

Statistik

**Für Studierende in den B.Sc. Studiengängen
Agrarbiologie (AB), Agrarwissenschaften (AW) und
Nachwachsende Rohstoffe (NaWaRo)**

an der Universität Hohenheim

1. Semester

Prof. Dr. Hans-Peter Piepho

Institut für Kulturpflanzenwissenschaften (340c)
FG Biostatistik
Universität Hohenheim
e-mail: piepho@uni-hohenheim.de

Hohenheim, im September 2015

Copyright: Hans-Peter Piepho
Nur für den internen Gebrauch, Vervielfältigung nur mit Genehmigung des Autors

A 1001

Glossar griechischer Buchstaben

Buchstabe	Umschrift	Verwendung
α	Alpha	Fehler 1. Art/Achsenabschnitt
β	Beta	Fehler 2. Art/Steigung
χ	Chi	Chi-Quadrat-Verteilung
δ	Delta	Lack-of-fit Effekt
γ	Gamma	Effekt im linearen Modell
η	Eta	Linearer Prädiktor, Erwartungswert, systematischer Teil des linearen Modells
λ	Lambda	Parameter der Poisson-Verteilung
μ	Mü	Mittelwert, Erwartungswert
π	Pi	Kreiskonstante/Binomialwahrscheinlichkeit
θ	Theta	Parameter
ρ	Rho	Korrelation
σ	Sigma	Standardabweichung
τ	Tau	Behandlungseffekt

Inhaltsverzeichnis

Seite

1. Datentypen	1
1.1 Kategoriale Daten	1
1.1.1 Nominale Daten	1
1.1.2 Ordinale Daten	2
1.2 Metrische Daten	5
1.3 Zur Wahl des Skalenniveaus	7
2. Beschreibende Statistik für metrische Daten	9
2.1 Histogramm	9
2.2 Statistische Maßzahlen	12
2.2.1 Quantile (Perzentile)	12
2.2.2 Lagemaße	16
2.2.3 Streuungsmaße	23
2.3 Box-Plot	28
3. Einführung in die schließende Statistik für normalverteilte Daten (univariat)	31
3.1 Das Histogramm und Wahrscheinlichkeiten	31
3.2 Normalverteilung	32
3.2.1 Berechnung von Wahrscheinlichkeiten	32
3.2.2 Quantile der Normalverteilung und Q-Q-Plots	41
3.3 Stichprobe, Grundgesamtheit	43
3.4 Zentraler Grenzwertsatz	44
3.5 Vertrauensintervall für einen Mittelwert	44
3.6 Vertrauensintervall für einen Mittelwert bei kleinen Stichproben	46
3.7 Stichprobenumfang zur Schätzung eines Mittelwertes	50
3.8 Vertrauensintervall für die Differenz von 2 Mittelwerten (verbundene Stichproben)	51
3.9 Vertrauensintervall für die Differenz von 2 Mittelwerten (unverbundene Stichproben)	54
3.10 Test zum Vergleich zweier verbundener Stichproben	56
3.11 Test zum Vergleich zweier unverbundener Stichproben	60
3.12 Verbundene oder unverbundene Stichprobe?	63
3.13 α -Fehler und β -Fehler	65
3.14 Stichprobenumfang für den unverbundenen t-Test	68
3.14.1 Post-hoc Berechnung der Teststärke	71
3.15 Interpretation von Computeroutput - Überschreitungswahrscheinlichkeit (p-Wert)	73
3.16 Allgemeine Konvention für Angabe von Signifikanzen	76
3.17 Test des Parameters μ	77
3.18 Vertrauensintervall für eine Varianz	78
3.19 Test zum Vergleich zweier unabhängiger Stichprobenvarianzen	79
3.20 Einseitige und zweiseitige Tests	82

	Seite
3.21 Äquivalenztest am Beispiel zweier unverbundener Stichproben	84
3.22 Was ein Signifikanztest nicht sagt	88
4. Die einfache Varianzanalyse	92
4.1 Notation und Urliste	92
4.2 Lineares Modell	94
4.3 Zwei Arten, die Fehlervarianz zu schätzen	98
4.4 Die Varianzanalyse-Tabelle	100
4.5 Multiple Mittelwertvergleiche	101
4.5.1 LSD-Test	102
4.5.2 Vergleichsbezogener vs. versuchsbezogener Fehler 1. Art	106
4.5.3 Tukey-Test	107
4.6 Ausgabe der Varianzanalyse mit einem Statistik-Paket	109
5. Einführung in die schließende Statistik für kategoriale Daten	111
5.1 Kombinatorik	111
5.2 Einige wichtige Grundregeln der Wahrscheinlichkeitsrechnung	115
5.3 Binomialverteilung	121
5.3.1 Mittelwert und Varianz einer Binomialverteilung	128
5.3.2 Schätzen des Parameters p der Binomialverteilung	129
5.3.3 Test für den Parameter p der Binomialverteilung	130
5.3.4 Vertrauensintervall für den Parameter p der Binomialverteilung	136
5.3.5 Vergleich von zwei Binomialwahrscheinlichkeiten - unverbundene Stichproben	140
5.3.6 Vergleich von zwei Binomialwahrscheinlichkeiten - verbundene Stichproben	142
5.3.7* Beispiel: Grenzwerte für GVO in Saatgut	145
5.3.8* Maximum-Likelihood Schätzung	152
5.4 Poissonverteilung	155
5.4.1 Eine interessante Beziehung zur Binomialverteilung	158
5.4.2 Parameterschätzung	160
5.4.3 Tests für den Vergleich zweier Parameter λ_1 und λ_2	164
5.4.4* Genbanken - ein einfaches, aber ausführliches Beispiel zu Poisson- und Binomialverteilung	165
5.5 Chi-Quadrat-Anpassungs-Test	171
5.5.1 Berechnung der erwarteten Häufigkeiten für eine Binomialverteilung	176
5.5.2 Was tun bei Überdispersion	180

	Seite
5.6 Test auf Unabhängigkeit in der 2 x 2 Feldertafel (4-Feldertafel)	182
5.6.1 Test bei großen Stichproben	182
5.6.2 Die Yates-Korrektur	187
5.6.3 Fishers exakter Test	187
5.6.4 Unterschied zwischen Test auf Unabhängigkeit in 2 x 2 Felder-Tafel und McNemar-Test	195
5.7 Test auf Unabhängigkeit in einer $r \times c$ Tafel	195
5.7.1 Exakter Test	199
5.8* Bemerkungen zur Chi-Quadrat-Verteilung	199
6. Korrelation und Regression	202
6.1 Die Pearsonsche Produkt-Moment Korrelation	204
6.2 Regression	211
6.2.1 Streuungszerlegung	217
6.2.2 t-Tests und Vertrauensintervalle	221
6.3 Vergleich von Korrelation und Regression	226
Tabellen wichtiger Verteilungen	229
I. Standardnormalverteilung - $P(Z > z)$	229
II. t-Verteilung (zweiseitig)	230
II(b). t-Verteilung (einseitig)	231
III. Standardnormalverteilung - Quantile	232
IV. Chi-Quadrat-Verteilung	233
V. F-Verteilung	234
VI. F-Verteilung	235
VII. Studentisierte Variationsbreiten	236
VIII. Chi-Quadrat-Verteilung	237
Anhang: Die Bonferroni-Methode	238
Statistik Übungsaufgaben	241

(*Abschnitte mit ausführlichen Beispielen bzw. Exkursen, die nicht in der Vorlesung behandelt werden)

Übungen

Begleitend zur Vorlesung werden Übungen angeboten. Zweck der Übungen ist, dass die Studierenden selbständig die gestellten Aufgaben lösen. Wenn Fragen auftreten, stehen Tutoren für die Beantwortung zur Verfügung. Das Bearbeiten der Übungsaufgaben stellt eine wesentliche Vorbereitung auf die Klausur dar. **In den Übungen werden keine Aufgaben an der Tafel vorgerechnet, es geht ums selber Rechnen. Das Vorrechnen von Beispielen passiert in der Vorlesung. Die Übungen ersetzen nicht die Vorlesung.**

Empfohlene Lehrbücher

- Bosch, K. 2007. Basiswissen Statistik. Oldenbourg, München.
- Dufner, J., Jensen, U., Schumacher, E. 2002. Statistik mit SAS. Teubner, Stuttgart.
- Eckhardt, K. 2013. Stochastik. Statistik und Wahrscheinlichkeitsrechnung in der Landwirtschaft. Ulmer, Stuttgart.
- Essl, A. 1987. Statistische Methoden in der Tierproduktion. Verlagsunion Agrar, Wien.
- Kesel, A.B., Junge, M.M., Nachtigall, W. 1999. Einführung in die Statistik für Biowissenschaftler. Birkhäuser, Basel.
- Köhler, W., Schachtel, G., Voleske, P. 1992. Biostatistik. Einführung in die Biometrie für Biologen und Agrarwissenschaftler. Springer-Verlag, Berlin.
- Linder, A., Berchtold, W. 1979. Elementare statistische Methoden. Birkhäuser, Basel.
- Lorenz, R.J. 1996. Grundbegriffe der Biometrie. Gustav Fischer Verlag, Stuttgart.
- Munzert, M. 2015. Landwirtschaftliche und gartenbauliche Versuche mit SAS. Springer, Berlin.
- Rasch, D. et al.: Verfahrensbibliothek. Oldenbourg, München. Band I: 1996. Band II: 1998.
- Rees, D.G. 1985. Essential statistics. Chapman and Hall, London.
- Sachs, L. 1997. Angewandte Statistik. 8. Auflage. Springer, Berlin.
- Stahel, W.A. 2002. Statistische Datenanalyse. 4. Auflage. Vieweg, Braunschweig.

Statistik-Software

Zur Vorlesung gibt es zwei Begleitskripte, welche die anhand der Beispiele der Vorlesung die Statistik-Pakete R und SAS einführen. Diese Skripte sind in ILIAS verfügbar (<https://ilias.uni-hohenheim.de/>).

Bemerkung zur Klausur

Ich weise ausdrücklich darauf hin, dass das primäre Lernen mit Altklausuren nicht zu empfehlen ist. Die Fragen ändern sich von Jahr zu Jahr. Konzentrieren Sie sich auf den Stoff der Vorlesung und die Übungen. In der Klausur geht es vor allem darum, das Verständnis der verschiedenen Methoden und die Fähigkeit zur richtigen Methodenwahl abzufragen, nicht um den alleinigen Nachweis, dass reine Rechentechnik beherrscht wird. Selbstverständlich müssen Sie in der Klausur zwar auch rechnen, aber nicht nur.

Kontaktzeit und Workload

Die Kontaktzeit für ein Modul an der Universität Hohenheim beträgt 56 Stunden. Die sogenannte Workload eines Moduls beträgt 150 bis 180 Stunden. Dies bedeutet, dass Sie für jede Stunde Vorlesung jeweils etwa zwei Stunden selbständig nacharbeiten müssen. Dies können sie anhand des Skriptes, ihrer Aufzeichnungen und mit den Übungsaufgaben tun. Darüber hinaus ist ihnen begleitend dringend die selbstständige Lektüre von Lehrbüchern empfohlen (siehe Liste mit empfohlenen Büchern).

1. Datentypen

Wenn ein Versuch oder eine Erhebung geplant wird, muss man sich Gedanken darüber machen, welche Merkmale untersucht werden sollen. Von besonderem Interesse ist hierbei der **Datentyp** oder das **Skalenniveau**. So kann man beispielsweise **qualitative** und **quantitative** Merkmale unterscheiden.

Qualitative (kategoriale) Merkmale:

- Blütenfarbe der japanischen Wunderblume (Ausprägungen: weiß, rosa, rot)
- Geschlecht (Ausprägungen: männlich, weiblich)
- Befall mit einem Pilz (Ausprägungen: ja, nein)
- Boniturnoten des Bundessortenamtes für Pilzbefall (Ausprägungen: 1-9)

Quantitative (metrische) Merkmale:

- Ertrag (dt/ha)
- Milchleistung (Liter/Laktation)
- Temperatur (°C)
- Pflanzenlänge
- Zahl der Blattläuse pro Pflanze

Für die statistische Auswertung ist es wichtig, welcher Datentyp vorliegt, da dies die Wahl eines geeigneten Auswertungsverfahrens bestimmt.

1.1 Kategoriale Daten

Kategoriale (qualitative) Merkmale zeichnen sich dadurch aus, dass es eine begrenzte Zahl von möglichen **Ausprägungen** gibt, sog. **Kategorien**. Falls es nur zwei Ausprägungen gibt, spricht man von **binären** oder **dichotomen** Merkmalen. Bei mehr als zwei Kategorien liegen **polytome** Daten vor. Wenn die Kategorien in einer Rangordnung stehen, spricht man von einer **Ordinalskala** oder von **geordneten Kategorien**. Liegt dagegen keine Rangordnung der Kategorien vor, so hat man eine **Nominalskala**. Bei kategorialen Daten ist es im Gegensatz zu metrischen (quantitativen) Daten in der Regel nicht sinnvoll oder möglich, "Abstände" zwischen den Kategorien zu quantifizieren. Arithmetische Operationen wie Mittelwertbildung machen bei kategorialen Daten daher meist keinen Sinn.

1.1.1 Nominale Daten

Die Nominalskala ist eine diskrete Skala, bei der die verschiedenen Ausprägungen (Klassen; Kategorien) qualitativ gleichwertig sind. Zur Auswertung werden die Häufigkeiten in den verschiedenen Klassen gezählt.

Beispiel: In einem Kreuzungsexperiment mit der japanischen Wunderblume (*Mirabilis jalapa*) wurde in der F2-Generation die Blütenfarbe der Einzelpflanzen bestimmt. Es ergab sich folgende Häufigkeitstabelle:

Ausprägung (Klasse)	Anzahl Pflanzen
Weiss	123
Rosa	279
Rot	173

[Berücksichtigt man in diesem Fall den genetischen Hintergrund, der bedingt, dass die Farben Rosa und Rot vom selben Farbstoff hervorgerufen werden, aber in unterschiedlicher Mengen, so kann man dieses Merkmal auch als ordinalskaliert auffassen, wobei die Rangordnung der Kategorien Rot > Rosa > Weiss ist.] ☐

Beispiel: In einer Parzelle mit 210 Rapspflanzen wurde der Befall mit dem Pilz *Phoma lingam* erfasst.

Befall	Anzahl Pflanzen	
ja	152	
nein	58	<input type="checkbox"/>

Beispiel: In einer Erhebung wurden landwirtschaftliche Betriebe klassifiziert nach ihrem dominierenden Produktionszweig.

Produktionszweig	Anzahl Betriebe	
Milchvieh	41	
Marktfrucht	37	
Schweinemast	79	
Sonstige	20	<input type="checkbox"/>

Beispiel: In einer Schafherde wurde eine Stichprobe von 50 Schafen auf Missbildungen der Klauen untersucht.

Missgebildete Klauen	Anzahl Tiere	
Ja	22	
Nein	28	<input type="checkbox"/>

1.1.2 Ordinale Daten

Die Ordinalskala ist ebenfalls eine kategoriale Skala. Im Unterschied zur Nominalskala besteht eine Rangordnung zwischen den verschiedenen Merkmalsausprägungen/Kategorien.

Beispiel: Das Bundessortenamt führt Bonituren für eine Vielzahl von Merkmalen auf einer Notenskala von 1 bis 9 durch.

Note	Bedeutung
1 =	fehlend oder sehr gering
2 =	sehr gering bis gering
3 =	gering
4 =	gering bis mittel
5 =	mittel
6 =	mittel bis stark
7 =	stark
8 =	stark bis sehr stark
9 =	sehr stark

(Quelle: Bundessortenamt 1988: Richtlinien für die Durchführung von landwirtschaftlichen Wertprüfungen und Sortenversuchen)

Nach diesem Grundschema sind zu bonitieren:

1. Krankheiten und Schädlingsbefall
2. Mängel
 - im Stand zu bestimmten Entwicklungsstadien
 - in der Formschönheit
 - im Decken von Stauden
 - im Schließen der Reihen
 - in der Viruserkennbarkeit
3. Auftreten von
 - Lager
 - Halm- und Ährenknicken
 - Zwiewuchs
 - Auswuchs
 - Platzen
 - Ausfall
 - Reifeverzögerung des Strohs
 - Kälte und Frostschäden
 - Bestockung
 - Blattabfall
 - Halsbildung
 - Hohlherzigkeit
 - Keimfreudigkeit im Lager
 - Losschaligkeit
 - Wachstumsrisse

(Quelle: Bundessortenamt 1988: Richtlinien für die Durchführung von landwirtschaftlichen Wertprüfungen und Sortenversuchen) □

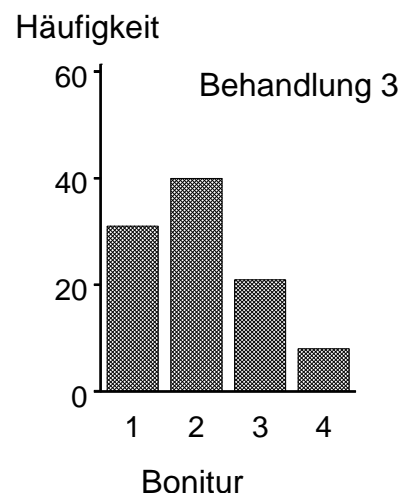
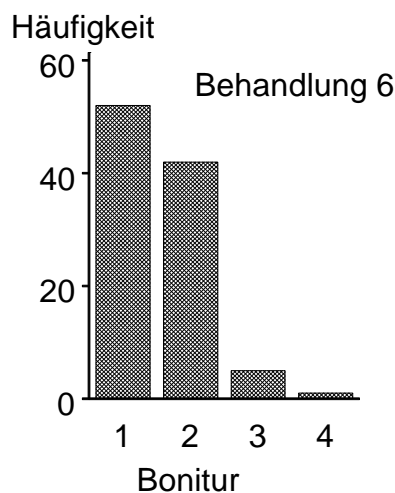
Beispiel: Die Leichtkalbigkeit bei Kühen kann auf einer Ordinalskala erfaßt werden, z.B.

Bonitur	Anzahl Tiere
1 = normale Geburt	167
2 = leichte Probleme	34
3 = starke Probleme	2

Beispiel: Zur Erfassung des Befalls von Erika mit dem Pilz *Phytophthora cinnamomi* wurden für 6 verschiedene Fungizidbehandlungen je 100 Pflanzen bonitiert nach folgendem Schlüssel:

- 1 = gesunde Pflanze
- 2 = schwach erkrankte Pflanze
- 3 = schwer erkrankte Pflanze
- 4 = dem Absterben nahe oder tote Pflanze

Nachfolgend ist die Häufigkeitsverteilung der Bonituren für zwei Behandlungen abgebildet.



Beispiel: Bei Befragungen wird die Ablehnung oder Zustimmung zu einer Aussage häufig auf einer sog. Likert-Skala erfasst. So kann beispielsweise die Zustimmung zu der Aussage "BSE ist ein sehr wichtiges Problem" auf der folgenden Skala beantwortet werden:

- 1 = starke Zustimmung (++)
- 2 = geringe Zustimmung (+)
- 3 = unentschieden (0)
- 4 = geringe Ablehnung (-)
- 5 = starke Ablehnung (--)

Beispiel: Beurteilung des Geschmacks von Quarkproben durch Testpersonen.
Notenskala: 1 = sehr gut bis 9 = sehr schlecht.

Viele Anwender statistischer Methoden behandeln Boniturnoten wie quantitative Daten und berechnen z.B. Mittelwerte. Dies ist problematisch und kann zu Fehlschlüssen führen.

Beispiel: Für den Befall mit einer Krankheit wird der folgende Boniturschlüssel verwendet:

Geschätzter Anteil befallener Oberfläche	Boniturwert
Kein Befall	1
Bis 10% Befall	2
11 bis 25% Befall	3
26 bis 50% Befall	4
Über 50% Befall	5
Stichprobe von 4 Pflanzen:	

Pflanze Nr.	Anteil befallener Oberfläche	Boniturwert
1	0%	1
2	25%	3
3	0%	1
4	20%	3
Mittelwert	11,25%	2

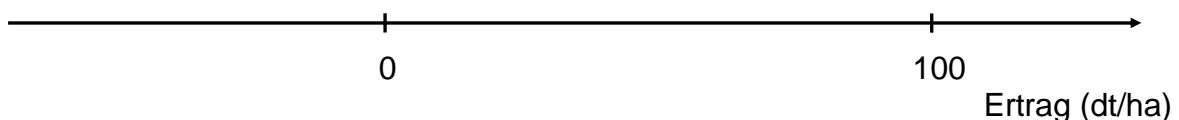
Der Mittelwert der Prozentwerte beträgt 11,25. Dies entspricht nach obiger Tabelle einer Note von 3. Der Mittelwert der Einzelnoten ergibt dagegen den Wert 2. □

1.2 Metrische Daten

Bei metrischen (quantitativen) Merkmalen lassen sich **Abstände** zweier Messwerte sinnvoll interpretieren. Die Messwerte können auf einem **Zahlenstrahl** abgebildet werden.

Beispiel:

Zahlenstrahl:



Zu unterscheiden ist zwischen stetigen und diskreten metrischen Merkmalen. Bei **stetigen** Daten gibt es prinzipiell unendlich viele mögliche Ausprägungen, da eine **kontinuierliche** Messskala vorliegt. Der Zahl der möglichen Merkmalsausprägungen ist nur durch die Genauigkeit der Messinstrumente Grenzen gesetzt.

Beispiel: Der Ertrag (dt/ha) ist ein stetiges Merkmal. □

Bei **diskreten** Daten gibt es nur bestimmte Ausprägungen; es liegt eine **diskontinuierliche** Messskala vor.

Beispiel: Die Zahl von Blattläusen pro Pflanze ist ein diskretes Merkmal. □

Wichtig ist bei metrischen Daten, dass Differenzen zwischen zwei Messwerten sinnvoll interpretiert werden können und dass Mittelwertbildung sinnvoll ist. Dies trifft für kategoriale (ordinale und nominalskalierte) Daten nicht zu.

Beispiel: Der Ertrag ist ein metrisches Merkmal. Die zwei Erträge $y_1 = 50$ dt/ha und $y_2 = 60$ dt/ha weisen eine Differenz von 10 dt/ha auf. Der Ertrag y_2 ist also um 10 dt/ha höher als der Ertrag y_1 . □

Beispiel: Die Entwicklungsstadien von Insekten werden in Kategorien eingeteilt, z.B.

Entwicklungsstadium	Kodierung
Ei	E
Larvenstadien	L1-L5
Puppe	P
Imago	J

(Hoffmann et al. 1985 Lehrbuch der Phytomedizin. Parey, Berlin, S. 138). Die Stadien spiegeln eine zeitliche Abfolge wider, so dass eine Ordinalskala vorliegt. Man könnte hier formal den Stadien E, L1, L2, L3, L4, L5, P und J die Zahlenwerte 1 bis 8 zuordnen. Aber es macht keinen Sinn, die Differenzen dieser Zahlen zu interpretieren, etwa durch eine Aussage wie: "Die Entwicklung von L1 zu L2 ist halb so weit wie die von L1 nach L3". □

Zählwerte sind metrische Daten, obschon sie nur diskrete Werte annehmen können. Auch für Zählwerte gilt, dass Differenzen sinnvoll interpretiert werden können.

Beispiele für Zählwerte:

- Zahl Blattläuse pro Pflanze
- Zahl der Unkräuter pro m²
- Zahl der Keime pro Liter Milch □

Zu den metrischen Daten gehören außerdem die Prozentzahlen.

Beispiel: In einer Parzelle mit 210 Rapspflanzen wurde der Befall mit dem Pilz *Phoma lingam* erfasst.

Befall	Anzahl Pflanzen
ja	152
nein	58

Damit sind $152/210 = 72,4\%$ der Pflanzen befallen. Diese Prozentzahl ist aus Zählwerten berechnet. Sie können bei einem begrenzten Stichprobenumfang (hier $n = 210$) nur diskrete Werte annehmen. Man beachte, dass diese Prozentzahl aus der kategorialen (dichotomen) Variable Befall ("ja"/"nein") abgeleitet wurde. □

Beispiel: Eine Probe von 1000 frisch geernteten Weizenkörnern wiegt 55 g. Nach einer Trocknung im Trockenschrank wiegen die Körner noch 46 g. Somit betrug der

Trockensubstanzgehalt der ungetrockneten Körner $46/55 = 83,6\%$. Diese Prozentzahl ist aus metrischen Daten berechnet. Es handelt sich um ein stetiges Merkmal. □

Beispiel: Mit dem Göttinger Schätzrahmen wird der Deckungsgrad des Bodens mit Unkräutern geschätzt. Hierbei handelt es sich um eine Bonitur, d.h. um geschätzte Prozente. Potentiell ist der Deckungsgrad ein stetiges Merkmal. In der Praxis nehmen die Schätzungen aber nur bestimmte diskrete Werte an (z.B. 1%, 2%, 5%, 7,5%, 10% etc). □

Bei metrischen Merkmalen kann unterschieden werden zwischen **Intervallskala** und **Verhältnisskala** (Köhler et al., 1992). Die Verhältnisskala hat im Gegensatz zur Intervallskala einen absoluten Nullpunkt. Bis auf wenige Ausnahmen ist diese Unterscheidung aber wenig relevant, da im wesentlichen dieselben statistischen Methoden anwendbar sind. Zu erwähnen ist hier, dass die Berechnung von Indizes (Preisindex, Selektionsindex, etc.) nur bei einer Verhältnisskala sinnvoll ist. Die meisten metrischen Merkmale haben einen absoluten Nullpunkt und sind daher verhältnisskaliert, so z.B.

- Ertrag (dt/ha)
- Milchleistung (Liter/Laktation)
- Pflanzenlänge (cm)

Bei diesen Merkmalen ist es sinnvoll, nicht nur die Abstände von Messwerten zu betrachten sondern auch deren Verhältnisse (Quotienten).

Beispiel: Pflanze 1 ist 30 cm lang, Pflanze 2 ist 60 cm lang. Pflanze 2 ist also doppelt so lang wie Pflanze 1, das Verhältnis der Längen ist $60/30 = 2$. Die Pflanzenlänge ist verhältnisskaliert. □

Beispiel: Die Temperatur in (°C; Grad Celsius) ist allerdings nur intervallskaliert, da kein absoluter Nullpunkt vorliegt. Wenn es an Tag 1 um 12 Uhr 14°C warm ist, während es an Tag 2 um 12 Uhr nur 7°C warm ist, ist es sinnlos und falsch, zu behaupten, an Tag 1 sei es doppelt so warm gewesen wie an Tag 2. □

Beispiel: Wird die Temperatur dagegen in Grad Kelvin gemessen, so hat die zugrunde gelegte Skala einen absoluten Nullpunkt. Es liegt eine Verhältnisskala vor. □

Abschließend sei darauf hingewiesen, dass bei Verhältnisskala und Intervallskala zusammenfassend von einer **Kardinalskala** (metrischen Skala) gesprochen wird.

1.3 Zur Wahl des Skalenniveaus

Der Informationsgehalt von Daten hängt vom Skalenniveau ab. Dabei gilt bezüglich des Informationsgehaltes folgende Ungleichung:

metrisch > ordinal > nominal

Es ist immer möglich, von einem höheren auf ein niedrigeres Skalenniveau überzugehen. So können ursprünglich metrische Daten durch Klassenbildung in ordinale

Daten umgewandelt werden. Ordinale Daten können prinzipiell auch in nominale umgewandelt werden, obwohl diese Umwandlung weniger häufig vorkommt. Die Umwandlung in entgegengesetzter Richtung (nominal \Rightarrow ordinal/metrisch oder ordinal \Rightarrow metrisch) ist dagegen nicht ohne weiteres möglich. Der Wechsel von einem höheren auf ein niedrigeres Skalenniveau ist immer mit einem Informationsverlust verbunden.

metrisch \Rightarrow ordinal \Rightarrow nominal
 $\xrightarrow{\hspace{1.5cm}}$ Informationsverlust

nominal \nRightarrow ordinal \nRightarrow metrisch

Der höhere Informationsgehalt metrischer Daten ist allerdings in der Regel auch mit einem höheren Messaufwand verbunden.

Beispiel: Oktavskala (ordinal) zur Erfassung des Deckungsgrades einer Pflanzenart (Gauch 1982, Multivariate analysis in community ecology. Cambridge University Press, Cambridge, S. 52):

Oktavskala	Metrische Skala
0	Nicht erhoben
1	0 bis <0,5% (Deckungsgrad) DG
2	0,5 bis <1% DG
3	1% bis <2% DG
4	2 bis <4% DG
5	4 bis <8% DG
6	8 bis <16% DG
7	16 bis <32% DG
8	32 bis <64% DG
9	64 bis <100% DG

(Der Name „Oktav-Skala“ kommt daher, dass sich die Werte der Klassengrenzen bei jedem Schritt verdoppeln, so wie bei zwei benachbarten Oktaven in der Musik das Frequenzverhältnis immer 2:1 ist). Es ist sehr viel aufwendiger, den prozentualen Deckungsgrad einer Pflanzenart exakt zu schätzen oder bildanalytisch (Fotografie mit anschließender Auswertung am PC) zu bestimmen (metrische Skala), als die oben angegebene Oktavskala zu verwenden, bei welcher der Betrachter einer Parzelle von 1 m² oder 0,5 m² per Augenschein lediglich die Deckungsgradklasse bestimmt. Liegen die Daten zum Deckungsgrad in einer Oktavskala vor, ist es nicht möglich, nachträglich den exakten Deckungsgrad zu bestimmen (ordinal \Rightarrow metrisch). Umgekehrt kann man bei Vorliegen des exakten Deckungsgrades nachträglich jederzeit die betreffende Klasse/Kategorie auf der Oktavskala ermitteln (metrisch \Rightarrow ordinal).

Eigentlich ist die Bezeichnung Oktavskala nicht ganz angemessen, da die Skala zehn, und nicht acht Klassen hat. Der Erfinder der Skala sieht dies auch so (Hugh Gauch, pers. Mitt., April 2005). Ursprünglich waren vielleicht einmal acht Klassen vorgesehen, und so wurde der Name geboren. Spätere Verfeinerungen der Skala haben dann wohl nicht mehr zur Verwerfung des Namens geführt.

2. Beschreibende Statistik für metrische Daten

2.1 Histogramm

Ein Histogramm ist eine graphische Darstellung der Häufigkeitsverteilung quantitativer Messwerte in einer Stichprobe.

Beispiel: Angenommen für ein Maisfeld mit 100.000 Maispflanzen wird an einer Stichprobe von $n = 50$ Maispflanzen die Pflanzenlänge (x) bestimmt. Dabei werden folgende Messungen (cm) gemacht:

175 172 179 167 163 154 163 164 157 177
 186 165 175 194 176 162 166 169 170 181
 168 166 180 164 179 170 150 192 170 173
 170 150 174 164 182 188 157 165 172 168
 179 179 164 162 178 162 182 171 182 183

Wir können die Häufigkeitsverteilung der Einzelmessungen graphisch in einem Balkendiagramm (Histogramm) darstellen. Dazu werden zunächst Größenklassen gebildet. Es liegen Werte zwischen 150 und 194 vor. Die **Variationsbreite** ist somit $194 - 150 = 44$. Wir wollen 5 Klassen der mit einer Breite von jeweils 10 cm bilden. Die Klassengrenzen legen wir bei 145, 155, 165, 175, 185 und 195 fest (wie man zu solchen Grenzen kommt, wird gleich besprochen). Die Zuordnung von Messwerten zu den Klassen muss auch dann eindeutig sein, wenn der Messwert genau auf die Klassengrenze fällt. Daher vereinbaren wir, dass der Wert der Klassengrenze immer der Klasse oberhalb der Grenze zufällt. So fällt der Wert 145 in die erste Klasse, der Wert 155 in die zweite Klasse, etc. Damit ist die Zuordnung an den Klassengrenzen eindeutig. Aus der obigen Urliste wird ausgezählt, wie viele Werte in die 1. Klasse fallen, wie viele in die 2. etc.

Höhe (cm)	Strichliste	Anzahl Pflanzen/ absolute Häufigkeit	Prozent/ relative Häufigkeit
145 – 154,9		3	6
155 – 164,9		11	22
165 – 174,9		17	34
175 – 184,9		15	30
185 – 194,9		4	8

Wir haben in obiger Tabelle die Klassengrenzen so bezeichnet, dass das Führen einer Strichliste besonders einfach ist. Die exakte Bezeichnung der ersten Klasse lautet $145 \leq x < 155$, die der zweiten $155 \leq x < 165$, etc. Bei dieser Bezeichnung taucht die Klassengrenze, z.B. der Wert 155, in zwei Klassen auf, was bei der Führung einer Strichliste leicht zu Flüchtigkeitsfehlern führt, obwohl die Bezeichnung eindeutig ist (155 fällt in die zweite und nicht in die erste Klasse). Daher hier die einfachere Bezeichnung $145 \leq x \leq 154,9$, $155 \leq x \leq 164,9$, etc. Eine graphische Darstellung der Häufigkeiten der obigen Tabelle ergibt das folgende Histogramm:

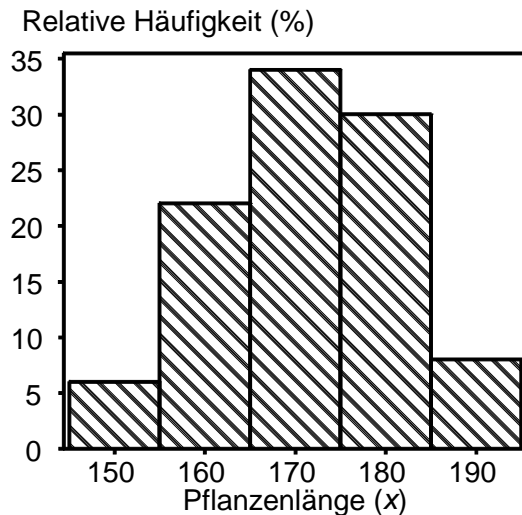


Abb.2.1.1: Histogramm für die Maisdaten.

Das Histogramm zeigt, dass die mittlere Klasse (Klassenmitte 170 cm) die größte Häufigkeit hat. Zu den Rändern hin nehmen die Häufigkeiten ab. Dieses Muster ist typisch für viele Merkmale, die in der Natur beobachtet werden können. Es deckt sich gut mit der Annahme einer Normalverteilung. □

Bei der Erstellung eines Histogramms stellt sich die Frage nach der Klassenbreite sowie nach der Zahl der Klassen. Je größer die Stichprobe, desto mehr Klassen können verwendet werden. Wählt man zu viele Klassen, so ist die Besetzung der Klassen so gering, dass das Histogramm wenig aussagekräftig ist. Im anderen Extrem ist die Zahl der Klassen zu klein, um die Form der Verteilung einschätzen zu können, z.B. wenn man nur zwei Klassen hat. Mit etwas Übung kann man schnell eine vernünftige Klassenzahl ermitteln. Es gibt außerdem zahlreiche Faustregeln für die Ermittlung der Zahl der Klassen, von denen hier 2 angegeben werden:

Bestimmung der Zahl der Klassen und der Klassenbreite

k = Zahl der Klassen

b = Klassenbreite

n = Stichprobenumfang

x_{max} = größter Messwert

x_{min} = kleinster Messwert

V = Variationsbreite = $x_{max} - x_{min}$

Zahl der Klassen:

$k \geq (2n)^{1/3}$ (Terrel GR, Scott DW 1985 Oversmoothed nonparametric density estimates. JASA 80, 209-214; SAS nutzt diese Regel)

oder

$k = 1 + 3,32 \log_{10}(n)$
 $\approx 1 + 1,44 \log_e(n)$ (Sturges' Formel)

Klassenbreite: $b > \frac{V}{k}$

Sturges' Formel liefert meist eine etwas größere Zahl von Klassen als die Formel von Terrel und Scott.

