hoch. Es besteht also keine Veranlassung für weiteres Handeln. Ein Ausschluss einer Variable käme außerdem sowieso nicht in Frage, da beide für die Produktionsfunktion relevant sind.

Aufgabe I7

- a) Konfidenzintervall:

$$\hat{\beta}_1 \pm t_{1-\alpha/2} \hat{\sigma}_{\hat{\beta}_1} = 0,6991 \pm 2,11 \cdot \sqrt{0,0485} = [0,2344; 1,1638]$$

Hypothesentest:

H_0 kann nicht abgelehnt werden, da der hypothetische Parameterwert im Konfidenzintervall liegt.

$$b) F = \frac{R^2 / K}{(1 - R^2) / (n - K - 1)} = \frac{0,9466 / 2}{(1 - 0,9466) / (20 - 2 - 1)} = 150,6760$$

Damit lässt sich H_0 bei einem Signifikanzniveau von 5 % ablehnen, da der vorliegende F-Wert deutlich größer als der kritische F-Wert der F(2;17)-Verteilung, d.h. 3,59, ist.

Aufgabe I9

Die Schätzung der polynomiauen Funktion liefert (Standardfehler in Klammern):

$$\hat{Y}_i = 134,66 + 57,97X_i + 11,03X_i^2 + 1,14X_i^3 \\ (44,80) \quad (29,97) \quad (5,76) \quad (0,34)$$

- a) Für eine lineare Kostenfunktion wäre $\beta_2 = \beta_3 = 0$ erforderlich, sodass wir die Hypothese $H_0: \beta_2 = \beta_3 = 0$ gegenüber $H_A: \text{"mindestens ein } \beta \text{ ist von Null verschieden"}$ testen müssen. Da $F = 62,10$ größer als der kritische F-Wert 3,40 ist, der sich bei einem Signifikanzniveau von 5 % und Freiheitsgraden von (2; 24) ergibt, können wir H_0 ablehnen und folgern, dass die vorliegenden Daten keine lineare Kostenfunktion unterstützen.
- b) In diesem Fall ist $H_0: \beta_3 = 0$ gegenüber $H_A: \beta_3 \neq 0$ zu testen. Da der t-Wert 3,40 hier größer ist als der kritische t-Wert 2,06, der sich bei einem Signifikanzniveau von 5 % und 24 Freiheitsgraden ergibt, können wir H_0 ablehnen. Die Daten unterstützen demnach auch keine quadratische Kostenfunktion.

- c) Die Schätzung der logarithmischen Kostenfunktion liefert (Standardfehler in Klammern):

$$\widehat{\ln Y_i} = 4,84 + 0,62 \ln X_i \\ (0,09) \quad (0,05)$$

Die RESET-Ergebnisse sind in folgender Tabelle dargestellt:

Ramsey RESET-Test: Polynomiales Modell			
F-Statistik (1 Proxy)	0,9874	p-Wert	0,3307
F-Statistik (2 Proxies)	0,5456	p-Wert	0,5871
Ramsey RESET-Test: Logarithmisches Modell			
F-Statistik (1 Proxy)	26,229	p-Wert	0,000027
F-Statistik (2 Proxies)	32,753	p-Wert	0,000000

Der Test des polynomiaalen Modells liefert p-Werte, die beide über einem Signifikanzniveau von 5 % liegen und damit nicht gegen das Modell sprechen. Das logarithmische Modell hingegen führt zu zwei sehr niedrigen p-Werten, die es als nicht adäquat einstufen lassen. Auf Basis der RESET-Ergebnisse würde man also das polynomiale Modell vorziehen.

Je nachdem ob $\beta_1 < 1$ oder $\beta_1 > 1$ gilt, ist die Grenzkostenfunktion im logarithmischen Modell entweder immer fallend oder immer steigend. Sie kann also anders als das polynomiale Modell keine Segmente steigender und fallender Grenzkosten haben, ist daher weniger flexibel und damit auch weniger realistisch als die polynomiale.

- d) Betrachtet man z.B. ein einfaches Modell $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, so würde eine Modifikation in der Form $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \gamma \hat{Y}_i + \varepsilon_i$, in welcher $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$ gilt, dazu führen, dass die modifizierte Modellgleichung nicht schätzbar ist. \hat{Y} wäre nämlich perfekt mit X_1 und X_2 korreliert.

Aufgabe D1

- a) Gerundet erhält man folgende Schätzergebnisse (Standardfehler in Klammern):

$$\hat{Y}_i = 24200 + 1643X_i + 5033D_{1i} \\ (1078) \quad (352) \quad (456)$$

Interpretation: Wird der Notendurchschnitt c.p. um eine Einheit erhöht (verbessert), nimmt das durchschnittliche Einstiegsgehalt um 1.643 US-Dollar zu. Studenten, die Ökonometrie belegt haben weisen im Durchschnitt ein um 5.033 US-Dollar höheres Einstiegsgehalt auf, als solche, die das Fach nicht belegt haben.

- b) Hinzufügen der Dummy-Variable D_2 liefert (Standardfehler in Klammern):

$$\hat{Y}_i = 24.242 + 1.658X_i + 5.024D_{1i} - 205D_{2i} \\ (1.091) \quad (356) \quad (460) \quad (420)$$

Interpretation: Diese Ergebnisse weisen darauf hin, dass das Einstiegsgehalt bei Frauen im Durchschnitt 205 US-Dollar niedriger ist als bei Männern. Aufgrund der Tatsache, dass der t-Wert des geschätzten Koeffizienten der zweiten Dum-

my-Variable jedoch sehr klein ist, kann die Hypothese, dass Männer und Frauen im Durchschnitt das gleiche Einstiegsgehalt aufweisen, nicht verworfen werden.

- c) Hinzufügen der Interaktionsvariable D_1D_2 liefert (Standardfehler in Klammern):

$$\hat{Y}_i = 24.222 + 1.675X_i + 4.924D_{1i} - 280D_{2i} + 274D_{1i}D_{2i}$$

$$(1.104) \quad (365) \quad (582) \quad (500) \quad (966)$$

Interpretation: Die Belegung des Fachs Ökonometrie scheint für Frauen einen höheren Wert als für Männer (Wert für Männer = 4.924, Wert für Frauen = 4.924 + 274 = 5.198) zu haben. Der geschätzte Koeffizient der Interaktionsvariablen ist jedoch nicht signifikant von Null verschieden. Die Hypothese, dass das Fach für beide Geschlechter gleich wertvoll ist, kann nicht abgelehnt werden.

Aufgabe D3

- a) Die Parameterschätzungen auf der Folgeseite weisen die ökonomisch erwarteten Vorzeichen auf und sind auf Basis einseitiger t-Tests für die erwarteten Vorzeichen auf einem 5 %-Niveau signifikant. Die Schätzung weist ein R^2 von 0,8706 bei einer F-Statistik von 1.113,02 auf, was auf eine gute Erklärungskraft des Modells hindeutet.

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	97.840,97	24.776,04	3,95	0,0001
D_1	109.851,30	33.699,85	3,26	0,0012
X_1	608,98	19,62	31,04	0,0000
X_1D_1	103,90	26,57	3,91	0,0001
X_2	-760,92	204,83	-3,71	0,0002
D_2	17.512,06	4.787,07	3,66	0,0003
D_3	6.610,86	3.888,06	1,70	0,0894

Für Häuser in Universitätsnähe ergibt sich die Schätzgleichung

$$\hat{Y}_i = \underbrace{207.692,27}_{97.840,97+109.851,30} + \underbrace{712,88}_{608,98+103,90} X_{1i} - 760,92 X_{2i} + 17.512,06 D_{2i} + 6.610,86 D_{3i}$$

für weiter entfernt liegende Häuser

$$\hat{Y}_i = 97.840,97 + 608,98 X_{1i} - 760,92 X_{2i} + 17.512,06 D_{2i} + 6.610,86 D_{3i}$$

Wir können also insgesamt folgern:

- Unabhängig von der Größe ist der Preis für in Universitätsnähe liegende Immobilien durchschnittlich 109.851,30 Euro höher. Zusätzlich steigt der m^2 -Preis bei Universitätsnähe von 608,98 Euro auf 712,88 Euro.
- Immobilien verlieren durchschnittlich 760,92 Euro pro Jahr an Wert.
- Ein Pool erhöht den Immobilienpreis durchschnittlich um 17.512,06 Euro.
- Ein Kamin erhöht den Immobilienpreis durchschnittlich um 6.610,86 Euro.

- b) Die Schätzung der linksseitigen Semi-Log-Form liefert folgende Ergebnisse:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	12,7577	0,0264	483,50	0,0000
D ₁	0,3335	0,0359	9,29	0,0000
X ₁	0,0007	2,09E-05	34,42	0,0000
X ₁ D ₁	-6,86E-05	2,83E-05	-2,43	0,0155
X ₂	-0,0009	0,0002	-4,15	0,0000
D ₂	0,0190	0,0051	3,73	0,0002
D ₃	0,0066	0,0041	1,59	0,1129

Ein zusätzlicher Quadratmeter erhöht den Preis im Durchschnitt um ungefähr 0,07 %. Eine Erhöhung des Gebäudealters um ein Jahr führt zu einer Reduzierung des Gebäudepreises um ca. 0,09 %. (Für eine exakte Berechnung vgl. die Ausführungen der nachfolgende Seite.)

Man beachte: D₃ ist im neuen Modell bei $\alpha = 0,05$ nicht mehr signifikant.

- c) Die Differenz der logarithmierten Preise von Häusern mit und ohne Pool entspricht dem Parameter β_5 der Dummy-Variablen D₂. Daraus ergibt sich

$$\ln Y_{\text{Pool}} - \ln Y_{\text{Kein Pool}} = \beta_5 \leftrightarrow \ln\left(\frac{Y_{\text{Pool}}}{Y_{\text{Kein Pool}}}\right) = \beta_5 .$$

Dies bedeutet, dass der Preis ungefähr um 1,90 % steigt. Exakter gilt

$$\frac{Y_{\text{Pool}}}{Y_{\text{Kein Pool}}} = e^{\beta_5} \leftrightarrow \frac{Y_{\text{Pool}} - Y_{\text{Kein Pool}}}{Y_{\text{Kein Pool}}} = e^{\beta_5} - 1 ,$$

d.h. die gesuchte Preisänderung ergibt sich zu $(e^{0,0190} - 1) \cdot 100 = 1,92 \%$.

- d) In diesem Fall ergibt sich die Differenz der logarithmierten Preise als

$$\ln Y_{\text{universitätsnah}} - \ln Y_{\text{universitätfern}} = 0,3335 - 0,0000686 \cdot 2,500 = 0,162 ,$$

d.h. die ungefähre Preissteigerung zu 16,2 %. Exakt ergibt sich die gesuchte Preisänderung als $(e^{0,162} - 1) \cdot 100 = 17,59 \%$.

Aufgabe HA1

Die OLS-Schätzung und die WLS-Schätzungen liefern (inkl. der p-Werte des White-Tests in seinen EViews-Standardeinstellungen) die auf der Folgeseite zu findenden Ergebnisse. Jede Spezifikation sieht eine andere Heteroskedastizitätsstruktur vor, wobei davon ausgegangen werden kann, dass bei $\sigma^2 \ln X_i$ die Heteroskedastizität am geringsten und bei $\sigma^2 X_i^2$ am stärksten ausgeprägt ist. Insgesamt reagieren die Schätzungen der Steigungsparameter und die geschätzten Standardfehler nicht allzu sensibel auf die Heteroskedastizitätsspezifikation. Geben wir eine Irrtumswahrscheinlichkeit von 5 % vor, so kann nur bei den Spezifikationen (ii) und (iii) die Nullhypothese der Homoskedastizität nicht mehr abgelehnt werden. Bei einer Irrtumswahrscheinlichkeit von 10 % erscheint (iii) als zu favorisierende Spezifikation.

Hinweis: Bei $\text{Var}(\epsilon_i) = \sigma^2 \ln X_i$ ist empirisch natürlich darauf zu achten, dass die Beobachtungen X_i natürlich größer als 1 sein müssen.

Schätzung	Var(ϵ_i)	WLS-Gewichte	$\hat{\beta}_0$	$\hat{\beta}_1$	p-Wert (White-Test)
OLS	-	-	40,768 (22,139)	0,128 (0,031)	0,0007
WLS (i)	$\sigma^2 \ln X_i$	$1/\sqrt{\ln X_i}$	39,550 (21,469)	0,130 (0,030)	0,0038
WLS (ii)	$\sigma^2 X_i$	$1/\sqrt{X_i}$	31,924 (17,986)	0,141 (0,027)	0,0710
WLS (iii)	$\sigma^2 X_i^2$	$1/X_i$	21,286 (14,038)	0,158 (0,023)	0,4300
WLS (iv)	$\sigma^2 \sqrt{X_i}$	$1/X_i^{0,25}$	36,753 (20,052)	0,134 (0,029)	0,0136

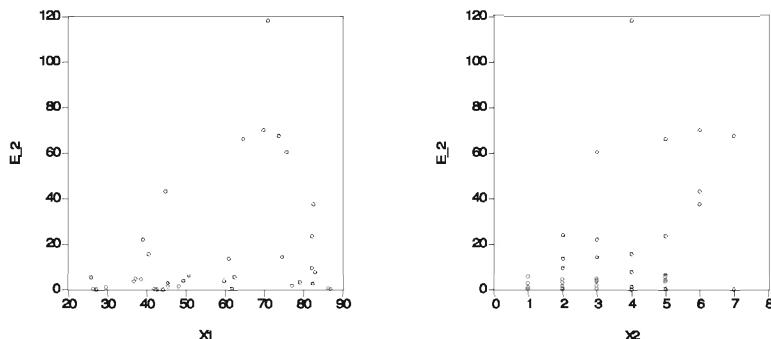
Aufgabe HA3

- a) Die OLS-Schätzung liefert folgende Ergebnisse:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	2,2432	2,6688	0,8405	0,4063
X_1	0,1645	0,0354	4,6454	0,0000
X_2	1,1451	0,4144	2,7630	0,0091

Die Schätzungen der Steigungskoeffizienten sind signifikant von Null verschieden. Eine Erhöhung des Einkommens um 1.000 Euro führt c.p. durchschnittlich zu einer Erhöhung der Lebensmittelausgaben um 164,50 Euro; eine zusätzliche Person im Haushalt zu einer Erhöhung um 1.145,10 Euro.

- b) Trägt man die quadrierten Residuen (Proxies für die Störtermvarianzen) gegenüber X_1 und X_2 in einem Koordinatensystem ab, steigen diese mit X_1 und X_2 . Die Störtermvarianz scheint also sowohl von X_1 als auch von X_2 abzuhängen.



Die Nullhypothese der Homoskedastizität kann beim White-Test (p-Wert 0,0347) und beim Breusch-Pagan LM-Test (p-Wert 0,0104) bei einer Irrtumswahrscheinlichkeit von 5 % abgelehnt werden. Da die erklärenden Variablen X_1 und X_2 der Testregression des Breusch-Pagan LM-Tests signifikanten Erklärungsgehalt besitzen, hängt die Störtermvarianz in der Tat von X_1 und X_2 ab.