Beispiel: Wir betrachten die oben angegebenen Pflanzenlängen von $n = 50$ Maispflanzen. Nach Sturges Formel ist

$$k = 1 + 3,32 \log_{10}(n) = 1 + 3,32 \log_{10}(50) = 6,64$$

Aufrunden ergibt $k = 7$. Nach der Terrel-Scott Formel dagegen ist

$$k \geq (2n)^{1/3} = (100)^{1/3} = 4,64$$

zu wählen, also $k = 5$. Wir entscheiden uns für $k = 5$. Weiterhin finden wir

$$V = \text{Variationsbreite} = x_{\max} - x_{\min} = 194 - 150 = 44.$$

Hieraus ergibt sich eine Klassenbreite von

$$b > V/k = 44/5 = 8,8$$

Wir entscheiden uns hier für $b = 10$. Der Einfachheit halber werden die Klassengrenzen 145, 155 etc. bis 195 gewählt (siehe oben). Das Ergebnis der weiteren Berechnungen findet sich in Abb. 2.1.1. \square

Stamm-und-Blatt Darstellung (Stem-and-leaf Plot): Eine dem Histogramm sehr ähnliche graphische Darstellung ist die Stamm-und-Blatt Darstellung. Dies hat den Vorteil, dass sie sehr einfach und schnell mit Papier und Bleistift zu erhalten ist. Im Mais-Beispiel schreibt man hierzu einen Baum der ersten beiden Ziffern wie folgt:

```

15 |
16 |
17 |
18 |
19 |

```

Der senkrechte Strich trennt den Stamm von den noch rechts davon einzutragenden Blättern. Sodann trägt man die jeweils dritte Ziffer einer jeden Zahl an die entsprechende Position des Stammes. Eintrag der ersten vier Zahlen (175, 172, 179, 167) liefert folgendes Bild:

```

15 |
16 | 7
17 | 529
18 |
19 |

```

```

15 | 47007
16 | 73345269864458422
17 | 529756090030429981
18 | 610282223
19 | 42

```

Die Zahl der Blätter hängt von der Zahl der am Stamm abzutragenden Ziffern ab. Daher ist es wichtig, wie viele Ziffern durch abgetragen werden und wie. In manchen Fällen ist es hilfreich, auch am Stamm Klassen zu bilden, um nicht zu viele und nicht zu wenige Blätter zu haben. Details kann man z.B. Bei Stahel (1999) nachlesen.

Daten liegen in der Regel in einer größeren Stichprobe vor. Es ist daher oft von Interesse, die Verteilung der Daten durch wenige Maßzahlen zu charakterisieren und zusammenzufassen. Insbesondere interessiert meistens, welche Werte die Stichprobe "im Durchschnitt" aufweist, d.h. welche Lage das Gros der Werte hat, und wie groß die "Streuung" der Daten ist. Bei entsprechenden Maßzahlen spricht man daher auch von **Lagemaßen** und **Streuungsmaßen**. Eine wichtige statistische Maßzahl, aus der sich sowohl Lage- als auch Streuungsmasse ableiten lassen, sind die sog. **Quantile** einer Verteilung, mit denen wir uns jetzt befassen.

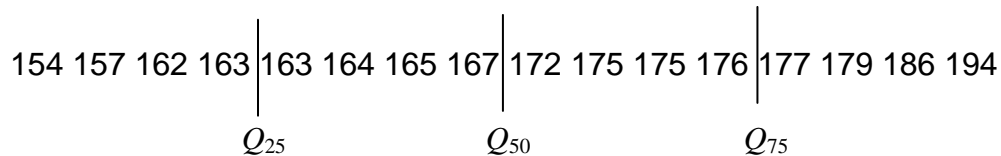
Eine Möglichkeit, eine Stichprobe von quantitativen Messwerten zu charakterisieren, besteht darin, Punkte (Werte) zu bestimmen oberhalb bzw. unterhalb denen ein bestimmter Prozentsatz der Werte der Stichprobe liegen. Bei solchen Punkten spricht man von **Quantilen** oder auch **Perzentilen**. Allgemein gilt für das $t\%$ -Quantil (Q_t), dass mindestens $t\%$ der Werte kleiner oder gleich Q_t sind und mindestens $(100-t)\%$ der Werte größer oder gleich Q_t sind. Von besonderem Interesse ist der sog. **Median**. Dies ist ein Punkt, unterhalb und oberhalb dessen gleich viele Werte liegen (Q_{50}). Der Median entspricht dem 50%-Quantil. Um die Quantile bestimmen zu können, muss die Stichprobe nach der Größe der Messwerte geordnet werden.

175 172 179 167 163 154 163 164 157 177 186 165 175 194 176 162

$$\begin{array}{cccccccc|cccccccc} 154 & 157 & 162 & 163 & 163 & 164 & 165 & 167 & 172 & 175 & 175 & 176 & 177 & 179 & 186 & 194 \\ & & & & & & & & \mathcal{O}_{50} & & & & & & & \end{array}$$

Jeder Wert zwischen 167 und 172 teilt die Stichprobe in zwei gleich große Hälften. Jeder dieser Punkte kann daher als Q_{50} , also als 50%-Quantil gelten. Man bezeichnet dieses Quantil auch als **Median**. Der Einfachheit halber wählen wir das arithmetische Mittel der Werte 167 und 172. Also ist der Median gleich $Q_{50} = (167+172)/2 = 169,5$.

Wir bestimmen außerdem das 25%- und das 75%-Quantil:



Wir finden:

$$Q_{25} = (163+163)/2 = 163 \text{ und}$$

$$Q_{75} = (176+177)/2 = 176,5.$$

Es liegen 25% der Werte unterhalb von $Q_{25} = 163$ und 75 unterhalb von $Q_{75} = 176,5$.

□

Im obigen Beispiel war es intuitiv klar, wie Q_{25} , Q_{50} und Q_{75} zu bestimmen sind, weil $n = 16$ durch 4 teilbar ist, so dass sich die geordnete Stichprobe in vier gleich große Abschnitte aufteilen lässt. Falls andere Quantile zu bestimmen sind, muss ein modifiziertes Verfahren verwendet werden. Grundlage der meisten Verfahren ist die einfache Überlegung, dass man für das $t\%$ -Quantil etwa den $n \cdot t/100$ -ten Wert der geordneten Stichprobe verwenden kann. Wenn z.B. $n = 10$ ist, so ergibt diese einfache Regel den 5ten Wert der geordneten Stichprobe. Besser ist das Mittel des 5. und 6. Wertes. Diese vereinfachte Regel muss also etwas verfeinert werden, um in allen Fällen möglichst gute Ergebnisse zu liefern. Insbesondere muss in den Fällen, die geordnete Stichprobe genau geteilt werden kann, so dass die „untere Hälfte“ $t\%$ umfasst, dieser Fall auch identifiziert werden und das Quantil als ein Wert zwischen dem größten der kleineren Hälfte und dem kleinsten der größeren Hälfte berechnet werden, so. z.B. im Falle des Median, wenn n eine gerade Zahl ist. Desweiteren ist es so, dass ein empirisches Quantil fast immer mit einer der beobachteten Werte übereinstimmt, damit die oben gegebene Definition für ein Quantil auch erfüllt ist. Nur in dem Fall, dass $n \cdot t/100$ eine ganze Zahl ist, kann das empirische Quantil zwischen dem $n \cdot t/100$ -ten Wert und dem $(n \cdot t/100 + 1)$ -ten Wert liegen. Eine Möglichkeit, die in Statistik-Programmen verwendet wird, ist im folgenden angegeben.

Eine allgemeine Regel zur Bestimmung der Quantile einer Stichprobe (SAS Institute: SAS Procedures Guide. Version 6. NC: Cary, S. 627; dort stehen fünf verschiedene Regeln!).

n = Stichprobenumfang

Für das $t\%$ -Quantil (Perzentil) definiere

$$p = t/100$$

Bezeichne die geordnete Stichprobe mit

$$x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]} \quad ([j] = \text{Ordnungszahl; } x_{[3]} \text{ ist z.B. der dritt-kleinste Wert})$$

Berechne

$$np = j + g$$

wobei

j = ganzzahliger Teil von np und

g = Rest von np nach Abzug von j

Das $t\%$ -Quantil ist gegeben durch

$$Q_t = (x_{[j]} + x_{[j+1]})/2 \quad \text{falls } g = 0$$

$$Q_t = x_{[j+1]} \quad \text{falls } g > 0$$

$x_{[0]}$ wird gleich $x_{[1]}$ gesetzt.

$x_{[n+1]}$ wird gleich $x_{[n]}$ gesetzt.

Beispiel: Gegeben sei wieder die Stichprobe von $n = 16$ Maispflanzen, deren Länge (x) gemessen wurde:

175 172 179 167 163 154 163 164 157 177 186 165 175 194 176 162

Sortieren ergibt:

154 157 162 163 163 164 165 167 172 175 175 176 177 179 186 194
 $x_{[1]} \quad x_{[2]} \quad x_{[3]} \quad x_{[4]} \quad x_{[5]} \quad x_{[6]} \quad x_{[7]} \quad x_{[8]} \quad x_{[9]} \quad x_{[10]} \quad x_{[11]} \quad x_{[12]} \quad x_{[13]} \quad x_{[14]} \quad x_{[15]} \quad x_{[16]}$

Wir wollen den Median bestimmen.

$$t = 50; p = t/100 = 0,5; np = 16 \cdot 0,5 = 8; j = 8; g = 0$$

$$\Rightarrow Q_{50} = (x_{[8]} + x_{[9]})/2 = (167 + 172)/2 = 169,5. \quad \square$$

Beispiel: Gegeben sind die Lebendgewichte einer Stichprobe von Milchkühen:

663 644 656 671 665 659 656 647 642 647 657 632 649

Es sollen die Quantile Q_{10} , Q_{20} , Q_{50} (Median) bestimmt werden.

Sortieren:

632 642 644 647 647 649 656 656 657 659 663 665 671
 $x_{[1]} \quad x_{[2]} \quad x_{[3]} \quad x_{[4]} \quad x_{[5]} \quad x_{[6]} \quad x_{[7]} \quad x_{[8]} \quad x_{[9]} \quad x_{[10]} \quad x_{[11]} \quad x_{[12]} \quad x_{[13]}$

Q_{10} :

$$t = 10; p = t/100 = 0,1; np = 13 \cdot 0,1 = 1,3; j = 1; g = 0,3$$

$$\Rightarrow Q_{10} = x_{[2]} = 642$$

Unterhalb dieses Wertes liegen hier 2 Werte (einschließlich des Wertes selber), also $2/13$ der Stichprobe = 15,4%, was zeigt, dass hier eine Näherung vorliegt. Je größer die Stichprobe, umso kleiner ist dieser Näherungsfehler. Die Berechnung von Quantilen ist also vor allem für größere Stichproben sinnvoll. Eine andere Betrachtungsweise in diesem Zusammenhang ist die, dass das Quantil der Stichprobe einen Schätzwert des Quantils in der Grundgesamtheit (100.000 Pflanzen) darstellt. Insofern ist es akzeptabel, dass das berechnete Q_t -Quantil nicht die Eigenschaft hat, dass immer genau $t\%$ der Werte unterhalb von Q_t liegen. Dies ist nur dann der Fall, wenn np eine ganze Zahl ist, also wenn $g = 0$ ist.

Q_{20} :

$$t = 20; p = t/100 = 0,2; np = 13 \cdot 0,2 = 2,6; j = 2; g = 0,6$$

$$\Rightarrow Q_{20} = x_{[3]} = 644$$

Unterhalb von Q_{20} liegen hier 3 Werte, also $3/13 = 23,1\%$ der Stichprobe statt 20%.

Q_{50} :

$$t = 50; p = t/100 = 0,5; np = 13 \cdot 0,5 = 6,5; j = 6; g = 0,5$$

$$\Rightarrow Q_{50} = x_{[7]} = 656. \quad \square$$

Beispiel: Für das vorhergehende Beispiel sollen Q_0 und Q_{100} bestimmt werden. Q_0 und Q_{100} sind immer der kleinste und der größte Wert der Stichprobe:

Q_0 :

$$t = 0; p = 0; np = 13 \cdot 0 = 0; j = 0; g = 0$$

$$\Rightarrow Q_0 = (x_{[0]} + x_{[1]})/2 = (632 + 632)/2 = 632 \quad (\text{beachte: } x_{[0]} = x_{[1]})$$

Q_{100} :

$$t = 100; p = 1,0; np = 13 \cdot 1,0 = 13; j = 13; g = 0$$

$$\Rightarrow Q_{100} = (x_{[13]} + x_{[14]})/2 = (671 + 671)/2 = 671 \quad (\text{beachte: } x_{[n+1]} = x_{[n]})$$

Bemerkung: Jeder Wert $x_{[j]}$ der geordneten Stichprobe repräsentiert auch ein Quantil der Stichprobe. Um zu bestimmen, für welches t ein Wert $x_{[j]}$ ein Quantil darstellt, wenden wir die oben im Kasten gegebene Regel zur Berechnung eines Quantils an. Aus dieser folgt, dass $x_{[j]}$ ein $t\%$ -Quantil ist für alle t mit $n \cdot t/100 = (j-1) + g$, wobei $0 < g < 1$ ist. Da g nicht eindeutig bestimmt ist, wählen wir hier einfach das Mittel, also $g = 0.5$. Somit setzen wir $n \cdot t/100 = (j-1) + 0.5$. Auflösen nach t liefert

$t = \frac{j - \frac{1}{2}}{n} 100$. Dies heißt: der j -te Wert in der Stichprobe repräsentiert das

$t = \frac{j - \frac{1}{2}}{n} 100$ %-Quantil der Stichprobe (Dies ist hilfreich für Q-Q-Plots; siehe unten).

Beispiel: Für $n = 13$ ist der siebte Wert der geordneten Stichprobe $x_{[j=7]}$ der Median,

da $t = \frac{7 - \frac{1}{2}}{13} 100 = 50$.

Q-Q-Plots, oder Quantil-Quantil-Plots, sind Grafiken zum Vergleich einer Stichprobenverteilung mit einer theoretischen Verteilung, z.B. einer Normalverteilung. Für die Stichprobe berechnet man für die Quantilprozent t der Werte der geordneten Stichprobe und trägt diese gegen die Quantile Q_t der theoretischen Verteilung ab (Siehe Abschnitt 3.2.2). Falls die Stichprobe mit der theoretischen Verteilung gut übereinstimmt, liegen die resultierenden n Punkte etwa auf einer Geraden. Q-Q-Plots werden daher zur Überprüfung von Verteilungsannahmen verwendet.

2.2.2 Lagemaße

Median

Ein gängiges Lagemaß ist der Median. Dieser entspricht, wie oben schon gesagt, dem 50% Quantil. Seine Berechnung ist sehr einfach, und es sind nur 2 Fälle zu unterscheiden. Daher wird die Berechnungsregel hier nochmals gesondert aufgeführt.

Falls n eine **ungerade** Zahl ist, ist der Median gleich dem mittleren Wert der geordneten Stichprobe $x_{[1]}, \dots, x_{[n]}$:

$$Q_{50} = x_{[(n+1)/2]}$$

Falls n eine **gerade** Zahl ist, ist der Median gleich dem arithmetischen Mittel der beiden mittleren Werte der geordneten Stichprobe:

$$Q_{50} = (x_{[n/2]} + x_{[n/2+1]})/2$$

Beispiel: Gegeben sei folgende Stichprobe von Pflanzenlängen:

23 78 41 79 48

Sortieren:

23 41 48 78 79

$n = 5$ ist ungerade $\Rightarrow Q_{50} = x_{[(n+1)/2]} = x_{[3]} = 48$

Beispiel: Gegeben sei folgende Stichprobe von Pflanzenlängen:

23 78 41 79 48 65

Sortieren:

23 41 48 65 78 79

$n = 6$ ist gerade $\Rightarrow Q_{50} = (x_{[n/2]} + x_{[n/2+1]})/2 = (x_{[3]} + x_{[4]})/2 = (48 + 65)/2 = 56,5$

Arithmetisches Mittel

Das gängigste Lagemaß ist das arithmetische Mittel. Dieses berechnet sich wie folgt:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Beispiel: Gegeben sei wieder die Stichprobe von $n = 16$ Maispflanzen, deren Länge (x) gemessen wurde:

175 172 179 167 163 154 163 164 157 177 186 165 175 194 176 162

Wir finden

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{175 + 172 + 179 + \dots + 194 + 176 + 162}{16} = 170,5625$$

Zum Vergleich: Der Median ist 169,5 und ist damit nur unwesentlich vom arithmetischen Mittel verschieden. Der Grund liegt darin, dass hier keine Ausreißer vorliegen und dass die Daten annähernd symmetrisch verteilt sind.

Beispiel: Angenommen, in der obigen Stichprobe ist bei der ersten Pflanze ein Tippfehler unterlaufen. Hierdurch liegt ein Ausreißer vor. Im vorliegenden Fall würden wir den Tippfehler wahrscheinlich bemerken. Dies kann aber in größeren, elektronisch gehaltenen Datensätzen schwieriger sein!

1175 172 179 167 163 154 163 164 157 177 186 165 175 194 176 162

$Q_{50} = 169,5$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1175 + 172 + 179 + \dots + 194 + 176 + 162}{16} = 233,0625$$

Der Median ist unverändert, während das arithmetische Mittel deutlich nach oben verschoben ist. Das Beispiel zeigt die große Robustheit des Median gegenüber

Ausreißern, während das arithmetische Mittel empfindlich gegenüber Ausreißern ist. Außerdem ist ein großer Unterschied zwischen Median und Mittelwert ein Hinweis auf Abweichung von einer symmetrischen Verteilung, entweder durch Ausreißer oder durch eine schiefe Verteilung.

Beispiel: Am Institut für Agrartechnik an der Universität Hohenheim wurden an einer Sämaschine mit einem sog. Optosensor die Abstände von ausgesäten Körnern in der Reihe gemessen (Müller, pers. Mitteilung, 2001). Der Stichprobenumfang betrug $n = 10971$ Kornabstände.

Arithmetisches Mittel (x): 0,028887 m = 2,8887 cm
Median (x): 0,020470 m = 2,0470 cm

Der Median liegt deutlich unter dem arithmetischen Mittel, was auf eine schiefe (asymmetrische) Verteilung hindeutet. Dies wird durch das untenstehende Histogramm bestätigt (obere Graphik). Interessanter Weise wird die Verteilung der Daten symmetrisch, wenn wir eine Potenz-Transformation der Daten durchführen: $y = x^{0,28602}$ (Box-Cox-Transformation), wie die untere Graphik zeigt. Für diese Daten finden wir:

Arithmetisches Mittel (y): 0,329257
Median (y): 0,328811

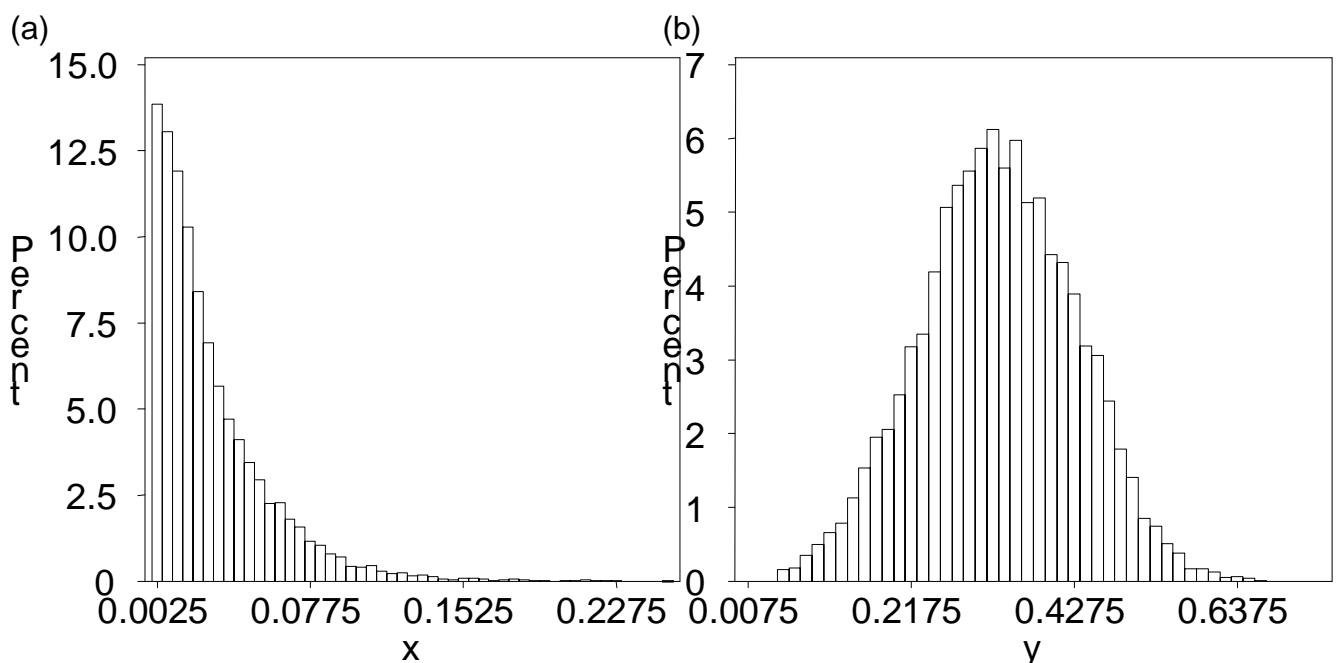


Abb. 2.2.2.1: Histogramm der Verteilung der Kornabstände. (a) Untransformierte Daten (x ; in m). (b) Transformierte Daten $y = x^{0,28602}$ (Box-Cox-Transformation).

Auf die Transformation wird später noch eingegangen. Die transformierten Daten folgen hier annähernd einer Normalverteilung, was im Rahmen eines Projektes für eine Simulation von Kornabständen mit dem Computer genutzt werden soll. Außerdem ist die Erfüllung der Annahme einer Normalverteilung Voraussetzung für die Anwendbarkeit verschiedener statistischer Verfahren (siehe unten; Kap. 4 ff.).

Geometrisches Mittel

$$G_x = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

Die praktische Berechnung wird durch Logarithmieren erleichtert:

$$\log(G_x) = \frac{\sum_{i=1}^n \log(x_i)}{n}$$

Das geometrische Mittel hat im landwirtschaftlichen Bereich eine sehr geringe Bedeutung.

Beispiel: Das geometrische Mittel ist für Wachstumsraten und Zinssätze ein sinnvolles Durchschnittsmaß. Über eine Periode von n Jahren wird die Preissteigerungsrate q_i eines Produktes in % erfaßt. Angenommen, der Preis des Produktes beträgt zu Beginn der Periode p_0 . Nach dem ersten Jahr ist der Preis

$$p_1 = p_0(1 + q_1/100)$$

Nach zwei Jahren ist der Preis

$$p_2 = p_1(1 + q_2/100) = p_0(1 + q_1/100)(1 + q_2/100)$$

und nach n Jahren

$$p_n = p_0 \prod_{i=1}^n (1 + q_i / 100)$$

Dies kann auch geschrieben werden als

$$p_n = p_0 \prod_{i=1}^n x_i$$

wobei $x_i = 1 + q_i/100$ die Preissteigerungsfaktoren sind. Wir können nun fragen, welche durchschnittliche Preissteigerungsrate dieselbe Preissteigerung über die n Jahre gebracht hätte wie die tatsächlichen Raten q_i . Hierzu können wir zunächst den Preissteigerungsfaktor G_x berechnen, für den dies der Fall ist:

$$p_n = p_0 \prod_{i=1}^n x_i = p_0 G_x^n \Leftrightarrow G_x = \sqrt[n]{\prod_{i=1}^n x_i}$$

Offenbar müssen wir hierfür das geometrische Mittel von x_i berechnen! Die durchschnittliche Preissteigerungsrate ist dann:

$$q = (G_x - 1)100$$

Die Preissteigerungen in vier Jahren seien: 2,3%, 3,8%, 2,4%, 3,3%. Daher ist $x_1 = 1,023$; $x_2 = 1,038$; $x_3 = 1,024$; $x_4 = 1,033$. Der mittlere Preissteigerungsfaktor ist

$$\log(G_x) = \frac{\sum_{i=1}^n \log(x_i)}{n} = \frac{\log(1,023) + \log(1,038) + \log(1,024) + \log(1,033)}{4} = 0,029055$$

$$G_x = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[4]{1,023 \cdot 1,038 \cdot 1,024 \cdot 1,033} = \exp[\log(G_x)] = \exp(0,029055) = 1,029481$$

und somit $q = 2,9481$. Die durchschnittliche Preissteigerung beträgt also 2,9481%. Zum Vergleich: das arithmetische Mittel beträgt 2,95 %, was ein sehr geringer Unterschied zum arithmetischen Mittel ist.

Beispiel: Preissteigerungsraten (%) über 10 Jahre.

q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}
11,2	2,8	6,5	18,7	12,3	8,4	6,6	3,5	2,6	3,6

Geometrisches Mittel (aus $x_i = 1+q_i/100$): 7,51; arithmetisches Mittel der q_i : 7,62.

Beispiel: Für Keimzahlen bei Milch wird in der Praxis oft das geometrische Mittel anstelle des arithmetischen Mittels berechnet. E. Renner (1970: Mathematisch-Statistische Methoden in der praktischen Anwendung. Parey Verlag, Berlin) gibt die Keimzahlen von 95 Milchproben (in 10^3) an:

5150	57400	14200	35800	12200
26900	61000	6750	51700	1390
285	3150	4000	9600	200
265	13100	5800	56800	440
4750	1170	4350	8200	47
60900	835	950	8800	270
1410	5150	3300	4900	995
3950	600	2550	3550	63000
2150	86200	4150	710	580
8250	975	1910	5400	1350
30500	16800	120	380	27500
295	965	32700	3030	56500
890	45700	980	595	765
1340	15700	12100	4550	20500
20500	52000	18700	4350	32900
910	2700	115	5900	33000
19100	1760	415	2450	70000
170	1820	21800	79800	230
9150	6800	7050	1940	2750

Das geometrische Mittel beträgt $G_x = 3996,83$. Zum Vergleich das arithmetische Mittel: $\bar{x} = 13954,76$. Dies ist mehr als das dreifache des geometrischen Mittels! Das arithmetische Mittel ist wegen der Schiefe der Verteilung ein problematisches Lagemaß. Der Median beträgt $Q_{50} = 4350$ und kommt dem geometrischen Mittel recht nahe.

Renner (1970) gibt als Grund für die Bevorzugung des geometrischen Mittels die Tatsache an, dass der Messbereich mit Werten zwischen 47 ($10^3/\text{ml}$) bis 86000 ($10^3/\text{ml}$) mehrere Zehnerpotenzen umspannt. Hiermit geht eine sehr schiefe Verteilung der Messwerte einher (siehe unten stehende Abbildung). Ich bin allerdings nicht sicher, warum gerade das geometrische Mittel hier besondere Vorteile bieten soll. Es hat aus meiner Sicht den Nachteil, dass die Interpretation des geometrischen Mittels in diesem Fall problematisch ist. Besser zu interpretieren ist der Median. Wir können sagen, dass in 50% der Milchproben die Keimzahl unter 4350 ($10^3/\text{ml}$) liegt. Eine solche Interpretation lassen weder das geometrische Mittel noch das arithmetische Mittel zu.

Eine weitere Begründung für die Berechnung des geometrischen Mittels ist, dass die Logarithmen der Keimzahlen oft näherungsweise einer Normalverteilung folgen (siehe Abb. 2.2.2.2). Somit folgen die Keimzahlen selbst näherungsweise einer Lognormalverteilung, welche sehr schief sein kann. Nun ist aber das geometrische Mittel kein guter Schätzer des Erwartungswertes einer Lognormalverteilung. Auch aus diesem Grund ist die Wahl dieses Lagemaßes bei Keimzahlen problematisch. Allerdings ist das geometrische Mittel ein guter Schätzer des Median, falls Lognormalverteilung gegeben ist. Insofern sind G_x und Q_{50} zwei alternative Lagemaße, die dasselbe schätzen. Der Median Q_{50} hat den Vorteil, dass er an keine besonderen Verteilungsannahmen gebunden ist.

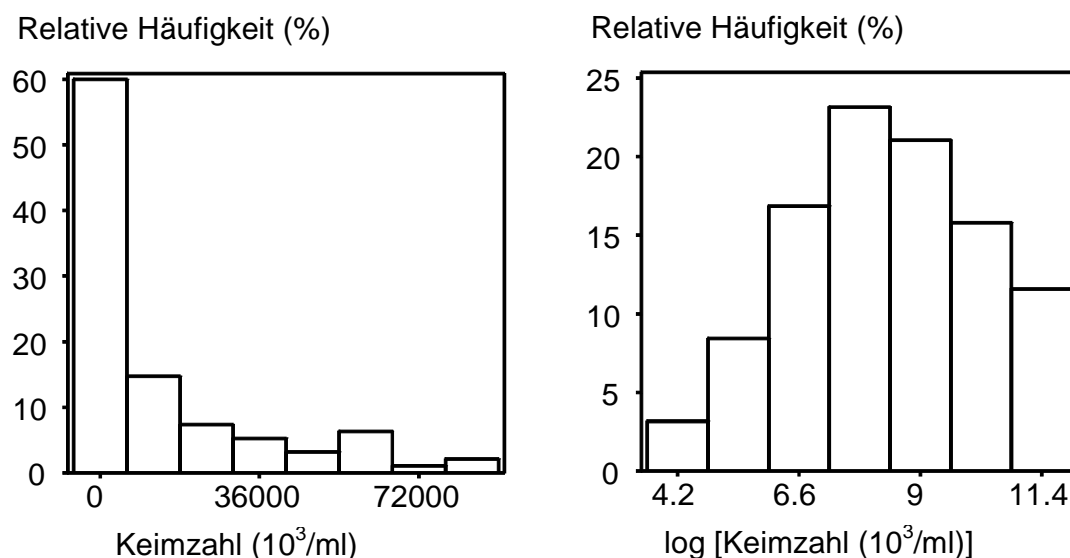


Abb 2.2.2.2: Verteilung der Keimzahldaten mit und ohne logarithmische Transformation

Harmonisches Mittel

Ein weiteres Lagemaß, welches relativ wenig gebräuchlich ist, hier der Vollständigkeit halber aber genannt werden soll, ist das harmonische Mittel.

$$H_x = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Dieses Lagemaß wird für die Mittlung etwa folgender Größen benötigt: Geschwindigkeiten, Dichten von Gasen, Flüssigkeiten etc., Überlebenszeiten (Sachs).

Beispiel: Ein Schlepper bearbeitet einen Hang. Hangaufwärts fährt er 4 km/h, auf der Rückfahrt hangabwärts fährt er 6 km/h. Wie groß ist seine Durchschnittsgeschwindigkeit? Antwort: $x_1 = 4$, $x_2 = 6$,

$$H_x = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{2}{\frac{1}{4} + \frac{1}{6}} = \frac{2}{\frac{6+4}{24}} = 48/10 = 4,8$$

Diese Berechnung kann man auch wie folgt nachvollziehen. Die Länge des Feldes in Hangrichtung sei 300 m. Die Zeit y_1 , die der Schlepper bergauf benötigt, ergibt sich durch einfachen Dreisatz:

$$(300 \text{ m}) / y_1 = 4000 \text{ m/h} = x_1 \Leftrightarrow y_1 = 300 \text{ m} / 4000 \text{ (m/h)} = 0,075 \text{ h}$$

Brauche 0,075 h, um 300 m bergauf zu fahren.

Allgemein für beliebige Geschwindigkeit x_1

$$y_1 = (300 \text{ m}) / x_1$$

Bergab ergibt sich

$$y_2 = (300 \text{ m}) / [6000 \text{ (m/h)}] = 0,05 \text{ h} = 300 \text{ m} / x_2$$

Die durchschnittliche Zeit für einen Weg (300m) beträgt (allgemein für beliebige Geschwindigkeiten x_1 und x_2)

$$\frac{y_1 + y_2}{2} = \frac{300/x_1 + 300/x_2}{2} = \frac{300}{2} \left(\frac{1}{x_1} + \frac{1}{x_2} \right) = 0,0625$$

Damit ist die durchschnittliche Geschwindigkeit (= Weg/Zeit) gleich

$$\frac{300}{\frac{300}{2} \left(\frac{1}{x_1} + \frac{1}{x_2} \right)} = \frac{2}{\left(\frac{1}{x_1} + \frac{1}{x_2} \right)} = \frac{300}{0,0625} = 4800 \text{ m/h} = 4,8 \text{ km/h}$$

Beispiel: Es werden 10 Bodenproben à 1 kg genommen. In jeder Probe wird die Dichte in kg/Liter bestimmt. Es ergaben sich folgende Werte: 1,3; 1,4; 1,3; 1,2; 1,1; 1,0; 1,5; 1,4; 1,4; 1,3. Die durchschnittliche Dichte beträgt

$$H_x = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{10}{\frac{1}{1,3} + \frac{1}{1,4} + \frac{1}{1,3} + \frac{1}{1,2} + \frac{1}{1,1} + \frac{1}{1,0} + \frac{1}{1,5} + \frac{1}{1,4} + \frac{1}{1,4} + \frac{1}{1,3}} = 1,27232$$

Wichtig ist, dass die Probemenge (kg) für alle Proben dieselbe ist, und nicht das Volumen. Hätten wir 10 Proben á 1 Liter genommen, wäre das arithmetische Mittel das richtige Maß gewesen!

Generell ist folgende Regel hilfreich für die Wahl zwischen harmonischem und arithmetischem Mittel:

Feste Mengenangabe [z.B. Dichte (kg/Liter) von 1 kg Boden; 1 kg hier fix] bezieht sich auf den

Zähler der Maßeinheit \Rightarrow harmonisches Mittel

Nenner der Maßeinheit \Rightarrow arithmetisches Mittel

Im obigen Beispiel ist die Mengenangabe der Proben in kg. Diese Einheit steht im Zähler der Zielgröße (Dichte), daher ist nach der o.g. Regel das harmonische Mittel angebracht. Im Beispiel der Schleppergeschwindigkeiten ist die Mengenangabe nicht so einfach zu erkennen. Die Probenahmeeinheiten sind hier die beiden Strecken hangaufwärts und hangabwärts. Die Mengenangabe ist die Länge dieser Strecke, und diese wird in km angegeben. Da dies im Zähler der Geschwindigkeit steht, ist das harmonische Mittel angemessen. Die Mengenangabe selbst wird übrigens nicht zur Berechnung des Mittels benötigt. Wichtig ist nur, dass für alle Proben die Menge dieselbe ist. Andernfalls müsste ein gewichtetes Mittel berechnet werden, worauf hier jedoch nicht näher eingegangen wird.

2.2.3 Streuungsmaße

Beispiel (Kontrolle): In On-Farm Versuchen mit Sorghum im subsaharischen Afrika wurden auf insgesamt 28 Farmen jeweils drei Düngermethoden untersucht (Bürkert, pers. Mitteilung, 2000):

1. Kein Dünger
2. NPK (Stickstoff-Phosphor-Kalium)
3. DAP (Di-Ammon-Phosphat)

Die Felder variierten in der Größe zwischen 400 und 5000 m². Es wurde der Ertrag pro Fläche (dt ha⁻¹) ermittelt. Die Daten, sortiert nach dem Ertrag der Kontrolle, sind wie folgt (Erträge in dt ha⁻¹).

Farm	Kein	NPK	DAP
1	0,30	0,80	1,64
2	0,34	1,12	1,38
3	0,39	1,12	1,70
4	0,40	1,60	2,80
5	0,40	2,80	2,40
6	0,42	1,14	1,56

7	0,48	3,20	1,92
8	0,54	1,34	1,46
9	0,56	1,20	1,66
10	0,58	1,22	1,60
11	0,62	1,40	2,30
12	0,68	2,24	2,76
13	0,74	1,54	1,66
14	0,74	1,52	2,42
15	0,78	1,46	1,80
16	0,82	1,60	2,50
17	0,96	1,60	2,06
18	1,02	1,74	2,16
19	1,06	1,40	1,74
20	1,10	1,44	1,74
21	1,44	4,16	3,84
22	1,60	2,00	2,40
23	1,68	4,80	2,56
24	2,40	4,48	3,84
25	2,40	9,60	3,84
26	2,56	5,28	3,24
27	3,60	4,80	5,60
28	4,50	5,50	6,75

Es ist nun von Interesse, ob sich die Düngermethoden in der Ertragsstabilität unterscheiden. Eine Möglichkeit, die Stabilität zu erfassen, ist die Messung der "Streuung" über die Farmen. Je kleiner die Streuung, desto größer die Stabilität. □

Im folgenden wird eine Reihe von Streuungsmaßen kurz vorgestellt und anhand des obigen Beispiels erläutert. Wir führen die Rechnungen exemplarisch für die Kontrolle ("Kein Dünger") durch.

Variationsbreite

$$V = Q_{100} - Q_0 = x_{max} - x_{min}$$

Je größer V , desto größer ist die Streuung.

Beispiel (Kontrolle): $x_{max} = 4,50$, $x_{min} = 0,30$, $V = 4,50 - 0,30 = 4,20$

Interquartilabstand

Bei den 25%- und 75%-Quantilen spricht man auch von **Quartilen**, weil diese sich auf das obere und das untere Viertel (engl.: quarter) der geordneten Stichprobe beziehen. Der Abstand dieser beiden Quartile ist ein weiteres Streuungsmaß.

$$R_{IQ} = Q_{75} - Q_{25}$$

Beispiel (Kontrolle):

$x_{[i]}$	$[i]$
0,30	1
0,34	2
0,39	3
0,40	4
0,40	5
0,42	6
0,48	7
0,54	8
0,56	9
0,58	10
0,62	11
0,68	12
0,74	13
0,74	14
0,78	15
0,82	16
0,96	17
1,02	18
1,06	19
1,10	20
1,44	21
1,60	22
1,68	23
2,40	24
2,40	25
2,56	26
3,60	27
4,50	28

Q_{25}

Q_{75}

Q_{75} :

$$t = 75; p = 0,75; n = 28; np = 21; j = 21; g = 0;$$

$$Q_{75} = (x_{[21]} + x_{[22]})/2 = (1,44 + 1,60)/2 = 1,52$$

Q_{25} :

$$t = 25; p = 0,25; n = 28; np = 7; j = 7; g = 0;$$

$$Q_{25} = (x_{[7]} + x_{[8]})/2 = (0,48 + 0,54)/2 = 0,51$$

$$R_{IQ} = 1,52 - 0,51 = 1,01 \quad \square$$

Varianz

Die Streuung der Daten hängt davon ab, wie weit die Einzelwerte um den Mittelwert streuen. Daher ist es sinnvoll, die Streuung auf Basis der Abweichungen von Mittelwert $(x_i - \bar{x}_\bullet)$ zu definieren. Die vielleicht zunächst naheliegende Idee, diese Abweichungen einfach zu mitteln, führt jedoch nicht zum Ziel, da diese Summe

immer exakt gleich Null ist, wie man leicht zeigen kann. Der Grund ist, dass sich positive und negative Abweichungen genau aufheben. Stattdessen kann man beispielsweise die Beträge der Abweichungen mitteln. Üblicher ist es jedoch, die Abweichungen zu quadrieren und dann zu mitteln. Die sich ergebende Größe heißt Varianz und misst die mittlere quadratische Abweichung. Aus einem Grund, der später im Zusammenhang mit der Varianzanalyse (Abschnitt. 4.4) und in Abschnitt 5.8 näher erläutert wird, wird hier nicht durch den Stichprobenumfang n , sondern durch die Freiheitsgrade $(n-1)$ geteilt.

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

Beispiel (Kontrolle):

x_i	x_i^2
0,30	0,0900
0,34	0,1156
0,39	0,1521
0,40	0,1600
0,40	0,1600
0,42	0,1764
0,48	0,2304
0,54	0,2916
0,56	0,3136
0,58	0,3364
0,62	0,3844
0,68	0,4624
0,74	0,5476
0,74	0,5476
0,78	0,6084
0,82	0,6724
0,96	0,9216
1,02	1,0404
1,06	1,1236
1,10	1,2100
1,44	2,0736
1,60	2,5600
1,68	2,8224
2,40	5,7600
2,40	5,7600
2,56	6,5536
3,60	12,9600
4,50	20,2500

$$\sum_{i=1}^n x_i \longleftarrow 33,11 \quad 68,2841 \longrightarrow \sum_{i=1}^n x_i^2$$

$$s_x^2 = \frac{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 / n}{n-1} = \frac{68,2841 - (33,11)^2 / 28}{27} = 1,079$$

Standardabweichung

Die Varianz hat den Nachteil, dass die Messeinheiten quadriert eingehen. Werden die Daten z.B. in t/ha gemessen, so hat die Varianz die Einheit (t²/ha²). Um wieder auf die ursprüngliche Messskala zu kommen, dann man das Quadrieren gewissermaßen rückgängig machen, indem man die Quadratwurzel zieht. Dies führt zur sog. Standardabweichung.

$$s_x = \sqrt{s_x^2}$$

Beispiel (Kontrolle): $s_x = \sqrt{1,079} = 1,039$

Variationskoeffizient

$$CV = \frac{s_x}{\bar{x}}$$

Der Variationskoeffizient ist nur dann ein sinnvolles Maß, wenn die Messwerte alle größer oder gleich Null sind. Es handelt sich um ein relatives Streuungsmaß, welches die Standardabweichung in Relation zum Mittelwert ausdrückt. Der Variationskoeffizient ist eine dimensionslose Maßzahl.

Beispiel (Kontrolle): $\bar{x} = 1,183$; $s_x = 1,039$; $CV = 1,039/1,183 = 0,8784 = 87,84\%$

Beispiel: Abschließend fassen wir alle Streuungsmaße für die drei Düngerbehandlungen zusammen.

Streuungsmaß	Kontrolle	NPK	DAP
Variationsbreite	4,20	8,80	5,37
Interquartilabstand	1,01	2,31	1,10
Varianz	1,079	3,995	1,588
Standardabweichung	1,039	1,999	1,260
Variationskoeffizient (%)	87,84	77,62	49,47
Arithmetisches Mittel	1,183	2,575	2,548

Es gibt bei allen Streuungsmaßen deutliche Unterschiede zwischen den Behandlungen. NPK hat eine höhere Streuung nach allen Maßzahlen, mit Ausnahme des Variationskoeffizienten. DAP hat mit Abstand den kleinsten Variationskoeffizienten. DAP und NPK haben etwa den gleichen mittleren Ertrag, aber DAP hat deutlich die geringere Streuung. Aus diesem Grund ist DAP sicher die zu bevorzugende Behandlung: Sie weist die größere Ertragsstabilität auf. Die höhere Streuung von DAP

relativ zu der Kontrolle wird mehr als ausgeglichen durch den höheren Mittelwert. Eine genauere Stabilitätsanalyse erfordert wahrscheinlichkeitstheoretische Aussagen, die später erörtert werden.

2.3 Box-Plot

Eine sehr sinnvolle Methode zur graphischen Darstellung der Verteilung einer Stichprobe sind sog. Box-and-Whiskers-Plots, oder kurz Box-Plot. Sie umfassen in der einfachsten Form vier Maßzahlen:

1. Arithmetisches Mittel
2. Median
3. Interquartilabstand (Box)
4. Variationsbreite (Whiskers)

Ab besten wird dies gleich an einem Beispiel erläutert.

Beispiel: Gegeben sei eine Stichprobe von $n = 16$ Maispflanzen, deren Länge (x) gemessen wurde:

154	157	162	163		163	164	165	167		172	175	175	176		177	179	186	194
					Q_{25}					Q_{50}					Q_{75}			

Wir finden:

$Q_0 = 154$	$Q_{75} = (176+177)/2 = 176,5$
$Q_{25} = (163+163)/2 = 163$	$Q_{100} = 194$
$Q_{50} = (167+172)/2 = 169,5$	$\bar{x} = 170,5625$

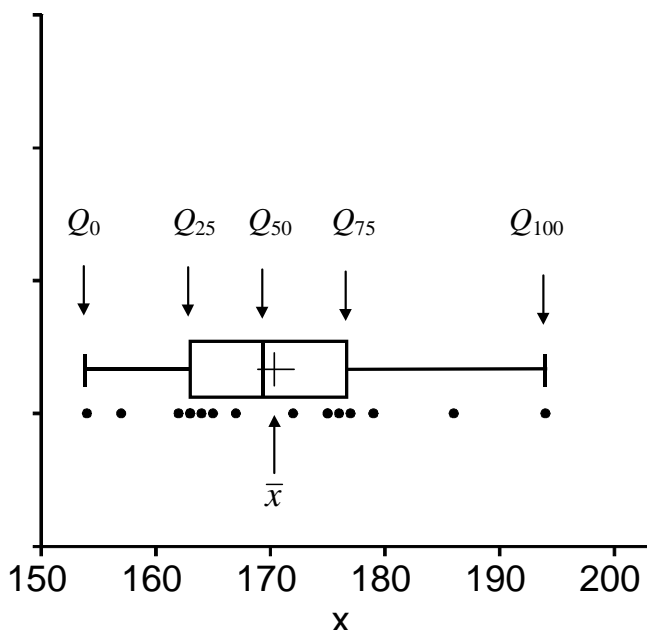
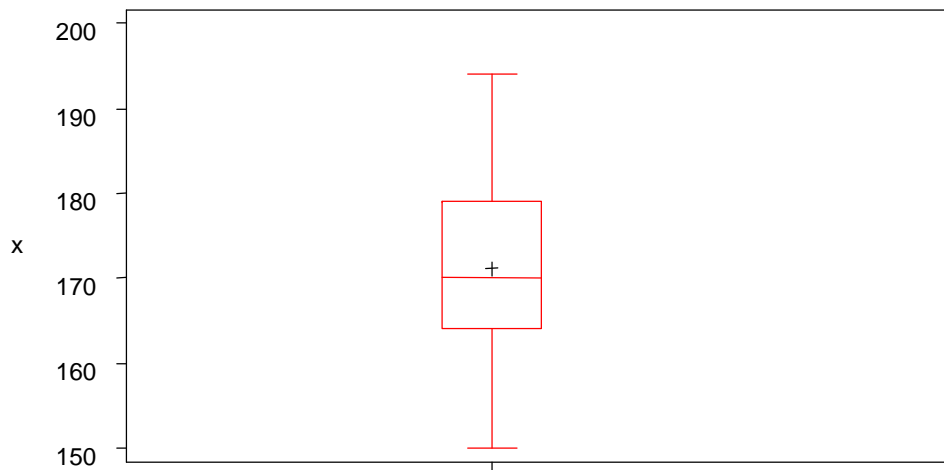


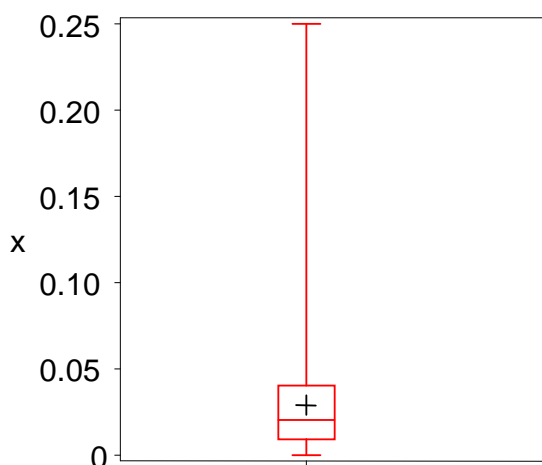
Abb. 2.3.1: Box-Plot für $n = 16$ Maispflanzen.

Das arithmetische Mittel wird durch ein Kreuz dargestellt. Um dieses Kreuz wird eine "Box" gezeichnet, die den Interquartilabstand wiedergibt. Durch einen Strich, der die Box in zwei Hälften teilt, wird der Median angezeigt. Der kleinste und der größte Wert werden durch einen von der Box ausgehenden Strich gekennzeichnet ("Whisker" = Barthaar). Median und arithmetisches Mittel liegen hier dicht beieinander, da die Verteilung der Daten annähernd symmetrisch ist.

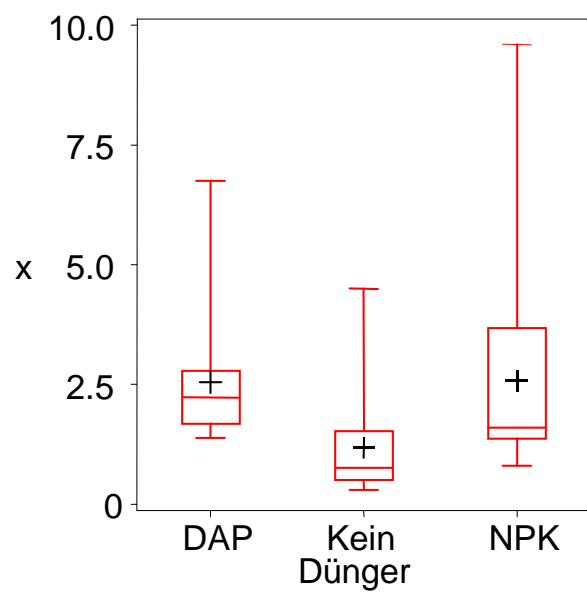
Beispiel: In Abschnitt 2.1 hatten wir einen Datensatz mit 50 Maispflanzenlängen vorgestellt. Der Box-Plot zeigt, dass die Verteilung der Daten annähernd symmetrisch ist.



Beispiel: Der Box-Plot der Kornabstände (x) zeigt eine deutlich schiefe Verteilung an (links). Die transformierten Kornabstände ($y = x^{0.28602}$) sind dagegen offensichtlich symmetrisch verteilt.



Beispiel: Der Box-Plot für die On-farm Ergebnisse zum Vergleich von drei Düngerformen zeigt deutlich, dass DAP etwa den gleichen mittleren Ertrag liefert wie NPK, aber eine kleinere Streuung aufweist. Zwar werden nicht die hohen Maximalerträge von NPK erreicht, dafür gibt es aber auch nicht ganz so niedrige Minimalerträge. Die Box-Plots bestätigen daher die Einschätzung, dass DAP eine bessere Stabilität aufweist als NPK. Deutlich wird ist außerdem, dass alle drei Verteilungen eine ausgeprägte Schiefe haben (arithmetisches Mittel \gg Median).



3. Einführung in die schließende Statistik für normalverteilte Daten (univariat)

3.1 Das Histogramm und Wahrscheinlichkeiten

Beispiel: In Abschnitt 2.1 hatten wir den Fall eines Maisfeldes mit 100.000 Maispflanzen betrachtet, für welches an einer Stichprobe von $n = 50$ Maispflanzen die Pflanzenlänge (x) bestimmt wurde. Hier nochmals die Daten:

175 172 179 167 163 154 163 164 157 177
 186 165 175 194 176 162 166 169 170 181
 168 166 180 164 179 170 150 192 170 173
 170 150 174 164 182 188 157 165 172 168
 179 179 164 162 178 162 182 171 182 183

Die Stichprobe hat einen Mittelwert von $\bar{x} = 171,2$ cm und eine Standardabweichung von $s_x = 10,0$ cm. Wir hatten mittels einer Strichliste die relativen Häufigkeiten in 5 Klassen bestimmt:

Höhe (cm)	Strichliste	Anzahl Pflanzen/ absolute Häufigkeit	Prozent/ relative Häufigkeit
145 – 154		3	6
155 – 164		11	22
165 – 174		17	34
175 – 184		15	30
185 – 194		4	8

Eine graphische Darstellung der Häufigkeiten ergab das folgende Histogramm:

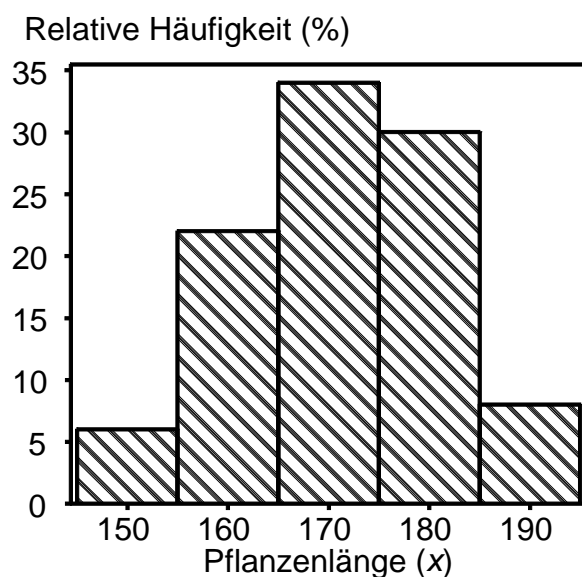


Abb. 3.1.1: Histogramm für die Maisdaten.

Mit Hilfe eines Histogramms lassen sich auch Wahrscheinlichkeitsaussagen treffen. Der erste Balken steht für die Klasse von Pflanzen mit einer Länge zwischen 145 cm und 154 cm. Der Anteil der Fläche dieser Klasse an der Gesamtfläche der Balken aller Klassen ist 6% ($= 3/50$). Dies entspricht der Wahrscheinlichkeit, dass eine Pflanze, die wir zufällig aus der Stichprobe von 50 Pflanzen ziehen, eine Länge zwischen 145 cm und 154 cm hat.

Man kann sich nun vorstellen, dass nicht eine Stichprobe von nur 50 Pflanzen untersucht wird, sondern alle 100.000 Pflanzen des Maisfeldes. Dann kann auch ein Histogramm erstellt werden, wobei allerdings wegen der großen Pflanzenzahl die Klassenbreiten sehr eng gewählt werden können (siehe Abschnitt 2.1). Die einzelnen Balken des Histogramms werden dann sehr eng, so eng, dass man einen einzelnen Balken nicht mehr graphisch darstellen kann. Statt dessen kann man noch die obere begrenzende Linie des Histogramms zeichnen. Dabei ergibt sich oft eine eingipfelige glockenartige Kurve, die einer Normalverteilung nahe kommen kann.

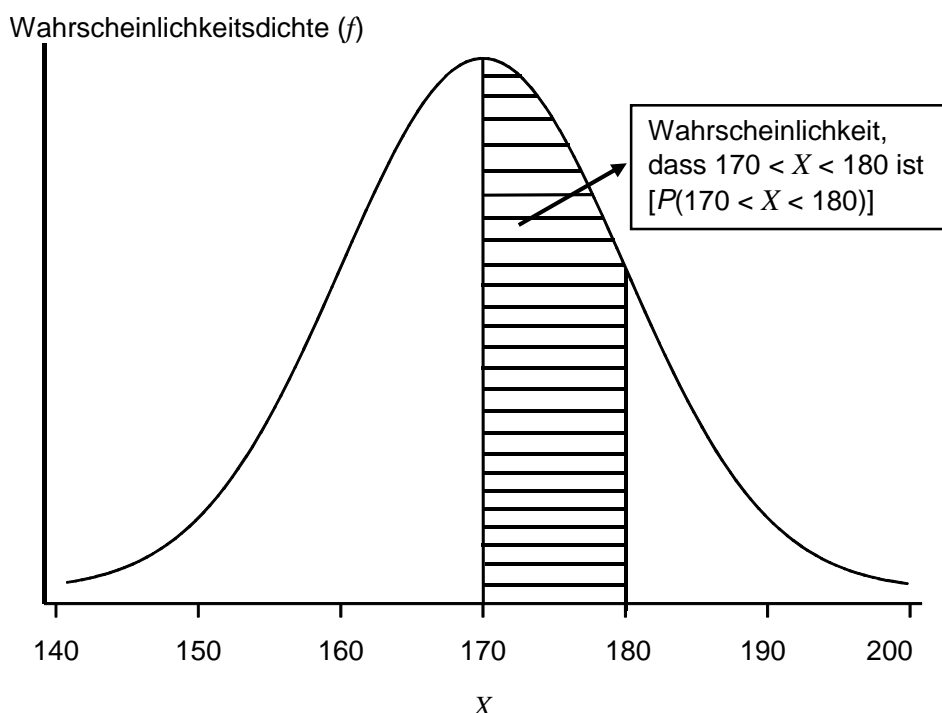


Abb. 3.1.2: Berechnung von Wahrscheinlichkeiten anhand einer Wahrscheinlichkeitsdichte.

Ein Beispiel einer solchen Kurve findet sich oben. Die Kurve kann so skaliert werden, dass die Gesamtfläche 1 beträgt. Dann kann die Kurve herangezogen werden, um Wahrscheinlichkeiten zu berechnen. Die schraffierte Fläche in Abb. 3.1.2 gibt z.B. die Wahrscheinlichkeit an, dass eine Pflanze der Feldes eine Länge zwischen 170 cm und 180 cm hat.

3.2 Normalverteilung

Oft kann eine Wahrscheinlichkeitsdichte wie sie eben für die Maisdaten beschrieben wurde, durch eine Normalverteilung angenähert werden. Eine Normalverteilung ist in der Abb. 3.2.1 wiedergegeben. Der Gipfel der Glockenkurve liegt genau über dem

Mittelwert μ . Die Kurve ist symmetrisch um diesen Mittelwert. Neben dem Mittelwert μ hat die Normalverteilung einen zweiten Parameter: die Varianz σ_x^2 . Je größer die Varianz, umso breiter ist die Kurve.

Die Dichte der Normalverteilung ist gegeben durch:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}}$$

wobei μ der Mittelwert (Erwartungswert) und σ_x^2 die Varianz ist. Eine Dichte wie die der Normalverteilung hat die Eigenschaft, dass die Fläche unter der Dichte immer gleich 1 ist:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Mittelwert μ und Varianz σ_x^2 der Normalverteilung können wie folgt definiert werden:

Der Mittelwert μ ist der Erwartungswert der Zufallsvariable X :

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = E(X)$$

Die Varianz σ_x^2 ist der Erwartungswert für die quadratische Abweichung der Zufallsvariable X von ihrem Erwartungswert μ , also der Erwartungswert von $(X-\mu)^2$:

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = E[(X - \mu)^2]$$

[Für diskrete Zufallsvariablen wie z.B. die Augenzahl eines Würfels, wird die Integration durch Summation ersetzt:

$$\mu = \sum_{k=1}^K x_k f(x_k) \quad \text{und} \quad \sigma_x^2 = \sum_{k=1}^K (x_k - \mu)^2 f(x_k),$$

wobei x_k ($k = 1, \dots, K$) die möglichen Realisationen der Zufallsvariablen sind und $f(x_k)$ die zugehörige Wahrscheinlichkeit.]

3.2.1 Berechnung von Wahrscheinlichkeiten

Mit Hilfe der Normalverteilung können einige Aussagen getroffen werden, die eine Vorstellung davon geben, warum die Standardabweichung σ_x (oder ihr Quadrat, die Varianz) ein aussagekräftiges Streuungsmaß ist.

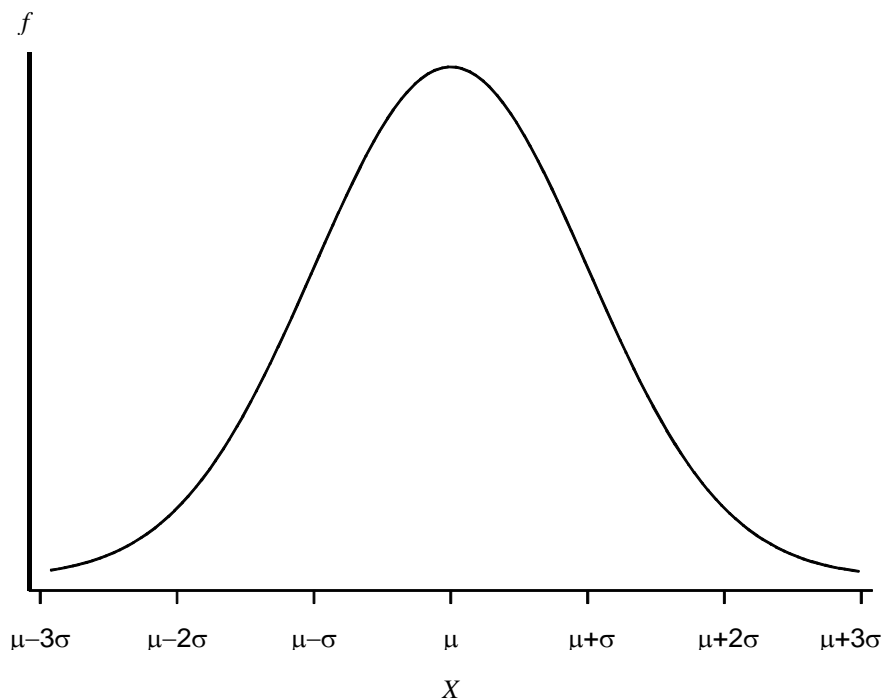


Abb. 3.2.1: Normalverteilung.

Die Aussagen sind wie folgt:

- (1) Etwa 68% der Fläche unter der Normalverteilungskurve liegen in den Grenzen von $\pm\sigma_x$ um den Mittelwert μ , also zwischen in den Grenzen $\mu - \sigma_x$ und $\mu + \sigma_x$.
- (2) Etwa 95% der Fläche liegen in Grenzen von $\pm 2\sigma_x$ um den Mittelwert μ (Genau 95% liegen innerhalb von $\pm 1,96\sigma_x$).
- (3) Etwa 99,7% der Fläche liegen in Grenzen von $\pm 3\sigma_x$ um den Mittelwert μ .

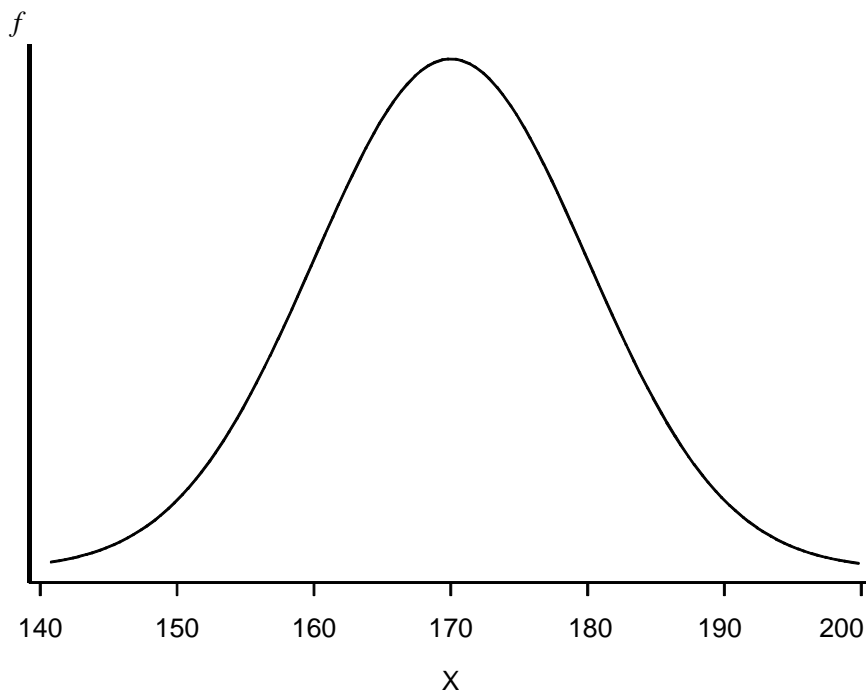


Abb. 3.2.2: Normalverteilung mit Mittelwert 170 und Standardabweichung 10.

Eine solche Verteilungskurve passt gut zu den Maisdaten, wie Abb. 3.2.3 zeigt:

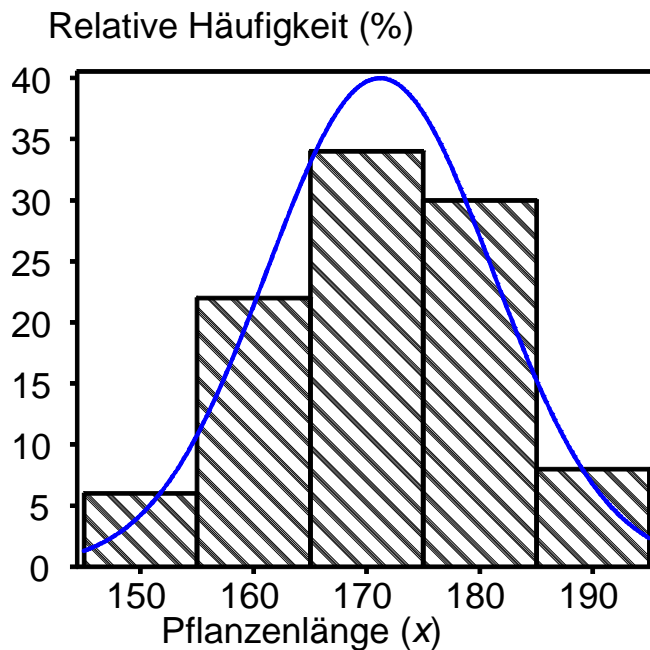


Abb. 3.2.3: Histogramm der Maisdaten, mit angepasster Normalverteilung.

Bezogen auf das Beispiel der Pflanzenlängen in einem Maisfeld mit Mittelwert $\mu = 170$ und Standardabweichung $\sigma_x = 10$ können folgende Aussagen getroffen werden:

- (1) Etwa 68% der Maispflanzen haben eine Länge zwischen $170 - 10$ und $170 + 10$ cm, also zwischen 160 und 180 cm.
- (2) Etwa 95% der Maispflanzen haben eine Länge zwischen $170 - 2 \cdot 10$ und $170 + 2 \cdot 10$ cm, also zwischen 150 und 190 cm.
- (3) Etwa 99,7% der Maispflanzen haben eine Länge zwischen $170 - 3 \cdot 10$ und $170 + 3 \cdot 10$ cm, also zwischen 140 und 200 cm.

Die erste Aussage ist gleichbedeutend mit folgender Aussage: Die Wahrscheinlichkeit, dass eine zufällig aus dem Feld gezogene Pflanze eine Länge zwischen 160 und 180 cm hat, beträgt etwa 68% (0,68). Analoge Aussagen lassen sich aus (2) und (3) ableiten. Anhand der Normalverteilung können auch Wahrscheinlichkeiten für andere Pflanzenlängen berechnet werden. Wir können beispielsweise fragen:

Frage 1: Wie groß ist die Wahrscheinlichkeit, dass eine zufällig aus dem Feld gezogene Pflanze $x = 185$ cm lang ist oder länger? Diese Frage lässt sich graphisch anhand der Normalverteilungskurve (Abb. 3.2.4) beantworten.

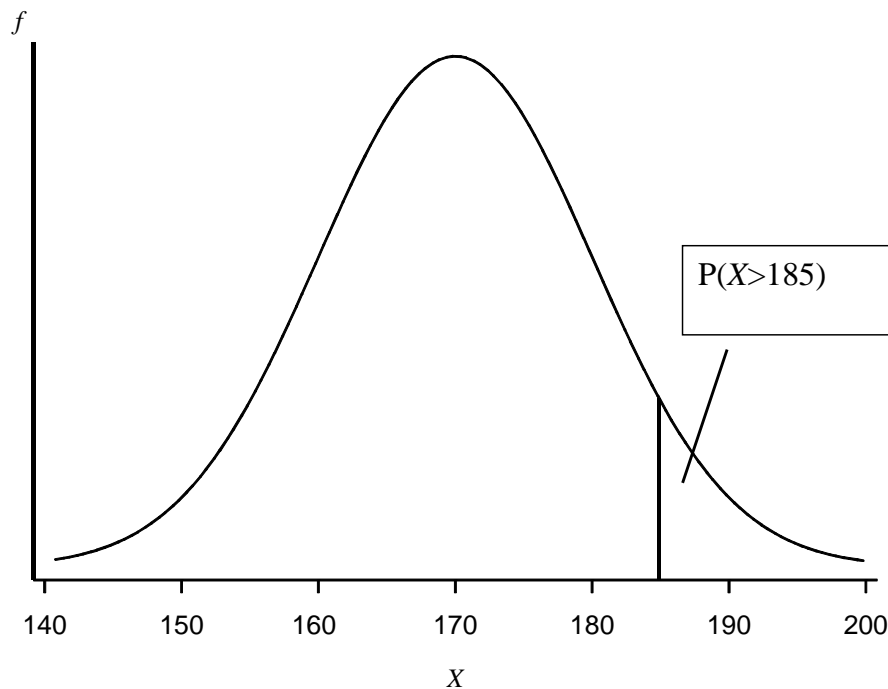


Abb. 3.2.4: Veranschaulichung der Wahrscheinlichkeit, dass eine zufällig bestimmte Pflanze eine Länge größer als 185 cm aufweist.

Die Fläche unter der Kurve rechts von $X = 185$ (Abb. 3.2.4) entspricht der Wahrscheinlichkeit, dass $X > 185$ ist [$P(X > 185)$].

Um diese Fläche zu berechnen, müssen wir das folgende Integral lösen:

$$P(X > 185) = \int_{185}^{\infty} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu)^2}{2\sigma_x^2}} dx = \int_{185}^{\infty} f(x) dx$$

$P(\cdot)$ bedeutet „Wahrscheinlichkeit“. Dieses Integral ist nicht analytisch zu berechnen. Es müssen numerische Integrationsverfahren verwendet werden. Die resultierenden Überschreitungswahrscheinlichkeiten können beispielsweise tabelliert werden, oder sie sind mit Computerroutinen zu berechnen.

Um die Berechnungen in der Praxis zu vereinfachen, ist es sinnvoll, die obige Verteilung in eine Standardnormalverteilung umzuwandeln. Die Standardnormalverteilung ist eine Normalverteilung mit Erwartungswert $\mu = 0$ (theoretischer Mittelwert) und Varianz $\sigma_x^2 = 1$. Eine Umformung zur Standardnormalverteilung ist deswegen notwendig, weil nur diese Verteilung tabelliert ist. Ohne diese Umformung müsste man für jeden Erwartungswert und jede Varianz neue Berechnungen von Integralen durchführen. Dieser Aufwand lässt sich umgehen.