- c) Unter Verwendung von White-Standardfehlern ergibt sich folgendes Ergebnis:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	2,2432	2,3653	0,9484	0,3494
X ₁	0,1645	0,0299	5,5039	0,0000
X ₂	1,1451	0,4540	2,5224	0,0164

Die White-Standardfehler geben Hinweise darauf, dass die Standardfehler aus a) die Präzision von $\hat{\beta}_1$ unter-, die von $\hat{\beta}_2$ überschätzen.

- d) Nach den Ergebnissen aus Teilaufgabe b) würde sich eine Modellierung in Abhängigkeit der Variablen X₁ und X₂ anbieten. Da Var(ϵ_t) jedoch auch von anderen (uns unbekannten) Variablen abhängen kann, besteht das Risiko einer falschen Spezifikation der Heteroskedastizitätsstruktur und damit auch in großen Stichproben die Gefahr nicht effizienter Schätzungen. Gerade daher empfiehlt sich ein Ausweichen auf White-Standardfehler.

Aufgabe HA5

- a) Die OLS-Schätzung des Erfahrungskurvenmodells liefert:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	5,8256	0,2762	21,091	0,0000
ln X _{t-1}	-0,3595	0,0364	-9,8884	0,0000

Die Parameter des Modells sind bei jeder Irrtumswahrscheinlichkeit signifikant von Null verschieden. Eine Erhöhung der kumulierten Produktion um 1 % führt zu einer Senkung der Stückkosten um 0,36 %.

- b) Die Durbin-Watson-Statistik nimmt einen Wert von d = 0,8244 an und bestätigt positive Autokorrelation erster Ordnung, da H₀: $\rho \leq 0$ gegenüber H_A: $\rho > 0$ aufgrund von $0 < d < 1,08$ (vgl. Tabellenwerk bei n = 15 und K = 1) abgelehnt werden kann. Der p-Wert des Breusch-Godfrey LM-Tests ist 0,1283. Mit diesem Test kann bei $\alpha = 0,05$ also keine Autokorrelation erster Ordnung festgestellt werden.

- c) Die GLS-Schätzung liefert:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	7,0419	1,1933	5,9011	0,0001
ln X _{t-1}	-0,5161	0,1522	-3,3907	0,0060
AR(1)	0,5947	0,2599	2,2880	0,0429

Die neue Schätzung liefert neben einem deutlich veränderten Steigungsparameter auch einen höheren Standardfehler und damit einen niedrigeren t-Wert. Dennoch bleibt der Einfluss von ln X auf einem Niveau von 1 % signifikant.

Eine erneute Betrachtung der Durbin-Watson Statistik liefert $d = 1,3660$, was nun in den Nichtablehnungsbereich $1,36 \leq d \leq (4 - 1,36 = 2,64)$ fällt, sodass positive Autokorrelation nicht mehr bestätigt werden kann. Der Breusch-Godfrey LM-Test weist nun mit einem p-Wert von 0,3802 noch deutlicher auf nicht vorliegende Autokorrelation hin.

- d) Unter Verwendung von Newey-West-Standardfehlern ergibt sich:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	5,8256	0,3981	14,632	0,0000
ln X _{t-1}	-0,3595	0,0521	-6,9021	0,0000

Mit einem kritischen t-Wert von 2,14 bei 14 Freiheitsgraden ergibt sich:

- i) OLS: $-0,4374 \leq \beta_1 \leq -0,2816$
- ii) Newey-West: $-0,4710 \leq \beta_1 \leq -0,2480$
- iii) GLS: $-0,8418 \leq \beta_1 \leq -0,1904$

Aufgrund der geringeren Intervallbreite erscheinen die Intervallschätzungen bei OLS präziser als sie wirklich sind. Erst die Berücksichtigung der Autokorrelation führt zur Verbreiterung des Intervalls. Diese fällt im GLS-Fall erheblich stärker aus als bei Newey-West-Standardfehlern.

- d) Der sehr geringe Stichprobenumfang von $n = 15$ ist in mehrerlei Hinsicht problematisch. Erstens ist der Breusch-Godfrey LM-Test ein asymptotischer Test, der daher für kleine Stichproben schlecht geeignet ist. Dies erklärt, warum er unter b) nicht in der Lage war, die Autokorrelation aufzudecken. Zweitens sind sowohl Newey-West-Standardfehler als auch eine GLS-Schätzung nur für große Stichproben vorgesehen und in kleinen Stichproben unpräzise. Die Ergebnisse aus d) sind daher mit Vorsicht zu betrachten.

Hinweis: Analog zu Aufgabe HA4 besteht darüber hinaus das Problem, dass die verwendeten Variablen nicht stationär sind, sodass die Schätzungen möglicherweise weniger signifikant sind, als sie hier erscheinen. Zunächst müsste daher dem Problem der Nicht-Stationarität begegnet werden (vgl. dazu Abschnitt IV 4.7) und erst im Anschluss daran noch auftretende Autokorrelation bekämpft werden. Der Stichprobenumfang ist im hier allerdings so gering, dass die dafür nötigen Methoden nicht anwendbar sind. Es ist also im gegebenen Beispiel einfach nicht möglich aussagekräftige Schlüsse aus dem Datenmaterial zu ziehen.

- e) Im Falle einer Fehlspezifikation in Form von vernachlässigten Variablen sind OLS-Schätzer verzerrt und inkonsistent. Festgestellte Autokorrelation ist ein deutlicher Hinweis auf eine derartige Problematik. Autokorrelation, die aus der Vernachlässigung von Variablen reduziert, darf nicht durch GLS-Schätzung oder Newey-West-Standardfehler begegnet werden. Ihre Bekämpfung muss durch Modellmodifikation erfolgen.

Bei der Erklärung von Stückkosten sind nicht nur Lerneffekte sondern ggf. auch Skaleneffekte oder Faktorpreise relevant. Es empfiehlt sich also zur Beseitigung der Fehlspezifikation (und damit der Autokorrelation) beispielsweise eine Spezifikation, die zusätzlich die kontemporäre Produktionsmenge berücksichtigt.

Aufgabe IV1

- a) Die OLS-Schätzung des Modells liefert:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	6,3019	1,7457	3,6099	0,0007
X	0,0207	0,0222	0,9286	0,3578

Der geschätzte Steigungsparameters ist wie erwartet positiv, jedoch sehr niedrig. Es zeigt sich außerdem, dass der Einfluss des Einkommens auf die Ersparnisse nicht signifikant ist. Wir können ausschließen, dass aufgrund von Autokorrelation oder Heteroskedastizität verzerrte Standardfehler für dieses Ergebnis verantwortlich sind, da entsprechende Tests keine Anzeichen für diese Annahmenverletzungen liefern.

- b) Die IV- bzw. TSLS-Schätzung mit Instrument Z liefert:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	2,2748	2,6343	0,8635	0,3921
X	0,0727	0,0338	2,1498	0,0366

Der geschätzte Steigungsparameter ist wieder positiv, jedoch nun deutlich höher. $H_0: \beta_1 \leq 0$ kann außerdem bei $\alpha = 0,05$ aufgrund von $p = 0,0366/2 < 0,05$ gegenüber $H_A: \beta_1 > 0$ abgelehnt werden.

- c) In einem ersten Schritt wird die potenziell endogene Variable X auf das Instrument Z regressiert und die Residuen dieser Regression festgehalten. Diese werden als zusätzliche Variable V in das Ursprungsmodell aufgenommen:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	2,2748	2,3848	0,9539	0,3450
X	0,0727	0,0306	2,3748	0,0217
V	-0,1005	0,0425	-2,3627	0,0223

Der geschätzte Koeffizient der Residuenvariable ist bei $\alpha = 0,05$ signifikant von Null verschieden. Wir können daher die Nullhypothese, dass X und ϵ nicht korreliert sind, ablehnen. Dieses Ergebnis spricht gegen die OLS-Schätzung aus a).

Aufgabe IV3

- a) Die OLS-Schätzung liefert:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	1,7623	1,0553	1,6698	0,0994
X ₁	0,1468	0,0634	2,3149	0,0235
X ₂	0,4380	0,1176	3,7253	0,0004
X ₃	0,2392	0,0998	2,3964	0,0192

Alle Parameterschätzungen weisen das erwartete Vorzeichen auf und sind auf einem Niveau $\alpha = 0,05$ signifikant von Null verschieden.

Die Stichprobenmittelwerte für X_2 und X_3 sind 7,8347 und 10,0467. Wir erhalten daraus die Punktprognose (NB = neue Brauerei)

$$\hat{Y}_{NB} = \hat{\beta}_0 + \hat{\beta}_1 X_{1,NB} + \hat{\beta}_2 X_{2,NB} + \hat{\beta}_3 X_{3,NB}$$

$$= 1,7623 + 0,1468 \cdot 10 + 0,4380 \cdot 7,8347 + 0,2392 \cdot 10,0467 = 9,0647.$$

- b) In einem ersten Schritt wird die potenziell endogene Variable X_1 auf das Instrument Z und die exogenen Variablen X_2 und X_3 regressiert und die Residuen dieser Regression festgehalten. Diese werden als zusätzliche Variable V in das Ursprungsmodell aufgenommen:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	-2,4867	2,1896	-1,1357	0,2600
X_1	0,5121	0,1773	2,8882	0,0052
X_2	0,3321	0,1242	2,6737	0,0093
X_3	0,2400	0,0972	2,4685	0,0160
V	-0,4158	0,1892	-2,1978	0,0313

Der Hausman-Test auf Korrelation zwischen X_1 und ϵ entspricht einem Signifikanztest für die Residuenvariable V. Da der p-Wert 0,0313 unter $\alpha = 0,05$ liegt, kann gefolgert werden, dass X_1 und ϵ korreliert sind.

- c) Die IV- bzw. TSLS-Schätzung liefert folgendes Ergebnis:

	Koeffizient	Std.-Fehler	t-Statistik	p-Wert
C	-2,4867	2,7226	-0,9133	0,3642
X_1	0,5121	0,2205	2,3227	0,0231
X_2	0,3321	0,1545	2,1502	0,0349
X_3	0,2400	0,1209	1,9852	0,0510

Wie bei a) haben die Schätzungen das erwartete Vorzeichen. Im Vergleich zu den OLS-Ergebnissen haben sich die geschätzten Parameter der Variablen X_1 und X_2 deutlich verändert, während der von X_3 fast gleich geblieben ist. Alle Schätzungen sind auf einem 10 %-Niveau signifikant (bzw. auf einem 5 %-Niveau, wenn nur ein einseitiger t-Test verwendet wird).

Unsere neue Prognose ist nun

$$\hat{Y}_{NB} = \hat{\beta}_0^{IV} + \hat{\beta}_1^{IV} X_{1,NB} + \hat{\beta}_2^{IV} X_{2,NB} + \hat{\beta}_3^{IV} X_{3,NB}$$

$$= 1,7623 + 0,1468 \cdot 10 + 0,4380 \cdot 7,8347 + 0,2392 \cdot 10,0467 = 7,6475.$$

Dies zeigt, dass ein Ignorieren der Korrelation zwischen X_1 und ϵ in unserem Fall zu einer Überschätzung des Bieroutputs führt.

Aufgabe ZR1

Die Dickey-Fuller Stationaritätstests liefern folgende Ergebnisse, wobei aus Gründen der Übersichtlichkeit auf die Abbildung der Zusatzterme zur Berücksichtigung von Autokorrelation verzichtet wird. Wir schließen Variablen nach dem sequenziellen Schema aus, wenn sie nicht mind. bei einem Niveau von 10 % signifikant sind.

	Y_1 (Stufe 1)		Y_2 (Stufe 1)	
	Koeffizient	p-Wert	Koeffizient	p-Wert
C	24,1291	0,2047	65,1034	0,6012
@TREND(1950)	4,3766	0,0737	9,6978	0,0446
$Y(-1)$	-0,0205	0,2540	-0,0566	0,0993
p-Wert (DF)	0,9103		0,7483	
Akaike-Lag-Zahl	1		1	

	Y_3 (Stufe 1)		Y_4 (Stufe 1)	
	Koeffizient	p-Wert	Koeffizient	p-Wert
C	0,3993	0,1424	-0,4730	0,6892
@TREND(1900)	0,0152	0,0083	0,1022	0,0165
$Y(-1)$	-0,0867	0,0255	-0,1910	0,0730
p-Wert (DF)	0,4470		0,6895	
Akaike-Lag-Zahl	3		11	

Zum Test $\vartheta = 0$ darf nicht der individuelle p-Wert verwendet werden. Stattdessen ist der p-Wert des Dickey-Fuller Stationaritätstests heranzuziehen, der bereits die speziellen kritischen Werte berücksichtigt. Bei allen Variablen ist lediglich die Schätzung der Stufe 1 des Prüfungsschemas erforderlich, um eine Nicht-Stationarität aufzuzeigen. Zur Bestimmung des Integrationsgrades prüfen wir die ersten Differenzen der Zeitreihen auf Stationarität. Wir erhalten die folgenden Ergebnisse, die uns zeigen, dass die ersten Differenzen aller Reihen stationär und damit alle Reihen I(1) sind. Abgesehen von Variable Y_3 ist dafür nur Stufe 1 zu schätzen.

	ΔY_1 (Stufe 1)		ΔY_2 (Stufe 1)	
	Koeffizient	p-Wert	Koeffizient	p-Wert
C	22,3596	0,2391	41,4173	0,1939
@TREND(1950)	1,7780	0,0355	1,9773	0,0630
D($Y(-1)$)	-0,5427	0,0005	-0,7294	0,0000
p-Wert (DF)	0,0283		0,0008	
Akaike-Lag-Zahl	0		0	

	ΔY_3 (Stufe 1)		ΔY_3 (Stufe 2)		ΔY_4 (Stufe 1)	
	Koeffizient	p-Wert	Koeffizient	p-Wert	Koeffizient	p-Wert
C	-0,0533	0,7454	0,1327	0,0986	-1,1355	0,2814
@TREND(1900)	0,0034	0,1967			0,0441	0,0126
D($Y(-1)$)	-0,7340	0,0000	-0,7017	0,0000	-3,4067	0,0000
p-Wert (DF)	0,0001		0,0000		0,0000	
Akaike-Lag-Zahl	3		3		7	

VI ANHANG

Hier im Anhang finden sich in der Statistik bedeutende Tafeln. Dazu gehören neben ausgewählten Binomialkoeffizienten (nicht jeder Taschenrechner unterstützt deren automatische Berechnung) vor allem Wahrscheinlichkeitsverteilungen bzw. konkret die

- Binomialverteilung (Verteilungsfunktion),
- Poissonverteilung (Verteilungsfunktion),
- Standardnormalverteilung (Verteilungsfunktion, Quantile),
- t-Verteilung (Quantile),
- Chi-Quadrat-Verteilung (Quantile) und
- F-Verteilung (Quantile).

Darüber hinaus sind die wichtigsten ökonometrischen Tabellen enthalten. Konkret liefern wir kritische Werte zum

- Durbin-Watson-Autokorrelationstest und
- Dickey-Fuller-Stationaritätstest.

Sämtliche Tabellen stellen wir zum besseren Einsatz in der Lehre auch in unserem Online-Service in digitaler Form zur Verfügung. Dort finden Sie auch Verteilungen (z.B. die Hypergeometrische Verteilung), die aus Platzgründen nicht im Buch abgedruckt wurden.