Um zu einer Standardnormalverteilung zu gelangen, müssen wir eine sog. z -Transformation durchführen. Hierzu berechnen wir

$$Z = (X - \mu) / \sigma_x$$

Man kann sich die z-Transformation Vorstellen als eine Verschiebung und Stauchung bzw. Streckung einer beliebigen Normalverteilung, so dass daraus eine Standardnormalverteilung mit Mittelwert = 0 und Varianz = 1 (Standardabweichung = 1) wird. Dies wird anhand der folgenden drei Abbildungen für eine Normalverteilung mit Mittelwert = 3 und Standardabweichung = 1,8 veranschaulicht.

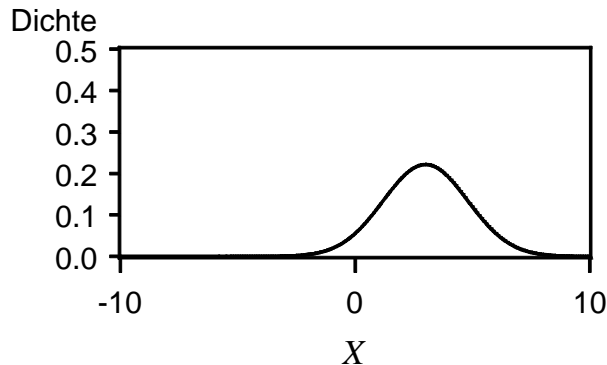


Abb 3.2.5: Wahrscheinlichkeitsdichte der Normalverteilung für Zufallsvariable X mit Mittelwert $\mu = 3$ und Standardabweichung $\sigma_X = 1,8$.

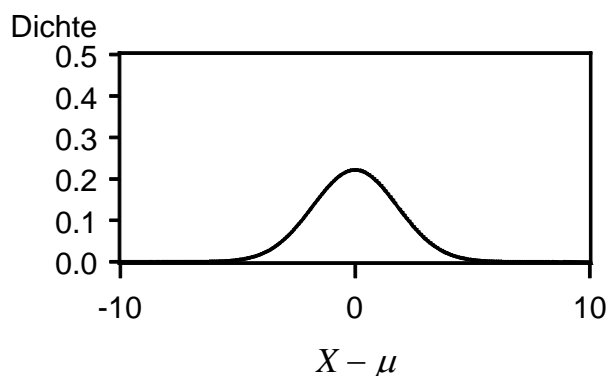


Abb 3.2.6: Wahrscheinlichkeitsdichte der Normalverteilung für Zufallsvariable $(X - \mu)$, wobei X normalverteilt ist mit Mittelwert $\mu = 3$ und Standardabweichung $\sigma_X = 1,8$.

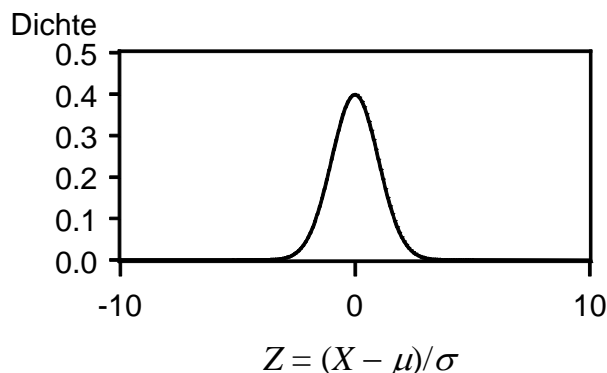


Abb 3.2.7: Wahrscheinlichkeitsdichte der Normalverteilung für Zufallsvariable $Z = (X - \mu)/\sigma_X$, wobei X normalverteilt ist mit Mittelwert $\mu = 3$ und Standardabweichung $\sigma_X = 1,8$. Z hat eine Standardnormalverteilung.

In unserem Fall ist $x = 185$ cm und somit

$$z = (x - \mu) / \sigma_x = (185 - 170) / 10 = 1,5$$

Dies bedeutet, x liegt 1,5 Standardabweichungen über μ . (Bemerkung zur Schreibweise: X groß geschrieben bezeichnet eine Zufallsvariable, x kleingeschrieben ist eine beobachtete **Realisation** dieser Zufallsvariable X).

Nachfolgend eine Abbildung der Standardnormalverteilung (Abb. 3.2.8)

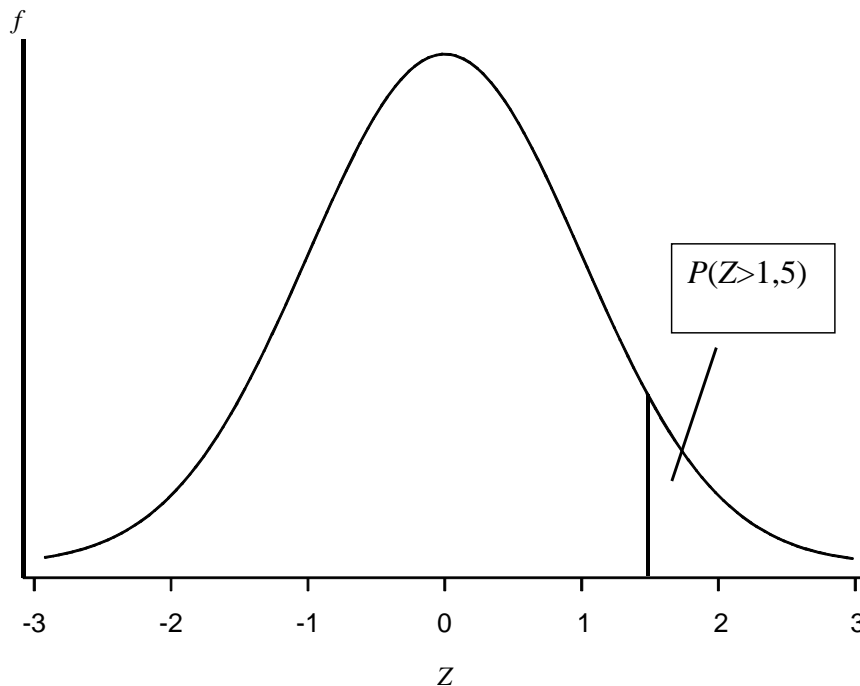


Abb. 3.2.8: Standardnormalverteilung.

Die Standardnormalverteilung hat folgende Dichtefunktion:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Die Fläche unter der Standardnormalverteilung rechts von $z = 1,5$ ist unsere gewünschte Wahrscheinlichkeit. Man spricht bei dieser Fläche auch von **Überschreitungswahrscheinlichkeit**. Denn die Fläche rechts von z gibt an, wie groß die Wahrscheinlichkeit ist, dass eine standardnormalverteilte Zufallsvariable den Wert z überschreitet. Sie ist für verschiedene Werte von z in Tab. I tabelliert (Anhang). Für $z = 1,5$ finden wir beispielsweise, dass die Fläche rechts von $z = 1,5$ gleich 0,0668 ist. Formal schreibt man auch

$$P(Z > 1,5) = \int_{1,5}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{1,5}^{\infty} f(z) dz = 0,0668$$

Dies können wir zurückbeziehen auf die ursprüngliche Frage: Es kann gesagt werden, dass die Wahrscheinlichkeit, eine Maispflanze mit einer Länge von größer oder gleich 185 cm zu beobachten, gleich 0,0668 oder 6,68% beträgt. Formal kann man auch schreiben:

$$P(X > 185) = P(Z > 1,5) = 0,0668$$

Man beachte auch, dass die Normalverteilungen für X und Z dieselbe Form haben.

Wichtiger Hinweis: Wir verwenden hier Überschreitungswahrscheinlichkeiten $P(Z > z)$. In vielen Lehrbüchern findet man dagegen sog. Unterschreitungswahrscheinlichkeiten $P(Z < z) = 1 - P(Z > z)$.

Der Grund für die z -Transformation ist, wie angedeutet, dass es nicht praktikabel ist, für alle möglichen Werte von Mittelwert μ und Varianz σ_x^2 die Überschreitungswahrscheinlichkeiten zu tabellieren. Praktischer ist es, eine gegebene Normalverteilung mit bestimmten Mittelwert μ und bestimmter Varianz σ_x^2 in eine Standardnormalverteilung umzuwandeln. Dann benötigt man lediglich eine einzige Tabelle. Die statistische Rechtfertigung für die Transformation ist noch einmal im folgenden Kasten zusammengefaßt:

Wenn X normalverteilt ist mit Mittelwert μ und Varianz σ_x^2 , dann ist

$$Z = (X - \mu) / \sigma_x$$

standardnormalverteilt (Mittelwert 0 und Varianz 1). Z ist eine standardnormalverteilte Zufallsvariable.

Überschreitungswahrscheinlichkeiten spielen eine große Rolle in der Statistik, z.B. bei der Durchführung von statistischen Tests mit Hilfe von Computerprogrammen (siehe Abschnitt 3.15). Der Umgang mit Überschreitungswahrscheinlichkeiten soll daher an einigen weiteren Beispielen verdeutlicht werden. Wir bleiben bei den Pflanzenlängen unseres Maisfeldes mit Mittelwert $\mu = 170$ und $\sigma_x = 10$.

Frage 2: Wie groß ist die Wahrscheinlichkeit, dass eine zufällig aus dem Feld gezogene Pflanze eine Länge zwischen 175 und 185 cm hat? Hierzu müssen wir die Fläche unter der Normalverteilung zwischen 175 und 185 cm berechnen. Da nur Überschreitungswahrscheinlichkeiten aus der Tabelle abgelesen werden können, d.h. Flächen rechts von einem bestimmten Wert, kann diese Wahrscheinlichkeit nicht direkt aus Tabellen abgelesen werden. Es ist ein Zwischenschritt nötig. Etwas Überlegung zeigt, dass gilt:

Fläche zwischen 175 und 185 = Fläche rechts von 175 - Fläche rechts von 185

Also benötigen wir die Flächen rechts von 175 und rechts von 185, um die gewünschte Wahrscheinlichkeit berechnen zu können. Zunächst müssen die z -Transformationen berechnet werden. Wir finden

$$z_1 = (185 - 170) / 10 = 1,5 \quad \text{und} \quad z_2 = (175 - 170) / 10 = 0,5$$

Aus Tab. I finden wir die Flächen rechts von z_1 und z_2 :

$$P(Z > 1,5) = 0,0668 \quad \text{und} \quad P(Z > 0,5) = 0,3085$$

Also ist die Fläche zwischen 175 und 185 gleich $0,3085 - 0,0668 = 0,2417$. Damit ist die Antwort auf unsere Frage: Die Wahrscheinlichkeit, dass eine zufällig gezogene Maispflanze eine Länge zwischen 175 und 185 cm hat, beträgt 0,2417. Wir können außerdem sagen, dass 24,17% der Maispflanzen in unserem Feld eine Länge zwischen 175 und 185 cm haben.

Frage 3: Wie groß ist die Wahrscheinlichkeit, dass eine Maispflanze kleiner als 145 cm ist? Die Antwort ist, dass diese Wahrscheinlichkeit der Fläche unter der Kurve links von $x = 145$ cm entspricht. Tabelliert sind aber nur Flächen rechts einer Schwelle. Diese muss außerdem rechts vom Mittelwert μ liegen. Wegen des symmetrischen Aufbaus der Normalverteilung kann aber trotzdem die gewünschte Wahrscheinlichkeit berechnet werden, wie wir gleich sehen. Zunächst berechnen wir die z -Transformation.

$$z = (145 - 170)/10 = -2,5 \quad ,$$

d.h., dass x mehr als 2,5 Standardabweichungen unterhalb von μ liegt. Wir suchen demnach die Fläche links von $z = -2,5$. Wegen der Symmetrie der Normalverteilung ist dies aber gleich der Fläche rechts von $z = +2,5$. Diese Fläche ist 0,0062 (Tab. I). Daher schließen wir, dass 0,62% der Pflanzen eine Länge kleiner als 145 cm haben. Die Wahrscheinlichkeit einer Pflanze kleiner 145 cm ist 0,0062.

Frage 4: Wie groß ist die Wahrscheinlichkeit, eine Pflanze zwischen 145 cm und 175 cm zu ziehen? Wir können diese Wahrscheinlichkeit auch formal schreiben:

$$P(145 \leq X \leq 175)$$

Die Gesamtfläche unter der Normalverteilung ist 1. Aus den vorhergehenden Antworten ergibt sich, dass die gesuchte Wahrscheinlichkeit

$$1 - 0,0062 - 0,3085 = 0,6853$$

beträgt.

Frage 5: Wie groß ist die Wahrscheinlichkeit, eine Pflanze zu beobachten, deren Länge 170 cm oder mehr beträgt? Da 170 der Mittelwert der Verteilung der Maispflanzen ist und die Verteilung symmetrisch um diesen Mittelwert ist, beträgt die Wahrscheinlichkeit 0,5. Wir könnten auch die z -Transformation berechnen:

$$z = (170 - 170)/10 = 0$$

Hierzu finden wir in einer Tabelle der Normalverteilung die Überschreitungswahrscheinlichkeit von 0,5.

Zur Berechnung der hier betrachteten Wahrscheinlichkeiten sind folgende Regeln hilfreich:

$P(X < x)$ und $z > 0 \Rightarrow$ berechne $1 - P(Z > z)$

$P(X < x)$ und $z < 0 \Rightarrow$ berechne $P(Z > -z)$

$P(X > x)$ und $z > 0 \Rightarrow$ berechne $P(Z > z)$

$P(X > x)$ und $z < 0 \Rightarrow$ berechne $1 - P(Z > -z)$

$P(x_1 < X < x_2) = 1 - P(X < x_1) - P(X > x_2)$

Außerdem ist folgendes zu beachten: Bei einer stetigen Verteilung kann man nicht die Wahrscheinlichkeit für einen bestimmten Wert angeben. Genauer gesagt beträgt diese Wahrscheinlichkeit Null: $P(X = x) = 0$ für jedes x . Es ist lediglich möglich, die Wahrscheinlichkeit anzugeben, daß X in ein bestimmtes Intervall fällt. Aus demselben Grund gilt für stetige Verteilungen: $P(X < x) = P(X \leq x)$ und $P(X > x) = P(X \geq x)$.

3.2.2 Quantile der Normalverteilung und Q-Q-Plots

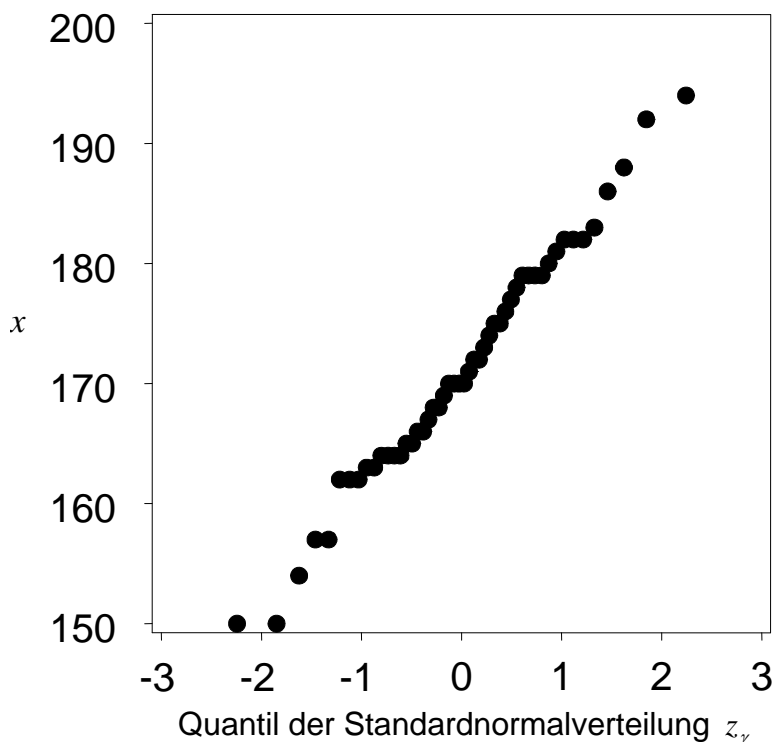


Abb. 3.2.9: Q-Q-Plot der Maisdaten aus Abschnitt 3.1. x = Pflanzenlänge.

Mit Hilfe der Überschreitungswahrscheinlichkeit lassen sich Quantile der Normalverteilung bestimmen. Im Falle der Standardnormalverteilung bestimmen wir für eine vorgegebene Wahrscheinlichkeit γ eine kritische Schwelle z_γ , so dass gilt

$$P(Z < z_\gamma) = \gamma.$$

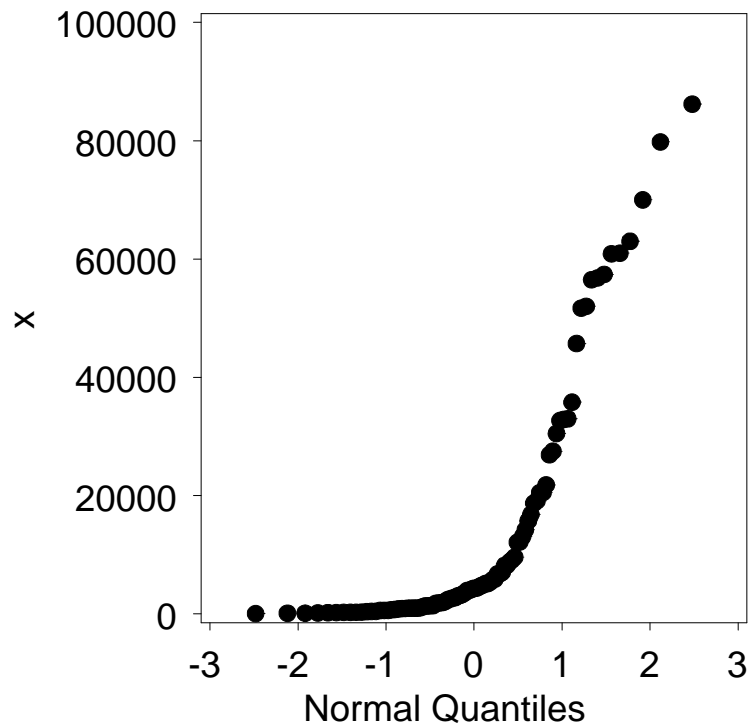


Abb. 3.2.10: Q-Q-Plot für Keimzahlen in der Milch aus Abschnitt 2.2.2.

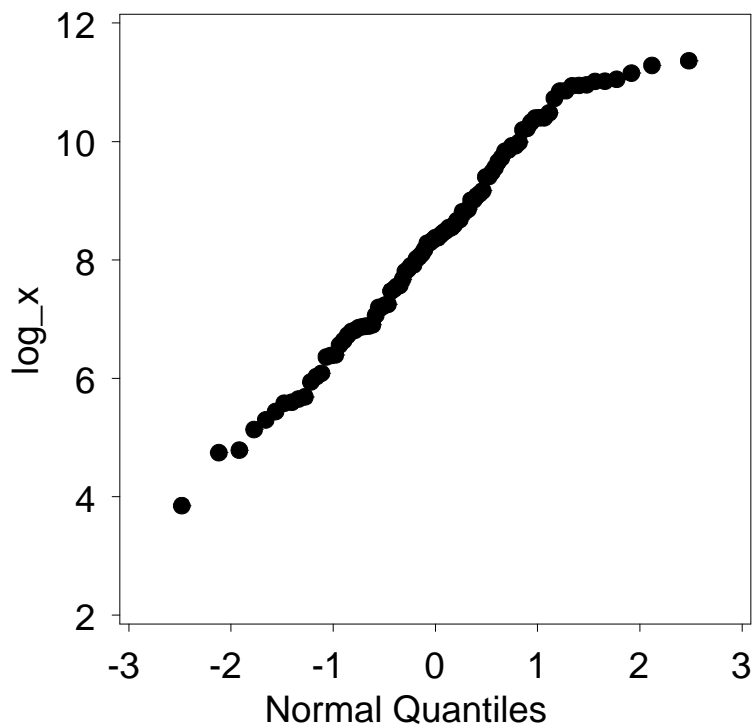


Abb. 3.2.11: Q-Q-Plot für logarithmierte Keimzahlen in der Milch aus Abschnitt 2.2.2.

Diese Schwelle ist das $\gamma_{100\%}$ -Quantil der Standardnormalverteilung, denn $\gamma_{100\%}$ der Dichte der Standardnormalverteilung liegen unterhalb von z_γ (vgl. Abschnitt 2.2.1).

Für eine gegebene Stichprobe kann man ebenfalls Quantile bestimmen (Abschnitt 2.2.1). Man bestimmt, welches Quantil Q_t die Beobachtungswerte der Stichprobe repräsentieren. Für die sich n ergebenden Quantilwahrscheinlichkeiten t bestimmt man dann die entsprechenden Quantile der Normalverteilung mit $\gamma = t/100$. Die Quantile der Stichprobe werden dann gegen die entsprechenden Quantile z_γ der Standardnormalverteilung abgetragen. Solch eine Abbildung heisst **Quantil-Quantil-Plot** oder **Q-Q-Plot**. Q-Q-Plots spielen eine große Rolle bei der Überprüfung der Normalverteilungsannahme in vielen statistischen Auswertungen.

Wenn die Stichprobe etwa einer Normalverteilung folgt, so liegen die Punkte etwa auf einer Geraden. Für das Beispiel der Maisdaten aus Abschnitt 3.1 ist dies in Abb. 3.2.9 gezeigt. Andernfalls zeigt sich eine deutliche Abweichung von der Linearität, und wir können schließen, dass die Annahme einer Normalverteilung nicht zutreffend ist. In Abb. 3.2.10 ist dies für Keimzahlen aus der Milch (Abschnitt 2.2.2) gezeigt. Eine logarithmische Transformation schafft etwas Abhilfe (Abb. 3.2.11).

3.3 Stichprobe, Grundgesamtheit

Im Rahmen der schließenden Statistik sind die folgenden zwei Begriffe von Bedeutung: **Stichprobe** und **Grundgesamtheit**. Um diese zu erklären, werden hier zunächst zwei Beispiele betrachtet.

Beispiel: Vor einer Bundestagswahl wird eine politische Meinungsumfrage durchgeführt. Man möchte das Wahlverhalten der wahlberechtigten Bürger untersuchen. Die Grundgesamtheit, über die hier eine Aussage getroffen werden soll, ist die Menge aller wahlberechtigten Bürger der Bundesrepublik Deutschland. Im Prinzip ist eine Totalerfassung möglich, aber diese würde die eigentliche Wahl vorwegnehmen, und der Aufwand wäre unverhältnismäßig hoch (kein Auftraggeber würde die Kosten bezahlen). Aus diesem Grund wird in der Regel nur eine Zufallsstichprobe von Personen befragt. Von dieser Stichprobe wird dann auf die Grundgesamtheit zurückgeschlossen.

Beispiel: In einem Maisfeld, welches 100.000 Pflanzen aufweist, soll die durchschnittliche Pflanzenlänge bestimmt werden. Da eine Messung aller 100.000 Pflanzen unpraktikabel ist, wird eine Zufallsstichprobe von 50 Pflanzen untersucht. Aus der durchschnittlichen Pflanzenlänge in der Stichprobe schließen wir auf die Grundgesamtheit von 100.000 Maispflanzen zurück.

Allgemein:

Die **Grundgesamtheit** ist eine genau begrenzte und eindeutig definierte Menge aller Objekte, über die eine Untersuchung Auskunft geben soll. Eine Untersuchung (Versuch, Erhebung) bleibt oft auf einen Teil der Grundgesamtheit, die **Stichprobe**, beschränkt. Wichtig ist, dass die Stichprobe zufällig ausgewählt wird.

Bis jetzt haben wir Grundgesamtheiten betrachtet, die endlich und konkret sind. Man kann sie "sehen und anfassen". In der Statistik wird der Begriff der Grundgesamtheit aber auch in etwas allgemeinerer Form gebraucht. Nehmen wir das Beispiel eines Versuches, bei dem der Ertrag einer Weizensorte auf 13 Parzellen getestet wird. Hierbei ist die (hypothetische) Grundgesamtheit aufzufassen als die Menge aller unter den gleichen Bedingungen eines Versuches denkbaren Ergebnisse. Wird ein Versuch auf n Parzellen durchgeführt, so stellt dieser Versuch eine Stichprobe vom Umfang n aus dieser hypothetischen Grundgesamtheit dar. Auf diese etwas abstrakte Vorstellung werden wir später zurückkommen.

3.4 Zentraler Grenzwertsatz

Die Summe bzw. der Mittelwert von vielen unabhängig verteilten Zufallsvariablen ist angenähert normalverteilt. Dies gilt relativ unabhängig von der Verteilung der Einzelwerte.

Dies ist eine vereinfachte umgangssprachliche Formulierung des **zentralen Grenzwertsatzes**.

Der zentrale Grenzwertsatz ist von großer Bedeutung in der Statistik.

Beachte: Viele Zufallsvariablen sind höchstens näherungsweise normalverteilt. Der Ertrag kann z.B. nicht kleiner als 0 sein, obwohl eine Normalverteilung für Werte kleiner 0 immer eine gewisse Wahrscheinlichkeit lässt. Daher kann der Ertrag nie exakt normalverteilt sein! Aber: Die Summe sowie der Mittelwert von n Einzelwerten sind oft in guter Näherung normalverteilt. Dies ist ein Grund, warum die Normalverteilung eine so große Rolle in der Statistik spielt.

Beispiel: Maisfeld mit 100.000 Einzelpflanzen. Es wird eine Stichprobe von 50 Pflanzen gezogen und die Länge gemessen. Der Mittelwert der Stichprobe ist angenähert normalverteilt, selbst dann, wenn die Einzelpflanzenlängen nicht normalverteilt sind!

3.5 Vertrauensintervall für einen Mittelwert

Hier wird das zuletzt genannte Beispiel aufgenommen: Wir haben ein Maisfeld mit 100.000 Einzelpflanzen und möchten wissen, wie groß die durchschnittliche Länge der Pflanzen dieses Feldes ist. Die Grundgesamtheit ist hier endlich und real beobachtbar. Der Durchschnitt aller Pflanzenlängen kann als wahrer Mittelwert μ bezeichnet werden. Im Prinzip könnten alle 100.000 Pflanzen vermessen werden, so dass der wahre Wert ermittelt werden kann. Aber dies ist nicht praktikabel. Daher wird man die Pflanzenlänge nur an einer Stichprobe messen. Aus der Stichprobe berechnen wir einen Stichprobenmittelwert \bar{x} . Dieser Stichprobenwert \bar{x} ist ein **Schätzwert** für den wahren Wert μ der Grundgesamtheit. Es wird selten bzw. nie der Fall sein, dass unser Schätzwert den wahren Wert genau trifft. Er wird mehr oder weniger vom wahren Wert abweichen. Theoretisch könnten wir die Stichprobennahme auch wiederholen und einen weiteren Mittelwert \bar{x} berechnen. Dieser wird in der Regel nicht

mit dem ersten Mittelwert und auch nicht mit dem wahren Wert μ übereinstimmen. Wie kann nun angegeben werden, wie nah der Schätzwert am wahren Wert liegt?

Eine mögliche Lösung besteht in der Berechnung eines sog. Vertrauensintervalls. Ein Vertrauensintervall gibt Grenzen an, die den wahren Wert μ mit einer vorgegebenen Wahrscheinlichkeit einschließen. Um ein solches Intervall zu berechnen, kann man sich folgende Tatsachen zu nutze machen:

1. In der Grundgesamtheit schwanken die Einzelwerte um den wahren Mittelwert μ . Sie weisen eine Varianz σ_x^2 auf.
2. Wenn die Stichprobennahme oft wiederholt wird, schwankt der Stichprobenmittelwert \bar{x} um den gleichen Wert μ wie die Einzelwerte der Stichprobe.
3. Die Varianz des Mittelwertes steht in direkter Beziehung zur Varianz der Einzelwerte σ_x^2 , und sie ist umgekehrt proportional zum Umfang der Stichprobe, aus welcher der Mittelwert berechnet wird. Es gilt:

$$\sigma_{\bar{x}}^2 = \sigma_x^2 / n$$

wobei n der Stichprobenumfang ist und $\sigma_{\bar{x}}^2$ die Varianz des Mittelwertes. Die Quadratwurzel aus der Varianz des Mittelwertes ist der sog. **Standardfehler**:

$$\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$$

4. Aufgrund des zentralen Grenzwertsatzes ist der Mittelwert angenähert normalverteilt, selbst dann, wenn die Einzelwerte der Grundgesamtheit nicht normalverteilt sind.

Zusammenfassend können wir sagen, dass der Mittelwert \bar{x} näherungsweise normalverteilt ist mit Erwartungswert μ und Varianz $\sigma_{\bar{x}}^2 = \sigma_x^2 / n$. Diese Tatsache kann ausgenutzt werden zur Berechnung eines Vertrauensintervalls.

Für die Normalverteilung gilt, dass 95% der Werte innerhalb der Grenzen von $\pm 1,96$ mal Standardabweichung um den Mittelwert μ liegen. Dies gilt auch für die Verteilung des Stichprobenmittelwertes. Da die Standardabweichung des Mittelwertes \bar{x} (der Standardfehler) gleich der Wurzel aus σ_x^2/n , also σ_x/\sqrt{n} ist, liegen 95% der Werte zwischen $\mu - 1,96\sigma_x/\sqrt{n}$ und $\mu + 1,96\sigma_x/\sqrt{n}$. Man kann auch schreiben

$$P(\mu - 1,96\sigma_x/\sqrt{n} \leq \bar{x} \leq \mu + 1,96\sigma_x/\sqrt{n}) = 0,95$$

$P(.)$ bedeutet Wahrscheinlichkeit. Die obige Gleichung besagt auch folgendes: Mit 95%iger Wahrscheinlichkeit liegt ein einzelner Stichprobenmittelwert innerhalb der Grenzen $\mu - 1,96\sigma_x/\sqrt{n}$ und $\mu + 1,96\sigma_x/\sqrt{n}$. Durch leichte Umformung erhält man aus obiger Ungleichung:

$$P(\bar{x} - 1,96\sigma_x/\sqrt{n} \leq \mu \leq \bar{x} + 1,96\sigma_x/\sqrt{n}) = 0,95$$

Diese Gleichung liefert ein Vertrauensintervall für μ , welches aus den Daten berechnet werden kann. Nach der obigen Gleichung sind die untere und obere Grenze dieses Intervalls gegeben durch $\bar{x} - 1,96\sigma_x/\sqrt{n}$ und $\bar{x} + 1,96\sigma_x/\sqrt{n}$. Mit 95% Wahrscheinlichkeit überdeckt dieses Intervall den wahren Wert μ .

In der Praxis ist σ_x nicht bekannt und muss durch die Standardabweichung s_x der gleichen Stichprobe ersetzt werden, an der auch der Mittelwert ermittelt wurde. Wenn die Stichprobe groß ist, ändert sich hierdurch nichts an der obigen Wahrscheinlichkeitsaussage. Wir können dann schreiben:

$$P(\bar{x} - 1,96s_x/\sqrt{n} \leq \mu \leq \bar{x} + 1,96s_x/\sqrt{n}) = 0,95$$

Diese Formel erlaubt folgende Aussage:

Wenn wir einen Stichprobenmittelwert \bar{x} und eine Standardabweichung s_x berechnen, so überdeckt das aus diesen Werten berechnete Vertrauensintervall mit Grenzen $\bar{x} - 1,96s_x/\sqrt{n}$ und $\bar{x} + 1,96s_x/\sqrt{n}$ den wahren Wert μ mit einer Vertrauenswahrscheinlichkeit von näherungsweise 95%.

Beispiel: An einer Stichprobe von $n = 50$ Maispflanzen eines Maisfeldes finden wir einen Mittelwert von $\bar{x} = 171,2$ und eine Standardabweichung von $s_x = 10$ cm. Dann ist ein 95%-Vertrauensintervall für die durchschnittliche Pflanzenhöhe in der Grundgesamtheit (Maisfeld) gegeben durch die Grenzen

$$171,2 - 1,96 * \frac{10}{\sqrt{50}} \quad \text{bis} \quad 171,2 + 1,96 * \frac{10}{\sqrt{50}}$$

$$168,4 \qquad \qquad \text{bis} \qquad \qquad 174,0$$

Wir können sagen, dass das Intervall von 168,4 bis 174,0 mit 95%er Wahrscheinlichkeit den wahren Mittelwert μ enthält.

Standardabweichung und Standardfehler: Wir haben als Streuungsmaß für den Mittelwert den Standardfehler eingeführt. Dieser ist klar zu trennen vom Begriff der Standardabweichung. Diese bezieht sich immer auf die Streuung einzelner Messwerte. Interessant ist die Tatsache, dass beide Größen in einer direkten Beziehung zueinander stehen nach

$$\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$$

wobei $\sigma_{\bar{x}}$ der Standardfehler und σ_x die Standardabweichung ist.

3.6 Vertrauensintervall für einen Mittelwert bei kleinen Stichproben

In Abschnitt 2.5 wurde bei der Berechnung eines 95% Vertrauensintervalls der Form $\mu \pm 1,96\sigma_x/\sqrt{n}$ die Standardabweichung σ_x durch einen Schätzwert s_x ersetzt. Dabei wurde so verfahren, also ob der Schätzwert genauso gut ist wie der wahre Wert. Für

große Stichproben ($n > 30$) ist dieses Vorgehen akzeptabel, bei kleineren Stichproben muss dagegen die Ungenauigkeit der Schätzung von σ_x durch s_x berücksichtigt werden. Dies geschieht dadurch, dass der Faktor 1,96 durch einen größeren Wert (t) ersetzt wird, der vom Stichprobenumfang abhängt. Allgemein können wir das Vertrauensintervall wie folgt schreiben:

$$\bar{x} - t s_x / \sqrt{n} \text{ bis } \bar{x} + t s_x / \sqrt{n}$$

Die Konstante wird mit t bezeichnet, weil sie mit der t -Verteilung in Verbindung steht. Die t -Verteilung ist eine Glockenkurve wie die Normalverteilung. Im Vergleich zur Normalverteilung verläuft die t -Verteilung jedoch etwas flacher, und zwar umso flacher, je geringer der Stichprobenumfang (n) ist, bzw. die Freiheitsgrade ($n - 1$). t -Werte können aus einer Tabelle der t -Verteilung abgelesen werden. Hierzu müssen zwei Parameter vorher bestimmt werden:

1. Die Zahl der Freiheitsgrade: $FG = (n - 1)$.
2. Der Wert für den α -Fehler. Dieser hängt ab von der gewünschten Vertrauenswahrscheinlichkeit. Bei einer Vertrauenswahrscheinlichkeit von 95% ist beispielsweise $\alpha = 1 - 0,95 = 0,05$.

Tab. II (siehe Anhang) enthält kritische t -Werte für verschiedene Irrtumswahrscheinlichkeiten α (zweiseitig). Angenommen, für die Maisdaten sind der Mittelwert $\bar{x} = 171,2$ und die Standardabweichung von $s_x = 10$ cm aus einer Stichprobe von nur 10 Pflanzen berechnet worden. Dann ist die Zahl der Freiheitsgrade gleich $(10 - 1) = 9$. Für ein 95% Intervall ist $\alpha = 1 - 0,95 = 0,05$. In einer t -Tabelle („zweiseitig“) finden wir einen t -Wert von $t = 2,262$. Hiermit ergeben sich folgende Vertrauensgrenzen:

$$171,2 - 2,262 * \frac{10}{\sqrt{10}} \quad \text{bis} \quad 171,2 + 2,262 * \frac{10}{\sqrt{10}}$$

$$164,0 \qquad \qquad \text{bis} \qquad \qquad 178,4$$

Bei einem Stichprobenumfang von $n = 10$ sind wir somit zu 95% sicher, dass der gesuchte Wert μ zwischen 164,0 und 178,4 liegt. Die Breite des Vertrauensintervalls ist jetzt deutlich größer als das oben für den Stichprobenumfang von $n = 50$ berechnete. Hierfür gibt es zwei Gründe: den geringeren Stichprobenumfang sowie die Verwendung der t -Verteilung anstelle der Normalverteilung.