1.

Statistische Tafeln

1.1 Binomialkoeffizienten

Definition für $N, n \in \mathbb{N}$ und $N \geq n$:

$$\binom{N}{n} = \binom{N}{N-n} = \frac{N \cdot (N-1) \cdot \dots \cdot (N-n+1)}{n!} = \frac{N!}{n! \cdot (N-n)!}$$

Sonderfälle:

$$\binom{N}{0} = 1, \quad \binom{N}{N} = 1, \quad \binom{N}{1} = N, \quad \binom{N}{N-1} = N$$

N	n	0	1	2	3	4	5	6	7	8	9	10
0	1	1										
1	1	1	1									
2	1	2	1									
3	1	3	3	1								
4	1	4	6	4	1							
5	1	5	10	10	5	1						
6	1	6	15	20	15	6	1					
7	1	7	21	35	35	21	7	1				
8	1	8	28	56	70	56	28	8	1			
9	1	9	36	84	126	126	84	36	9	1		
10	1	10	45	120	210	252	210	120	45	10	1	
11	1	11	55	165	330	462	462	330	165	55	11	
12	1	12	66	220	495	792	924	792	495	220	66	
13	1	13	78	286	715	1287	1716	1716	1287	715	286	
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003	
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008	
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448	
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758	
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378	
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756	
21	1	21	210	1330	5985	20349	54264	116280	203490	293930	352716	
22	1	22	231	1540	7315	26334	74613	170544	319770	497420	646646	
23	1	23	253	1771	8855	33649	100947	245157	490314	817190	1144066	
24	1	24	276	2024	10626	42504	134596	346104	735471	1307504	1961256	
25	1	25	300	2300	12650	53130	177100	480700	1081575	2042975	3268760	
26	1	26	325	2600	14950	65780	230230	657800	1562275	3124550	5311735	
27	1	27	351	2925	17550	80730	296010	888030	2220075	4686825	8436285	
28	1	28	378	3276	20475	98280	376740	1184040	3108105	6906900	13123110	
29	1	29	406	3654	23751	118755	475020	1560780	4292145	10015005	20030010	
30	1	30	435	4060	27405	142506	593775	2035800	5852925	14307150	30045015	

1.2 Binomialverteilung - Verteilungsfunktion

Ablesebeispiel:

Nehmen wir an, wir interessieren uns für die Wahrscheinlichkeit, dass eine $B(7;0,1)$ -verteilte Zufallsvariable X einen Wert von höchstens 2 annimmt, so erhalten wir diesen aus der untenstehenden Tabelle zu $P(X \leq 2) = 0,9743$. Die Wahrscheinlichkeit $P(X = 2)$ erhalten wir über $P(X = 2) = P(X \leq 2) - P(X \leq 1) = 0,9743 - 0,8503 = 0,1240$.

Hinweis:

Für $p > 0,5$ kann der gesuchte Wert über $F_{B(n,p)}(x) = 1 - F_{B(n,1-p)}(n-x-1)$ bestimmt werden.

n	x	p									
		0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
8	0	0,6634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0319	0,0168	0,0084	0,0039
	1	0,9428	0,8131	0,6572	0,5033	0,3671	0,2553	0,1691	0,1064	0,0632	0,0352
	2	0,9942	0,9619	0,8948	0,7969	0,6785	0,5518	0,4278	0,3154	0,2201	0,1445
	3	0,9996	0,9950	0,9786	0,9437	0,8862	0,8059	0,7064	0,5941	0,4770	0,3633
	4	1,0000	0,9996	0,9971	0,9896	0,9727	0,9420	0,8939	0,8263	0,7396	0,6367
	5	1,0000	1,0000	0,9998	0,9988	0,9958	0,9887	0,9747	0,9502	0,9115	0,8555
	6	1,0000	1,0000	1,0000	0,9999	0,9996	0,9987	0,9964	0,9915	0,9819	0,9648
	7	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9993	0,9983	0,9961
9	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	0	0,6302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0207	0,0101	0,0046	0,0020
	1	0,9288	0,7748	0,5995	0,4362	0,3003	0,1960	0,1211	0,0705	0,0385	0,0195
	2	0,9916	0,9470	0,8591	0,7382	0,6007	0,4628	0,3373	0,2318	0,1495	0,0898
	3	0,9994	0,9917	0,9661	0,9144	0,8343	0,7297	0,6089	0,4826	0,3614	0,2539
	4	1,0000	0,9991	0,9944	0,9804	0,9511	0,9012	0,8283	0,7334	0,6214	0,5000
	5	1,0000	0,9999	0,9994	0,9969	0,9900	0,9747	0,9464	0,9006	0,8342	0,7461
	6	1,0000	1,0000	1,0000	0,9997	0,9987	0,9957	0,9888	0,9750	0,9502	0,9102
	7	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9986	0,9962	0,9909	0,9805
10	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9992	0,9980
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	0	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010
	1	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,0860	0,0464	0,0233	0,0107
	2	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2616	0,1673	0,0996	0,0547
	3	0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5138	0,3823	0,2660	0,1719
	4	0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7515	0,6331	0,5044	0,3770
	5	1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9051	0,8338	0,7384	0,6230
	6	1,0000	1,0000	0,9999	0,9991	0,9965	0,9894	0,9740	0,9452	0,8980	0,8281
	7	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984	0,9952	0,9877	0,9726	0,9453
	8	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9983	0,9955	0,9893
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990
11	10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	0	0,5688	0,3138	0,1673	0,0859	0,0422	0,0198	0,0088	0,0036	0,0014	0,0005
	1	0,8981	0,6974	0,4922	0,3221	0,1971	0,1130	0,0606	0,0302	0,0139	0,0059
	2	0,9848	0,9104	0,7788	0,6174	0,4552	0,3127	0,2001	0,1189	0,0652	0,0327
	3	0,9984	0,9815	0,9306	0,8389	0,7133	0,5696	0,4256	0,2963	0,1911	0,1133
	4	0,9999	0,9972	0,9841	0,9496	0,8854	0,7897	0,6683	0,5328	0,3971	0,2744
	5	1,0000	0,9997	0,9973	0,9883	0,9657	0,9218	0,8513	0,7535	0,6331	0,5000
	6	1,0000	1,0000	0,9997	0,9980	0,9924	0,9784	0,9499	0,9006	0,8262	0,7256
	7	1,0000	1,0000	1,0000	0,9998	0,9988	0,9957	0,9878	0,9707	0,9390	0,8867
	8	1,0000	1,0000	1,0000	1,0000	0,9999	0,9994	0,9980	0,9941	0,9852	0,9673
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9978	0,9941
	10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9995
	11	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
12	0	0,5404	0,2824	0,1422	0,0687	0,0317	0,0138	0,0057	0,0022	0,0008	0,0002
	1	0,8816	0,6590	0,4435	0,2749	0,1584	0,0850	0,0424	0,0196	0,0083	0,0032
	2	0,9804	0,8891	0,7358	0,5583	0,3907	0,2528	0,1513	0,0834	0,0421	0,0193
	3	0,9978	0,9744	0,9078	0,7946	0,6488	0,4925	0,3467	0,2253	0,1345	0,0730
	4	0,9998	0,9957	0,9761	0,9274	0,8424	0,7237	0,5833	0,4382	0,3044	0,1938
	5	1,0000	0,9995	0,9954	0,9806	0,9456	0,8822	0,7873	0,6652	0,5269	0,3872
	6	1,0000	0,9999	0,9993	0,9961	0,9857	0,9614	0,9154	0,8418	0,7393	0,6128
	7	1,0000	1,0000	0,9999	0,9994	0,9972	0,9905	0,9745	0,9427	0,8883	0,8062

n	x	p										
			0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
18	17	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	18	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
19	0	0,3774	0,1351	0,0456	0,0144	0,0042	0,0011	0,0003	0,0001	0,0000	0,0000	0,0000
	1	0,7547	0,4203	0,1985	0,0829	0,0310	0,0104	0,0031	0,0008	0,0002	0,0000	0,0000
	2	0,9335	0,7054	0,4413	0,2369	0,1113	0,0462	0,0170	0,0055	0,0015	0,0004	0,0000
	3	0,9868	0,8850	0,6841	0,4551	0,2631	0,1332	0,0591	0,0230	0,0077	0,0022	0,0000
	4	0,9980	0,9648	0,8556	0,6733	0,4654	0,2822	0,1500	0,0696	0,0280	0,0096	0,0000
	5	0,9998	0,9914	0,9463	0,8369	0,6678	0,4739	0,2968	0,1629	0,0777	0,0318	0,0000
	6	1,0000	0,9983	0,9837	0,9324	0,8251	0,6655	0,4812	0,3081	0,1727	0,0835	0,0000
	7	1,0000	0,9997	0,9959	0,9767	0,9225	0,8180	0,6656	0,4878	0,3169	0,1796	0,0000
	8	1,0000	1,0000	0,9992	0,9933	0,9713	0,9161	0,8145	0,6675	0,4940	0,3238	0,0000
	9	1,0000	1,0000	0,9999	0,9984	0,9911	0,9674	0,9125	0,8139	0,6710	0,5000	0,0000
	10	1,0000	1,0000	1,0000	0,9997	0,9977	0,9895	0,9653	0,9115	0,8159	0,6762	0,0000
	11	1,0000	1,0000	1,0000	1,0000	0,9995	0,9972	0,9886	0,9648	0,9129	0,8204	0,0000
	12	1,0000	1,0000	1,0000	1,0000	0,9999	0,9994	0,9969	0,9884	0,9658	0,9165	0,0000
	13	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9993	0,9969	0,9891	0,9682	0,0000
	14	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9994	0,9972	0,9904	0,0000
	15	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9978	0,0000
	16	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,0000
	17	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,0000
	18	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,0000
	19	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,0000
20	0	0,3585	0,1216	0,0388	0,0115	0,0032	0,0008	0,0002	0,0000	0,0000	0,0000	0,0000
	1	0,7358	0,3917	0,1756	0,0692	0,0243	0,0076	0,0021	0,0005	0,0001	0,0000	0,0000
	2	0,9245	0,6769	0,4049	0,2061	0,0913	0,0355	0,0121	0,0036	0,0009	0,0002	0,0000
	3	0,9841	0,8670	0,6477	0,4114	0,2252	0,1071	0,0444	0,0160	0,0049	0,0013	0,0000
	4	0,9974	0,9568	0,8298	0,6296	0,4148	0,2375	0,1182	0,0510	0,0189	0,0059	0,0000
	5	0,9997	0,9887	0,9327	0,8042	0,6172	0,4164	0,2454	0,1256	0,0553	0,0207	0,0000
	6	1,0000	0,9976	0,9781	0,9133	0,7858	0,6080	0,4166	0,2500	0,1299	0,0577	0,0000
	7	1,0000	0,9996	0,9941	0,9679	0,8982	0,7723	0,6010	0,4159	0,2520	0,1316	0,0000
	8	1,0000	0,9999	0,9987	0,9900	0,9591	0,8867	0,7624	0,5956	0,4143	0,2517	0,0000
	9	1,0000	1,0000	0,9998	0,9974	0,9861	0,9520	0,8782	0,7553	0,5914	0,4119	0,0000
	10	1,0000	1,0000	1,0000	0,9994	0,9961	0,9829	0,9468	0,8725	0,7507	0,5881	0,0000
	11	1,0000	1,0000	1,0000	0,9999	0,9991	0,9949	0,9804	0,9435	0,8692	0,7483	0,0000
	12	1,0000	1,0000	1,0000	1,0000	0,9998	0,9987	0,9940	0,9790	0,9420	0,8684	0,0000
	13	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9985	0,9935	0,9786	0,9423	0,0000
	14	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9984	0,9936	0,9793	0,0000
	15	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9985	0,9941	0,0000
	16	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9987	0,0000
	17	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,0000
	18	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,0000
	19	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,0000
	20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,0000
25	0	0,2774	0,0718	0,0172	0,0038	0,0008	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000
	1	0,6424	0,2712	0,0931	0,0274	0,0070	0,0016	0,0003	0,0001	0,0000	0,0000	0,0000
	2	0,8729	0,5371	0,2537	0,0982	0,0321	0,0090	0,0021	0,0004	0,0001	0,0000	0,0000
	3	0,9659	0,7636	0,4711	0,2340	0,0962	0,0332	0,0097	0,0024	0,0005	0,0001	0,0000
	4	0,9928	0,9020	0,6821	0,4207	0,2137	0,0905	0,0320	0,0095	0,0023	0,0005	0,0000
	5	0,9988	0,9666	0,8385	0,6167	0,3783	0,1935	0,0826	0,0294	0,0086	0,0020	0,0000
	6	0,9998	0,9905	0,9305	0,7800	0,5611	0,3407	0,1734	0,0736	0,0258	0,0073	0,0000
	7	1,0000	0,9977	0,9745	0,8909	0,7265	0,5118	0,3061	0,1536	0,0639	0,0216	0,0000

n	x	p									
		0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
25	8	1,0000	0,9995	0,9920	0,9532	0,8506	0,6769	0,4668	0,2735	0,1340	0,0539
	9	1,0000	0,9999	0,9979	0,9827	0,9287	0,8106	0,6303	0,4246	0,2424	0,1148
	10	1,0000	1,0000	0,9995	0,9944	0,9703	0,9022	0,7712	0,5858	0,3843	0,2122
	11	1,0000	1,0000	0,9999	0,9985	0,9893	0,9558	0,8746	0,7323	0,5426	0,3450
	12	1,0000	1,0000	1,0000	0,9996	0,9966	0,9825	0,9396	0,8462	0,6937	0,5000
	13	1,0000	1,0000	1,0000	0,9999	0,9991	0,9940	0,9745	0,9222	0,8173	0,6550
	14	1,0000	1,0000	1,0000	1,0000	0,9998	0,9982	0,9907	0,9656	0,9040	0,7878
	15	1,0000	1,0000	1,0000	1,0000	1,0000	0,9995	0,9971	0,9868	0,9560	0,8852
	16	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9992	0,9957	0,9826	0,9461
	17	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9988	0,9942	0,9784
	18	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9997	0,9984	0,9927
	19	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9980
	20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995
	21	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
	22	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
30	0	0,2146	0,0424	0,0076	0,0012	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000
	1	0,5535	0,1837	0,0480	0,0105	0,0020	0,0003	0,0000	0,0000	0,0000	0,0000
	2	0,8122	0,4114	0,1514	0,0442	0,0106	0,0021	0,0003	0,0000	0,0000	0,0000
	3	0,9392	0,6474	0,3217	0,1227	0,0374	0,0093	0,0019	0,0003	0,0000	0,0000
	4	0,9844	0,8245	0,5245	0,2552	0,0979	0,0302	0,0075	0,0015	0,0002	0,0000
	5	0,9967	0,9268	0,7106	0,4275	0,2026	0,0766	0,0233	0,0057	0,0011	0,0002
	6	0,9994	0,9742	0,8474	0,6070	0,3481	0,1595	0,0586	0,0172	0,0040	0,0007
	7	0,9999	0,9922	0,9302	0,7608	0,5143	0,2814	0,1238	0,0435	0,0121	0,0026
	8	1,0000	0,9980	0,9722	0,8713	0,6736	0,4315	0,2247	0,0940	0,0312	0,0081
	9	1,0000	0,9995	0,9903	0,9389	0,8034	0,5888	0,3575	0,1763	0,0694	0,0214
	10	1,0000	0,9999	0,9971	0,9744	0,8943	0,7304	0,5078	0,2915	0,1350	0,0494
	11	1,0000	1,0000	0,9992	0,9905	0,9493	0,8407	0,6548	0,4311	0,2327	0,1002
	12	1,0000	1,0000	0,9998	0,9969	0,9784	0,9155	0,7802	0,5785	0,3592	0,1808
	13	1,0000	1,0000	1,0000	0,9991	0,9918	0,9599	0,8737	0,7145	0,5025	0,2923
	14	1,0000	1,0000	1,0000	0,9998	0,9973	0,9831	0,9348	0,8246	0,6448	0,4278
	15	1,0000	1,0000	1,0000	0,9999	0,9992	0,9936	0,9699	0,9029	0,7691	0,5722
	16	1,0000	1,0000	1,0000	0,9998	0,9979	0,9876	0,9519	0,8644	0,7077	
	17	1,0000	1,0000	1,0000	0,9999	0,9999	0,9994	0,9955	0,9788	0,9286	0,8192
	18	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9986	0,9917	0,9666	0,8998
	19	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9996	0,9971	0,9862	0,9506
	20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9991	0,9950	0,9786
	21	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998	0,9984	0,9919
	22	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9996	0,9974
	23	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9993
	24	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9998
	25	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	26	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
50	0	0,0769	0,0052	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	1	0,2794	0,0338	0,0029	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	2	0,5405	0,1117	0,0142	0,0013	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000
	3	0,7604	0,2503	0,0460	0,0057	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000
	4	0,8964	0,4312	0,1121	0,0185	0,0021	0,0002	0,0000	0,0000	0,0000	0,0000
	5	0,9622	0,6161	0,2194	0,0480	0,0070	0,0007	0,0001	0,0000	0,0000	0,0000
	6	0,9882	0,7702	0,3613	0,1034	0,0194	0,0025	0,0002	0,0000	0,0000	0,0000
	7	0,9968	0,8779	0,5188	0,1904	0,0453	0,0073	0,0008	0,0001	0,0000	0,0000
	8	0,9992	0,9421	0,6681	0,3073	0,0916	0,0183	0,0025	0,0002	0,0000	0,0000
	9	0,9998	0,9755	0,7911	0,4437	0,1637	0,0402	0,0067	0,0008	0,0001	0,0000

1.3 Poissonverteilung - Verteilungsfunktion

Ablesebeispiel:

Interessieren wir uns für die Wahrscheinlichkeit, dass eine Po(1)-verteilte Zufallsvariable X einen Wert von höchstens 3 annimmt, so erhalten wir diese als $P(X \leq 3) = 0,9810$. Die Wahrscheinlichkeit $P(X = 3)$ erhalten wir als $P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0,9810 - 0,9197 = 0,0613$.

1.4 Standardnormalverteilung - Verteilungsfunktion

Ablesebeispiel:

Interessieren wir uns für die Wahrscheinlichkeit, dass eine standardnormalverteilte Zufallsvariable Z einen Wert von höchstens 0,15 annimmt, so erhalten wir diese konkret als $P(Z \leq 0,15) = \Phi(0,15) = 0,5596$.

Hinweis:

Es gilt $\Phi(-z) = 1 - \Phi(z)$.

1.5 Standardnormalverteilung - Wichtige Quantile

Es gilt:

$$P(Z \leq z_\alpha) = \Phi(z_\alpha) = \alpha$$

α	z_α
0,005	-2,58
0,01	-2,33
0,025	-1,96
0,05	-1,64
0,1	-1,28
0,2	-0,84
0,25	-0,67
0,275	-0,60
0,3	-0,52
0,35	-0,39
0,375	-0,32
0,4	-0,25
0,45	-0,13
0,475	-0,06
0,5	0,00
0,55	0,13
0,575	0,19
0,6	0,25
0,65	0,39
0,675	0,45
0,7	0,52
0,75	0,67
0,775	0,76
0,8	0,84
0,85	1,04
0,875	1,15
0,9	1,28
0,95	1,64
0,975	1,96
0,99	2,33
0,995	2,58

1.6 Chi-Quadrat-Verteilung - Quantile

Ablesebeispiel:

Interessieren wir uns für das 40 %-Quantil der Chi-Quadrat-Verteilung mit $w = 2$ Freiheitsgraden, so erhalten wir dieses als $\chi^2_{2;0,40} = 1,02$. Es gilt somit für dieses Quantil $P(\chi^2_2 \leq 1,02) = F_{\chi^2_2}(1,02) = 0,40$.

w	Wahrscheinlichkeit									
	0,001	0,005	0,010	0,025	0,050	0,100	0,250	0,300	0,400	0,500
1	0,00	0,00	0,00	0,00	0,00	0,02	0,10	0,15	0,27	0,45
2	0,00	0,01	0,02	0,05	0,10	0,21	0,58	0,71	1,02	1,39
3	0,02	0,07	0,11	0,22	0,35	0,58	1,21	1,42	1,87	2,37
4	0,09	0,21	0,30	0,48	0,71	1,06	1,92	2,19	2,75	3,36
5	0,21	0,41	0,55	0,83	1,15	1,61	2,67	3,00	3,66	4,35
6	0,38	0,68	0,87	1,24	1,64	2,20	3,45	3,83	4,57	5,35
7	0,60	0,99	1,24	1,69	2,17	2,83	4,25	4,67	5,49	6,35
8	0,86	1,34	1,65	2,18	2,73	3,49	5,07	5,53	6,42	7,34
9	1,15	1,73	2,09	2,70	3,33	4,17	5,90	6,39	7,36	8,34
10	1,48	2,16	2,56	3,25	3,94	4,87	6,74	7,27	8,30	9,34
11	1,83	2,60	3,05	3,82	4,57	5,58	7,58	8,15	9,24	10,34
12	2,21	3,07	3,57	4,40	5,23	6,30	8,44	9,03	10,18	11,34
13	2,62	3,57	4,11	5,01	5,89	7,04	9,30	9,93	11,13	12,34
14	3,04	4,07	4,66	5,63	6,57	7,79	10,17	10,82	12,08	13,34
15	3,48	4,60	5,23	6,26	7,26	8,55	11,04	11,72	13,03	14,34
16	3,94	5,14	5,81	6,91	7,96	9,31	11,91	12,62	13,98	15,34
17	4,42	5,70	6,41	7,56	8,67	10,09	12,79	13,53	14,94	16,34
18	4,90	6,26	7,01	8,23	9,39	10,86	13,68	14,44	15,89	17,34
19	5,41	6,84	7,63	8,91	10,12	11,65	14,56	15,35	16,85	18,34
20	5,92	7,43	8,26	9,59	10,85	12,44	15,45	16,27	17,81	19,34
21	6,45	8,03	8,90	10,28	11,59	13,24	16,34	17,18	18,77	20,34
22	6,98	8,64	9,54	10,98	12,34	14,04	17,24	18,10	19,73	21,34
23	7,53	9,26	10,20	11,69	13,09	14,85	18,14	19,02	20,69	22,34
24	8,08	9,89	10,86	12,40	13,85	15,66	19,04	19,94	21,65	23,34
25	8,65	10,52	11,52	13,12	14,61	16,47	19,94	20,87	22,62	24,34
26	9,22	11,16	12,20	13,84	15,38	17,29	20,84	21,79	23,58	25,34
27	9,80	11,81	12,88	14,57	16,15	18,11	21,75	22,72	24,54	26,34
28	10,39	12,46	13,56	15,31	16,93	18,94	22,66	23,65	25,51	27,34
29	10,99	13,12	14,26	16,05	17,71	19,77	23,57	24,58	26,48	28,34
30	11,59	13,79	14,95	16,79	18,49	20,60	24,48	25,51	27,44	29,34
40	17,92	20,71	22,16	24,43	26,51	29,05	33,66	34,87	37,13	39,34
50	24,67	27,99	29,71	32,36	34,76	37,69	42,94	44,31	46,86	49,33
60	31,74	35,53	37,48	40,48	43,19	46,46	52,29	53,81	56,62	59,33
70	39,04	43,28	45,44	48,76	51,74	55,33	61,70	63,35	66,40	69,33
80	46,52	51,17	53,54	57,15	60,39	64,28	71,14	72,92	76,19	79,33
90	54,16	59,20	61,75	65,65	69,13	73,29	80,62	82,51	85,99	89,33
100	61,92	67,33	70,06	74,22	77,93	82,36	90,13	92,13	95,81	99,33

W	Wahrscheinlichkeit									
	0,600	0,700	0,750	0,900	0,950	0,975	0,980	0,990	0,995	0,999
1	0,71	1,07	1,32	2,71	3,84	5,02	5,41	6,63	7,88	10,83
2	1,83	2,41	2,77	4,61	5,99	7,38	7,82	9,21	10,60	13,82
3	2,95	3,66	4,11	6,25	7,81	9,35	9,84	11,34	12,84	16,27
4	4,04	4,88	5,39	7,78	9,49	11,14	11,67	13,28	14,86	18,47
5	5,13	6,06	6,63	9,24	11,07	12,83	13,39	15,09	16,75	20,51
6	6,21	7,23	7,84	10,64	12,59	14,45	15,03	16,81	18,55	22,46
7	7,28	8,38	9,04	12,02	14,07	16,01	16,62	18,48	20,28	24,32
8	8,35	9,52	10,22	13,36	15,51	17,53	18,17	20,09	21,95	26,12
9	9,41	10,66	11,39	14,68	16,92	19,02	19,68	21,67	23,59	27,88
10	10,47	11,78	12,55	15,99	18,31	20,48	21,16	23,21	25,19	29,59
11	11,53	12,90	13,70	17,28	19,68	21,92	22,62	24,73	26,76	31,26
12	12,58	14,01	14,85	18,55	21,03	23,34	24,05	26,22	28,30	32,91
13	13,64	15,12	15,98	19,81	22,36	24,74	25,47	27,69	29,82	34,53
14	14,69	16,22	17,12	21,06	23,68	26,12	26,87	29,14	31,32	36,12
15	15,73	17,32	18,25	22,31	25,00	27,49	28,26	30,58	32,80	37,70
16	16,78	18,42	19,37	23,54	26,30	28,85	29,63	32,00	34,27	39,25
17	17,82	19,51	20,49	24,77	27,59	30,19	31,00	33,41	35,72	40,79
18	18,87	20,60	21,60	25,99	28,87	31,53	32,35	34,81	37,16	42,31
19	19,91	21,69	22,72	27,20	30,14	32,85	33,69	36,19	38,58	43,82
20	20,95	22,77	23,83	28,41	31,41	34,17	35,02	37,57	40,00	45,31
21	21,99	23,86	24,93	29,62	32,67	35,48	36,34	38,93	41,40	46,80
22	23,03	24,94	26,04	30,81	33,92	36,78	37,66	40,29	42,80	48,27
23	24,07	26,02	27,14	32,01	35,17	38,08	38,97	41,64	44,18	49,73
24	25,11	27,10	28,24	33,20	36,42	39,36	40,27	42,98	45,56	51,18
25	26,14	28,17	29,34	34,38	37,65	40,65	41,57	44,31	46,93	52,62
26	27,18	29,25	30,43	35,56	38,89	41,92	42,86	45,64	48,29	54,05
27	28,21	30,32	31,53	36,74	40,11	43,19	44,14	46,96	49,65	55,48
28	29,25	31,39	32,62	37,92	41,34	44,46	45,42	48,28	50,99	56,89
29	30,28	32,46	33,71	39,09	42,56	45,72	46,69	49,59	52,34	58,30
30	31,32	33,53	34,80	40,26	43,77	46,98	47,96	50,89	53,67	59,70
40	41,62	44,16	45,62	51,81	55,76	59,34	60,44	63,69	66,77	73,40
50	51,89	54,72	56,33	63,17	67,50	71,42	72,61	76,15	79,49	86,66
60	62,13	65,23	66,98	74,40	79,08	83,30	84,58	88,38	91,95	99,61
70	72,36	75,69	77,58	85,53	90,53	95,02	96,39	100,43	104,21	112,32
80	82,57	86,12	88,13	96,58	101,88	106,63	108,07	112,33	116,32	124,84
90	92,76	96,52	98,65	107,57	113,15	118,14	119,65	124,12	128,30	137,21
100	102,95	106,91	109,14	118,50	124,34	129,56	131,14	135,81	140,17	149,45

1.7 t-Verteilung - Quantile

Ablesebeispiel:

Interessieren wir uns für das 99 %-Quantil einer t-Verteilung mit $w = 10$ Freiheitsgraden, so erhalten wir dieses als $t_{10; 0,99} = 2,76$. Es gilt somit $P(T_{10} \leq 2,76) = F_{T_{10}}(2,76) = 0,99$.

Hinweis:

Es gilt $F_T(-t_\alpha) = 1 - F_T(t_\alpha)$.

w	Wahrscheinlichkeit									
	0,600	0,700	0,750	0,800	0,900	0,950	0,975	0,990	0,995	0,999
1	0,32	0,73	1,00	1,38	3,08	6,31	12,71	31,82	63,66	318,29
2	0,29	0,62	0,82	1,06	1,89	2,92	4,30	6,96	9,92	22,33
3	0,28	0,58	0,76	0,98	1,64	2,35	3,18	4,54	5,84	10,21
4	0,27	0,57	0,74	0,94	1,53	2,13	2,78	3,75	4,60	7,17
5	0,27	0,56	0,73	0,92	1,48	2,02	2,57	3,36	4,03	5,89
6	0,26	0,55	0,72	0,91	1,44	1,94	2,45	3,14	3,71	5,21
7	0,26	0,55	0,71	0,90	1,41	1,89	2,36	3,00	3,50	4,79
8	0,26	0,55	0,71	0,89	1,40	1,86	2,31	2,90	3,36	4,50
9	0,26	0,54	0,70	0,88	1,38	1,83	2,26	2,82	3,25	4,30
10	0,26	0,54	0,70	0,88	1,37	1,81	2,23	2,76	3,17	4,14
11	0,26	0,54	0,70	0,88	1,36	1,80	2,20	2,72	3,11	4,02
12	0,26	0,54	0,70	0,87	1,36	1,78	2,18	2,68	3,05	3,93
13	0,26	0,54	0,69	0,87	1,35	1,77	2,16	2,65	3,01	3,85
14	0,26	0,54	0,69	0,87	1,35	1,76	2,14	2,62	2,98	3,79
15	0,26	0,54	0,69	0,87	1,34	1,75	2,13	2,60	2,95	3,73
16	0,26	0,54	0,69	0,86	1,34	1,75	2,12	2,58	2,92	3,69
17	0,26	0,53	0,69	0,86	1,33	1,74	2,11	2,57	2,90	3,65
18	0,26	0,53	0,69	0,86	1,33	1,73	2,10	2,55	2,88	3,61
19	0,26	0,53	0,69	0,86	1,33	1,73	2,09	2,54	2,86	3,58
20	0,26	0,53	0,69	0,86	1,33	1,72	2,09	2,53	2,85	3,55
21	0,26	0,53	0,69	0,86	1,32	1,72	2,08	2,52	2,83	3,53
22	0,26	0,53	0,69	0,86	1,32	1,72	2,07	2,51	2,82	3,50
23	0,26	0,53	0,69	0,86	1,32	1,71	2,07	2,50	2,81	3,48
24	0,26	0,53	0,68	0,86	1,32	1,71	2,06	2,49	2,80	3,47
25	0,26	0,53	0,68	0,86	1,32	1,71	2,06	2,49	2,79	3,45
26	0,26	0,53	0,68	0,86	1,31	1,71	2,06	2,48	2,78	3,43
27	0,26	0,53	0,68	0,86	1,31	1,70	2,05	2,47	2,77	3,42
28	0,26	0,53	0,68	0,85	1,31	1,70	2,05	2,47	2,76	3,41
29	0,26	0,53	0,68	0,85	1,31	1,70	2,05	2,46	2,76	3,40
30	0,26	0,53	0,68	0,85	1,31	1,70	2,04	2,46	2,75	3,39
40	0,26	0,53	0,68	0,85	1,30	1,68	2,02	2,42	2,70	3,31
50	0,25	0,53	0,68	0,85	1,30	1,68	2,01	2,40	2,68	3,26
60	0,25	0,53	0,68	0,85	1,30	1,67	2,00	2,39	2,66	3,23
70	0,25	0,53	0,68	0,85	1,29	1,67	1,99	2,38	2,65	3,21
80	0,25	0,53	0,68	0,85	1,29	1,66	1,99	2,37	2,64	3,20
90	0,25	0,53	0,68	0,85	1,29	1,66	1,99	2,37	2,63	3,18
100	0,25	0,53	0,68	0,85	1,29	1,66	1,98	2,36	2,63	3,17

1.8 F-Verteilung - Quantile

Ablesebeispiel:

Interessieren wir uns für das 90 %-Quantil einer F-Verteilung mit $u = 5$ (Zählerfreiheitsgrade) und $w = 10$ (Nennerfreiheitsgrade) Freiheitsgraden, so erhalten wir dieses konkret als $f_{0,95; 5; 10} = 2,52$. Es gilt also $P(F_{5;10} \leq 2,52) = F_{F_{5;10}}(2,52) = 0,90$.