Wenn n sehr groß wird, so nähert sich der tabellierte t -Wert immer näher dem entsprechenden Wert für die Normalverteilung. Für eine Vertrauenswahrscheinlichkeit von 95% finden wir beispielsweise einen Wert von $t = 1,96$ für große n . Man mag sich fragen, wie groß der Stichprobenumfang sein muss, damit man den Wert 1,96 verwenden kann. Diese Frage kann nicht eindeutig beantwortet werden. Es kann nur gesagt werden, dass der richtige Wert um so näher an 1,96 liegt, je größer die Stichprobe ist. Bei $n = 30$ liegt der kritische t -Wert beispielsweise ca. 10% über den der Normalverteilung. In der Praxis ist es am sichersten, unabhängig vom Stichprobenumfang immer mit der t -Verteilung anstelle der Normalverteilung zu

arbeiten. Die Schritte für die Berechnung eines Vertrauensintervalls sind im folgenden Kasten noch einmal zusammengefasst:

Fragestellung: Welches Intervall enthält den wahren Mittelwert μ mit der Sicherheitswahrscheinlichkeit $(1 - \alpha)$?

Voraussetzung: Die Grundgesamtheit ist zufallsverteilt mit unbekanntem Mittelwert μ und unbekannter Varianz σ_x^2 . Der Mittelwert einer Stichprobe vom Umfang n aus der Grundgesamtheit ist angenähert normalverteilt.

Rechenweg:

- (1) Berechne den Mittelwert \bar{x} und die Standardabweichung s_x der Stichprobe.
- (2) Lies in t -Tabelle (Tab. II, zweiseitig) den Wert $t_{Tab}(FG; \alpha)$ ab, wobei α das Signifikanzniveau und $FG = n - 1$ die Freiheitsgrade sind.
- (3) Das $(1 - \alpha)100\%$ -Vertrauensintervall ist dann

$$[\bar{x} - t_{Tab} s_x / \sqrt{n} ; \bar{x} + t_{Tab} s_x / \sqrt{n}]$$

Weiteres Beispiel: In einem Versuch wurde eine Weizensorte auf 13 Parzellen getestet. Mit diesem Versuch sollte das Leistungspotential μ der Sorte geschätzt werden. Es ergaben sich folgende Erträge (dt/ha): 44, 40, 18, 20, 45, 26, 55, 55, 20, 46, 15, 8, 41. Mit diesen Ergebnissen soll ein Vertrauensintervall für das Leistungspotential μ berechnet werden. Hieraus berechnen wir zunächst den Mittelwert $\bar{x} = 33,31$ und die Standardabweichung $s_x = 16,01$. Für $FG = n - 1 = 12$ Freiheitsgrade und $\alpha = 5\%$ finden wir einen t -Wert von $t_{Tab} = 2,179$. Ein 95% Vertrauensintervall berechnet sich hiermit zu

$$33,31 - 2,179 \cdot 16,01 / \sqrt{13} \quad \text{bis} \quad 33,31 + 2,179 \cdot 16,01 / \sqrt{13}$$

$$23,63 \quad \quad \quad \text{bis} \quad \quad \quad 42,98$$

Beachte: In diesem Beispiel bezieht sich μ auf eine hypothetische Grundgesamtheit (siehe Abschnitt 3.3)!

Zum Begriff der Freiheitsgrade: Eine häufig gestellte Frage ist: was sind eigentlich die Freiheitsgrade. Hierauf gibt es verschiedene Antworten. Hier sollen zwei gegeben werden.

(1) Wenn der Mittelwert einer Stichprobe \bar{x} gegeben ist, dann können wir $n-1$ der n Einzelwerte frei wählen. Der n -te Wert liegt damit fest. Dies ist natürlich eine konstruierte Betrachtung, weil wir ja die beobachteten Werte tatsächlich nicht frei wählen, sondern diese beobachten bzw. messen.

(2) Bei der Schätzung der Varianz dividieren wir die Quadratsumme

$$SQ = \sum_{i=1}^n (x_i - \bar{x})^2 \text{ durch die Freiheitsgrade } n-1, \text{ und nicht etwa durch } n, \text{ um eine}$$

unverzerrte Schätzung der wahren Varianz in der Grundgesamtheit zu erhalten. Dass dies tatsächlich zu einer unverzerrten Schätzung führt, sei am Beispiel eines fairen Würfels erläutert. Hierzu betrachten wir alle 36 möglichen Stichproben von jeweils $n = 2$ Würfeln des Würfels (siehe Tab 3.6.1).

Tabelle 3.6.1: 36 mögliche Würfe von zwei Würfeln mit den dazugehörigen Varianzschätzungen.

1. Wurf	2. Wurf	Mittelwert $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	Quadratsumme $SQ = \sum_{i=1}^n (x_i - \bar{x})^2$	Varianzschätzungen	
				$s_x^2 = \frac{SQ}{n-1}$	$\frac{SQ}{n}$
1	1	1,0	0,0	0,0	0,00
1	2	1,5	0,5	0,5	0,25
1	3	2,0	2,0	2,0	1,00
1	4	2,5	4,5	4,5	2,25
1	5	3,0	8,0	8,0	4,00
1	6	3,5	12,5	12,5	6,25
2	1	1,5	0,5	0,5	0,25
2	2	2,0	0,0	0,0	0,00
2	3	2,5	0,5	0,5	0,25
2	4	3,0	2,0	2,0	1,00
2	5	3,5	4,5	4,5	2,25
2	6	4,0	8,0	8,0	4,00
3	1	2,0	2,0	2,0	1,00
3	2	2,5	0,5	0,5	0,25
3	3	3,0	0,0	0,0	0,00
3	4	3,5	0,5	0,5	0,25
3	5	4,0	2,0	2,0	1,00
3	6	4,5	4,5	4,5	2,25
4	1	2,5	4,5	4,5	2,25
4	2	3,0	2,0	2,0	1,00
4	3	3,5	0,5	0,5	0,25
4	4	4,0	0,0	0,0	0,00
4	5	4,5	0,5	0,5	0,25
4	6	5,0	2,0	2,0	1,00
5	1	3,0	8,0	8,0	4,00
5	2	3,5	4,5	4,5	2,25
5	3	4,0	2,0	2,0	1,00
5	4	4,5	0,5	0,5	0,25
5	5	5,0	0,0	0,0	0,00
5	6	5,5	0,5	0,5	0,25
6	1	3,5	12,5	12,5	6,25
6	2	4,0	8,0	8,0	4,00
6	3	4,5	4,5	4,5	2,25
6	4	5,0	2,0	2,0	1,00
6	5	5,5	0,5	0,5	0,25
6	6	6,0	0,0	0,0	0,00
Erwartungswerte:				2,91 $\bar{6}$	1,458 $\bar{3}$

Anhand dieser Stichprobe soll die wahre Varianz σ_x^2 der Augenzahl eines fairen Würfels geschätzt werden. Hierzu verwenden wir zum einen die Stichprobenvarianz s_x^2 , berechnet als Quotient der Summe der Quadrate (SQ), dividiert durch die Freiheitsgrade, $(n - 1)$, und zum anderen die Summe der Quadrate, dividiert durch n . Mitteln wir die Varianzschätzungen über die 36 Möglichen Fälle, so erhalten wir die Erwartungswerte der beiden Varianzschätzer, weil jeder der Fälle gleich wahrscheinlich ist. Die Erwartungswerte unterscheiden sich um den Faktor Zwei.

Der Erwartungswert der Augenzahl beträgt

$$\mu = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3,5.$$

Die wahre Varianz lässt sich wie folgt berechnen:

$$\sigma_x^2 = \frac{1}{6}[(1 - \mu)^2 + (2 - \mu)^2 + (3 - \mu)^2 + (4 - \mu)^2 + (5 - \mu)^2 + (6 - \mu)^2] = 2,91\bar{6}$$

Das Beispiel zeigt, dass $s_x^2 = \frac{SQ}{n - 1}$ ein unverzerrter Schätzer der wahren Varianz σ_x^2 ist,

während $\frac{SQ}{n}$ eine Verzerrung nach unten aufweist.

3.7 Stichprobenumfang zur Schätzung eines Mittelwertes

Um vor einem Experiment den Stichprobenumfang zu planen, benötigt man eine Vorinformation über die Varianz der Einzelwerte der Grundgesamtheit (σ_x^2). Diese Information benötigt man deshalb, weil die Genauigkeit des Stichprobenmittelwertes \bar{x} um so größer ist, je kleiner die Varianz in der Grundgesamtheit ist. Denn die Varianz des Mittelwertes ($\sigma_{\bar{x}}^2$) hängt direkt von der Varianz der Einzelwerte und vom Stichprobenumfang (n) ab:

$$\sigma_{\bar{x}}^2 = \sigma_x^2 / n$$

Je größer der Stichprobenumfang ist, um so kleiner ist die Varianz des Mittelwertes. Dementsprechend hatten wir in den Abschnitten 3.5 und 3.6 gesehen, dass ein Vertrauensintervall um so enger wird, je größer der Stichprobenumfang ist.

Für eine Stichprobenplanung muss nun eine Genauigkeitsvorgabe gemacht werden. Dies kann beispielsweise in Form einer Vorgabe für die Varianz des Mittelwertes ($\sigma_{\bar{x}}^2$) geschehen. Anschaulicher ist oft eine Vorgabe über die gewünschte maximale Breite eines Vertrauensintervalls. Diese Möglichkeit soll hier betrachtet werden. Die halbe Breite eines Vertrauensintervalls ist gegeben durch $t_{Tab} s_x / \sqrt{n}$. Für ein 95% Vertrauensintervall ist der tabellarische t -Wert bei nicht zu kleinen Stichproben etwa 2. Also ist die halbe Breite des Intervalls etwa $2s_x / \sqrt{n}$.

Angenommen, im Fall des Maisfeldes haben wir eine Schätzung für die Standardabweichung von $s_x = 10$ und wollen eine Stichprobenplanung für die Schätzung des Mittelwertes der Pflanzenlängen im Maisfeld machen. Wir möchten, dass das Vertrauensintervall nicht breiter als 2 cm wird. Somit ist unsere Vorgabe, dass die halbe Breite des Vertrauensintervalls gleich 1 cm wird, also

$$1 = 2s_x / \sqrt{n}.$$

Einsetzen unseres Schätzwertes für die Standardabweichung liefert $1 = 20/\sqrt{n}$. Durch Auflösen dieser Gleichung nach n erhalten wir $n = 20^2 = 400$. Der notwendige Stichprobenumfang beträgt also $n = 400$ Pflanzen.

Der Stichprobenumfang für ein 95% Vertrauensintervall berechnet sich nach

$$n = \frac{4s_x^2}{(HB)^2}$$

wobei HB die gewünschte halbe Breite des Vertrauensintervalls ist.

3.8 Vertrauensintervall für die Differenz von 2 Mittelwerten (verbundene Stichproben)

In vielen Versuchen hat man es nicht nur mit einem Mittelwert, sondern mit zwei oder mehreren Mittelwerten zu tun. So wird in einem Feldversuch in der Regel nicht nur eine Behandlung geprüft, sondern zwei oder mehr als zwei.

Beispiel: In einem on-farm Versuch werden zwei Maissorten geprüft: Eine traditionelle Sorte (LM) und eine verbesserte Sorte (CCA). Die Versuche wurden auf insgesamt 14 Farmen in 2 Dörfern des Phalombe Project in Südost-Malawi durchgeführt (*Agronomy Journal* 76 (1984): 271-274). Die Erträge (t/ha) sind wie folgt:

Farm	Sorte	
	LM	CCA
1	2,2	3,5
2	2,2	2,0
3	1,9	2,9
4	1,2	0,4
5	1,3	0,6
6	0,9	0,5
7	1,0	0,6
8	0,5	0,3
9	1,8	2,2
10	1,1	0,7
11	1,6	0,9
12	1,0	0,3
13	1,6	1,1
14	0,6	0,3
Mittelwert	$\bar{x}_1 = 1,35$	$\bar{x}_2 = 1,16$

Es soll geprüft werden, welche der zwei Sorten das höhere Ertragspotential der Region hat. Die Ertragspotentiale der Sorten können mit μ_1 und μ_2 bezeichnet werden. μ_1 und μ_2 können auch als Durchschnittserträge der beiden Sorten in der gesamten Projektregion Phalombe betrachtet werden. Dementsprechend sind die 14 Farmen, auf denen die On-farm Versuche durchgeführt wurden, eine Stichprobe aus der Grundgesamtheit „Projektregion“. Die Schätzwerte für μ_1 und μ_2 sind die Stichprobenmittelwerte ($\bar{x}_1 = 1,35$ und $\bar{x}_2 = 1,16$). Für den Vergleich der beiden Sorten interessiert die Differenz der beiden Sorten: $\delta = \mu_1 - \mu_2$. Diese Differenz kann aus den Stichprobenmittelwerten für beide Sorten geschätzt werden. Wir erhalten eine Differenz von $\bar{d} = \bar{x}_1 - \bar{x}_2 = 1,35 - 1,16 = 0,19$. Für die Berechnung eines Vertrauensintervalls ist allerdings eine andere Berechnung hilfreich, die zum gleichen Ergebnis führt. Wir können auf jeder der Farmen die Differenz der Sortenerträge berechnen. Mitteln wir diese Differenzen, erhalten wir ebenfalls einen Wert von 0,19.

Farm	Sorte		Differenzen (d) (LM – CCA)
	LM	CCA	
1	2,2	3,5	-1,3
2	2,2	2,0	+0,2
3	1,9	2,9	-1,0
4	1,2	0,4	+0,8
5	1,3	0,6	+0,7
6	0,9	0,5	+0,4
7	1,0	0,6	+0,4
8	0,5	0,3	+0,2
9	1,8	2,2	-0,4
10	1,1	0,7	+0,4
11	1,6	0,9	+0,7
12	1,0	0,3	+0,7
13	1,6	1,1	+0,5
14	0,6	0,3	+0,3
Mittelwert	$\bar{x}_1 = 1,35$	$\bar{x}_2 = 1,16$	$\bar{d} = +0,19$

Um nun ein Vertrauensintervall für die tatsächliche Differenz $\delta = \mu_1 - \mu_2$ zu berechnen, können wir die Streuung der Differenzen auf den Farmen heranziehen. Wir können die beobachteten Differenzen als Stichprobe aus einer Grundgesamtheit von Differenzen betrachten, deren Mittel wir schätzen wollen. Um hierfür ein Vertrauensintervall zu berechnen, können wir dieselbe Methode verwenden wie in Abschnitt 3.6 für die mittlere Pflanzenlänge eines Maisfeldes, mit dem Unterschied, dass hier Differenzen betrachtet werden.

Fragestellung: Es soll ein $(1-\alpha)100\%$ -Vertrauensintervall für die Differenz δ der Erwartungswerte zweier Stichproben (Behandlungen) 1 und 2 berechnet werden.

Voraussetzungen: Beide Grundgesamtheiten sind normalverteilt. Die beiden Stichproben sind verbunden und haben die Erwartungswerte (wahren Mittelwerte) μ_1 und μ_2 .

Rechenweg:

Ein $(1-\alpha)100\%$ -Vertrauensintervall für die Differenz $\delta = \mu_1 - \mu_2$ kann berechnet werden nach

$$\bar{d} \pm t_{Tab} s_d / \sqrt{n}$$

wobei

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} \quad (\text{Stichprobenmittelwert der Differenzen}),$$

$d_i = x_{1i} - x_{2i}$ ($i = 1 \dots, n$) (paarweise Differenzen der Messwerte der 1. und 2. Behandlung)

x_{1i} = Messwerte der 1. Behandlung,

x_{2i} = Messwerte der 2. Behandlung,

$$s_d = \sqrt{\frac{\sum_{i=1}^n d_i^2 - \frac{\left(\sum_{i=1}^n d_i\right)^2}{n}}{n-1}} \quad (\text{Standardabweichung der Differenzen}), \text{ und}$$

$t_{Tab}(FG; \alpha)$ = t -Wert für eine vorgegebene Irrtumswahrscheinlichkeit α
(Tab. II, zweiseitig)

$FG = n-1$ Freiheitsgrade.

Im vorliegenden Fall beträgt der Mittelwert der Differenzen $\bar{d} = 0,19$ und die Standardabweichung der Differenzen beträgt $s_d = 0,643$. Für $FG = n - 1 = 13$ Freiheitsgrade und $\alpha = 5\%$ finden wir einen t -Wert von $t_{Tab} = 2,16$. Damit ergibt sich das folgende Vertrauensintervall:

$$0,19 - 2,16 \cdot 0,643 / \sqrt{14} \quad \text{bis} \quad 0,19 + 2,16 \cdot 0,643 / \sqrt{14}$$

$$-0,18 \quad \text{bis} \quad 0,56$$

Mit 95% Wahrscheinlichkeit überdecken die Grenzen von $-0,18$ bis $0,56$ die Differenz $\delta = \mu_1 - \mu_2$ zwischen den Sorten. Zwar ist der Schätzwert der Differenz, \bar{d} , positiv, was auf einen Vorteil der traditionellen Sorte hinweist. Das Vertrauensintervall überdeckt jedoch die Null, so dass der gefundenen Differenz nicht allzu viel Bedeutung beigemessen werden kann.

Im hier verwendeten Beispiel liegt eine **verbundene** Stichprobe vor. Dies bedeutet, dass je zwei Beobachtungen (eine für jede der beiden Sorten) miteinander verbunden sind, indem sie auf der gleichen Farm gemacht wurden. Hätten wir die eine Sorte auf 14 Farmen und die andere Sorte auf 14 anderen Farmen getestet, so läge eine **unverbundene** Stichprobe vor. Es ist wichtig zwischen verbundenen und unverbundenen Stichproben zu unterscheiden, weil das Auswertungsverfahren (Vertrauensintervall oder Test) von der Art der Stichprobe abhängt.

3.9 Vertrauensintervall für die Differenz von 2 Mittelwerten (unverbundene Stichproben)

Im Beispiel aus Abschnitt 3.8 hatten wir es mit einer verbundenen Stichprobe zu tun. In diesem Abschnitt geht es um unverbundene Stichproben. Um auf das Maisfeld zurückzukommen, soll angenommen werden, dass wir nicht nur an einem Maisfeld interessiert sind, sondern an zwei Feldern. Die zwei Felder sollen bezüglich der durchschnittlichen Pflanzenlänge verglichen werden. Die Ergebnisse für das erste Maisfeld können wie folgt zusammengefasst werden:

$$\bar{x}_1 = 171,2, \quad s_1 = 10, \quad n_1 = 50$$

Für das zweite Feld wurden 40 Pflanzen gemessen. Die Ergebnisse sind:

$$\bar{x}_2 = 173,7, \quad s_2 = 9,2, \quad n_2 = 40$$

Die Indizes 1 und 2 werden hier verwendet, um die beiden Felder zu unterscheiden. Die Daten sind unverbunden in dem Sinn, dass keine Maispflanze des einen Feldes irgendwie mit einer bestimmten Pflanze des anderen Feldes zusammenhängt. Die Daten sollen verwendet werden, um die wahren Mittelwerte μ_1 und μ_2 der beiden Felder zu vergleichen.

Fragestellung: Es soll ein $(1-\alpha)100\%$ -Vertrauensintervall für die Differenz der Erwartungswerte zweier Stichproben 1 und 2 berechnet werden.

Voraussetzungen: Beide Grundgesamtheiten sind normalverteilt mit gleicher, unbekannter Varianz. Die beiden Stichproben sind unabhängig (unverbunden) und haben die Erwartungswerte (wahren Mittelwerte) μ_1 und μ_2 .

Rechenweg:

Ein $(1-\alpha)100\%$ -Vertrauensintervall für die Differenz $\delta = \mu_1 - \mu_2$ ist

$$(\bar{x}_1 - \bar{x}_2) \pm t_{Tab} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

wobei

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1}, \quad \bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2},$$

x_{1i} ($i = 1, \dots, n_1$) = Messwerte der 1. Behandlung

x_{2i} ($i = 1, \dots, n_2$) = Messwerte der 2. Behandlung

n_1 und n_2 = Umfänge der Stichproben 1 und 2

$t_{Tab}(FG; \alpha)$ = t -Wert für eine vorgegebene Irrtumswahrscheinlichkeit α
(Tab. II, zweiseitig)

$FG = n_1 + n_2 - 2$ Freiheitsgrade, sowie

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \text{ mit}$$

$$s_1^2 = \frac{\sum_{i=1}^{n_1} x_{1i}^2 - \frac{\left(\sum_{i=1}^{n_1} x_{1i}\right)^2}{n_1}}{n_1 - 1} \text{ und } s_2^2 = \frac{\sum_{i=1}^{n_2} x_{2i}^2 - \frac{\left(\sum_{i=1}^{n_2} x_{2i}\right)^2}{n_2}}{n_2 - 1}.$$

Die letzte Formel zeigt, dass s^2 ein gewichtetes Mittel der Varianzen der beiden Stichproben (Maisfelder) ist. In unserem Beispiel finden wir

$$s^2 = \frac{(50 - 1) * 10^2 + (40 - 1) * 9,2^2}{50 + 40 - 2} = 93,19.$$

und $s = 9,65$. Für $FG = 50 + 40 - 2 = 88$ Freiheitsgrade und $\alpha = 5\%$ finden wir einen t -Wert von $t_{Tab} = 1,99$. Somit ist das Vertrauensintervall für die Differenz ($\delta = \mu_1 - \mu_2$)

$$(171,2 - 173,7) \pm 1,99 * 9,65 * \sqrt{\frac{1}{50} + \frac{1}{40}}$$

$$-2,5 \quad \pm \quad 4,1$$

also $-6,6$ bis $+1,6$

Mit 95%iger Wahrscheinlichkeit überdeckt das Intervall zwischen $-6,6$ und $+1,6$ die wahre Differenz ($\delta = \mu_1 - \mu_2$).

3.10 Test zum Vergleich zweier verbundener Stichproben

In den vorangegangenen Abschnitten ging es um die Schätzung von Parametern und die Berechnung von Vertrauensintervallen um diese Schätzwerte. Vertrauensintervalle sind ein wichtiger Bereich der schließenden Statistik. Ein anderer wichtiger Bereich ist die Prüfung von Hypothesen mittels statistischer Signifikanztests. Um solche Tests soll es im folgenden gehen. Die Durchführung eines Signifikanztests kann in sechs Schritten allgemein beschrieben werden:

- (i) Formulierung einer Nullhypothese (H_0).
- (ii) Formulierung einer Alternativhypothese (H_1).
- (iii) Wahl des Signifikanzniveaus α .
- (iv) Berechnung einer Teststatistik.
- (v) Bestimmung eines kritischen Wertes mittels einer geeigneten Tabelle.
- (vi) Vergleich der berechneten Teststatistik mit dem kritischen Tabellenwert. Dieser Vergleich führt zu einer Entscheidung, ob die Nullhypothese (H_0) verworfen werden kann oder nicht.

Ganz wichtig ist es, unter (vi) die Asymmetrie zwischen H_0 und H_1 zu erkennen. Kann man H_0 verwerfen, so hat man einen starken Hinweis für die Gültigkeit von H_1 . Kann H_0 dagegen nicht verworfen werden, so hat man H_0 keineswegs bewiesen. Man ist in einer eher schwachen Position, weil die Daten keine deutliche Entscheidung erlauben. Zusammengefasst eignet sich ein wie oben konstruierter Test zum Nachweis von H_1 , jedoch nicht zum Nachweis von H_0 . [Will man H_0 nachweisen, muss man eine andere Art von Test verwenden, einen sog. Äquivalenz-Test (Wellek S 2003 Testing statistical hypotheses of equivalence. Chapman & Hall, London; siehe Abschnitt 3.21).]

Beispiel: Ohne hier eine exakte Herleitung des Tests zu geben, soll das Vorgehen beim Testen am Beispiel der Maissortenversuch aus Malawi (Abschnitt 3.8) beschrieben werden. In diesem Versuch wurden zwei Sorten (LM und CCA) auf 14 Farmen getestet.

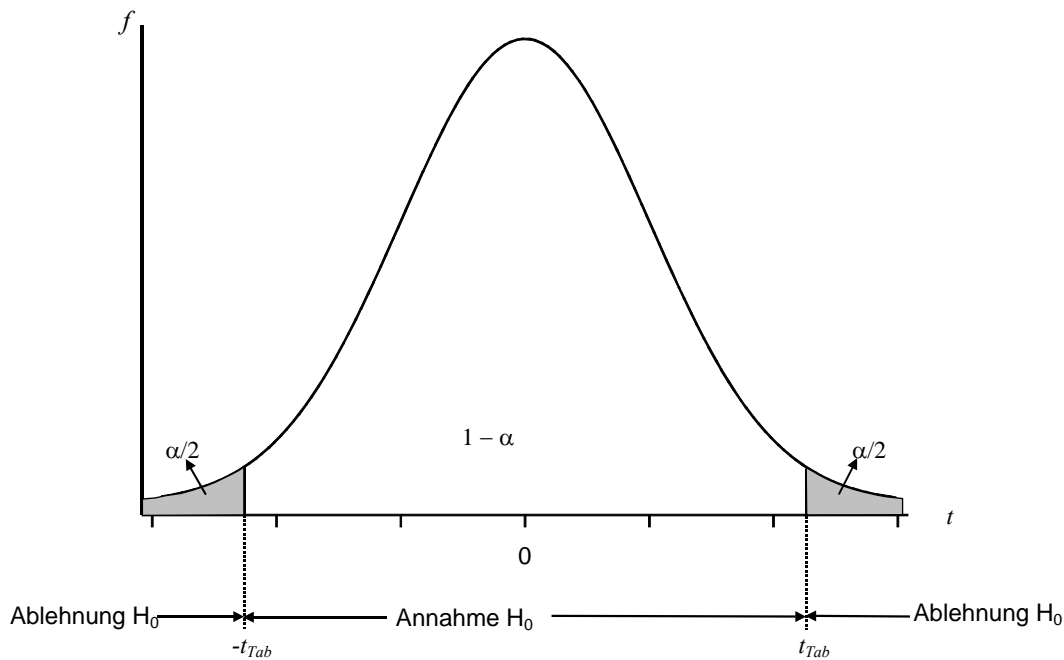


Abb. 3.10.1: Annahme- und Ablehnungsbereich für einen zweiseitigen t-Test.

(i) Wir können hier die Nullhypothese formulieren, dass die beiden Sorten sich nicht in ihrem Ertragspotential unterscheiden:

$$H_0 : \mu_1 = \mu_2$$

(ii) Die Alternativhypothese (H_1) lautet, dass sich die beiden Sorten unterscheiden, also

$$H_1 : \mu_1 \neq \mu_2$$

(iii) Das Signifikanzniveau α ist die Wahrscheinlichkeit, die Nullhypothese zu verwerfen, obwohl sie zutrifft. Meistens wird $\alpha = 5\%$ gewählt. Diese Festlegung ist relativ willkürlich, hat sich aber in der Praxis durchgesetzt. Andere Signifikanzniveaus können auch verwendet werden, z.B. 1% oder 10%. Wir wollen hier mit $\alpha = 5\%$ arbeiten.

(iv) Im vorliegenden Fall hat die Teststatistik folgende Form:

$$t_{Vers} = \frac{|\bar{d}|}{s_d} \sqrt{n}$$

(Die Schreibweise $|\quad|$ bedeutet „Betrag von“. Der Betrag eines Wertes ist immer positiv. Hat der Wert ein negatives Vorzeichen, wird das Vorzeichen weggelassen, um den Betrag zu erhalten. Der Betrag von 3 ist gleich 3. Der Betrag von -2 ist 2). Für die Sortenversuchsdaten finden wir

$$t_{Vers} = (0,19/0,643) \cdot \sqrt{14} = 1,11$$

(v) In einer t -Tabelle (Tab. II) finden wir für $FG = n - 1 = 13$ Freiheitsgrade und $\alpha = 5\%$ einen kritischen t -Wert von $t_{Tab} = 2,16$.

(vi) Wir stellen fest, dass t_{Vers} den Tabellenwert t_{Tab} nicht überschreitet. Daher besteht kein signifikanter Unterschied zwischen den beiden Sorten und wir können die Nullhypothese H_0 nicht verwerfen.

Die Begründung für das in den Schritten (i) bis (vi) beschriebene Vorgehen für einen statistischen Test soll nun anhand der Abbildung 3.10.1 gegeben werden. Die Kurve gibt die Verteilung der Test-Statistik $t = \frac{\bar{d}}{s_d} \sqrt{n}$ unter der Nullhypothese wieder. Diese

"Nullverteilung" ist symmetrisch um die Null. Dies bedeutet, dass die Test-Statistik um die Null herum verteilt ist. Unter der Alternative (Mittelwerte unterscheiden sich) ist die Verteilung der Test-Statistik nach links oder rechts verschoben, je nachdem ob $\mu_1 > \mu_2$ ist, oder $\mu_1 < \mu_2$. Je stärker die Alternative von der Nullhypothese abweicht, desto stärker wird die Teststatistik von der Null abweichen ("nach unten" oder "nach oben"). Um nun zu einem Test zu kommen, wird eine untere und eine obere Schwelle definiert, deren Überschreiten zu einer Ablehnung der Nullhypothese führt. Diese Schwellen sind in Abb. 3.10.1. mit $-t_{Tab}$ und t_{Tab} bezeichnet. Die Abbildung zeigt, dass es bei Gültigkeit der Nullhypothese mit einer gewissen Wahrscheinlichkeit zum Verwerfen der Nullhypothese und damit zu einem Fehler kommt (α -Fehler, Fehler 1. Art). Diese Wahrscheinlichkeit ist gegeben durch die Fläche unter der Dichte im Ablehnungsbereich, also unterhalb von $-t_{Tab}$ und oberhalb von t_{Tab} . Die Grenzen werden nun gerade so gewählt, dass die Irrtumswahrscheinlichkeit den vorgegebenen Wert α annimmt. Die Wahrscheinlichkeit, die Nullhypothese anzunehmen, vorausgesetzt sie ist wahr, beträgt demnach $1 - \alpha$, was auch in Abb. 3.10.1 angedeutet ist.

Abb. 3.10.1 zeigt, dass der α -Fehler umso kleiner wird, je weiter die kritischen Werte von der Null entfernt liegen. Dies könnte nahelegen, dass die Grenzen möglichst weit von Null entfernt liegen sollten. Dies ist aber ein Trugschluss. In der Konsequenz müsste man unendlich große Grenzen wählen, mit der Folge, dass man nie die Nullhypothese verwerfen würde, auch dann nicht, wenn in Wirklichkeit Unterschiede bestehen, also die Alternative zutrifft. Auch dies ist ein Fehler, und zwar ein Fehler 2. Art oder β -Fehler. Legen wir den Fehler 1. Art also bei 0% fest, so ist der Fehler 2. Art immer 100%, so dass wir mit dem Test zu überhaupt keiner praktikablen Entscheidung kommen! Diese Überlegung macht klar, dass wir einen von Null verschiedenen Wert für den Fehler 1. Art vorgeben müssen, um zu einer sinnvollen Entscheidungsregel zu kommen.

Abb. 3.10.1 zeigt, dass es einen zweiseitigen Ablehnungsbereich gibt. Voraussetzung für eine Ablehnung ist, dass die Statistik $t = \frac{\bar{d}}{s_d} \sqrt{n}$ vom Betrag her groß ist,

und zwar unabhängig vom Vorzeichen. Um die praktische Durchführung des Tests zu vereinfachen, berechnet man einfach gleich den Betrag dieser Statistik,

$t_{Vers} = \frac{|\bar{d}|}{s_d} \sqrt{n}$, und fordert, dass t_{Vers} einen kritischen Wert t_{Tab} **überschreitet**. Diese

vereinfachte "einseitige" Durchführung des Tests sollte aber nicht darüber hinweg täuschen, dass der Test in Wirklichkeit einen zweiseitigen Ablehnungsbereich beinhaltet. Denn sowohl ein großer negativer Wert als auch ein großer positiver Wert für

$$t = \frac{\bar{d}}{s_d} \sqrt{n} \text{ führen zu einer Ablehnung der Nullhypothese.}$$

An dieser Stelle ist eine Warnung angebracht. Im obigen Beispiel haben wir die Nullhypothese nicht verwerfen können. Das heißt allerdings nicht, dass wir schließen können, dass beide Sorten den gleichen Mittelwert haben ($\mu_1 = \mu_2$). Es gibt zwei mögliche Gründe, warum wir die Nullhypothese nicht verwerfen konnten: (1) Es besteht tatsächlich kein Unterschied oder (2) der Stichprobenumfang war nicht groß genug, um einen Unterschied festzustellen. Oft ist es so, dass man von vornherein davon ausgehen kann, dass die Nullhypothese nicht zutrifft, dass also Unterschiede zwischen den Sorten bestehen müssen. Wenn man genau darüber nachdenkt, ist es so gut wie ausgeschlossen, dass zwei Sorten exakt das gleiche Ertragspotential haben. Was man wissen möchte, ist, welche der beiden Sorten besser ist. In anderen Worten, man möchte wissen ob $\mu_1 > \mu_2$ oder $\mu_1 < \mu_2$ ist. Wenn ein Test die Nullhypothese verwirft, dann können wir diese Frage beantworten (im Rahmen der Irrtumswahrscheinlichkeit α). Falls $\bar{x}_1 < \bar{x}_2$, schließen wir, dass $\mu_1 < \mu_2$ ist. Falls $\bar{x}_1 > \bar{x}_2$, schließen wir, dass $\mu_1 > \mu_2$ ist. Zusammenfassend kann gesagt werden, dass wir in einer günstigeren Position sind, falls der Test signifikant ist. Ist er nicht signifikant, so können wir eigentlich keine Aussagen machen, außer der, dass wir keine Unterschiede nachweisen konnten.

In diesem Zusammenhang ist es wichtig, zwischen Relevanz und Signifikanz zu unterscheiden. Man kann durch Wahl eines sehr hohen Stichprobenumfanges einen noch so kleinen Unterschied signifikant nachweisen. Ob dieser dann auch relevant ist, ist eine ganz andere Frage, die nur aus Fachwissenschaftlicher Sicht beantwortet werden kann. Um die Relevanz einzuschätzen, bietet es sich in jedem Fall an, das Vertrauensintervall zu betrachten. Dies liefert Auskunft über die Größe des Unterschiedes zwischen zwei Mittelwerten, während der Test nur einen Aussage über die Signifikanz zulässt.

Das praktische Vorgehen beim Test zum Vergleich zweier verbundener Stichproben ist im folgenden Kasten noch einmal zusammengefasst.

Fragestellung: Sind die Mittelwerte μ_1 und μ_2 zweier Grundgesamtheiten 1 und 2, aus denen eine Stichprobe gezogen wurde, signifikant verschieden?

Voraussetzungen: Die Stichproben sind verbunden. Die Differenzen sind normalverteilt mit unbekanntem Mittelwert δ .