Hinweis:

Es gilt $f_{\alpha, u, w} = 1 / f_{1-\alpha, w, u}$.

Wahrscheinlichkeit 0,90

w	u	1	2	3	4	5	6	7	8	9	10
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73	
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	
70	2,78	2,38	2,16	2,03	1,93	1,86	1,80	1,76	1,72	1,69	
80	2,77	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71	1,68	
90	2,76	2,36	2,15	2,01	1,91	1,84	1,78	1,74	1,70	1,67	
100	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69	1,66	

Wahrscheinlichkeit 0,90

w u	11	12	13	14	15	20	30	40	50	100
1	60,47	60,71	60,90	61,07	61,22	61,74	62,26	62,53	62,69	63,01
2	9,40	9,41	9,41	9,42	9,42	9,44	9,46	9,47	9,47	9,48
3	5,22	5,22	5,21	5,20	5,20	5,18	5,17	5,16	5,15	5,14
4	3,91	3,90	3,89	3,88	3,87	3,84	3,82	3,80	3,80	3,78
5	3,28	3,27	3,26	3,25	3,24	3,21	3,17	3,16	3,15	3,13
6	2,92	2,90	2,89	2,88	2,87	2,84	2,80	2,78	2,77	2,75
7	2,68	2,67	2,65	2,64	2,63	2,59	2,56	2,54	2,52	2,50
8	2,52	2,50	2,49	2,48	2,46	2,42	2,38	2,36	2,35	2,32
9	2,40	2,38	2,36	2,35	2,34	2,30	2,25	2,23	2,22	2,19
10	2,30	2,28	2,27	2,26	2,24	2,20	2,16	2,13	2,12	2,09
11	2,23	2,21	2,19	2,18	2,17	2,12	2,08	2,05	2,04	2,01
12	2,17	2,15	2,13	2,12	2,10	2,06	2,01	1,99	1,97	1,94
13	2,12	2,10	2,08	2,07	2,05	2,01	1,96	1,93	1,92	1,88
14	2,07	2,05	2,04	2,02	2,01	1,96	1,91	1,89	1,87	1,83
15	2,04	2,02	2,00	1,99	1,97	1,92	1,87	1,85	1,83	1,79
16	2,01	1,99	1,97	1,95	1,94	1,89	1,84	1,81	1,79	1,76
17	1,98	1,96	1,94	1,93	1,91	1,86	1,81	1,78	1,76	1,73
18	1,95	1,93	1,92	1,90	1,89	1,84	1,78	1,75	1,74	1,70
19	1,93	1,91	1,89	1,88	1,86	1,81	1,76	1,73	1,71	1,67
20	1,91	1,89	1,87	1,86	1,84	1,79	1,74	1,71	1,69	1,65
21	1,90	1,87	1,86	1,84	1,83	1,78	1,72	1,69	1,67	1,63
22	1,88	1,86	1,84	1,83	1,81	1,76	1,70	1,67	1,65	1,61
23	1,87	1,84	1,83	1,81	1,80	1,74	1,69	1,66	1,64	1,59
24	1,85	1,83	1,81	1,80	1,78	1,73	1,67	1,64	1,62	1,58
25	1,84	1,82	1,80	1,79	1,77	1,72	1,66	1,63	1,61	1,56
26	1,83	1,81	1,79	1,77	1,76	1,71	1,65	1,61	1,59	1,55
27	1,82	1,80	1,78	1,76	1,75	1,70	1,64	1,60	1,58	1,54
28	1,81	1,79	1,77	1,75	1,74	1,69	1,63	1,59	1,57	1,53
29	1,80	1,78	1,76	1,75	1,73	1,68	1,62	1,58	1,56	1,52
30	1,79	1,77	1,75	1,74	1,72	1,67	1,61	1,57	1,55	1,51
40	1,74	1,71	1,70	1,68	1,66	1,61	1,54	1,51	1,48	1,43
50	1,70	1,68	1,66	1,64	1,63	1,57	1,50	1,46	1,44	1,39
60	1,68	1,66	1,64	1,62	1,60	1,54	1,48	1,44	1,41	1,36
70	1,66	1,64	1,62	1,60	1,59	1,53	1,46	1,42	1,39	1,34
80	1,65	1,63	1,61	1,59	1,57	1,51	1,44	1,40	1,38	1,32
90	1,64	1,62	1,60	1,58	1,56	1,50	1,43	1,39	1,36	1,30
100	1,64	1,61	1,59	1,57	1,56	1,49	1,42	1,38	1,35	1,29

Wahrscheinlichkeit 0,95

w u	1	2	3	4	5	6	7	8	9	10
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93

Wahrscheinlichkeit 0,95

w u	11	12	13	14	15	20	30	40	50	100
1	242,98	243,90	244,69	245,36	245,95	248,02	250,10	251,14	251,77	253,04
2	19,40	19,41	19,42	19,42	19,43	19,45	19,46	19,47	19,48	19,49
3	8,76	8,74	8,73	8,71	8,70	8,66	8,62	8,59	8,58	8,55
4	5,94	5,91	5,89	5,87	5,86	5,80	5,75	5,72	5,70	5,66
5	4,70	4,68	4,66	4,64	4,62	4,56	4,50	4,46	4,44	4,41
6	4,03	4,00	3,98	3,96	3,94	3,87	3,81	3,77	3,75	3,71
7	3,60	3,57	3,55	3,53	3,51	3,44	3,38	3,34	3,32	3,27
8	3,31	3,28	3,26	3,24	3,22	3,15	3,08	3,04	3,02	2,97
9	3,10	3,07	3,05	3,03	3,01	2,94	2,86	2,83	2,80	2,76
10	2,94	2,91	2,89	2,86	2,85	2,77	2,70	2,66	2,64	2,59
11	2,82	2,79	2,76	2,74	2,72	2,65	2,57	2,53	2,51	2,46
12	2,72	2,69	2,66	2,64	2,62	2,54	2,47	2,43	2,40	2,35
13	2,63	2,60	2,58	2,55	2,53	2,46	2,38	2,34	2,31	2,26
14	2,57	2,53	2,51	2,48	2,46	2,39	2,31	2,27	2,24	2,19
15	2,51	2,48	2,45	2,42	2,40	2,33	2,25	2,20	2,18	2,12
16	2,46	2,42	2,40	2,37	2,35	2,28	2,19	2,15	2,12	2,07
17	2,41	2,38	2,35	2,33	2,31	2,23	2,15	2,10	2,08	2,02
18	2,37	2,34	2,31	2,29	2,27	2,19	2,11	2,06	2,04	1,98
19	2,34	2,31	2,28	2,26	2,23	2,16	2,07	2,03	2,00	1,94
20	2,31	2,28	2,25	2,22	2,20	2,12	2,04	1,99	1,97	1,91
21	2,28	2,25	2,22	2,20	2,18	2,10	2,01	1,96	1,94	1,88
22	2,26	2,23	2,20	2,17	2,15	2,07	1,98	1,94	1,91	1,85
23	2,24	2,20	2,18	2,15	2,13	2,05	1,96	1,91	1,88	1,82
24	2,22	2,18	2,15	2,13	2,11	2,03	1,94	1,89	1,86	1,80
25	2,20	2,16	2,14	2,11	2,09	2,01	1,92	1,87	1,84	1,78
26	2,18	2,15	2,12	2,09	2,07	1,99	1,90	1,85	1,82	1,76
27	2,17	2,13	2,10	2,08	2,06	1,97	1,88	1,84	1,81	1,74
28	2,15	2,12	2,09	2,06	2,04	1,96	1,87	1,82	1,79	1,73
29	2,14	2,10	2,08	2,05	2,03	1,94	1,85	1,81	1,77	1,71
30	2,13	2,09	2,06	2,04	2,01	1,93	1,84	1,79	1,76	1,70
40	2,04	2,00	1,97	1,95	1,92	1,84	1,74	1,69	1,66	1,59
50	1,99	1,95	1,92	1,89	1,87	1,78	1,69	1,63	1,60	1,52
60	1,95	1,92	1,89	1,86	1,84	1,75	1,65	1,59	1,56	1,48
70	1,93	1,89	1,86	1,84	1,81	1,72	1,62	1,57	1,53	1,45
80	1,91	1,88	1,84	1,82	1,79	1,70	1,60	1,54	1,51	1,43
90	1,90	1,86	1,83	1,80	1,78	1,69	1,59	1,53	1,49	1,41
100	1,89	1,85	1,82	1,79	1,77	1,68	1,57	1,52	1,48	1,39

Wahrscheinlichkeit 0,99

w u	1	2	3	4	5	6	7	8	9	10
1	4052,18	4999,34	5403,53	5624,26	5763,96	5858,95	5928,33	5980,95	6022,40	6055,93
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50

Wahrscheinlichkeit 0,99

w u	11	12	13	14	15	20	30	40	50	100
1	6083,40	6106,68	6125,77	6143,00	6156,97	6208,66	6260,35	6286,43	6302,26	6333,92
2	99,41	99,42	99,42	99,43	99,43	99,45	99,47	99,48	99,48	99,49
3	27,13	27,05	26,98	26,92	26,87	26,69	26,50	26,41	26,35	26,24
4	14,45	14,37	14,31	14,25	14,20	14,02	13,84	13,75	13,69	13,58
5	9,96	9,89	9,82	9,77	9,72	9,55	9,38	9,29	9,24	9,13
6	7,79	7,72	7,66	7,60	7,56	7,40	7,23	7,14	7,09	6,99
7	6,54	6,47	6,41	6,36	6,31	6,16	5,99	5,91	5,86	5,75
8	5,73	5,67	5,61	5,56	5,52	5,36	5,20	5,12	5,07	4,96
9	5,18	5,11	5,05	5,01	4,96	4,81	4,65	4,57	4,52	4,41
10	4,77	4,71	4,65	4,60	4,56	4,41	4,25	4,17	4,12	4,01
11	4,46	4,40	4,34	4,29	4,25	4,10	3,94	3,86	3,81	3,71
12	4,22	4,16	4,10	4,05	4,01	3,86	3,70	3,62	3,57	3,47
13	4,02	3,96	3,91	3,86	3,82	3,66	3,51	3,43	3,38	3,27
14	3,86	3,80	3,75	3,70	3,66	3,51	3,35	3,27	3,22	3,11
15	3,73	3,67	3,61	3,56	3,52	3,37	3,21	3,13	3,08	2,98
16	3,62	3,55	3,50	3,45	3,41	3,26	3,10	3,02	2,97	2,86
17	3,52	3,46	3,40	3,35	3,31	3,16	3,00	2,92	2,87	2,76
18	3,43	3,37	3,32	3,27	3,23	3,08	2,92	2,84	2,78	2,68
19	3,36	3,30	3,24	3,19	3,15	3,00	2,84	2,76	2,71	2,60
20	3,29	3,23	3,18	3,13	3,09	2,94	2,78	2,69	2,64	2,54
21	3,24	3,17	3,12	3,07	3,03	2,88	2,72	2,64	2,58	2,48
22	3,18	3,12	3,07	3,02	2,98	2,83	2,67	2,58	2,53	2,42
23	3,14	3,07	3,02	2,97	2,93	2,78	2,62	2,54	2,48	2,37
24	3,09	3,03	2,98	2,93	2,89	2,74	2,58	2,49	2,44	2,33
25	3,06	2,99	2,94	2,89	2,85	2,70	2,54	2,45	2,40	2,29
26	3,02	2,96	2,90	2,86	2,81	2,66	2,50	2,42	2,36	2,25
27	2,99	2,93	2,87	2,82	2,78	2,63	2,47	2,38	2,33	2,22
28	2,96	2,90	2,84	2,79	2,75	2,60	2,44	2,35	2,30	2,19
29	2,93	2,87	2,81	2,77	2,73	2,57	2,41	2,33	2,27	2,16
30	2,91	2,84	2,79	2,74	2,70	2,55	2,39	2,30	2,25	2,13
40	2,73	2,66	2,61	2,56	2,52	2,37	2,20	2,11	2,06	1,94
50	2,63	2,56	2,51	2,46	2,42	2,27	2,10	2,01	1,95	1,82
60	2,56	2,50	2,44	2,39	2,35	2,20	2,03	1,94	1,88	1,75
70	2,51	2,45	2,40	2,35	2,31	2,15	1,98	1,89	1,83	1,70
80	2,48	2,42	2,36	2,31	2,27	2,12	1,94	1,85	1,79	1,65
90	2,45	2,39	2,33	2,29	2,24	2,09	1,92	1,82	1,76	1,62
100	2,43	2,37	2,31	2,27	2,22	2,07	1,89	1,80	1,74	1,60

2.