Rechenweg:

(1) Berechne $t_{\text{vers}} = \frac{|\bar{d}|}{s_d} \sqrt{n}$, wobei

n = Stichprobenumfang

$d_i = x_{1i} - x_{2i} = i$ -te Messwertdifferenz

x_{1i} = Messwerte der 1. Behandlung,

x_{2i} = Messwerte der 2. Behandlung,

$$s_d = \sqrt{\frac{\sum_i d_i^2 - \frac{(\sum_i d_i)^2}{n}}{n-1}}$$

(2) Lies in der t -Tabelle (Tab. II, zweiseitig) $t_{Tab}(FG; \alpha)$ ab, wobei:

α = Signifikanzniveau

$FG = n - 1$ = Freiheitsgrade

(3) Vergleiche t_{Vers} mit t_{Tab} :

Falls $t_{Vers} \leq t_{Tab} \Rightarrow H_0 (\delta = 0)$ (Differenz nicht signifikant von Null verschieden)

Falls $t_{Vers} > t_{Tab} \Rightarrow H_1 (\delta \neq 0)$ (Differenz von Null verschieden)

Abschließend sei darauf hingewiesen, dass ein enger Zusammenhang besteht zwischen dem eben beschriebenen Test und dem Vertrauensintervall für die Differenz verbundener Stichproben (Abschnitt 3.8). Für den Sortenversuch aus Malawi hatten wir für die Differenz der beiden Sorten, $\delta = \mu_1 - \mu_2$, die 95% Vertrauensgrenzen $-0,18$ bis $0,56$ berechnet. Es fällt auf, dass das Intervall die Null mit einschließt. Anhand dieses Ergebnisses können wir zunächst intuitiv schließen, dass die Differenz nicht signifikant von Null verschieden ist. Tatsächlich kann gezeigt werden, dass der t -Test immer genau (und nur) dann nicht signifikant ist, wenn das Vertrauensintervall die Null einschließt. Ebenso ist der t -Test immer dann signifikant, wenn das Vertrauensintervall die Null nicht mit einschließt. Im Vergleich zum Test liefert das Vertrauensintervall mehr Information. Während der Test nur anzeigt, ob ein Unterschied nachgewiesen werden kann oder nicht, zeigt das Vertrauensintervall an, in welchem Bereich die tatsächliche Differenz wahrscheinlich liegt.

3.11 Test zum Vergleich zweier unverbundener Stichproben

Das Vorgehen bei einem Vergleich von unverbundenen Stichproben ist im folgenden Kasten beschrieben.

Fragestellung: Sind die Mittelwerte μ_1 und μ_2 zweier Grundgesamtheiten 1 und 2, aus denen jeweils eine Stichprobe gezogen wurde, signifikant verschieden?

Voraussetzungen: Beide Grundgesamtheiten sind normalverteilt mit gleicher, unbekannter Varianz. Die Stichproben sind unabhängig.

Rechenweg:

(1) Berechne

$$t_{Vers} = \frac{|\bar{x}_1 - \bar{x}_2|}{s \sqrt{1/n_1 + 1/n_2}} \quad \text{wobei}$$

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1}, \quad \bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2},$$

x_{1i} ($i = 1, \dots, n_1$) = Messwerte der 1. Behandlung

x_{2i} ($i = 1, \dots, n_2$) = Messwerte der 2. Behandlung

n_1 und n_2 = Umfänge der Stichproben 1 und 2

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad \text{mit}$$

$$s_1^2 = \frac{\sum_{i=1}^{n_1} x_{1i}^2 - \frac{\left(\sum_{i=1}^{n_1} x_{1i}\right)^2}{n_1}}{n_1 - 1} \quad \text{und} \quad s_2^2 = \frac{\sum_{i=1}^{n_2} x_{2i}^2 - \frac{\left(\sum_{i=1}^{n_2} x_{2i}\right)^2}{n_2}}{n_2 - 1}.$$

(2) Lies aus der t -Tabelle $t_{Tab}(FG; \alpha)$ ab (Tab. II, zweiseitig), wobei:

α = Signifikanzniveau

$FG = n_1 + n_2 - 2$

(3) Vergleiche t_{Vers} und t_{Tab} :

$t_{Vers} \leq t_{Tab} \Rightarrow H_0 (\mu_1 = \mu_2)$ (keine Unterschiede)

$t_{Vers} > t_{Tab} \Rightarrow H_1 (\mu_1 \neq \mu_2)$ (es bestehen Unterschiede)

Als Beispiel für unverbundene Stichproben werden die Ergebnisse für die Pflanzenlängen in zwei Maisfeldern verwendet (siehe Abschnitt 3.9). Dort hatten wir folgende Werte zur Verfügung:

$$\bar{x}_1 = 171,2, \quad s_1 = 10, \quad n_1 = 50$$

$$\bar{x}_2 = 173,7, \quad s_2 = 9,2, \quad n_2 = 40$$

Wir wollen nun prüfen, ob sich die durchschnittliche Pflanzenlänge der beiden Maisfelder unterscheidet. Zunächst berechnen wir aus den Standardabweichungen für die beiden Stichproben eine gemeinsame Varianz:

$$s^2 = \frac{(50 - 1) * 10^2 + (40 - 1) * 9,2^2}{50 + 40 - 2} = 93,19$$

Somit ist $s = 9,65$. Hieraus berechnet sich folgender Wert für t_{Vers} :

$$t_{Vers} = \frac{|171,2 - 173,7|}{9,65\sqrt{(1/50 + 1/40)}} = 1,22$$

Der kritische t -Wert für $\alpha = 5\%$ und $FG = 50 + 40 - 2 = 88$ Freiheitsgrade ist $t_{Tab} = 1,99$. Da $t_{Vers} < t_{Tab}$ ist, können wir die Nullhypothese nicht verwerfen. Auch dieses Ergebnis stimmt wieder mit der Schlussfolgerung des Vertrauensintervalls überein (siehe Abschnitt 3.9): Das Intervall schließt die Null mit ein, woraus sich ebenfalls ergibt, dass keine signifikanten Unterschiede bestehen.

Beispiel: In einem Feldversuch werden Tomatenerträge bei zwei verschiedenen Düngerbehandlungen A und B verglichen. Dünger A wird auf 5 Parzellen getestet, Dünger B auf 6. Die Verteilung der Behandlungen A und B auf die insgesamt 11 Parzellen ist dabei zufällig (Randomisation). Die Parzellenerträge sind wie folgt:

	A	B
	29,9	26,6
	11,4	23,7
	25,3	28,5
	16,5	14,2
	21,1	17,9
		24,3
$\sum x$	104,2	135,2
$\sum x^2$	2381,51	3194,04
	$\bar{x}_1 = 20,84$	$\bar{x}_2 = 22,53$
	$n_1 = 5$	$n_2 = 6$

$$s_1^2 = (2381,51 - (104,2)^2/5)/(5 - 1) = 209,99/4 = 52,50$$

$$s_2^2 = (3194,04 - (135,2)^2/6)/(6 - 1) = 147,53/5 = 29,51$$

$$s^2 = \frac{(5 - 1) * 52,50 + (6 - 1) * 29,51}{5 + 6 - 2} = 39,73$$

$$t_{Vers} = \frac{|20,84 - 22,53|}{\sqrt{39,73}\sqrt{(1/5 + 1/6)}} = 0,44$$

Für $FG = 5 + 6 - 2 = 9$ Freiheitsgrade und $\alpha = 5\%$ finden wir $t_{Tab} = 2,262$.

$t_{Vers} = 0,44 < t_{Tab} = 2,262 \Rightarrow H_0$ kann nicht verworfen werden.

In diesem Beispiel wird übrigens wieder eine hypothetische (unendliche) Grundgesamtheit aller unter den verwendeten Versuchsbedingungen möglichen Erträge angenommen (siehe Abschnitt 3.3).

3.12 Verbundene oder unverbundene Stichprobe?

Es wurde bereits darauf hingewiesen, und dies ergibt sich auch aus den vorhergehenden Abschnitten, dass das Auswertungsverfahren davon abhängt, ob verbundene oder unverbundene Stichproben vorliegen. Bei der Planung eines Versuches muss entschieden werden, ob eine verbundene oder unverbundene Stichprobe zur Beantwortung der Versuchsfrage herangezogen werden soll. In vielen Fällen hat man allerdings keine Wahl. Im Fall der beiden Maisfelder, deren Pflanzenlängen verglichen werden sollen, ist von der Natur der Sache her vorgegeben, dass die Stichproben unverbunden sind. Beim Sortenversuch, der auf 14 verschiedenen Farmen durchgeführt wird, ist die Situation anders: Hier haben wir die Wahl, ob jede der beiden Sorten auf verschiedenen Farmen geprüft wird (Sorte LM auf 14 Farmen, Sorte CCA auf 14 anderen Farmen \Rightarrow unverbundene Stichprobe), oder ob beide Sorten jeweils auf den selben Farmen getestet werden sollen (verbundene Stichprobe). Oft ist es von Vorteil, eine verbundene Stichprobe zu verwenden. Der Grund ist, dass die Differenz zwischen zwei Behandlungen/Stichproben bei verbundenen Stichproben genauer gemessen werden kann, so dass Unterschiede leichter nachzuweisen sind. Um dies zu verdeutlichen, wird hier ein nicht-landwirtschaftliches Beispiel herangezogen.

Es sollen 2 verschiedene Materialien A und B für Schuhsohlen von Schuhen für Jungen verglichen werden. Dazu soll jedes Material von jeweils 10 Jungen getestet werden. Bei der Versuchsdurchführung gibt es verschiedene Möglichkeiten. Zum Beispiel können 10 Jungen das eine Material testen und 10 andere Jungen das 2. Material (unverbundene Stichprobe). Eine Alternative hierzu ist, dass jeder Junge beide Materialien testet (verbundene Stichprobe). In der folgenden Tabelle sind die Ergebnisse einer Untersuchung wiedergegeben, bei der verbundene Stichproben verwendet wurden. Jeder Junge bekam einen Schuh mit Material A und einen Schuh mit Material B. Dabei wurde per Münzwurf entschieden, ob Material A links (L) oder rechts (R) getestet wird (Randomisation). Von der verbundenen Stichprobe versprach man sich eine höhere Aussagekraft als bei unverbundener Stichprobe. Nach einer gewissen Zeit wurde für jeden Schuh ein Abnutzungsindex berechnet (Tab. 3.12.1).

Die Tabelle 3.12.1 sowie Abb. 3.12.1 zeigen, dass die Abnutzung sich zwischen den Versuchspersonen relativ stark unterscheidet. Vergleicht man dazu die Differenz d zwischen beiden Materialien bei jedem Jungen, so sieht man, dass die Differenzen viel weniger von Junge zu Junge schwanken, als die eigentlichen Messwerte. Dies bedeutet, dass die Differenzen zwischen beiden Materialien genauer gemessen werden können, als die Mittelwerte für jedes der beiden Materialien. Dies kann wie folgt erklärt werden: Die Jungen unterscheiden sich wahrscheinlich deutlich in ihrem Freizeitverhalten. Die einen gehen nach der Schule gleich raus auf die Straße und spielen Fußball oder ähnliches, wobei die Schuhe starken Belastungen ausgesetzt sind. Andere Jungen bleiben dagegen vielleicht lieber zu Hause, schauen Fernsehen oder machen ihre Hausaufgaben und bewegen sich wenig. Bei diesen Jungen ist die Abnutzung gering. Aus diesem Grund ist eine starke Schwankung der Abnutzung zwischen den Jungen zu erwarten. Andererseits ist der Unterschied zwischen beiden Materialien nicht so starken Schwankungen ausgesetzt. Angenommen, ein Junge spielt viel Fußball. Bei ihm haben Material A und B Abnutzungswerte von 13 und 14. Die Differenz beträgt also 1. Ein anderer Junge ist viel zu Hause. Bei ihm sind die Werte z.B. 6 und 7. Auch hier ist die Differenz 1. In diesem Beispiel ist also die

Differenz für beide Jungen gleich (nämlich 1), während die Messwerte selbst sich stark zwischen beiden Jungen unterscheiden. Somit kann die Differenz der Materialien A und B relativ genau gemessen werden, trotz der hohen Schwankungen zwischen den Jungen. Dementsprechend finden wir auch einen signifikanten Unterschied:

$$s_d^2 = 0,149$$

$$t_{Vers} = \frac{|\bar{d}|}{s_d} \sqrt{n} = \frac{|0,41|}{\sqrt{0,149}} \sqrt{10} = 3,4 > t_{Tab} = 2,62$$

Da $t_{Vers} > t_{Tab}$ ist, wird die Nullhypothese, dass keine Unterschiede bestehen, verworfen. Diesen Unterschied hätten wir bei unverbundenen Stichproben wahrscheinlich nicht gefunden.

Tab. 3.12.1: Abnutzungsindex bei Schuhen von Jungen. L = linker Schuh, R = rechter Schuh.

Junge	Material		Differenz d = B - A
	A	B	
1	13,2 (L)	14,0 (R)	0,8
2	8,2 (L)	8,8 (R)	0,6
3	10,9 (R)	11,2 (L)	0,3
4	14,3 (L)	14,2 (R)	-0,1
5	10,7 (R)	11,8 (L)	1,1
6	6,6 (L)	6,4 (R)	-0,2
7	9,5 (L)	9,8 (R)	0,3
8	10,8 (L)	11,3 (R)	0,5
9	8,8 (R)	9,3 (L)	0,5
10	13,3 (L)	13,6 (R)	0,3
Durchschnitt:			0,41

(a) Abnutzungsindex

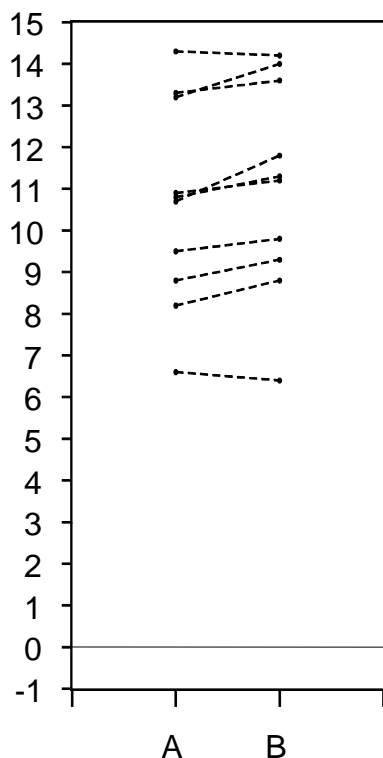
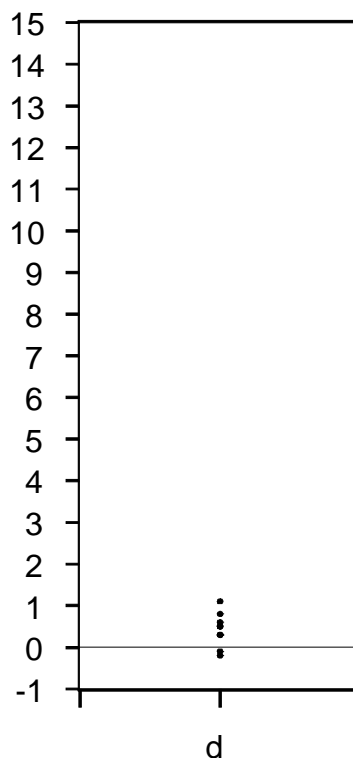
(b) Differenzen (d)

Abb. 3.12.1: Abnutzungsindizes für 10 Jungen. (a) Werte für die Materialien A und B; Werte vom selben Jungen sind durch eine gestrichelte Linie verbunden. (b) Differenzen (d) für die 10 Jungen.

Wie erkenne ich eine verbundene Stichprobe?

Es ist oft nicht ganz einfach, festzustellen, ob eine verbundene oder eine unverbundene Stichprobe vorliegt. Eine einfaches Kriterium ist die Antwort auf die folgende Frage: **Ist es beim verwendeten Versuchsplan möglich, den Stichprobenumfang für die eine Behandlung zu erhöhen, ohne gleichzeitig den Stichprobenumfang der anderen Behandlung zu erhöhen?** Falls die Antwort ja lautet, wie im Fall der beiden Maisfelder, so liegt eine unverbundene Stichprobe vor. Im Fall des Versuches mit den beiden Maissorten lautet die Antwort nein: Erhöht man die Zahl der Felder für die Landsorte, so ist dies beim verwendeten Design gleichbedeutend mit einer Erhöhung der Zahl der Betriebe. Weiterhin bedeutet dies, falls man bei dem ursprünglichen Design bleibt, dass man gleichzeitig die Zahl der Felder für die neue Sorte genauso erhöht wie für die Landsorte.

3.13 α -Fehler und β -Fehler

Bisher hatten wir immer mit einem α -Fehler von 5% gearbeitet (Fehler 1. Art). Der α -Fehler gibt die Wahrscheinlichkeit an, dass die Nullhypothese verworfen wird, obwohl sie zutrifft. Das fälschliche Verwerfen der Nullhypothese ist aber nur ein möglicher Fehler, der begangen werden kann. Ein anderer möglicher Fehler tritt auf, wenn die Alternativhypothese zutrifft, wenn also beispielsweise Unterschiede zwischen zwei

Sorten bestehen, diese Unterschiede aber nicht nachgewiesen werden können. In diesem Fall spricht man von einem Fehler 2. Art oder auch vom β -Fehler. α -Fehler und β -Fehler hängen eng zusammen.

Wir könnten fordern, dass der α -Fehler nicht 5% betragen soll, sondern 1% oder 0,1% oder noch weniger. Für sehr kleines α wird der kritische Wert, also t_{Tab} im Fall eines t -Tests, relativ groß. Je kleiner das α , um so größer der kritische Wert. Je größer aber der kritische Wert, um so schwieriger ist es, tatsächliche Unterschiede nachzuweisen. Wenn tatsächliche Unterschiede nicht nachgewiesen werden können, begehen wir einen β -Fehler. Je kleiner der α -Fehler gewählt wird, um so größer wird der β -Fehler. Daher ist es nicht sinnvoll, den α -Fehler sehr klein zu wählen. Dies ist einer der Gründe, warum sich ein α -Fehler von 5% eingebürgert hat.

Entscheidung des Tests	Wirklichkeit	
	H_0 wahr	H_0 falsch
H_0 abgelehnt	α -Fehler	richtige Entscheidung
H_0 beibehalten	richtige Entscheidung	β -Fehler

Wahrscheinlichkeiten:

Entscheidung des Tests	Wirklichkeit	
	H_0 wahr	H_0 falsch
H_0 abgelehnt	α	$1 - \beta$
H_0 beibehalten	$1 - \alpha$	β

Die Wahl von α hat auch eine Bedeutung für Vertrauensintervalle. Die Vertrauenswahrscheinlichkeit ($1 - \alpha$) ist die Wahrscheinlichkeit, dass das angegebene Intervall den wahren Wert überdeckt. Wollen wir z.B. ganz sicher sein, dass das Intervall den wahren Mittelwert enthält, so müssen wir die Grenzen von minus bis plus Unendlich wählen. Ein solches Intervall hat natürlich keine Aussagekraft. Um überhaupt eine Aussage treffen zu können, müssen wir eine gewisse Wahrscheinlichkeit zulassen, dass unsere Aussage falsch ist. Die Wahrscheinlichkeit, dass die Aussage „Dieses Intervall überdeckt den wahren Wert“ falsch ist, entspricht der Irrtumswahrscheinlichkeit α . Man ist natürlich daran interessiert, ein möglichst enges Intervall angeben zu können. Je enger bei gegebenen Daten allerdings das Intervall gemacht wird, um so größer wird die Wahrscheinlichkeit α , dass das Intervall den gesuchten Wert nicht enthält.

Abb. 3.13.1 zeigt α - und β -Fehler für einen gepaarten t -Test (zweiseitig). Die Abbildung zeigt die Verteilung der mittleren Stichprobendifferenz \bar{d} bei einem Stichprobenumfang $n = n_0$. Es wird angenommen, dass die wahre Differenz unter der Alternative den Wert δ hat. Abb. 3.13.2 zeigt denselben Sachverhalt bei einer um den Faktor vier kleineren Stichprobengröße ($n = n_0/4$). Die Verteilungen sind wegen des kleineren Stichprobenumfanges "breiter". Während der α -Fehler unverändert ist, ist der β -Fehler deutlich höher. Übrigens hätte eine Vervielfachung der Fehlervarianz dieselbe Konsequenz wie die Viertelung des Stichprobenumfanges, nämlich eine flachere Wahrscheinlichkeitsdichte und damit einen größeren β -Fehler bei gegebenem α -Fehler.

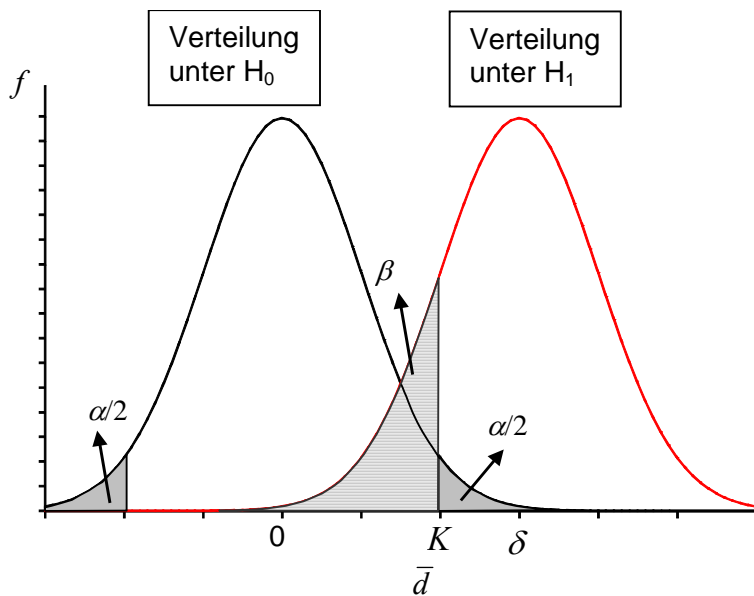


Abb. 3.13.1: Zweiseitiger t -Test - α - und β -Fehler. Stichprobenumfang $n = n_0$. K = kritische Schwelle.

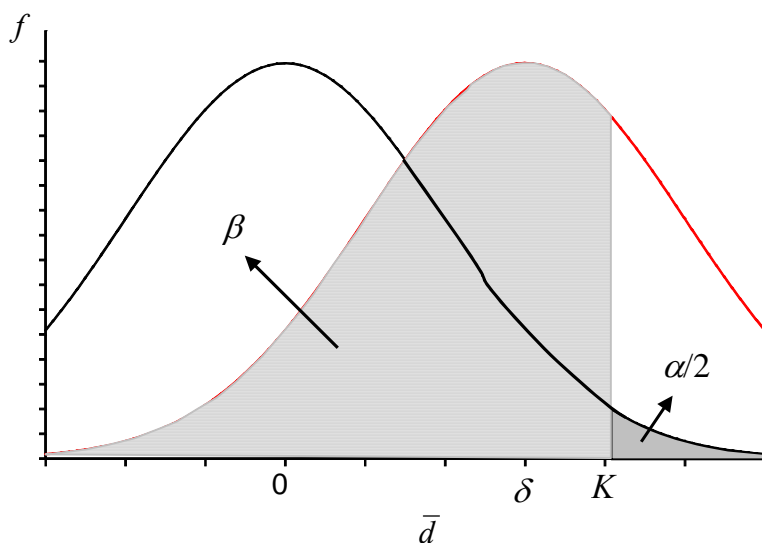


Abb. 3.13.2: Zweiseitiger t -Test - α - und β -Fehler. Viertelung des Stichprobenumfanges ($n = n_0/4$) bedeutet einen größeren β -Fehler. K = kritische Schwelle.

Neben dem Stichprobenumfang hat die Größe des Unterschiedes der Erwartungswerte δ einen maßgeblichen Einfluss auf den β -Fehler. Je größer δ ist, umso deutlicher sind die beiden Normalverteilungen getrennt und umso kleiner wird der β -Fehler. Für die Versuchsplanung hat dies die Konsequenz, dass man die zu prüfenden Behandlungen möglichst unterschiedlich wählen sollte, so dass δ einen großen Wert annimmt. Man kann dann Unterschiede relativ leicht nachweisen.

3.14 Stichprobenumfang für den unverbundenen t -Test

In Abschnitt 3.7 hatten wir die Stichprobenplanung anhand der gewünschten Breite eines Vertrauensintervalls für einen Mittelwert besprochen. Ein ähnliches Verfahren kann man verwenden, wenn die Differenz von zwei Mittelwerten geschätzt werden soll (hier nicht näher behandelt). Die Stichprobenplanung hängt immer vom Auswertungsverfahren ab. Für statistische Tests ist ein etwas anderes Verfahren zu verwenden. Die Abbildungen 3.13.1 und 3.13.2 sind der Schlüssel für eine Stichprobenplanung für den t -Test. Zur Planung des Stichprobenumfanges müssen folgende Größen vorgegeben werden:

- (1) α (Niveau/Irrtumswahrscheinlichkeit des Tests)
- (2) $(1 - \beta)$ (Güte)
- (3) δ (Kleinste nachzuweisende Differenz)
- (4) σ^2 (Fehlervarianz)

Aus diesen Vorgaben ergibt sich der

notwendige Stichprobenumfang für den unverbundenen t -Test:

$$n = 2 \frac{\sigma^2}{\delta^2} (z_{1-\alpha/2} + z_{1-\beta})^2$$

wobei z_γ das γ -Quantil der Standardnormalverteilung ist.

Die Quantile der Standardnormalverteilung sind wie folgt definiert:

$$P(Z < z_\gamma) = \gamma$$

Entsprechende Werte für verschiedene Wahrscheinlichkeiten γ sind in Tab. III angegeben (Siehe Anhang).

Beispiel: Wir möchten wissen, welcher Stichprobenumfang notwendig ist, um Sortendifferenzen von 5 dt/ha oder größer mit einer Wahrscheinlichkeit von mindestens 90% nachzuweisen, falls ein t -Test verwendet wird. Die Varianz ist aus einem Vorversuch bekannt und beträgt $\sigma^2 = 7,733$. Mit $z_{0,975} = 1,96$ und $z_{0,90} = 1,28$ finden wir:

$$n = 2 \frac{7,733}{5^2} (1,96 + 1,28)^2 = 6,5$$

Differenzen ≥ 5 werden mit einer Wahrscheinlichkeit von mindestens 90% nachgewiesen, wenn wir $n = 7$ wählen, wobei eine Irrtumswahrscheinlichkeit eines Fehlers 1. Art von $\alpha = 5\%$ eingehalten wird. \square

Bei der Stichprobenplanung kann sich herausstellen, dass der für eine vorgegebene Genauigkeit notwendige Stichprobenumfang nicht realisiert werden kann. Es bestehen dann verschiedene Möglichkeiten:

- Der Versuch wird nicht durchgeführt

Dies ist sicher die radikalste Reaktion. Oft ist es aber akzeptabel, die Genauigkeitsvorgaben so zu modifizieren, dass ein geringerer Stichprobenumfang resultiert. Die Möglichkeiten sind wie folgt:

- Die kleinste nachzuweisende Differenz δ wird erhöht
- Es wird versucht, die Versuchsbedingungen so zu verbessern, dass die Fehlervarianz σ^2 sinkt.
- Wir akzeptieren einen höheren Fehler 1. Art (α -Fehler)
- Wir akzeptieren einen höheren Fehler 2. Art (β -Fehler)/eine geringere Güte
- Kombination aus den vorangegangenen vier Komponenten

Auf den ersten Blick mag die Wahl einer dieser Möglichkeiten wie Selbstbetrug erscheinen. Tatsächlich ist es ein Missbrauch des Verfahrens, wenn wir mit den Anforderungen so lange herunter gehen, bis ein uns genehmer Stichprobenumfang herauskommt. Wir müssen uns bei jedem Schritt kritisch fragen, ob uns die erreichbare Genauigkeit noch ausreicht. Ist dies nicht der Fall, sollte der Versuch nicht durchgeführt werden.

Wir haben hier exemplarisch nur die Formel für den unverbundenen t-Test angegeben. Die Formel für den verbundenen t-Test ist ähnlich. Eine Sammlung von Formeln für die Stichprobenplanung bei einer Reihe von Auswertungsverfahren findet sich in: Rasch D et al., 1998, Verfahrensbibliothek, Verlag Oldenbourg. Außerdem gibt es Software, z.B. CADEMO, mit der sich die Stichprobenplanung für die verschiedensten Verfahren leicht durchführen lässt.

Erklärung der Formel: Um die Formel für den Stichprobenumfang zu verstehen, betrachten wir die Verteilung von \bar{d} unter H_0 und unter H_1 (siehe Abb. 3.13.1). Für die folgende Betrachtung ist es hilfreich, die beiden Verteilungen zu standardisieren. Im Fall einer unverbundenen Stichprobe hat die mittlere Differenz \bar{d} die Varianz

$$\text{var}(\bar{d}) = \text{var}(\bar{x}_1) + \text{var}(\bar{x}_2) = \frac{2\sigma^2}{n}$$

wobei n der Stichprobenumfang in beiden Stichproben ist und σ^2 die Varianz. Wir betrachten nun die Verteilung der standardisierten mittleren Differenz

$$z(\bar{d}) = \frac{\bar{d}}{\sqrt{\frac{2\sigma^2}{n}}}$$

Unter der Nullhypothese ist diese verteilt mit Varianz Eins und Erwartungswert

$$z(0) = \frac{0}{\sqrt{\frac{2\sigma^2}{n}}} = 0$$

Unter der Alternative ist die Varianz ebenfalls Eins, der Erwartungswert dagegen gleich

$$z(\delta) = \frac{\delta}{\sqrt{\frac{2\sigma^2}{n}}}$$

Die beiden standardisierten Verteilungen sind in Abb. 13.4.1 veranschaulicht. Man beachte, dass die Graphik bis auf die Standardisierung identisch mit Abb. 3.13.1 ist.

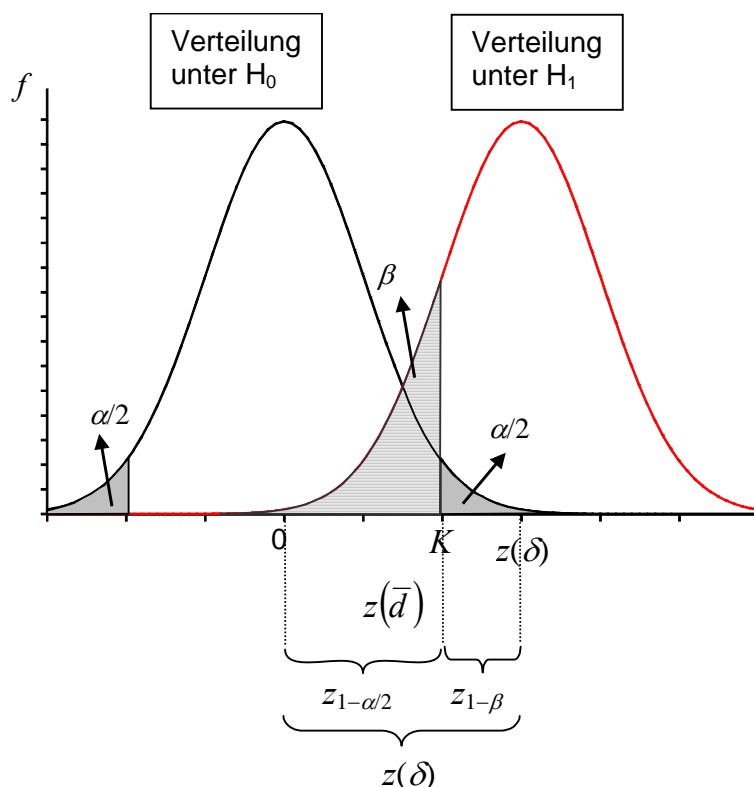


Abb. 3.14.1: Zweiseitiger t-Test - α - und β -Fehler. Standardisierte Differenz. K = kritische Schwelle.

Wegen der Standardisierung muss der Abstand vom Mittelwert Null unter H_0 und der kritischen Schwelle gleich $z_{1-\alpha/2}$ sein, während der Abstand von der kritischen Schwelle bis zum Mittelwert unter der Alternative gleich $z_{1-\beta}$ sein muss. Die Summe

der beiden Abstände muss aber gerade gleich $z(\delta)$ sein. Hieraus folgt eine Beziehung, aus der sich n für gegebenes α , β , δ und σ^2 gewinnen lässt:

$$z_{1-\alpha/2} + z_{1-\beta} = z(\delta) = \frac{\delta}{\sqrt{\frac{2\sigma^2}{n}}}$$

Quadrieren beider Seiten und Auflösen nach n liefert den gewünschten Stichprobenumfang (siehe Formel im obigen Kasten).

Relative Genauigkeitsvorgaben: Manchmal ist es leichter, Genauigkeitsvorgaben relativ auszudrücken. Dazu lässt sich die Formel leicht umstellen. Seien μ_1, μ_2 die wahren Mittelwerte, also $\delta = \mu_1 - \mu_2$. Dann ist

Gesamtmittel: $\bar{\mu}_\bullet = \frac{\mu_1 + \mu_2}{2}$

Variationskoeffizient: $v = \frac{\sigma}{\bar{\mu}_\bullet}$

Relative Differenz: $\delta_r = \frac{|\delta|}{\bar{\mu}_\bullet}$

Notwendiger Stichprobenumfang: $n = 2 \frac{v^2}{\delta_r^2} (z_{1-\alpha/2} + z_{1-\beta})^2$

Beispiel: Aus Vorversuchen ist bekannt, dass der Variationskoeffizient 7% beträgt. Um eine relative Differenz von 5% nachzuweisen, ist der nötige Stichprobenumfang gleich

$$n = 2 \frac{7^2}{5^2} (1,96 + 1,28)^2 \approx 41$$

3.14.1 Post-hoc Berechnung der Teststärke

Wenn der Versuch ohne Stichprobenplanung durchgeführt wird, weil die für die Planung notwendige Vorinformation über die Varianz fehlte, so kann man mit der oben angegebenen Formel im Nachhinein (*post hoc*) berechnen, welche Teststärke der Test hat. Wir schätzen die Varianz aus dem durchgeführten Versuch und berechnen dann die Teststärke für vorgegebene Werte von α und δ . Hierzu ist eine kleine Umstellung der Formel nach $z_{1-\beta}$ notwendig:

$$z_{1-\beta} = \sqrt{\frac{n\delta^2}{2\sigma^2}} - z_{1-\alpha/2}$$

Der Definition der Quantile z entnehmen wir:

$$P(Z < z_{1-\beta}) = 1 - \beta = \text{Teststärke}$$

Diese Wahrscheinlichkeit kann mit Hilfe von Tab. I (Anhang) leicht berechnet werden.

Post-hoc Berechnung der Teststärke

Fragestellung: Berechnung der Teststärke für einen durchgeführten Versuch.

Gegeben: Signifikanzniveau (α), Nachzuweisende Differenz (δ), Schätzwert der Varianz aus vorliegendem Versuch σ^2 .

Berechne:

$$z_{1-\beta} = \sqrt{\frac{n\delta^2}{2\sigma^2}} - z_{1-\alpha/2}$$

$$\text{Teststärke} = 1 - \beta = P(Z < z_{1-\beta})$$

$P(Z < z)$ kann mit Hilfe der Tab. I berechnet werden.

Voraussetzung: δ/σ ist nicht zu klein.

Beispiel: Ein Versuch zum Vergleich zweier Weizensorten wurde mit $n = 4$ Wiederholungen durchgeführt. Die Varianz wurde mit $7,733 \text{ (dt}^2/\text{ha}^2\text{)}$ geschätzt. Gesucht ist die Teststärke für eine nachzuweisende Differenz von $\delta = 5 \text{ dt/ha}$ bei einem Signifikanzniveau von $\alpha = 5\%$. Mit $z_{0,975} = 1,96$, $\sigma^2 = 7,733 \text{ (dt}^2/\text{ha}^2\text{)}$, $n = 4$ und $\delta = 5 \text{ dt/ha}$ finden wir:

$$z_{1-\beta} = \sqrt{\frac{4 * 5^2}{2 * 7,733}} - 1,96 = 0,58$$

$$\text{Teststärke} = 1 - \beta = P(Z < z_{1-\beta}) = P(Z < 0,58) = 1 - P(Z > 0,58) = 1 - 0,2810 = 0,7190$$

Die Wahrscheinlichkeit, eine Differenz von 5 dt/ha nachzuweisen, beträgt 72%. In der Abb. 3.14.1 ist die Teststärke für verschiedene nachzuweisende Differenzen δ wiedergegeben.

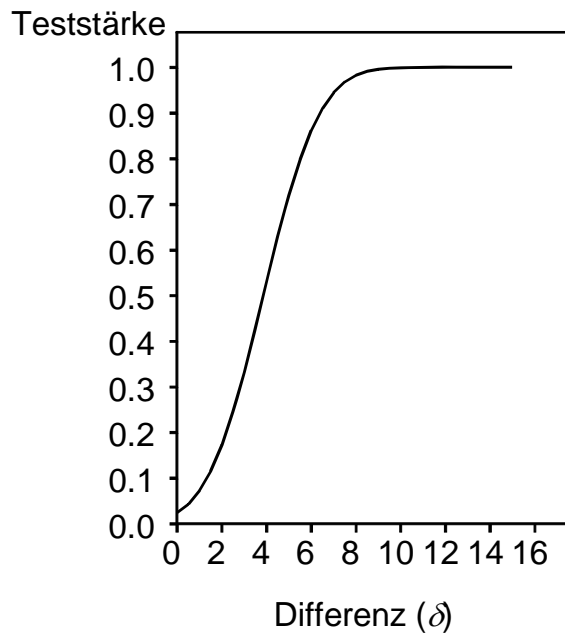


Abb. 3.14.1: Teststärke (approximativ) für $\alpha = 5\%$, $\sigma^2 = 7,733$, $n = 4$.

Man beachte, dass die Teststärke nahe $\delta = 0$ nahe bei 3% liegt. Der exakte Wert müsste 5% betragen. Dies liegt daran, dass die Approximation der t -Verteilung durch die Normalverteilung wegen des kleinen Stichprobenumfanges ungenau ist. Für eine Grobabschätzung reicht die Berechnung aber aus.

Obwohl diese Berechnung nichts im Sinne der nicht durchgeführten Versuchsplanung retten kann, so ermöglicht sie doch eine realistische Einschätzung der Aussagefähigkeit des Versuches.

3.15 Interpretation von Computeroutput – Überschreitungswahrscheinlichkeit (p-Wert)

Bisher wurden Vertrauensintervalle und Tests von Hand berechnet. Die Verfahren und Beispiele waren bisher auch relativ einfach. Oft hat man jedoch größere Datenmengen, und die Auswertungsverfahren können auch komplexer sein, als die hier beschriebenen. Dann wird man in der Regel die Auswertung mit dem Computer durchführen. Hier soll am Beispiel des Tomaten-Datensatzes aus Abschnitt 3.11 die Interpretation von Computeroutput erläutert werden. Die folgenden Anweisungen werden für das Statistikpaket SAS (Statistical Analysis System) benötigt, um für die Tomatendaten einen t -Test für unverbundene Stichproben durchzuführen. SAS ist ein Computerprogramm, das verschiedene statistische Verfahren umfasst. Es erfordert die Eingabe von Programmanweisungen in einer textorientierten Syntax.

```
data;
input duenger$ y;
datalines;
a      29.9
a      11.4
a      25.3
```

```

a      16.5
a      21.1
b      26.6
b      23.7
b      28.5
b      14.2
b      17.9
b      24.3
;
proc ttest;
class duenger;
var y; run;

```

Der Output ist wie folgt:

The TTEST Procedure							
Variable: y							
duenger		N	Mean	Std Dev	Std Err	Minimum	Maximum
a		5	20.8400	7.2456	3.2403	11.4000	29.9000
b		6	22.5333	5.4320	2.2176	14.2000	28.5000
Diff (1-2)			-1.6933	6.3028	3.8165		
duenger	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
a		20.8400	11.8435	29.8365	7.2456	4.3410	20.8205
b		22.5333	16.8328	28.2339	5.4320	3.3907	13.3226
Diff (1-2)	Pooled	-1.6933	-10.3269	6.9402	6.3028	4.3353	11.5064
Diff (1-2)	Satterthwaite	-1.6933	-10.8923	7.5056			
Equality of Variances							
Method		Variances	DF	t Value	Pr > t		
Pooled		Equal	9	-0.44	0.6677		
Satterthwaite		Unequal	7.3369	-0.43	0.6787		
Equality of Variances							
Method		Num DF	Den DF	F Value	Pr > F		
Folded F		4	5	1.78	0.5400		

Wir bekommen hier eine ganze Reihe von Informationen. Zunächst werden für beide Behandlungen Mittelwert (**mean**), Standardabweichung (standard deviation, **Std Dev**) sowie Standardfehler (standard error, **Std Err**) angegeben (zu den Begriffen Standardabweichung und Standardfehler siehe die Abschnitte 2.2.3 und 3.5). Der t -Test findet sich in der Tabelle am Ende des Output. Die erste Spalte dieser Tabelle ist mit **Variances** überschrieben. Die Einträge sind **Unequal** und **Equal**. Hierzu ist eine Erläuterung nötig. Bei dem in Abschnitt 3.11 besprochenen Test wurde die Annahme gemacht, dass die beiden Grundgesamtheiten sich nicht in ihrer Varianz unterscheiden. Diese Annahme ist für das Tomaten-Beispiel gerechtfertigt: Ein F -Test zeigt, dass keine signifikanten Unterschiede in der Varianz bestehen (siehe 3.19). Falls die Varianzen nicht gleich sind, muss der t -Test modifiziert werden. Der obige Output enthält beide Varianten des t -Tests. Für uns interessant ist hier die Zeile, die mit **Equal** beginnt. Dort ist das Ergebnis für den einfachen t -Test aus Abschnitt 3.11 aufgeführt.

Unter der Spaltenüberschrift **T** finden wir $t = -0,4437$. Da wir für den Test den Betrag verwenden, interpretieren wird dies als $t_{Vers} = 0,4437$. Unter **DF** stehen die Freiheitsgrade $n_1 + n_2 - 2 = 9$. Die Abkürzung **DF** steht für degrees of freedom (Freiheitsgrade). Wir würden nun als nächstes den kritischen t -Wert (t_{Tab}) im Output erwarten. Statt dessen bekommen wir eine Spalte mit der Überschrift **Prob>|T|**. Hierbei handelt es sich um eine sog. **Überschreitungswahrscheinlichkeit**, oft kurz als **p -Wert** bezeichnet. Der p -Wert ist wie folgt zu interpretieren: Unser Test ist signifikant, wenn der p -Wert kleiner als das vorgegebene α ist. Der p -Wert beträgt hier 0,6677. Wenn wir $\alpha = 5\% = 0,05$ wählen, dann ist der p -Wert $> \alpha$ und der Test ist nicht signifikant. Wir kommen hier natürlich zu demselben Ergebnis wie in Abschnitt 3.11. Dort mussten wir zunächst einen kritischen t -Wert (t_{Tab}) bestimmen und mit t_{Vers} vergleichen. Der Computer dagegen berechnet einen p -Wert, anhand dessen die Signifikanz beurteilt werden kann.

Der p -Wert ist die Wahrscheinlichkeit, einen t -Wert zu beobachten, der größer als der aus den Daten berechnete Wert (t_{Vers}) ist, unter der Voraussetzung, dass die Nullhypothese zutrifft. Dies ist in der Abb. 3.15.1 veranschaulicht. Dort ist $t_{Vers} > t_{Tab}$, so dass $p < \alpha$, und die Nullhypothese wird verworfen. Kleine p -Werte bedeuten, dass der t -Wert unwahrscheinlich ist bei Gültigkeit der Nullhypothese; damit ist der t -Wert schlecht mit der Nullhypothese vereinbar, so dass diese verworfen wird.

Die hier beschriebene Interpretation von p -Werten ist ganz allgemein gültig. Statistikprogramme geben die Ergebnisse eines Tests immer in Form von p -Werten (Überschreitungswahrscheinlichkeiten) an. Daher ist es wichtig, sich mit der Interpretation von p -Werten vertraut zu machen. Das Vorgehen bei der Interpretation von p -Werten ist im folgenden Kasten zusammengefasst:

- (1) Lege für den durchzuführenden Test ein Signifikanzniveau α fest
- (2) Führe den Test mit dem Computer durch. Lies den p -Wert (Überschreitungswahrscheinlichkeit) aus dem Output ab.
- (3) Falls $p\text{-Wert} < \alpha$, verwerfe H_0 des Tests
Falls $p\text{-Wert} \geq \alpha$, behalte H_0 des Tests bei

Bemerkung: Diese Regel zur Interpretation von p -Werten gilt für jeden mittels Computer durchgeführten Signifikanztest, nicht nur für den hier betrachteten t -Test. Generell werden Ergebnisse von Signifikanztests mit dem Computer in der hier beschriebenen Form ausgegeben. In Kurzform lautet die Entscheidungsregel:

$p\text{-Wert} < \alpha \Rightarrow \text{signifikant}$

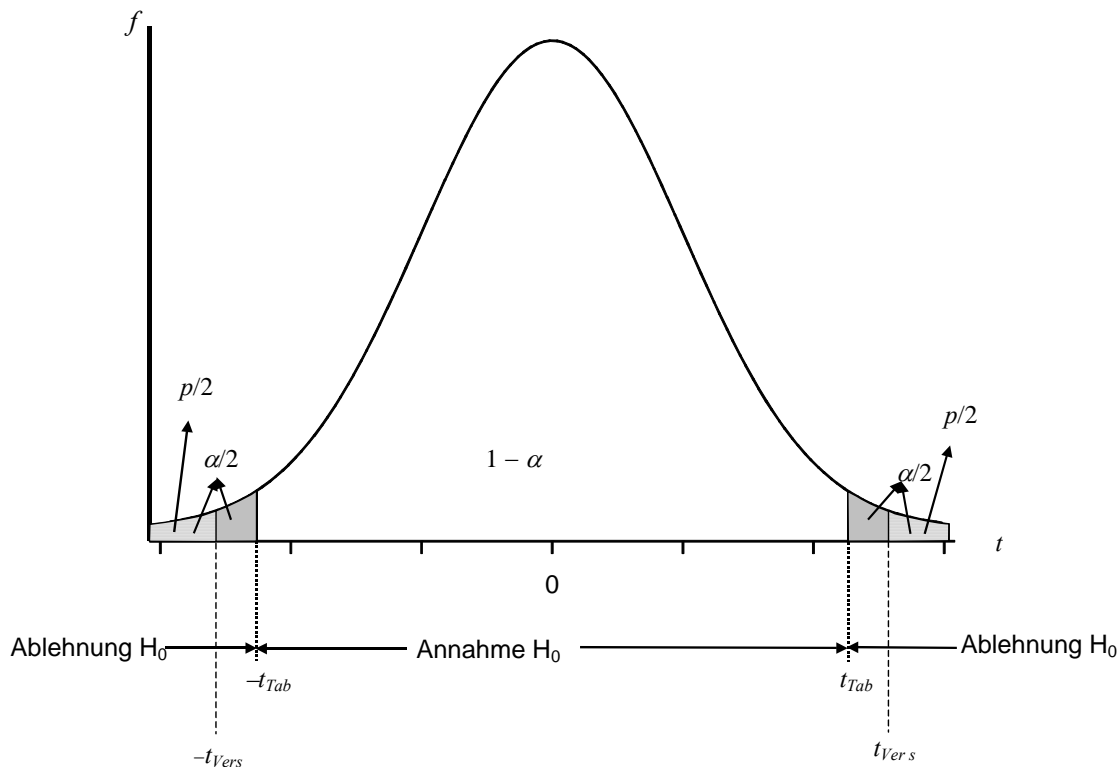


Abb. 3.15.1: p -Wert für einen zweiseitigen t -Test. t_{Vers} hier ohne Betragszeichen im Zähler berechnet, so dass auch negative Werte zu betrachten sind (zweiseitige Betrachtung). Für die praktische Durchführung ist die Berechnung mit Betragszeichen einfacher (siehe 3.10 und 3.11), Ergebnis des Tests ist aber im Ergebnis identisch.

3.16 Allgemeine Konvention für Angabe von Signifikanzen

Das Ergebnis eines Signifikanztests kann mit Hilfe hochgestellter Symbole, welche an den Wert der Teststatistik geschrieben werden, gekennzeichnet werden. Hierfür hat sich folgende Konvention eingebürgert:

$p < 0,001 \Rightarrow$	***	= höchstsignifikant
$0,001 \leq p < 0,01 \Rightarrow$	**	= hochsignifikant
$0,01 \leq p < 0,05 \Rightarrow$	*	= signifikant
$p \geq 0,05 \Rightarrow$	ns	= nicht signifikant

(p = Überschreitungswahrscheinlichkeit)

Beispiel: Für das Beispiel in Abschnitt 3.14 hatten wir einen t -Wert von $t_{Vers} = -0,4437$ ermittelt, welcher nicht signifikant ist. Dies kann durch die Schreibweise

$$t_{Vers} = -0,4437^{ns}$$

kenntlich gemacht werden.

3.17 Test des Parameters μ

Fragestellung: Ist der Mittelwert μ der Grundgesamtheit, aus der eine Stichprobe gezogen wurde, signifikant von einem theoretischen Wert μ_0 verschieden?

$H_0: \mu = \mu_0$ (zweiseitige Fragestellung)

$H_1: \mu \neq \mu_0$

Voraussetzungen: Die Grundgesamtheit ist normalverteilt mit unbekannter Varianz.

Rechenweg:

(1) Berechne

$$t_{Vers} = \frac{|\bar{x} - \mu_0|}{\sqrt{s_x^2/n}} \quad \text{wobei}$$

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

x_i ($i = 1, \dots, n$) = Messwerte

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

(2) Lies aus der t -Tabelle $t_{Tab}(FG; \alpha)$ ab (Tab. II, zweiseitig), wobei α das Signifikanzniveau und $FG = n - 1$ ist.

(3) Vergleiche t_{Vers} und t_{Tab} :

$t_{Vers} \leq t_{Tab} \Rightarrow H_0 (\mu = \mu_0)$ (keine signifikanter Unterschied)

$t_{Vers} > t_{Tab} \Rightarrow H_1 (\mu \neq \mu_0)$ (signifikanter Unterschied)

Beispiel: In einer Umfrage unter Studenten an der GhK in Witzenhausen wurde anhand einer Stichprobe von 80 Studenten das durchschnittliche verfügbare Einkommen erhoben (Finkenzeller, Mai 2001, pers. Mitteilung). Die Stichprobe ergab einen Mittelwert von 978 DM/Monat und eine Standardabweichung von 221 DM/Monat. Der Bundesdurchschnitt des verfügbaren Einkommens pro Student liegt bei 1300 DM/Monat. Weicht Witzenhausen signifikant vom Bundesdurchschnitt ab?

$$t_{Vers} = \frac{|\bar{x} - \mu_0|}{s_x} \sqrt{n} = \frac{|978 - 1300|}{221} \sqrt{80} = 13,03$$

$$t_{Tab} = 1,99 < t_{Vers}$$

⇒ Das Einkommen der Studenten in Witzenhausen liegt signifikant unter dem Bundesdurchschnitt.

3.18 Vertrauensintervall für eine Varianz

Wie der Stichprobenmittelwert \bar{x} , so hat auch die Stichprobenvarianz s_x^2 eine statistische Verteilung. Wenn x_i eine normalverteilte Zufallsvariable mit Mittelwert μ und Varianz σ_x^2 ist, dann hat die Größe

$$\chi^2 = \frac{(n-1)s_x^2}{\sigma_x^2}$$

eine χ^2 -Verteilung mit $\nu = n - 1$ Freiheitsgraden. Da sowohl die Stichprobenvarianz s_x^2 als auch die theoretische Varianz σ_x^2 immer positive Größen sind, ist χ^2 ebenfalls immer positiv. Außerdem ist die χ^2 -Verteilung nicht symmetrisch, wie die nachfolgende Graphik für $\nu = 4$ Freiheitsgrade zeigt (Abb. 3.18.1).

Wahrscheinlichkeitsdichte

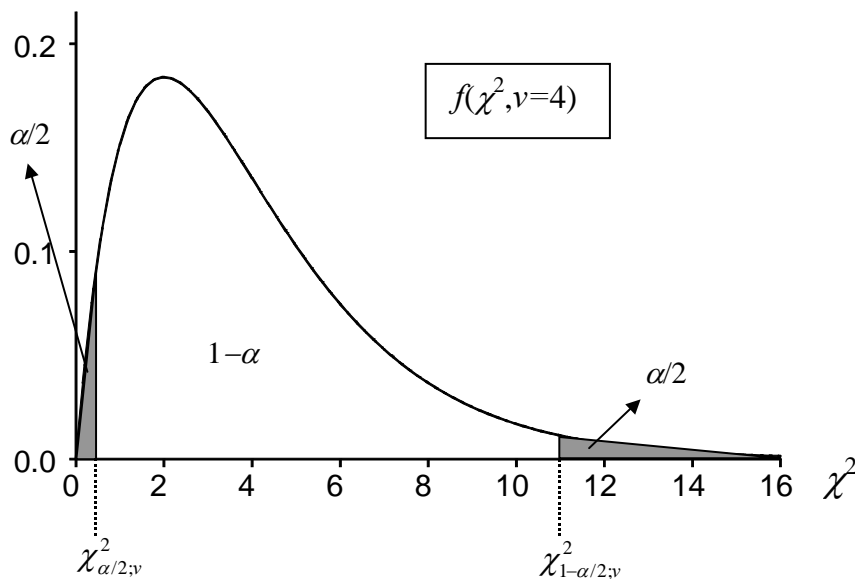


Abb. 3.18.1: Wahrscheinlichkeitsdichte der χ^2 -Verteilung mit $\nu = 4$ FG.

Kritische Werte $\chi_{\gamma;\nu}^2$ der χ^2 -Verteilung, so dass $P(\chi^2 < \chi_{\gamma;\nu}^2) = \gamma$, sind in Tab. IV (Anhang) zu finden.

Es gilt:

$$P\left(\chi_{\alpha/2;\nu}^2 \leq \frac{(n-1)s_x^2}{\sigma_x^2} \leq \chi_{1-\alpha/2;\nu}^2\right) = 1 - \alpha$$

Hieraus ergibt sich durch algebraische Umformung das

(1- α)100%-Vertrauensintervall für eine Stichprobenvarianz

$$P\left(\frac{(n-1)s_x^2}{\chi_{1-\alpha/2;v}^2} \leq \sigma_x^2 \leq \frac{(n-1)s_x^2}{\chi_{\alpha/2;v}^2}\right) = 1 - \alpha$$

Beispiel: Für die Stichprobe von $n = 50$ Maispflanzen hatten wir eine Varianz der Pflanzenlängen von $s_x^2 = 100$ gefunden. Hierfür soll ein 95%iges Vertrauensintervall berechnet werden.

$$\alpha = 0,05$$

$$n = 50$$

$$v = 49$$

$$\chi_{\alpha/2;v}^2 = \chi_{0,025;v=49}^2 = 31,555; \chi_{1-\alpha/2;v}^2 = \chi_{0,975;v=49}^2 = 70,222$$

$$P\left(\frac{(n-1)s_x^2}{\chi_{1-\alpha/2;v}^2} \leq \sigma_x^2 \leq \frac{(n-1)s_x^2}{\chi_{\alpha/2;v}^2}\right) = P\left(\frac{49 \cdot 100}{70,222} \leq \sigma_x^2 \leq \frac{49 \cdot 100}{31,555}\right) = P(69,78 \leq \sigma_x^2 \leq 155,28) = 0,95$$

Mit 95% Vertrauenswahrscheinlichkeit wird die wahre Varianz vom Intervall mit den Grenzen 69,78 und 155,28 überdeckt.

3.19 Test zum Vergleich zweier unabhängiger Stichprobenvarianzen

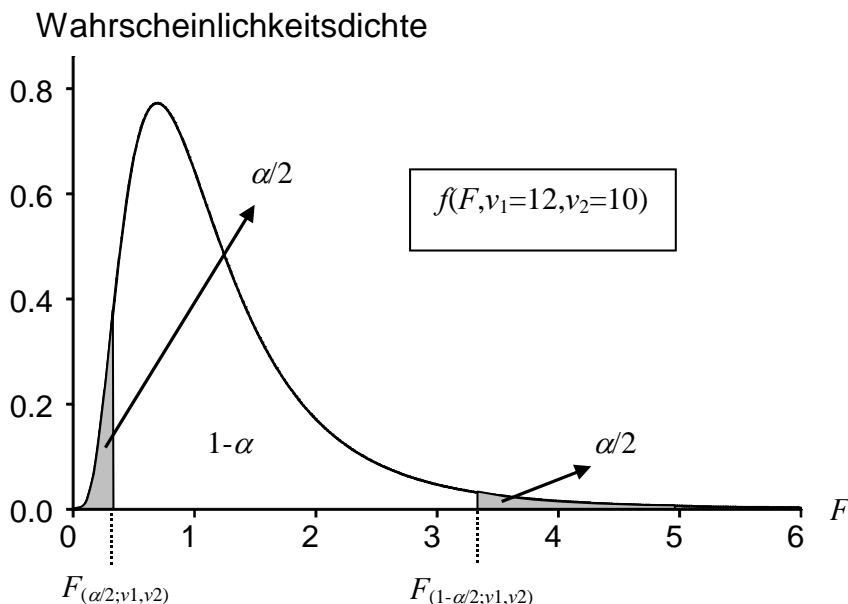


Abb. 3.19.1: F-Verteilung mit $v_1 = 12$, $v_2 = 10$ Freiheitsgraden.

Wenn x_{i1} eine normalverteilte Zufallsvariable mit Mittelwert μ_1 und Varianz σ_x^2 ist, und wenn x_{i2} eine normalverteilte Zufallsvariable mit Mittelwert μ_2 und derselben (!)

Varianz σ_x^2 ist (Nullhypothese), dann hat die Größe

$$F = \frac{s_1^2}{s_2^2}$$

eine F -Verteilung mit $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ Freiheitsgraden, wobei n_1 und n_2 die Umfänge der beiden Stichproben x_{i1} und x_{i2} sind und s_1^2 und s_2^2 die beiden Stichprobenvarianzen. Die Wahrscheinlichkeitsdichte einer F -Verteilung ist nachstehend abgebildet (Abb. 3.19.1).

Es kann gezeigt werden, dass für den Test der Nullhypothese, dass beide Varianzen gleich groß sind, es ausreichend ist, in der F -Statistik die größere Stichprobenvarianz in den Zähler zu schreiben. Hierdurch muss nur ein kritischer Wert bestimmt werden anstatt zwei. Dies ergibt sich aus der Tatsache, dass sowohl F als auch $1/F$ einer F -Verteilung folgen. Kritische Werte $F_{(0,975; \nu_1, \nu_2)}$ sind in Tab. V (Anhang) tabelliert.

Voraussetzung: Gegeben sind zwei unabhängige Stichproben aus normalverteilten Grundgesamtheiten.

H_0 : Beide Stichproben haben dieselbe Varianz.

s_1^2, s_2^2 = Stichprobenvarianzen

$\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ = Freiheitsgrade von s_1^2 und s_2^2

n_1, n_2 = Stichprobenumfänge

Berechne:

$$F_{Vers} = \frac{s_1^2}{s_2^2}$$

wobei s_1^2 die größere Stichprobenvarianz ist (!).

Bestimme:

$$F_{Tab} = F_{(1-\alpha/2; \nu_1, \nu_2)} \quad (\text{Tab. V})$$

Falls $F_{Vers} \leq F_{Tab} \Rightarrow$ Behalte H_0 bei

Falls $F_{Vers} > F_{Tab} \Rightarrow$ Verwerfe H_0

Beispiel: In einem Feldversuch werden Tomatenerträge bei zwei verschiedenen Düngerbehandlungen A und B verglichen. Dünger A wird auf 5 Parzellen getestet, Dünger B auf 6. Die Verteilung der Behandlungen A und B auf die insgesamt 11 Parzellen ist dabei zufällig (Randomisation). Die Parzellenerträge sind wie folgt:

	A	B
	29,9	26,6
	11,4	23,7
	25,3	28,5
	16,5	14,2
	21,1	17,9
		24,3
$\sum x$	104,2	135,2
$\sum x^2$	2381,51	3194,04
	$\bar{x}_1 = 20,84$	$\bar{x}_2 = 22,53$
	$n_1 = 5$	$n_2 = 6$

$$s_1^2 = (2381,51 - (104,2)^2/5)/(5 - 1) = 209,99/4 = 52,50$$

$$s_2^2 = (3194,04 - (135,2)^2/6)/(6 - 1) = 147,53/5 = 29,51$$

Wir testen die Nullhypothese, dass beide Sorten dieselbe Varianz haben, bei einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$.

$$F_{Tab} = F_{(0,975; 4, 5)} = 7,39$$

$$F_{Vers} = \frac{52,50}{29,51} = 1,78 < F_{Tab} = 7,39 \Rightarrow H_0$$

Die Nullhypothese gleicher Varianzen wird nicht verworfen.

Wozu ein Vergleich der Varianzen? Zum einen ist Gleichheit der Varianzen Voraussetzung vieler anderer Tests, z.B. für den t -Test bei unabhängigen Stichproben. Ist die Voraussetzung verletzt, hält dieser Test nicht die vorgegebene Irrtumswahrscheinlichkeit ein und ist damit ungültig. Es ist generell bei jedem Test wichtig, sich von der Gültigkeit der gemachten Voraussetzungen zu überzeugen. Der hier gezeigte F-Test ist eine Möglichkeit, die Annahme der Varianzhomogenität zu prüfen (siehe Abschnitt 3.15).

In anderen Fällen stehen die Varianzen im Zentrum der Fragestellung. So kann es beispielsweise von Interesse sein, zu ermitteln, ob ein Produktionssystem A eine größere "Stabilität", und damit eine geringere Varianz hat als ein System B.

Der F -Test zum Vergleich zweier Varianzen ist sehr empfindlich gegenüber einer Verletzung der Voraussetzung der Normalverteilung. Es gibt zahlreiche alternative Tests.

Für den Vergleich von Varianzen in **verbundenen Stichproben** gibt es ebenfalls zahlreiche Tests [Piepho, H.P. (1997): Tests for equality of dispersion in correlated samples - Review and empirical comparison. *Journal of Statistical Computation and Simulation* **56**, 353-372.], auf die hier aber nicht eingegangen wird.

3.20 Einseitige und zweiseitige Tests

Bei der Diskussion statistischer Tests haben wir bisher immer zweiseitige Ablehnungsbereiche gehabt. So haben wir bei Vergleich zweier Mittelwerte die Nullhypothese

$$H_0: \mu_1 = \mu_2$$

gegen die Alternativhypothese (H_1)

$$H_1: \mu_1 \neq \mu_2$$

geprüft. Die Nullhypothese ist verletzt, wenn $\mu_1 > \mu_2$ ist oder $\mu_1 < \mu_2$. Dementsprechend ist der Ablehnungsbereich zweiseitig (Abb. 3.10.1 in Abschnitt 3.10).

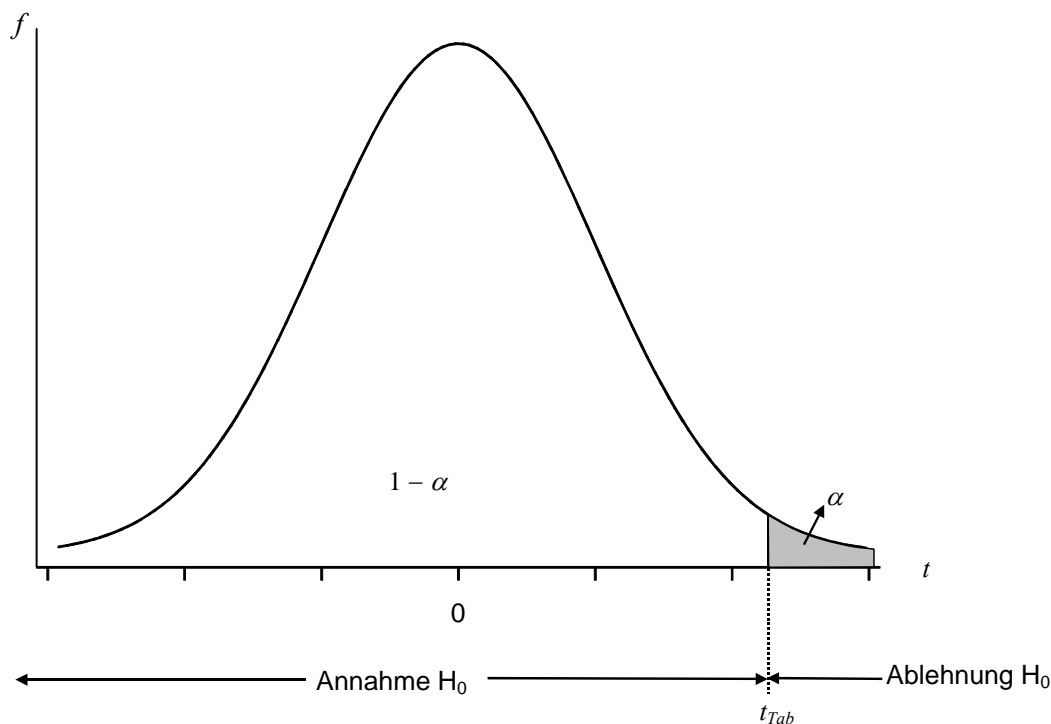


Abb. 3.20.1: Annahme- und Ablehnungsbereich bei einseitiger Fragestellung.

In manchen Fällen liegt dagegen eine einseitige Fragestellung vor (Abb. 3.20.1).

Beispiel: Ein Pflanzenschutzmittelhersteller hat eine neue Verbindung synthetisiert, die gegen Mehltau wirken soll. Der Befall von Getreide bei Behandlung mit diesem neuen Mittel sei μ_1 . Das Mittel ist nur dann von Interesse, wenn der Befall niedriger ist als bei Behandlung mit dem bisher üblichen Mittel. Letzterer Befall sei mit μ_2 bezeichnet. Das neue Mittel soll also nur dann vermarktet werden, wenn $\mu_1 < \mu_2$. In diesem Fall kann die Nullhypothese formuliert werden als

$$H_0: \mu_1 \geq \mu_2 \text{ (Das neue Mittel ist nicht besser)} \quad \text{bzw.} \quad \delta = \mu_2 - \mu_1 \leq 0$$

mit Alternative

$H_1: \mu_1 < \mu_2$ (Das neue Mittel ist besser) bzw. $\delta = \mu_2 - \mu_1 > 0$

In diesem Fall hat der t-Test einen einseitigen Ablehnungsbereich, wie Abb. 3.20.1 zeigt. Entsprechende kritische t-Werte sind gesondert tabelliert [Tab. II(b)]. Benutzt man die t-Werte aus Tab. II, so können die dort angegebenen α -Werte halbiert werden, um die für den einseitigen Test zutreffenden α -Werte zu erhalten.

Wegen der einseitigen Fragestellung ist ein kleinerer kritischer t-Wert zu wählen, und es wird leichter, signifikante Ergebnisse zu bekommen. Dies mag wie Zauberei erscheinen. Ist es aber nicht. Wir zahlen einen Preis: wir können bei Beibehaltung der Nullhypothese keine Aussage darüber treffen ob das neue Mittel genau so gut ist wie das alte, oder ob es schlechter ist. Eine solche Aussage erfordert einen zweiseitigen Test. Das Beispiel macht auch klar, dass es von der Fragestellung abhängt, ob einseitig oder zweiseitig zu testen ist. Im obigen Beispiel wird aus Sicht des Herstellers eine einseitige Fragestellung vorliegen. Ein Wissenschaftler wird dagegen an der zweiseitigen Fragestellung interessiert sein. Nach meinem Dafürhalten wird im wissenschaftlichen Bereich oft eher eine zweiseitige Fragestellung vorliegen. Wenn man im Zweifel ist, sollte man auf jeden Fall zweiseitig testen. In Kap. 3.17 hatten wir übrigens ein Beispiel behandelt, bei dem zweiseitig getestet wurde (ohne das dies dort explizit so benannt wurde), weil Abweichungen in beiden Richtungen von Interesse waren.

Als Beispiel für einen einseitigen Test geben wir hier den Rechengang für das Einstichprobenproblem (Mittelwert) an. Der Rechengang für einseitige Tests zum Vergleich zweier Stichproben und zum Vergleich von Varianzen ist analog.

Fragestellung: Ist der Mittelwert μ der Grundgesamtheit, aus der eine Stichprobe gezogen wurde, signifikant größer als ein theoretischer Wert μ_0 ?

$H_0: \mu \leq \mu_0$ (einseitige Fragestellung)

$H_1: \mu > \mu_0$

Voraussetzungen: Die Grundgesamtheit ist normalverteilt mit unbekannter Varianz.

Rechenweg:

(1) Berechne

$$t_{Vers} = \frac{\bar{x} - \mu_0}{\sqrt{s_x^2/n}} \quad \text{wobei}$$

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

x_i ($i = 1, \dots, n$) = Messwerte

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

(2) Lies aus der t -Tabelle $t_{Tab}(FG; \alpha)$ ab (Tab. II(b), einseitig!), wobei α das Signifikanzniveau und $FG = n - 1$ ist.

(3) Vergleiche t_{Vers} und t_{Tab} :

$t_{Vers} \leq t_{Tab} \Rightarrow H_0 (\mu \leq \mu_0)$ (nicht signifikant größer)

$t_{Vers} > t_{Tab} \Rightarrow H_1 (\mu > \mu_0)$ (signifikant größer)

Beispiel: Im Hessischen Ried sind Schnaken ein großes Problem. In stehenden Gewässern werden regelmäßig Schöpfproben à 350 ml genommen. Wenn mehr als 10 Schnakenlarven je 350 ml vorliegen, wird mit einem Insektizid bekämpft (Frankfurter Rundschau vom 22. Juni 2001, S. 29). In einem Tümpel wurden 20 Schöpfproben genommen. Die durchschnittliche Zahl von Larven je 350 ml betrug 11, die Standardabweichung war 5,5 Larven je 350 ml. Muss hier bekämpft werden?

$$\mu_0 = 10$$

$$H_0: \mu \leq \mu_0$$

$$H_1: \mu > \mu_0$$

$$\bar{x} = 11$$

$$s_x = 5,5$$

$$n = 20$$

$$t_{Tab}(FG = 19, \alpha = 5\%) = 1,729$$

$$t_{Vers} = \frac{11 - 10}{5,5 / \sqrt{20}} = 0,813 < t_{Tab}$$

H_0 wird beibehalten. Es muss nicht bekämpft werden, da die Bekämpfungsschwelle nicht signifikant überschritten wurde.