Ökonometrische Tafeln

2.1 Kritische Werte der Durbin-Watson-Statistik

$\alpha = 0,01$

n	K = 1		K = 2		K = 3		K = 4		K = 5		K = 6		K = 7	
	d _u	d _o												
6	0,39	1,14	—	—	—	—	—	—	—	—	—	—	—	—
7	0,44	1,04	0,29	1,68	—	—	—	—	—	—	—	—	—	—
8	0,50	1,00	0,35	1,49	0,23	2,10	—	—	—	—	—	—	—	—
9	0,55	1,00	0,41	1,39	0,28	1,88	0,18	2,43	—	—	—	—	—	—
10	0,60	1,00	0,47	1,33	0,34	1,73	0,23	2,19	0,15	2,69	—	—	—	—
11	0,65	1,01	0,52	1,30	0,40	1,64	0,29	2,03	0,19	2,45	0,12	2,89	—	—
12	0,70	1,02	0,57	1,27	0,45	1,58	0,34	1,91	0,24	2,28	0,16	2,67	0,11	3,05
13	0,74	1,04	0,62	1,26	0,50	1,53	0,39	1,83	0,29	2,15	0,21	2,49	0,14	2,84
14	0,78	1,05	0,66	1,25	0,55	1,49	0,44	1,76	0,34	2,05	0,26	2,35	0,18	2,67
15	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,97	0,30	2,24	0,23	2,53
16	0,84	1,09	0,74	1,25	0,63	1,45	0,53	1,66	0,44	1,90	0,35	2,15	0,27	2,42
17	0,87	1,10	0,77	1,26	0,67	1,43	0,57	1,63	0,48	1,85	0,39	2,08	0,31	2,32
18	0,90	1,12	0,81	1,26	0,71	1,42	0,61	1,60	0,52	1,80	0,44	2,02	0,36	2,24
19	0,93	1,13	0,84	1,27	0,74	1,42	0,65	1,58	0,56	1,77	0,48	1,96	0,40	2,17
20	0,95	1,15	0,86	1,27	0,77	1,41	0,69	1,57	0,60	1,74	0,52	1,92	0,44	2,11
21	0,98	1,16	0,89	1,28	0,80	1,41	0,72	1,55	0,63	1,71	0,55	1,88	0,47	2,06
22	1,00	1,17	0,91	1,28	0,83	1,41	0,75	1,54	0,67	1,69	0,59	1,85	0,51	2,02
23	1,02	1,19	0,94	1,29	0,86	1,41	0,78	1,53	0,70	1,67	0,62	1,82	0,55	1,98
24	1,04	1,20	0,96	1,30	0,88	1,41	0,81	1,53	0,73	1,66	0,65	1,80	0,58	1,94
25	1,06	1,21	0,98	1,31	0,91	1,41	0,83	1,52	0,76	1,65	0,68	1,78	0,61	1,92
26	1,07	1,22	1,00	1,31	0,93	1,41	0,86	1,52	0,78	1,64	0,71	1,76	0,64	1,89
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,52	0,81	1,63	0,74	1,74	0,67	1,87
28	1,10	1,24	1,04	1,33	0,97	1,42	0,90	1,51	0,83	1,62	0,76	1,73	0,70	1,85
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,86	1,61	0,79	1,72	0,72	1,83
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61	0,81	1,71	0,75	1,81
31	1,15	1,27	1,09	1,35	1,02	1,43	0,96	1,51	0,90	1,60	0,83	1,70	0,77	1,80
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60	0,86	1,69	0,79	1,79
33	1,17	1,29	1,11	1,36	1,06	1,43	1,00	1,51	0,94	1,59	0,88	1,68	0,82	1,78
34	1,18	1,30	1,13	1,36	1,07	1,44	1,01	1,51	0,95	1,59	0,90	1,68	0,84	1,77
35	1,20	1,31	1,14	1,37	1,09	1,44	1,03	1,51	0,97	1,59	0,91	1,67	0,86	1,76
36	1,21	1,32	1,15	1,38	1,10	1,44	1,04	1,51	0,99	1,59	0,93	1,67	0,88	1,75
37	1,22	1,32	1,17	1,38	1,11	1,45	1,06	1,51	1,00	1,59	0,95	1,66	0,90	1,74
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,59	0,97	1,66	0,91	1,74
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58	0,98	1,66	0,93	1,73
40	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58	1,00	1,65	0,95	1,72
45	1,29	1,38	1,25	1,42	1,20	1,47	1,16	1,53	1,11	1,58	1,07	1,64	1,02	1,70
50	1,32	1,40	1,29	1,45	1,25	1,49	1,21	1,54	1,16	1,59	1,12	1,64	1,08	1,69
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59	1,17	1,64	1,13	1,69
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60	1,21	1,64	1,18	1,68
65	1,41	1,47	1,38	1,50	1,35	1,53	1,32	1,57	1,28	1,60	1,25	1,64	1,22	1,68
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61	1,28	1,65	1,25	1,68
75	1,45	1,50	1,42	1,53	1,40	1,56	1,37	1,59	1,34	1,62	1,31	1,65	1,28	1,68
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62	1,34	1,65	1,31	1,68
85	1,48	1,53	1,46	1,55	1,44	1,58	1,41	1,60	1,39	1,63	1,36	1,66	1,34	1,69
90	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64	1,38	1,66	1,36	1,69
95	1,51	1,55	1,49	1,57	1,47	1,60	1,45	1,62	1,43	1,64	1,40	1,67	1,38	1,69
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65	1,42	1,67	1,40	1,69
150	1,61	1,64	1,60	1,65	1,58	1,67	1,57	1,68	1,56	1,69	1,54	1,71	1,53	1,72
200	1,66	1,68	1,65	1,69	1,64	1,70	1,63	1,72	1,62	1,73	1,61	1,74	1,60	1,75

$\alpha = 0,05$

n	K = 1		K = 2		K = 3		K = 4		K = 5		K = 6		K = 7	
	d _u	d _o												
6	0,61	1,40	—	—	—	—	—	—	—	—	—	—	—	—
7	0,70	1,36	0,47	1,90	—	—	—	—	—	—	—	—	—	—
8	0,76	1,33	0,56	1,78	0,37	2,29	—	—	—	—	—	—	—	—
9	0,82	1,32	0,63	1,70	0,46	2,13	0,30	2,59	—	—	—	—	—	—
10	0,88	1,32	0,70	1,64	0,53	2,02	0,38	2,41	0,24	2,82	—	—	—	—
11	0,93	1,32	0,66	1,60	0,60	1,93	0,44	2,28	0,32	2,65	0,20	3,01	—	—
12	0,97	1,33	0,81	1,58	0,66	1,86	0,51	2,18	0,38	2,51	0,27	2,83	0,17	3,15
13	1,01	1,34	0,86	1,56	0,72	1,82	0,57	2,09	0,45	2,39	0,33	2,69	0,23	2,99
14	1,05	1,35	0,91	1,55	0,77	1,78	0,63	2,03	0,51	2,30	0,39	2,57	0,29	2,85
15	1,08	1,36	0,95	1,54	0,81	1,75	0,69	1,98	0,56	2,22	0,45	2,47	0,34	2,73
16	1,11	1,37	0,98	1,54	0,86	1,73	0,73	1,94	0,62	2,16	0,50	2,39	0,40	2,62
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,66	2,10	0,55	2,32	0,45	2,54
18	1,16	1,39	1,05	1,54	0,93	1,70	0,82	1,87	0,71	2,06	0,60	2,26	0,50	2,46
19	1,18	1,40	1,07	1,54	0,97	1,69	0,86	1,85	0,75	2,02	0,65	2,21	0,55	2,40
20	1,20	1,41	1,10	1,54	1,00	1,68	0,89	1,83	0,79	1,99	0,69	2,16	0,60	2,34
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96	0,73	2,12	0,64	2,29
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94	0,77	2,09	0,68	2,25
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92	0,80	2,06	0,72	2,21
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90	0,84	2,04	0,75	2,17
25	1,29	1,45	1,21	1,55	1,12	1,65	1,04	1,77	0,95	1,89	0,87	2,01	0,78	2,14
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,87	0,90	1,99	0,82	2,12
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,75	1,00	1,86	0,93	1,97	0,85	2,09
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85	0,95	1,96	0,87	2,07
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84	0,98	1,94	0,90	2,05
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83	1,00	1,93	0,93	2,03
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83	1,02	1,92	0,95	2,02
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82	1,04	1,91	0,97	2,00
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81	1,06	1,90	0,99	1,99
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,14	1,81	1,08	1,89	1,02	1,98
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80	1,10	1,88	1,03	1,97
36	1,41	1,53	1,35	1,59	1,30	1,65	1,24	1,72	1,18	1,80	1,11	1,88	1,05	1,96
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80	1,13	1,87	1,07	1,95
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,20	1,79	1,15	1,86	1,09	1,94
39	1,44	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79	1,16	1,86	1,10	1,93
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79	1,18	1,85	1,12	1,92
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78	1,24	1,84	1,19	1,90
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77	1,29	1,82	1,25	1,88
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,37	1,77	1,33	1,81	1,29	1,86
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77	1,37	1,81	1,34	1,85
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77	1,40	1,81	1,37	1,84
70	1,58	1,64	1,55	1,67	1,53	1,70	1,49	1,74	1,46	1,77	1,43	1,80	1,40	1,84
75	1,60	1,65	1,57	1,68	1,54	1,71	1,52	1,74	1,49	1,77	1,46	1,80	1,43	1,83
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77	1,48	1,80	1,45	1,83
85	1,62	1,67	1,60	1,70	1,58	1,72	1,55	1,75	1,53	1,77	1,50	1,80	1,47	1,83
90	1,64	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78	1,52	1,80	1,49	1,83
95	1,65	1,69	1,62	1,71	1,60	1,73	1,58	1,76	1,56	1,78	1,54	1,80	1,51	1,83
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78	1,55	1,80	1,53	1,83
150	1,72	1,75	1,71	1,76	1,69	1,77	1,68	1,79	1,67	1,80	1,65	1,82	1,64	1,83
200	1,76	1,78	1,75	1,79	1,74	1,80	1,73	1,81	1,72	1,82	1,71	1,83	1,70	1,84

Zeichenerklärung:

n = Stichprobengröße

K = Anzahl erklärender Variablen im Modell (ohne Konstante)

Tabellenquelle:

Savin, N.E., White, K.J. (1977) i.V.m. Farebrother, R.W. (1980)

2.2 Kritische Dickey-Fuller t-Werte

n	t_{nc}^*		t_c^*		t_{ct}^*	
	1 %	5 %	1 %	5 %	1 %	5 %
25	-2,66	-1,95	-3,75	-3,00	-4,38	-3,60
50	-2,62	-1,95	-3,58	-2,93	-4,15	-3,50
100	-2,60	-1,95	-3,51	-2,89	-4,04	-3,45
250	-2,58	-1,95	-3,46	-2,88	-3,99	-3,43
500	-2,58	-1,95	-3,44	-2,87	-3,98	-3,42
∞	-2,58	-1,95	-3,43	-2,86	-3,96	-3,41

Zeichenerklärung:

n = Stichprobenumfang

t = kritische Werte zum angegebenen Signifikanzniveau, wobei die Subskripte nc, c und ct bedeuten, dass die Dickey-Fuller-Testgleichung keine Konstante, eine Konstante bzw. eine Konstante und einen deterministischen Trend enthält.

Tabellenquelle:

Fuller, W.A. (1976) sowie Dickey, D.A., Fuller, W.A. (1981)

Literaturverzeichnis

- Akaike, H. (1981):** "Likelihood of a Model and Information Criteria", *Journal of Econometrics* 16 (1), 3-14.
- Albrecht, P. (1980):** "On the Correct Use of the Chi-square Goodness-of-fit Test", *Scandinavian Actuarial Journal* 7, 149-160.
- Allen, R.C., Stone, J.H. (2005):** "The Neglected Constant Coefficient: A Pedagogical Note on the Textbook Literature", *Journal of Economic Education*.
- Anscombe, F.J. (1973):** "Graphs in Statistical Analysis", *American Statistican* 27 (1), 17-21.
- Asteriou, D., Hall, S.G. (2007):** "Applied Econometrics: A Modern Approach Using EViews and Microfit", Revised Edition, Palgrave, New York.
- Auer, B.R., Schuhmacher, F. (2009):** "Portfoliooptimierung", WISU, Heft 5, Mai, 679-687.
- Auer, B.R., Seitz, F. (2008):** "Performancemessung - Theoretische Maße und empirische Umsetzung mit VBA", Weidener Diskussionspapiere Nr. 12.
- Auer, B.R., Seitz, F. (2009):** "Grundkurs Wirtschaftsmathematik", 2. Auflage, Gabler, Wiesbaden.
- Bamberg, G., Baur, F. (2004):** "Statistik Arbeitsbuch", 7. Auflage, Oldenbourg, München, Wien.
- Bamberg, G., Baur, F. (2002):** "Statistik", 12. Auflage, Oldenbourg, München, Wien.
- Bartlett, M.S. (1946):** "On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series", *Journal of the Royal Statistical Society* 8 (1), 27-41.
- Bera, A.K., Jarque, C.M. (1982):** "Model Specification Tests: A Simultaneous Approach", *Journal of Econometrics* 20 (1), 59-82.
- Berndt, E.R. (1991):** "The Practice of Econometrics: Classic and Contemporary", Addison-Wesley.
- Betancourt, R., Kelejian, H. (1981):** "Lagged Endogenous Variables and Cochrane-Orcutt Procedure", *Econometrica* 49 (4), 1073-1078.
- Bleymüller, J., Gehlert, G. (2003):** "Statistische Formeln, Tabellen und Programme", 10. Auflage, Vahlen, München.
- Bleymüller, J., Gehlert, G., Gülicher, H. (2004):** "Statistik für Wirtschaftswissenschaftler", 14. Auflage, Vahlen, München.
- Bodie, Z., Kane, A., Marcus, A. J. (2005):** "Investments", 6. Auflage, McGraw-Hill.
- Bound, J., Jaeger, D.A., Baker, R.M. (1995):** "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and Endogenous Explanatory Variable is Weak", *Journal of the American Statistical Association* 90 (430), 443-450.

- 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* 65 (332), 1509-1526.
- Breusch, T., Pagan, A. (1979):** "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica* 47 (5), 1287-1294.
- Breusch, T.S. (1978):** "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economic Papers* 17 (31), 334-355.
- Brown, R., Durbin, J., Evans, J. (1975):** "Techniques for Testing the Constancy of Regression Relationships over Time", *Journal of the Royal Statistical Society, B*37, 149-63.
- Büning, H., Trenkler, G. (1998):** "Nonparametric statistical Methods", 2. Auflage, Gruyter, Berlin.
- Campbell, J.Y., Lo, A.W., MacKinlay, A.C. (1997):** "The Econometrics of Financial Markets", Princeton University Press, Princeton.
- Campbell, J.Y., Perron, P. (1991):** "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots", *NBER Macroeconomics Annual* 6, 141-201.
- Caulkins, J.P., Padman, R. (1993):** "Quantity Discounts and Quality Premia for Illicit Drugs", *Journal of the American Statistical Association* 88 (423), 748-757.
- Chen, W.W., Deo, R.S. (2004):** "Power Transformations to Induce Normality and their Applications", *Journal of the Royal Statistical Society B* 66 (1), 117-130.
- Chow, C. (1960):** "Tests of the Equality Between Sets of Coefficients in Two linear Regressions", *Econometrica* 28, 591-605.
- Cochrane, D., Orcutt, G.H. (1949):** "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms", *Journal of the American Statistical Association* 44 (245), 32-61.
- Davidson, R., MacKinnon, J. G. (1989):** "Testing for Consistency using Artificial Regressions", *Econometric Theory* 5 (3), 363-384.
- Dickey, D.A., Fuller, W.A. (1979):** "Distribution of the Estimators for Autoregressive Time-Series with a Unit Root", *Journal of the American Statistical Association* 74 (366), 427-431.
- Dickey, D.A., Fuller, W.A. (1981):** "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica* 49 (4), 1057-1072.
- Diebold, F.X. (2004):** "Elements of Forecasting", 3. Auflage, South-Western, Cincinnati.
- Döpke, J. (2004):** "Zur Qualität von Konjunkturprognosen", WiSt, Heft 1, Januar, 8-13.
- Dougherty, C. (1992):** "Introduction to Econometrics", Oxford University Press, Oxford.
- Durbin, J., Watson, G.S. (1951):** "Testing for Serial Correlation in Least-Squares Regression, II", *Biometrika* 38 (1-2), 159-178.

- Elton, E.J., Gruber, M.J., Brown, S.J., Goetzmann, W.N. (2007):** "Modern Portfolio Theory and Investment Analysis", Wiley & Sons, New York.
- Enders, W. (2004):** "Applied Econometric Time Series", 2. Auflage, Wiley & Sons, Hoboken.
- Engle, R.F., Granger, C.W.J. (1987):** "Co-Integration and Error Correction: Representation, Estimation and Testing", *Econometrica* 55 (2), 251-276.
- Evans, M., Hastings, N., Peacock, B. (2000):** "Statistical Distributions", 3. Auflage, Wiley-Interscience, New York.
- Fahrmeir, L., Künstler, R., Pigeot, I., Tutz, G. (2007):** "Statistik - Der Weg zur Datenanalyse", 6. Auflage, Springer, Berlin, Heidelberg.
- Fama, E.F., French, K.R. (1993):** "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics* 33 (1), 3-56.
- Fama, E.F., French, K.R. (2004):** "The Capital Asset Pricing Model: Theory and Evidence", *Journal of Economic Perspectives* 18 (3), 25-46.
- Farebrother, R.W. (1980):** "The Durbin-Watson Test for Serial Correlation When There Is No Intercept in the Regression", *Econometrica* 48, 1553-1563.
- Feldstein, M. (1971):** "The Error of Forecast in Econometric Models When the Forecast-Period Exogenous Variables are Stochastic", *Econometrica*, 39, 55-60.
- Frenkel, M., Funke, K., Koske, I. (2003a):** "Zeitreihenanalyse - Teil 1: Stationarität und Integration", WiSt, Heft 9, September, 545-550.
- Frenkel, M., Funke, K., Koske, I. (2003b):** "Zeitreihenanalyse - Teil 2: Kointegration und Fehlerkorrekturmodelle", WiSt, Heft 12, Dezember, 735-740.
- Fuller, W.A. (1976):** "Introduction to Statistical Time Series", Wiley & Sons, New York.
- Godfrey, L.G. (1978):** "Testing for Multiplicative Heteroscedasticity", *Journal of Econometrics* 8 (2), 227-236.
- Godfrey, L.G. (1978):** "Testing Against General Autoregressive and Moving Average Error Models When the Regressors include Lagged Dependent Variables", *Econometrica* 46 (6), 1293-1302.
- Goldfeld, S.M., Quandt, R.E. (1972):** "Nonlinear Methods in Econometrics", North-Holland, Amsterdam, London.
- Granger, C.W.J., Newbold, P. (1974):** "Spurious Regressions in Econometrics", *Journal of Econometrics* 2 (2), 111-120.
- Greene, W.H. (2008):** "Econometric Analysis", 6. Auflage, Pearson, Upper Saddle River.
- Gujarati, D.N. (2003):** "Basic Econometrics", 14. Auflage, McGraw-Hill.
- Hamilton, J.D. (1999):** "Time Series Analysis", Princeton University Press, Princeton.
- Hanson, G.H., Spilimbergo, A. (1999):** "Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the U.S.-Mexico Border", *American Economic Review* 89 (5), 1337-1357.

- Harris, R. (1995):** "Using Cointegration Analysis in Econometrics Modelling", Prentice Hall & Harvester Wheatsheaf.
- Hartung, J., Elpelt, B., Klösener, K.-H. (2005):** "Statistik - Lehr- und Handbuch der angewandten Statistik", 14. Auflage, Oldenbourg, München, Wien.
- Hausman, J. (1978):** "Specification Tests in Econometrics", *Econometrica* 46 (6), 1251-1271.
- Hendry, D.F. (1980):** "Econometrics - Alchemy or Science?", *Economica* 47 (188), 387-406.
- Hill, R.C., Griffiths, W.E., Judge, G.G. (2001):** "Undergraduate Econometrics", 2. Auflage, Wiley & Sons, New York.
- Hochstädter, D., Kaiser, U. (1988):** "Varianz- und Kovarianzanalyse", Harry Deutsch, Frankfurt am Main.
- Hogg, R.V., McKean, J.W., Craig, A.T. (1965):** "Introduction to Mathematical Statistics", 6. Auflage, Macmillan, New York.
- Jarque, C.M., Bera, A.K. (1987):** "A Test for Normality of Observations and Regression Residuals", *International Statistical Review* 55 (2), 163-172.
- Jensen, M. (1967):** "The Performance of Mutual Funds in the Period 1945-1964", *Journal of Finance* 23 (2), 389-416.
- Jevons, W.S. (1975):** "Influence of the Sun-Spot Period on the Price of Corn", *Nature*.
- Jevons, W.S. (1978):** "Commercial Crises and Sun-Spots", *Nature* 19 (472), 33-37.
- Johansen, S. (1991):** "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica* 59 (6), 1551-1580.
- Johansen, S. (1995):** "Likelihood-based Inference in Cointegrated Vector Autoregressive Models", Oxford University Press, Oxford.
- Johnson, N.L., Kotz, S., Balakrishnan, N. (1993):** "Continuous Univariate Distributions - Volume 1", Wiley-Interscience, New York.
- Johnson, N.L., Kotz, S., Balakrishnan, N. (1994):** "Continuous Univariate Distributions -Volume 2", Wiley-Interscience, New York.
- Johnson, N.L., Kotz, S., Kemp, A.W. (1993):** "Univariate Discrete Distributions", Wiley-Interscience, New York.
- Johnston, J., Dinardo, J. (1997):** "Econometric Methods", 4. Auflage, McGraw-Hill, New York.
- Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H., Lee, T.-C. (1985):** "Introduction to the Theory and Practice of Econometrics", 2. Auflage, Wiley & Sons, New York.
- Kanji, G.K. (2006):** "100 Statistical Tests", 3. Auflage, SAGE Publications, London.
- Karlis, D., Xekalaki, E. (2000):** "A Simulation Comparison of Several Procedures for testing the Poisson Assumption", *The Statistician* 49 (3), 355-382.
- Kennedy, P. (2003):** "A Guide to Econometrics", 5. Auflage, MIT Press, Cambridge.

- Kleinbaum, D.G., Kupper, L.L, Müller, K.E. (1998):** "Applied Regression Analysis and other Multivariate Methods", 2. Auflage, PWS-Kent, Boston.
- Klien, L.R. (1962):** "An Introduction to Econometrics", Prentice-Hall, New York.
- Koenker, R. (1981):** "A Note on Studentizing a Test for Heteroscedasticity," Journal of Econometrics 17 (1), 107-112.
- Koyck, L.M. (1954):** "Distributed Lags and Investment Analysis", North-Holland, Amsterdam.
- Lintner, J. (1965):** "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", Review of Economics and Statistics 47 (1), 13–37.
- Ljung, G.M., Box, G.E.P. (1978):** "On a Measure of Lack of Fit in Time Series Models", Biometrika 66, 66-72.
- MacKinnon, J., White, H., Davidson, R. (1983):** "Tests for Model Specification in the Presence of Alternative Hypotheses - Some Further Results", Journal of Econometrics 21 (1), 53-70.
- MacKinnon, J.G. (1991):** "Critical Values of Cointegration Tests", in Egle, R., Granger, C.W.J. (1991): "Long-Run Economic Relationships: Readings in Cointegration", Oxford University Press, Oxford.
- Malinvaud, E. (1970):** "Statistical Methods of Econometrics", North Holland, Amsterdam.
- Markowitz, H. (1952):** "Portfolio Selection", Journal of Finance 7 (1), 77-91.
- Markowitz, H. (1959):** "Portfolio Selection: Efficient Diversification of Investments", Wiley & Sons, New York.
- Marsaglia, G., Tsang, W., Wang, J. (2003):** "Evaluating Kolmogorov's Distribution", Journal of Statistical Software 8 (18).
- Massey, F. J. (1951):** "The Kolmogorov-Smirnov Test for Goodness of Fit", Journal of the American Statistical Association 46 (253), 68-78.
- Mickey, R.M., Dunn, O.J., Clark, V.A. (2004):** "Applied Statistics: Analysis of Variance and Regression", 3. Auflage, Wiley-Interscience, New York.
- Miller, L. H. (1956):** "Table of Percentage Points of Kolmogorov Statistics", Journal of the American Statistical Association 51 (273), 111-121.
- Mossin, J. (1966):** "Equilibrium in a Capital Asset Market", Econometrica 34 (4), 768–783.
- Nelson, C.R., Plosser, C.I. (1982):** "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implication", Journal of Monetary Economics 10, 169-182.
- Newey, W.K., West, K.D. (1987):** "A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", Econometrica 55 (3), 703-708.
- Noether, G.E. (1971):** "Introduction to Statistics - A Fresh Approach", Houghton Mifflin, Boston.

- Pfanzagl, J. (1991):** "Elementare Wahrscheinlichkeitsrechnung", 2. Auflage, de Gruyter, Berlin, New York.
- Phillips, P.C.B., Ouliaris, S. (1990):** "Asymptotic Properties of Residual Based Tests for Cointegration", *Econometrica* 58 (1), 165-193.
- Pindyck, R.S., Rubinfeld, D.L. (1998):** "Econometric Models and Economic Forecasts", 4. Auflage, McGraw-Hill, Singapur.
- Ramsey, J.B. (1969):** "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis", *Journal of the Royal Statistical Society B* 31 (2), 350-371.
- Rao, P., Miller, R.L. (1971):** "Applied Econometrics", Wadsworth, Belmont.
- Rau-Bredow, H. (2002):** "Kreditrisikomodelle und Diversifikation", *Zeitschrift für Bankrecht und Bankwirtschaft*, 14. Jahrgang, 9-17.
- Rolfes, B. (2008):** "Gesamtbanksteuerung - Risiken ertragsorientiert steuern", 2. Auflage, Schäffer-Poeschel, Stuttgart.
- Roll, R. (1997):** "A Critique of the Asset Pricing Theory's Tests' Part I: On Past and Potential Testability of the Theory", *Journal of Financial Economics* 4 (2), 129-176.
- Rottmann, H., Franz, T. (2007):** "Die Performance deutscher Aktienfonds - Lassen sich Selektions- und Timingfähigkeiten nachweisen?", WiSt, Heft 1, ..., 16-24.
- Rottmann, H., Franz, T. (2008):** "Die Performance deutscher Aktienfonds", Weidener Diskussionspapiere Nr. 5.
- Sargan, J.D. (1964):** "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", in Hart, P.E., Mills, G., Whitaker, J.K. (Hrsg.): "Econometric Analysis für National Economic Planing", Butterworths, London.
- Sargan, J.D. (1984):** "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", in Wallis, K.F., Hendry, D.F.: "Quantitative Economics and Econometric Analysis", Oxford.
- Savin, N.E., White, K.J. (1977):** "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors", *Econometrica* 45 (8), 1989-1996.
- Scheffé, H. (1970):** "Practical Solutions of the Behrens-Fisher Problem", *Journal of the American Statistical Association* 65 (332), 1501-1509.
- Schlittgen, R. (2003):** "Einführung in die Statistik", 10. Auflage, Oldenbourg, München, Wien.
- Schmid, F., Trede, M. (2005):** "Finanzmarktstatistik", Springer, Berlin, Heidelberg.
- Schwarz, G. (1978):** "Estimating the Dimension of a Model", *The Annals of Statistics* 6 (2), 461-464.
- Sharpe, W. (1964):** "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", *Journal of Finance* 19 (3), 425-442.
- Stephens, M. A. (1970):** "Use of the Kolmogorov-Smirnov, Cramer-Von Mises and Related Statistics Without Extensive Tables", *Journal of the Royal Statistical Society B* 32 (1), 115-122.

- Stock, J.H., Watson, M.W. (1993):** "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica* 61 (4), 783-820.
- Stock, J.H., Watson, M.W. (2003):** "Introduction to Econometrics", Addison Wesley, Boston.
- Studenmund, A. H. (2007):** "Using Econometrics - A Practical Guide", 5. Auflage, Addison Wesley, Boston.
- Theil, H. (1966):** "Applied Economic Forecasting", North-Holland, Amsterdam.
- Thursby J.G. (1982):** "A Test Strategy for Discriminating between Autocorrelation and Misspecification in Regression Analysis: Reply", *Review of Economics and Statistics* 67 (1), 177-178.
- Treynor, J.L., Mazuy, K.K. (1966):** "Can Mutual Funds Outguess the Market?", *Harvard Business Review* 44, 131-136.
- Tsay, R.S. (2005):** "Analysis of Financial Time Series", 2. Auflage, Wiley & Sons, Hoboken.
- von Auer, L. (2007):** "Ökonometrie - Eine Einführung", 4. Auflage, Springer, Heidelberg.
- Welch, B.L. (1936):** "The Specification of Rules for Rejecting Too Variable A Product, with Particular Reference to An Electric Lamp Problem", Supplement to the *Journal of the Royal Statistical Society* 3 (1), 29-48.
- Welch, B.L. (1938):** "The Significance of the Difference Between Two Means When the Population Variances Are Unequal", *Biometrika* 29 (3/4), 350-362.
- White, H. (1980):** "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica* 48 (4), 817-838.
- Wolters, J., Hassler, U. (2006):** "Unit Root Testing", *Allgemeines Statistisches Archiv* 90, 43-58.
- Wooldridge, J.M. (2002):** "Econometric Analysis of Cross Section and Panel Data", MIT Press.
- Wooldridge, J.M. (2007):** "Introductory Econometrics - A Modern Approach", 4. Auflage, South-Western.
- Yule, G.U. (1926):** "Why Do We Sometimes Get Nonsense Correlations Between Time Series?", *Journal of the Royal Statistical Society* 89 (1), 1-63.
- Zöfel, P. (1992):** "Univariate Varianzanalysen", Gustav Fischer, Stuttgart.

Stichwortverzeichnis

Ablehnungsbereich.....	356	Breusch-Pagan LM-Test.....	523
Abweichung		BUE	457
–, mittlere absolute	47	Capital Asset Pricing Model.....	597
–, mittlere quadratische	49	Chi-Quadrat-Anpassungstest.....	386
Additionssatz.....	157	Chi-Quadrat-Homogenitätstest	397
ADF-Test.....	588	Chi-Quadrat-Koeffizient	102
Aggregatform	114, 116	Chi-Quadrat-Unabhängigkeitstest	392
Akaike-Informationskriterium	482	Chi-Quadrat-Verteilung	278, 321
Aktienindex, deutscher	132	Chow-Test	471
Annualisierung.....	58	Cochrane-Orcutt-Methode	550
ANOVA.....	469	Deflation.....	118
Approximation.....	257, 259, 261, 268, 284, 285, 286, 287, 290	Dezil.....	28
AR(1)-Prozess.....	534	Dichte, bedingte	239
AR(s)-Prozess	538	Dichtefunktion	201
arithmetisches Mittel.....	35	–, gemeinsame.....	239
–, bedingtes	88	Dickey-Fuller-Test.....	585
–, gewogenes.....	36	Differenz.....	153
Ausreißer	32, 449	–, erste	514
Autokorrelation	448, 534, 571, 615, 617	Differenzentest	376
–, negative	536	Differenzstationarität	576
–, positive.....	536, 545	Disinflation	118
Autokorrelationsfunktion	583	Dummy	
Autokorrelationskoeffizient	535, 543	–, Achsenabschnitts-.....	495
Autokovarianz	535	–, saisonaler	497
Basiswert	109	–, Steigungs-	499
Basiszeit.....	109	Durbin-Watson d-Test	543, 579
Baumdiagramm.....	178	Durchschnitt	152
Behrens-Fisher-Problem.....	370	Durchschnittsrendite	
Bereichshypothese	350	–, arithmetische	41
Berichtszeit.....	109	–, geometrische	41
Bernoulli-Experiment	245	Effizienz	328
Bestandsgleichung.....	5	Einstichprobentest	353
Bestandsgröße	5	für das arithmetische Mittel	361
Bestimmtheitsmaß	429, 468, 529	für die Varianz.....	366
–, angepasstes.....	433	für den Anteilswert	354
Bindung.....	99	Elementarereignis	147
Binomialkoeffizient	168	Endlichkeitskorrekturfaktor	255, 316, 320
Binomialverteilung	247, 257, 259, 285	Ereignis	149
BLUE	456	–, äquivalentes.....	151
Box-Whisker-Plot.....	61	–, disjunktes	152
Breusch-Godfrey LM-Test.....	546, 571		

–, komplementäres.....	153	Gini-Koeffizient.....	73
–, sicheres	151	–, normierter	73
–, unmögliches	151	Gleichverteilung	
Ereignisraum	147	–, diskrete	262
Ergebnismenge	147	–, stetige	261
Ergebnisraum	147	GLS-Schätzung	549, 572
Erwartungstreue	326	Grundgesamtheit.....	4
Erwartungswert.....	205, 232, 237, 239	–, hypothetische	5
–, bedingter.....	233, 437, 443	–, reale	5
EViews	427, 461	Gütefunktion	381
Exponentialindex.....	69		
Exponentialverteilung	264, 334		
Fakultät	167	Häufigkeit	
Fehler 1. Art	350	–, absolute	14
Fehler 2. Art	351	–, bedingte	86
Fehler		–, gemeinsame absolute	82
–, mittlerer.....	611	–, gemeinsame relative	82
–, mittlerer absoluter	612	–, kumulierte	18
–, mittlerer quadratischer.....	330, 612	–, kumulierte gemeinsame	83
Fehlerkorrekturmodell	594	–, relative	14
Fehlervarianz.....	612	Häufigkeitsdichte	23
FGLS-Schätzung	550	Häufigkeitspolygon.....	16
Fisherscher Idealindex	117	Häufigkeitstabelle	14
fit	428	–, klassierte	22
Formel von Bayes.....	185	Häufigkeitsverteilung	
Fraktil.....	28	–, eindimensionale	14
F-Test	467	–, klassierte	22
–, Zweistichproben-	373	–, zweidimensionale	82
funktionale Form		Hausman-Test.....	564
–, Doppel-Log-	489	Herfindahl-Index.....	68
–, inverse	493	Heteroskedastizität.....	447, 516
–, lineare	488	Hilfsregression.....	507
–, linksseitige Semi-Log-	491	Histogramm	23
–, Polynom-	492	Homoskedastizität.....	400, 446
–, rechtsseitige Semi-Log-	490	hypergeometrische Verteilung	253, 261, 286
F-Verteilung	281, 324		
FWLS-Schätzung.....	531		
Gauß-Markov-Theorem.....	456	Identifikation	561
Gaußtest		Index	112
–, approximativer	354, 362	Inflation	118
–, Einstichproben-	362, 381	Inflationsbereinigung	120
–, Zweistichproben-	370, 372	Instrument	558
Gegenhypothese.....	350	Instrumentenvariablenenschätzer	558
geometrische Verteilung	266	Instrumentenvariablenenschätzung	571
geometrisches Mittel.....	38	Integration	578
Gesetz der großen Zahlen	163, 314, 318	Interaktionsvariable	499
		Intervallschätzung	335
		Intervallskala	7
		Irrtumswahrscheinlichkeit	335

Jarque-Bera-Test	473	Lijung-Box-Statistik	585
Jensen-Alpha	601	Linearkombination	105, 237
Jensensche Ungleichung	210	Lognormalverteilung	276
Kardinalskala	8	Lorenz-Kurve	70
Kaufkraftvergleich	121	Mächtigkeit	148
Kausalität	106, 417	Maßzahl	31
Klasse	21	Maximum-Likelihood-Methode	332
Klassenbreite	21	Median	34, 218
Klassenhäufigkeit	21	Mengenindex nach Laspeyres	126
Klassenmitte	22	Mengenindex nach Paasche	126
Klassenobergrenze	21	Merksmal	4
Klassenuntergrenze	21	–, diskretes	10
Klassierung	10	–, qualitatives	11
Kointegration	591	–, quantitatives	11
Kointegrationstest	591	–, quasistetiges	10
Kollinearität	450	–, stetiges	10
Kombinatorik	167	Merkmalsausprägung	4
Komplement	153	Merkmalsträger	4
Konfidenzintervall	335	Messfehler	555
für das arithm. Mittel	338, 340, 341, 346	Messzahl	109
für den Anteilswert	343, 347, 358	Methode der kleinsten Quadrate	331, 418
für die Varianz	344, 347	–, gewichtete	528
für einen Regressionsparameter	465	–, verallgemeinerte	549
Konfidenzniveau	335	–, zweistufige	560
Konsistenz	328	Modalklasse	33
Kontingenzkoeffizient	103	Modell	
Kontingenztabelle	82	–, autoregressives	571
–, hypothetische	102	–, autoregressives distributives Lag-	571
Kontingenztest	392	–, distributives Lag-	485
Konzentration	65	–, dynamisches	484
–, absolute	66	–, einfaches Lag-	484
–, relative	70	–, Mehrgleichungs-	557
Konzentrationsfläche	73	–, simultanes	557
Konzentrationsrate	66	–, statisches	483
Korrelation	92, 417	Modus	32
Korrelationskoeffizient	94, 235, 379, 431	Moment	63, 216
Korrelationstest	379	Multikollinearität	450
Korrelogramm	583	–, unvollkommene	504
Kovarianz	92, 234, 239	–, vollkommene	503
Koyck-Transformation	570	Multiplikationssatz	177, 180, 236
Kreissektoriendiagramm	16	Neubasierung	124
kritischer Anteilswert	358	Newey-West Standardfehler	552
kritischer Wert	356	Nichtablehnungsbereich	356
Lageparameter	31	Nominalskala	6
Langfristbeziehung	594	Normalgleichung	420
Laplace-Experiment	160	Normalitätstest	391

Normalverteilung	268, 315, 316, 319, 320, 322, 323, 334	
Nullhypothese	350	
OLS-Schätzung	420	
–, restringierte	513	
Operationscharakteristik	250, 382	
Ordinalskala	7	
Parameterhypothese	350	
Parameterlinearität	426, 443	
Permutation	172	
Perzentil	28	
Piktogramm	16	
Poissonverteilung	257, 287, 333	
Portfoliorendite		
–, Erwartungswert der	240	
–, Varianz der	240	
Preisindex nach Drobisch	117	
Preisindex nach Laspeyres	114	
Preisindex nach Paasche	116	
Prognose		
–, bedingte	606	
–, ex ante	606	
–, ex post	606	
–, in-sample-	605	
–, Intervall-	605, 610	
–, out-of-sample-	606	
–, Punkt-	605, 610	
–, Trend-	618	
–, unbedingte	607, 620	
Prognosefehler	606, 608	
Prognosegüte	611	
Prüfplan, statistischer	250	
Punkthypothese	350	
Punktschätzung	310	
p-Wert	463	
Q-Statistik	584	
Qualitätskontrolle, statistische	365	
Quantifizierung	11	
Quantil	28, 219, 274	
Quantilsverhältnis	61	
Quartil	28	
Quartilsabstand	46	
Quartilskoeffizient	61	
Querschnittsregression	422, 431, 518	
Quotenauswahl	310	
Ramsey RESET	481	
Randdichte	239	
Randhäufigkeit	85	
Random Walk mit Drift	574, 576	
Random Walk ohne Drift	573, 576	
Randverteilung	85, 228	
Rang	98	
Rangkorrelationskoeffizient	99, 380	
Rangskala	7	
Rating-Skala	7	
Realisation	192	
Referenzkategorie	496	
Regressand	417	
Regression		
–, einfache	419	
–, multiple	426	
Regressionsfunktion		
der Grundgesamtheit	437	
der Stichprobe	440	
Regressionsgerade, geschätzte	419	
Regressionskoeffizient	419, 437	
Regressionskonstante	419, 486	
Regressionsmodell		
–, einfaches lineares	437	
–, klassisches lineares	443	
–, multiples	440	
Regressionsparameter	419, 437	
Regressor	417	
–, deterministischer	451	
–, stochastischer	451	
Rendite	221	
–, diskrete	40	
–, stetige	41	
Reproduktivitätseigenschaft	250, 259, 275, 279	
Residuum	419	
Risiko	56	
Roulette	222	
Sargan-Test	567	
Satz der totalen Wahrscheinlichkeit	183	
Säulendiagramm	16	
Schätzer	311	
Schätzfehler	326	
Schätzintervall	335	
Schätzung	311	
Scheinkorrelation	107, 572	
Scheinregression	572, 578	

Schiefe	216	Summenhäufigkeit	18
Schiefekoeffizient	63	Summenpolygon	26, 30
Schiefequartilskoeffizient	65		
Schwarz-Kriterium	482	Teilereignis	151
Semi-Volatilität	57	Teilerhebung	4
Signifikanzniveau	351	Test	349
Simultanität	557	–, einseitiger	350
Skala	6	–, idealer	382
Skalenerträge		–, linksseitiger	350
–, Test auf konstante	470	–, rechtsseitiger	350
Spannweite	45	–, robuster	401
Stabdiagramm	16	–, unverfälschter	382
Standardabweichung	49, 212	–, verteilungsfreier	388
Standardfehler der Regression	454	–, zweiseitiger	350
Standardfehler des OLS-Schätzers	453	Teststatistik	352
Standardfehler, robuster	532, 552	Theilscher Ungleichheitskoeffizient	613
Standardisierung	54, 215	Totalerhebung	4
Standardnormalverteilung	271	Transformation	
Stata	427, 461	–, lineare	9, 37, 48, 53, 104, 208, 213
Stationarität	534, 572	–, proportionale	9
Statistik		–, skalenerhaltende	8
–, deskriptive	3	–, skalensenkende	9
–, induktive	3	–, streng monotone	8
Steigungsparameter	419	Transformationseigenschaft	272
Stetigkeitskorrektur	285, 287, 319, 389	Trend	107
Stichprobe		–, deterministischer	576
–, abhängige	375	–, stochastischer	576
–, einfache	310	Trendstationarität	577
–, unabhängige	353, 375	Treynor-Mazuy-Ansatz	602
–, unverbundene	375	t-Test	362, 459, 464
–, verbundene	353, 375	t-Verteilung	280, 315
Stichprobenanteilswert	311, 318		
Stichprobenergebnis	310	Umbasierung	111, 123
Stichprobenerhebung	4	Unabhängigkeit	
Stichprobenmittel	311, 313, 327	–, statistische	89
Stichprobenumfang	310, 346	–, stochastische	180, 231, 234, 236, 237
Stichprobenvariable	310	Ungleichung von Tschebyscheff	219
Stichprobenvarianz	311, 321, 327	Unit Root	576
–, externe	400	Urliste	13, 81
–, interne	400	Urnenmodell	167, 309
Stichprobenverteilung	311		
Störterm	437	Variable	
Streudiagramm	81	–, abhängige	417
Streuungsparameter	45	–, deterministische	554
Streuungszerlegung	56, 400	–, endogene	417, 446, 557
Stromgröße	5	–, erklärende	417
Studentverteilung	280	–, erklärte	417
Stützbereich	605	–, exogene	417, 445, 557

–, redundante	511
–, stochastische.....	554
–, überflüssige	479
–, vernachlässigte	475, 539, 555
–, verzögert endogene.....	556
–, verzögerte	484
Variablenlinearität.....	424
Varianz.....	49, 211, 232, 237, 239
–, bedingte	89, 233
–, externe	56
–, interne	56
Varianzanalyse	400
Variationskoeffizient.....	59
Venn-Diagramm.....	151
Verbraucherpreisindex	128
–, harmonisierter	131
Vereinigung	152
Verhältnisskala	8
Verknüpfung	124
Verschiebungssatz	50, 94
Verteilung	
–, bedingte	87, 229
–, linksschiefe	27, 44
–, multimodale.....	27
–, rechtsschiefe	27, 44
–, symmetrische.....	27, 44
–, unimodale	27
Verteilungsfunktion	196, 200
Verteilungsfunktion	
–, empirische	18, 25
–, gemeinsame.....	228, 239
–, gemeinsame empirische	83
Verteilungshypothese	349
Verwerfungsbereich	352
Vierfeldertafel.....	180
Volatilität.....	56
Vollerhebung	4
Wachstumsfaktor	39
Wachstumsrate	39
Wahrscheinlichkeit	155
–, a-posteriori	186
–, a-priori-	186
–, bedingte	176
–, totale.....	183
Wahrscheinlichkeitsaxiome	156
Wahrscheinlichkeitsdefinition	
–, klassische	159
–, statistische	162
–, subjektive	164
Wahrscheinlichkeitsfunktion.....	195
–, gemeinsame	226
Wahrscheinlichkeitsintervall	336
Wahrscheinlichkeitssätze.....	156
Warenkorb.....	112
Wechselkurs	121
Weißes Rauschen.....	573
Welch-Test.....	370
Wertindex	127
Wettquotienten.....	165
White Noise	573
White Standardfehler	532
White-Test.....	525
WLS-Schätzung	528
Wölbung	216
Yates-Korrektur	389
Zeitreihenregression	422, 430, 518
Zentraleigenschaft	37, 207
Zentraler Grenzwertsatz	268, 283, 315, 316, 319
Zentrierung	54
Zerlegung	158
Ziehen mit Zurücklegen	169, 174, 182, 310
Ziehen ohne Zurücklegen	171, 173, 310
Zufallsauswahl	167, 309
Zufallsexperiment	147
Zufallsstichprobe	309
Zufallsvariable	192
–, dichotome	245
–, diskrete	193, 194, 206, 211
–, stetige	193, 200, 206, 211
–, zweidimensionale.....	226
Zufallsvorgang	147
Zweistichprobentest	353, 354, 368
für Anteilswerte	372
für arithmetische Mittel	369
für Varianzen	373