3.21 Äquivalenztest am Beispiel zweier unverbundener Stichproben

Beispiel: Eine ingezüchtete Mais-Linie, in die ein Gen für das Toxin von *Bacillus thuringiensis* (bt) eingebaut wurde, soll auf ihre ökologische Verträglichkeit hin untersucht werden. Das bt-Toxin soll den Maiszünsler, einen wichtigen Schädling an Mais, abtöten. Allerdings besteht der Verdacht, dass auch andere Schmetterlingsarten in Mitleidenschaft gezogen werden. Man muss daher den Nachweis führen, dass die Verwendung von "bt-Mais" nicht zu einer Reduzierung oder Erhöhung der Anzahl anderer Schmetterlingsarten kommt. Hierzu wird bt-Mais mit einer isogenen Mais-Linie verglichen, die dasselbe Genom hat, mit der Ausnahme, dass das Gen für das bt-Toxin fehlt. Beide Linien werden auf jeweils fünf zufällig ausgewählten Feldern angebaut. Auf jedem Feld wird die

Populationsdichte verschiedener Schmetterlingsarten mit Hilfe von Fangkäfigen erfasst. Die Rohdaten (Anzahl/ 100 m²) und die beiden Mittelwerte sind im folgenden wiedergegeben:

Bt-Linie	Isogene Linie	
12	10	
15	17	
13	12	
10	9	
8	16	
11,6	12,8	(Mittelwert)

Der Mittelwert für bt-Mais ist etwas reduziert. Ein t-Test der Nullhypothese H_0 : "Beide Mittelwerte sind gleich" für unverbundene Stichproben (siehe Abschnitt 3.11) liefert $t_{\text{Vers}} = 0,60$, was zum Niveau $\alpha = 5\%$ nicht signifikant ist [$t_{\text{Tab}}(\text{FG}=8, \alpha = 5\%) = 2,306$]. Dies ist allerdings keinesfalls ein Nachweis, dass die Nullhypothese zutrifft (siehe Abschnitt 3.10). Wenn das so einfach wäre, könnte man die Gleichheit von zwei Behandlungen immer ganz leicht nachweisen, indem man ein Experiment möglichst schlecht durchführt, so dass der Versuchsfehler möglichst groß ist und die Stichprobe möglichst klein. Die Teststärke (Abschnitte 3.13 und 3.14) wäre dann minimal, und man würde H_0 so gut wie nie verwerfen.

Um hier weiter zukommen, ist es notwendig, die Rolle von H_0 und Alternativhypothese (H_A) im Prinzip umzudrehen. Hierbei muss die Nullhypothese die Nicht-Gleichheit der beiden Mittelwerte umfassen. Allerdings ergibt sich bei der zugehörigen Alternative $H_A: \mu_1 - \mu_2 = 0$ das Problem, dass diese nur einen Punkt auf der reellen Achse umfasst. Es ist praktisch unmöglich, nachzuweisen, dass die Differenz einen ganz bestimmten Wert annimmt, denn dazu müsste die Stichprobe unendlich groß sein. Eine praktikable Lösung für dieses Problem ist dann möglich, wenn man den Begriff der Gleichheit durch den der Äquivalenz ersetzt. Zwei Mittelwerte sind dann als äquivalent zu betrachten, wenn deren Differenz sich innerhalb enger Grenzen um die Null herum bewegt, dass also

$$-\delta < \mu_1 - \mu_2 < \delta$$

für ein festzulegendes $\delta > 0$ ist. Diese Definition von Äquivalenz ist in Abb. 3.21.1 veranschaulicht.

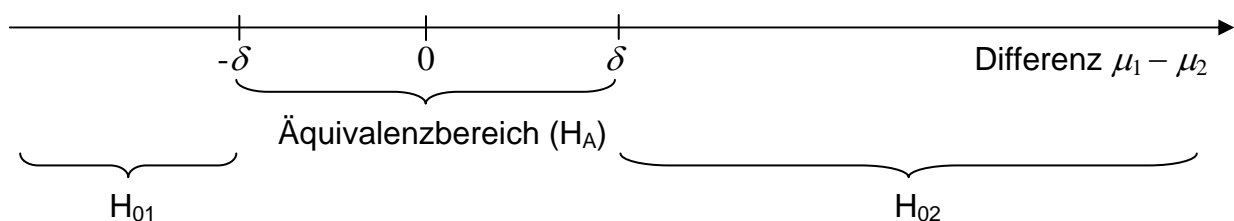


Abb. 3.21.1: Veranschaulichung des Begriffs der Äquivalenz.

Ein Test muss nun so formuliert werden, dass die Äquivalenz der Alternative (H_A) entspricht, also

$$H_A: -\delta < \mu_1 - \mu_2 < \delta$$

$$H_0: \mu_1 - \mu_2 < -\delta \text{ oder } \mu_1 - \mu_2 > \delta$$

Die Nullhypothese fällt hier nun in zwei getrennte Bereiche, nämlich

$$H_{01}: \mu_1 - \mu_2 < -\delta$$

$$H_{02}: \mu_1 - \mu_2 > \delta$$

Ein gültiger Test kann nun erhalten werden, indem diese beiden einfachen Nullhypothesen gegen die jeweils zugehörigen einfachen Alternativen

$$H_{A1}: \mu_1 - \mu_2 > -\delta$$

$$H_{A2}: \mu_1 - \mu_2 < \delta$$

Getestet werden. Es sind also folgende beiden einseitigen Tests durchzuführen:

$$H_{01}: \mu_1 - \mu_2 < -\delta \text{ gegen } H_{A1}: \mu_1 - \mu_2 > -\delta \quad \text{und}$$

$$H_{02}: \mu_1 - \mu_2 > \delta \text{ gegen } H_{A2}: \mu_1 - \mu_2 < \delta$$

Man beachte, dass die **Schnittmenge** (engl. **intersection**) der beiden einfachen Alternativen H_{A1} und H_{A2} die Alternative H_A ergibt. Denn wenn **sowohl** H_{A1} **als auch** H_{A2} gilt, so ist dies gleichbedeutend mit H_A . Man kann dies auch formal schreiben als

$$H_A = H_{A1} \cap H_{A2} \quad .$$

[Das Symbol \cap stammt aus der Mengenlehre und steht für die Bildung einer Schnittmenge.] Man beachte weiterhin, dass die Nullhypothese H_0 sich als **Vereinigung** (engl. **union**) der beiden einfachen Nullhypothesen H_{01} und H_{02} ergibt. Denn wenn **entweder** H_{01} **oder** H_{02} gilt, und H_{01} und H_{02} sich gegenseitig ausschließen, so ist dies gleichbedeutend mit H_0 . Dies kann auch wie folgt ausgedrückt werden

$$H_0 = H_{01} \cup H_{02} \quad .$$

[Das Symbol \cup steht für die Vereinigung zweier Mengen.] Man kann daher H_0 vs. H_A testen, indem man zwei einfache Tests für H_{01} gegen H_{A1} und H_{02} gegen H_{A2} durchführt. Nur wenn beide Nullhypothesen (H_{01} und H_{02}) abgelehnt werden, ist H_A (Äquivalenz) nachgewiesen. Wegen der Beziehung der Hypothesen der einfachen Tests zu den kombinierten Hypothesen bezeichnet man das hier verwendete Testprinzip auch als **Intersection-Union-Prinzip**. Die beiden einseitigen Tests kann man mit dem einfachen t-Test aus Abschnitt 3.20 durchführen, wie im folgenden Kasten erläutert. Da sich H_{01} und H_{02} gegenseitig ausschließen, darf man die zwei einseitigen Tests jeweils zum Niveau α durchführen und kontrolliert dabei trotz des multiplen Testens das Niveau insgesamt beim Niveau α .

Fragestellung: Sind die Mittelwerte μ_1 und μ_2 zweier Grundgesamtheiten 1 und 2, aus denen jeweils eine Stichprobe gezogen wurde, äquivalent?

$$H_A: -\delta < \mu_1 - \mu_2 < \delta \quad \quad \quad (\text{Die Mittelwerte sind äquivalent})$$

$H_0: \mu_1 - \mu_2 < -\delta$ oder $\mu_1 - \mu_2 > \delta$ (Die Mittelwerte sind nicht äquivalent)

Voraussetzungen: Beide Grundgesamtheiten sind normalverteilt mit gleicher, unbekannter Varianz. Die Stichproben sind unabhängig.

Rechenweg:

(1) Lege die Äquivalenzgrenzen $-\delta$ und δ fest

(2) Berechne

$$t_{Vers}^{(1)} = \frac{\bar{x}_1 - \bar{x}_2 + \delta}{s\sqrt{1/n_1 + 1/n_2}} \quad \text{und} \quad t_{Vers}^{(2)} = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{s\sqrt{1/n_1 + 1/n_2}} \quad \text{wobei}$$

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1}, \quad \bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2},$$

x_{1i} ($i = 1, \dots, n_1$) = Messwerte der 1. Behandlung

x_{2i} ($i = 1, \dots, n_2$) = Messwerte der 2. Behandlung

n_1 und n_2 = Umfänge der Stichproben 1 und 2

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad \text{mit}$$

$$s_1^2 = \frac{\sum_{i=1}^{n_1} x_{1i}^2 - \frac{\left(\sum_{i=1}^{n_1} x_{1i}\right)^2}{n_1}}{n_1 - 1} \quad \text{und} \quad s_2^2 = \frac{\sum_{i=1}^{n_2} x_{2i}^2 - \frac{\left(\sum_{i=1}^{n_2} x_{2i}\right)^2}{n_2}}{n_2 - 1}.$$

(3) Lies aus der t -Tabelle $t_{Tab}(FG; \alpha)$ ab (Tab. II(b), einseitig!!!), wobei:

α = Signifikanzniveau

$FG = n_1 + n_2 - 2$

(4) Vergleiche $t_{Vers}^{(1)}$ und $t_{Vers}^{(2)}$ mit t_{Tab} :

$$t_{Vers}^{(1)} > t_{Tab} \quad \text{und} \quad t_{Vers}^{(2)} < -t_{Tab} \Rightarrow H_A: -\delta < \mu_1 - \mu_2 < \delta \quad (\text{Äquivalenz})$$

Andernfalls ist keine Äquivalenz nachgewiesen (H_0)

Beispiel: Für die bt-Maisdaten legen wir fest, das Äquivalenz gegeben ist, wenn sich die wahren Abundanzen in bt-Mais und Kontrolle um weniger als $\delta = 1$ Insekt/100 m² unterscheiden. Wir finden

$$\bar{x}_1 = 11,6 \quad s_1 = 2,7019 \quad n_1 = 5 \quad s = 3,1623$$

$$\bar{x}_2 = 12,8 \quad s_2 = 3,5637 \quad n_2 = 5$$

$$FG = 8 \quad t_{Tab} = 1,860$$

$$t_{Vers}^{(1)} = \frac{\bar{x}_1 - \bar{x}_2 + \delta}{s\sqrt{1/n_1 + 1/n_2}} = \frac{11,6 - 12,8 + 1}{3,1623\sqrt{1/5 + 1/5}} = -0,10 < t_{Tab}$$

$$t_{Vers}^{(2)} = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{s\sqrt{1/n_1 + 1/n_2}} = \frac{11,6 - 12,8 - 1}{3,1623\sqrt{1/5 + 1/5}} = -1,10 > -t_{Tab}$$

Die Nullhypothese kann nicht abgelehnt werden, es kann also keine Äquivalenz nachgewiesen werden.

Man kann den Äquivalenz-Test auch sehr einfach mit Hilfe des Vertrauensintervalls für die Differenz der beiden Mittelwerte durchführen (siehe Abschnitt 3.10), wobei man anstatt des zweiseitigen t-Wertes aus Tab. II den einseitigen t-Wert aus Tab. II(b) verwenden muss. Äquivalenz ist genau dann nachgewiesen, wenn das Vertrauensintervall vollständig in den Grenzen $-\delta$ und δ enthalten ist.

Man kann leicht nachweisen, dass dieses Verfahren zur selben Testentscheidung führt wie das oben im Kasten beschriebene.

Beispiel: Mit $t_{Tab} = 1,860$ finden wir mit dem Verfahren aus Abschnitt 3.10 das Vertrauensintervall $(-4,919; 2,519)$. Dieses Intervall passt nicht in das Intervall $(-\delta; \delta) = (-1; 1)$, so dass keine Äquivalenz nachgewiesen werden kann.

3.22 Was ein Signifikanztest nicht sagt

Signifikanztests werden häufig fehl- bzw. überinterpretiert. Hier wird die häufigste unzulässige Interpretation angegeben und gesagt, warum sie nicht zulässig ist.

Falsche Aussage: "Zwei Mittelwerte sind bei $\alpha = 5\%$ signifikant verschieden. Daher unterscheiden sich die Mittelwerte mit einer Wahrscheinlichkeit von 95%."

Der erste Teil der Aussage ist richtig. Der zweite ist unzulässig. Um dies zu sehen, muss man zunächst berücksichtigen, dass die Irrtumswahrscheinlichkeit α eine sog. **bedingte Wahrscheinlichkeit** ist: Es ist die Wahrscheinlichkeit, die Nullhypothese zu verwerfen, **unter der Bedingung** dass in Wirklichkeit die Nullhypothese zutrifft. Man kann dies auch formal durch folgende Formel ausdrücken:

$$\alpha = P(\text{Test lehnt } H_0 \text{ ab} \mid H_0 \text{ trifft zu}) \quad .$$

Der vertikale Strich steht für "unter der Bedingung dass". Eben so kann man schreiben

$$1 - \alpha = P(\text{Test nimmt } H_0 \text{ an} \mid H_0 \text{ trifft zu}) \quad .$$

Diese Wahrscheinlichkeit beträgt im Fall $\alpha = 5\%$ gerade 95%, und mit diesem berechneten Wert operiert die *Falsche Aussage*. Die *Falsche Aussage* meint aber in Wirklichkeit eine ganz andere bedingte Wahrscheinlichkeit, nämlich die Wahrscheinlichkeit, dass H_A zutrifft, unter der Bedingung dass der Test H_0 ablehnt, also

$P(H_A \text{ trifft zu} \mid \text{Test lehnt } H_0 \text{ ab})$.

In der Regel stimmen diese beiden Wahrscheinlichkeiten keineswegs überein:

$$P(H_A \text{ trifft zu} \mid \text{Test lehnt } H_0 \text{ ab}) \neq 1 - \alpha = P(\text{Test nimmt } H_0 \text{ an} \mid H_0 \text{ trifft zu})$$

Man beachte, dass bei den beiden Wahrscheinlichkeiten die Testentscheidung (signifikanter oder nicht signifikanter Unterschied) gewissermaßen die Seite gewechselt hat: einmal steht sie hinter dem vertikalen Strich, einmal davor. Dasselbe gilt umgekehrt für die Aussage über die Wirklichkeit (H_0 oder H_A trifft zu). Da die Wahrscheinlichkeiten nicht identisch sind, kann man aus dem Ergebnis eines Tests sowie dem Signifikanzniveau nicht auf die Wahrscheinlichkeit für die Gültigkeit von H_0 oder H_A schließen. Ein statistischer Test sagt somit tatsächlich weniger aus, als gemeinhin vermutet wird. Ein signifikanter Unterschied gibt lediglich einen begründeten Hinweis darauf, dass tatsächlich ein Unterschied besteht, nicht mehr. Insbesondere sind keine Wahrscheinlichkeitsaussagen bzgl. der Gültigkeit von H_A möglich.

Um die gewünschte bedingte Wahrscheinlichkeit $P(H_A \text{ trifft zu} \mid \text{Test lehnt } H_0 \text{ ab})$ zu berechnen, müsste man unabhängig vom gerade betrachteten Versuch sog. *a priori* Aussagen darüber treffen, wie wahrscheinlich H_0 und H_A sind. Solche Aussagen sind in der Regel notwendigerweise subjektiv. Es gibt nur wenige Ausnahmen, für die diese ***a priori* Wahrscheinlichkeiten** objektiv ermittelt werden können. Solche Beispiele finden sich z.B. in der Medizin und in der Populationsgenetik. Wahrscheinlichkeiten wie $P(H_A \text{ trifft zu} \mid \text{Test lehnt } H_0 \text{ ab})$ kann man dann aus der *a priori* Wahrscheinlichkeit sowie den Versuchsergebnissen mit Hilfe des sog. **Satz von Bayes** berechnen, der hier aber nicht vertieft werden soll (Bosch, 2007: Basiswissen Statistik). Stattdessen betrachten wir ein einfaches Zahlenbeispiel.

Beispiel (hypothetisch, aber realistisch; modifiziert nach Kursunterlagen von Ulrike-Semmler Busch): Für eine Stoffwechselstörung wurde eine neue Diagnosemöglichkeit entdeckt: Die Blutkonzentration des K-Proteins ist bei Kranken von (genau) 50 auf 56 erhöht. Leider sind die Messungen noch recht ungenau: Die Varianz der Einzelwerte beträgt $\sigma^2 = 9$ (bei Normalverteilung). Wegen dieser großen Streuung werden bei einer zu untersuchenden Person jeweils 4 Proben untersucht. Für eine gesunde Person beträgt die tatsächliche Konzentration der K-Proteins $\mu = 50$. Wegen der bei Erkrankung zu erwartenden Verschiebung des wahren Wertes auf $\mu = 56$ führen wir einen einseitigen Test durch. Der Test hat folgende Null- und Alternativhypothese:

H_0 : Person ist gesund ($\mu = 50$)

H_A : Person ist krank ($\mu = 56$)

Wir lehnen H_0 : $\mu = 50$ bei $\alpha = 5\%$ ab, wenn

$$z_{\text{Vers}} = \frac{\bar{x} - 50}{\sqrt{9/4}} > z_{1-\alpha} = 1,64$$

ist, wobei \bar{x} der Mittelwert aus 4 Messwiederholungen der zu untersuchenden Person ist. Die Irrtumswahrscheinlichkeit unter der Alternative $\mu = 56$ ist mit den Methoden aus Abschnitt 3.2 wie folgt zu berechnen:

$$\begin{aligned}\beta &= P\left(\frac{\bar{x} - 50}{\sqrt{9/4}} < z_{1-\alpha} = 1,64 \mid \mu = 56\right) = P\left(\bar{x} < 1,64 * \sqrt{9/4} + 50 \mid \mu = 56\right) \\ &= P\left(z = \frac{\bar{x} - 56}{\sqrt{9/4}} < \frac{1,64 * \sqrt{9/4} - 56 + 50}{\sqrt{9/4}} = 1,64 - 4 = -2,36\right) = 0,0093\end{aligned}$$

Wir arbeiten also mit einem Test, der eine Wahrscheinlichkeit $\alpha = 5\%$ für einen Fehler 1. Art (gesunde Person fälschlicherweise als krank eingestuft) und eine Wahrscheinlichkeit von $\beta = 0,93\%$ für einen Fehler 2. Art (kranke Person fälschlicherweise als gesund eingestuft). Das Testergebnis (signifikanter Unterschied von $\mu = 50$ oder kein signifikanter Unterschied) führt zu einer **Diagnose**, ob die betreffende Person krank oder gesund ist.

Nun sei angenommen, dass die Einwohnerzahl eines Landes 10.010.000 Menschen beträgt. Von diesen haben 10.000 die Krankheit, 10.000.000 dagegen sind gesund. Der Anteil kranker Personen, die sog. **Prävalenz**, beträgt also nur ca. 0,1%. In anderen Worten, die sog. **a priori Wahrscheinlichkeit**, dass eine zufällig ausgewählte Person krank ist, beträgt $10.000/10.010.000 \approx 0,1\%$. Bei den gesunden Personen diagnostiziert der Test in 5% der Fälle fälschlicherweise die Krankheit, das sind 500.000, während bei 9.500.000 die Diagnose richtig ist ("gesund"). Bei den in Wirklichkeit Kranken 10.000 Personen erwarten wir, dass 0,0093*10.000, also 93 nicht erkannt werden, während 9.907 richtigerweise als krank diagnostiziert werden. Diese vier Häufigkeiten fassen wir in einer Tabelle zusammen.

		Wirklichkeit		
		Gesund	Krank	
Diagnose	Gesund	9.500.000	93	9.500.093
	Krank	500.000	9.907	509.907
		10.000.000	10.000	

Insgesamt würden 509.907 Personen als krank diagnostiziert, wenn man alle Personen der Bevölkerung untersuchen würde. Von diesen sind aber nur 9.907 wirklich krank! Das sind nur etwas mehr als 1,94%. Die Wahrscheinlichkeit, dass eine als krank diagnostizierte Person tatsächlich krank ist, beträgt somit nur 1,94%:

$$P(H_A \text{ trifft zu} \mid \text{Test lehnt } H_0 \text{ ab}) = P(\text{Person krank} \mid \text{Diagnose "krank"}) = 1,94\%$$

Man kann also selbst bei der Diagnose "krank" ziemlich sicher sein, dass man in Wirklichkeit gesund ist. Der Wert von 1,94% ist weit entfernt von 95%, also von

$$1-\alpha = P(\text{Test nimmt } H_0 \text{ an} \mid H_0 \text{ trifft zu}) = P(\text{Diagnose "gesund"} \mid \text{Person gesund})$$

Diese Wahrscheinlichkeit wird aus den **Häufigkeiten in der 1. Spalte** berechnet, während $P(\text{Person krank} \mid \text{Diagnose "krank"})$ aus den **Häufigkeiten in der 2. Zeile** berechnet wurde. Beide Wahrscheinlichkeiten werden also ganz anders berechnet.

Die *a posteriori* Wahrscheinlichkeit von 1,94% ist allerdings deutlich größer als die *a priori* Wahrscheinlichkeit von ca. 0,1% (und könnte eine genauere Untersuchung angeraten sein lassen), während $P(\text{Person krank} \mid \text{Diagnose "gesund"}) = 93/9.500.000 = 0,00098\%$ viel kleiner als 0,1% ist (und eine weitere Untersuchung unnötig macht).

Man sieht an diesem Beispiel, dass das Signifikanzniveau α in Verbindung mit der Testentscheidung keine Aussage über die Wahrscheinlichkeit der Gültigkeit von H_0 und H_A erlaubt. Eine solche Aussage ist nur unter der Berücksichtigung von *a priori* Wahrscheinlichkeiten möglich, wie wir es hier gesehen haben. Übrigens folgen die hier gemachten Wahrscheinlichkeitsberechnungen auch direkt aus dem Satz von Bayes, ohne dass wir diesen hier explizit verwendet hätten.

In dem gewählten Beispiel ergibt sich die *a priori* Wahrscheinlichkeit aus der Prävalenz der Stoffwechselstörung und kann prinzipiell objektiv ermittelt werden. In vielen anderen Situationen ist dies nicht möglich. Man kann lediglich subjektive Einschätzungen für die Wahrscheinlichkeit von H_0 und H_A abgeben. Die auf solchen subjektiven Einschätzungen beruhenden Test-Aussagen sind dann notwendigerweise auch subjektiv, was ein wesentlicher Kritikpunkt sogenannter **Bayes-Verfahren** ist. Wer diese Problematik näher interessiert, dem seien wärmstens die beiden folgenden leicht verständlichen (keine Formeln!) Bücher der Autoren Hans-Peter Beck-Bornholdt und Hans-Hermann Dubben empfohlen: "Der Hund, der Eier legt" und "Der Schein der Weisen" (beide als rororo Taschenbuch erschienen). Eine elementare Einführung in bedingte Wahrscheinlichkeiten und den Satz von Bayes findet man z.B. bei Stahel (2002). Eine sehr umfassende Darstellung von Bayes-Verfahren gibt das Buch "Bayesian data analysis" von A. Gelman, J.B. Carlin und H.S. Stern (2nd edition. CRC Press, 2003).

4. Die einfache Varianzanalyse

In Kapitel 3 hatten wir den t -Test zum Vergleich von zwei Stichproben (abhängig oder unabhängig) besprochen. In vielen Fällen sind mehr als zwei Stichproben zu vergleichen. In diesem Fall können die Daten mit Hilfe einer sog. Varianzanalyse ausgewertet werden. Die Varianzanalyse testet die globale Nullhypothese, dass alle Stichproben (Behandlungen) denselben Mittelwert (Erwartungswert) haben. Im Anschluss an die Varianzanalyse können paarweise Mittelwertvergleiche z.B. mittels t -Test durchgeführt werden.

4.1 Notation und Urliste

Im folgenden verwenden wir die "Punkt und Strich" Notation. Hierbei wird die Summation über einen Index durch einen Punkt gekennzeichnet, die Mittelwertbildung durch einen Querstrich. Die Verwendung dieser Notation soll anhand des Beispiels aus Abschnitt 3.1 erläutert werden.

Beispiel: Im folgenden sind die Erträge (dt/ha) von 5 Sorten (A, B, C, D, E) eines Feldversuches in einem Lageplan eingezeichnet. Der Versuch war in 4 Wiederholungen angelegt. Die Behandlungen wurden zufällig den Parzellen (Feldstücken) zugeordnet (randomisiert). Die Parzellen sind von 1 bis 20 durchnummeriert (Nummer oben links). Die Erträge stehen jeweils unten rechts.

1 B 21	2 D 34	3 D 32	4 E 24
5 C 27	6 E 23	7 A 31	8 B 23
9 A 32	10 C 29	11 E 27	12 A 37
13 D 31	14 D 27	15 A 32	16 B 25
17 B 19	18 C 34	19 E 26	20 C 34

Zur weiteren Berechnung ist es hilfreich, die Rohdaten in eine Urliste der folgenden Struktur eingetragen.

Wiederholungen	Behandlung					
	1	2 <i>i</i>	<i>t</i>	
1	y_{11}	y_{21}	y_{i1}	y_{t1}	
2	y_{12}	y_{22}	y_{i2}	y_{t2}	
.	
.	
.	
<i>j</i>	y_{1j}	y_{2j}	y_{ij}	y_{tj}	
.	
.	
.	
<i>r</i>	y_{1r}	y_{2r}	y_{ir}	y_{tr}	
Summe	$y_{1\bullet}$	$y_{2\bullet}$	$y_{i\bullet}$	$y_{t\bullet}$	$y_{\bullet\bullet}$
Mittelwert	$\bar{y}_{1\bullet}$	$\bar{y}_{2\bullet}$	$\bar{y}_{i\bullet}$	$\bar{y}_{t\bullet}$	$\bar{y}_{\bullet\bullet}$

In dieser Tabelle haben wir die Punkt-Strich-Notation verwendet:

Beobachtung: y_{ij} = Messwert *j*-te Wiederholung der *i*-ten Behandlung

$$(i = 1, \dots, t; j = 1, \dots, r)$$

Behandlungssummen: $y_{i\bullet} = \sum_{j=1}^r y_{ij}$

Behandlungsmittelwerte: $\bar{y}_{i\bullet} = \sum_{j=1}^r y_{ij} / r = y_{i\bullet} / r$

Gesamtsumme: $y_{\bullet\bullet} = \sum_{i=1}^t \sum_{j=1}^r y_{ij}$

Gesamtmittelwert: $\bar{y}_{\bullet\bullet} = \sum_{i=1}^t \sum_{j=1}^r y_{ij} / (rt)$

t = Zahl der Behandlungen (treatments)

r = Zahl der Wiederholungen (replications)

Beispiel: Für den Sortenversuch finden wir:

Meßwerte	Sorte					
	A	B	C	D	E	
1	$y_{11} = 31$	$y_{21} = 21$	$y_{31} = 27$	$y_{41} = 34$	$y_{51} = 24$	
2	$y_{12} = 32$	$y_{22} = 23$	$y_{32} = 29$	$y_{42} = 32$	$y_{52} = 23$	
3	$y_{13} = 37$	$y_{23} = 25$	$y_{33} = 34$	$y_{43} = 31$	$y_{53} = 27$	
4	$y_{14} = 32$	$y_{24} = 19$	$y_{34} = 34$	$y_{44} = 27$	$y_{54} = 26$	
Summe	$y_{1\bullet} = 132$	$y_{2\bullet} = 88$	$y_{3\bullet} = 124$	$y_{4\bullet} = 124$	$y_{5\bullet} = 100$	$y_{\bullet\bullet} = 568$
Mittelwert	$\bar{y}_{1\bullet} = 33$	$\bar{y}_{2\bullet} = 22$	$\bar{y}_{3\bullet} = 31$	$\bar{y}_{4\bullet} = 31$	$\bar{y}_{5\bullet} = 25$	$\bar{y}_{\bullet\bullet} = 28,4$

Die Sortenmittelwerte in der Urliste zeigen numerische Unterschiede. Es lässt sich jedoch nicht ohne weiteres sagen, ob diese Unterschiede zufallsbedingt sind oder ob sich dahinter wirkliche genetische Unterschiede verbergen. Um diese Frage zu beantworten, ist eine Varianzanalyse mit anschließenden Mittelwertvergleichen notwendig (siehe unten). □

4.2 Lineares Modell

Für eine Varianzanalyse können wir das folgende Modell ansetzen:

$$y_{ij} = \mu_i + e_{ij} \quad , \quad i = 1, \dots, t; \quad j = 1, \dots, r$$

wobei

y_{ij} = Messwert der j -ten Wiederholung der i -ten Behandlung

μ_i = theoretischer Mittelwert (Erwartungswert) der i -ten Behandlung

e_{ij} = Zufallsabweichung der Beobachtung y_{ij} vom Behandlungsmittel

Eine alternative Schreibweise ergibt sich, wenn wir den Behandlungsmittelwert μ_i ersetzen durch $\mu + \tau_i$, so dass das Modell lautet:

$$y_{ij} = \mu + \tau_i + e_{ij} \quad , \quad i = 1, \dots, t; \quad j = 1, \dots, r \quad ,$$

wobei

μ = Konstante

τ_i = Effekt der i -ten Behandlung

In dieser Schreibweise ist $\mu + \tau_i$ der wahre Mittelwert der i -ten Behandlung. Es gibt eine Reihe von Gründen, den Erwartungswert in eine Konstante und einen „Effekt“ aufzuteilen. Auf Details soll hier nicht eingegangen werden, sondern nur soviel gesagt werden: Die Konstante und die Effekte haben für sich genommen keine direkte Interpretation. Die entscheidende und direkt interpretierbare Größe bleibt der Erwartungswert, gegeben durch die Summe $\mu + \tau_i$. Differenzen der Erwartungswerte entsprechen Differenzen der Behandlungseffekte. So gilt z.B.

$$\mu_1 - \mu_2 = \mu + \tau_1 - (\mu + \tau_2) = \tau_1 - \tau_2.$$

Falls es keine Unterschiede in den Behandlungen gibt, kann man ein gültiges Modell für die Daten durch Weglassen des Behandlungseffektes τ_i erhalten. Dann hat μ die Bedeutung des für alle Behandlungen gleichermaßen gültigen Erwartungswertes.

Die Effekte des Modells sowie die Konstante können geschätzt werden. Eine mögliche Schätzung ist gegeben durch

$$\hat{\mu} = \bar{y}_{..} \text{ und} \\ \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..},$$

wobei die „Hut-Notation“ (^) bedeutet, dass ein Modellparameter geschätzt wird. Diese Schätzung des Effekts einer Behandlung mißt hier die Differenz zum Gesamtmittel aller Beobachtung. Für unser Beispiel finden wir

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..} = 28,4 \\ \hat{\tau}_1 &= \bar{y}_{1.} - \bar{y}_{..} = 33 - 28,4 = 4,6. \\ \hat{\tau}_2 &= \bar{y}_{2.} - \bar{y}_{..} = 22 - 28,4 = -6,4 \\ \hat{\tau}_3 &= \bar{y}_{3.} - \bar{y}_{..} = 31 - 28,4 = 2,6 \\ \hat{\tau}_4 &= \bar{y}_{4.} - \bar{y}_{..} = 31 - 28,4 = 2,6 \\ \hat{\tau}_5 &= \bar{y}_{5.} - \bar{y}_{..} = 25 - 28,4 = -3,4 \end{aligned}$$

Addieren wir die Konstante und die jeweiligen Effekte, so reproduziert dies die geschätzten Behandlungsmittelwerte:

$$\begin{aligned} \hat{\mu}_1 &= \hat{\mu} + \hat{\tau}_1 = 28,4 + 4,6 = 33 \\ \hat{\mu}_2 &= \hat{\mu} + \hat{\tau}_2 = 28,4 - 6,4 = 22 \\ \hat{\mu}_3 &= \hat{\mu} + \hat{\tau}_3 = 28,4 + 2,6 = 31 \\ \hat{\mu}_4 &= \hat{\mu} + \hat{\tau}_4 = 28,4 + 2,6 = 31 \\ \hat{\mu}_5 &= \hat{\mu} + \hat{\tau}_5 = 28,4 - 3,4 = 25 \end{aligned}$$

Es ist wichtig, darauf hinzuweisen, dass die gezeigten Schätzwerte nur eine von unendlich vielen möglichen Schätzwerten darstellen. So läßt sich leicht nachvollziehen, dass man dieselben Schätzungen der Mittelwerte der Behandlungen erhält, wenn man von allen oben angegebenen Schätzungen der

Behandlungseffekte einen beliebigen konstanten Wert abzieht und denselben Wert zur Schätzung der Konstante μ hinzuaddiert, z.B.

$$\hat{\mu} = 28,4 + 10 = 38,4$$

$$\hat{\tau}_1 = 4,6 - 10 = -5,4$$

$$\hat{\tau}_2 = -6,4 - 10 = -16,4$$

$$\hat{\tau}_3 = 2,6 - 10 = -7,4$$

$$\hat{\tau}_4 = 2,6 - 10 = -7,4$$

$$\hat{\tau}_5 = -3,4 - 10 = -13,4$$

Dies führt zu folgenden (identischen) Schätzungen der Behandlungsmittelwerte:

$$\hat{\mu}_1 = \hat{\mu} + \hat{\tau}_1 = 38,4 - 5,6 = 33$$

$$\hat{\mu}_2 = \hat{\mu} + \hat{\tau}_2 = 38,4 - 16,4 = 22$$

$$\hat{\mu}_3 = \hat{\mu} + \hat{\tau}_3 = 38,4 - 7,4 = 31$$

$$\hat{\mu}_4 = \hat{\mu} + \hat{\tau}_4 = 38,4 - 7,4 = 31$$

$$\hat{\mu}_5 = \hat{\mu} + \hat{\tau}_5 = 38,4 - 13,4 = 25$$

Dieses Beispiel zeigt, dass die Effektschätzungen nur bedingt eine direkte Interpretation haben. Tatsächlich verwenden Computerprogramme noch andere Schätzungen als die hier gezeigten, und es ist wichtig sich dieser Einschränkung bewußt zu sein. Unabhängig von der gewählten Schätzung ergeben sich jedoch immer dieselben Mittelwerte und Mittelwertdifferenzen, und geschätzte Mittelwertdifferenzen entsprechen immer den Differenzen der geschätzten Effekte. So ist

$$\hat{\mu}_1 - \hat{\mu}_2 = 33 - 22 = 11$$

sowie bei der ersten Schätzung

$$\hat{\tau}_1 - \hat{\tau}_2 = 4,6 - (-6,4) = 11$$

und bei der zweiten Schätzung

$$\hat{\tau}_1 - \hat{\tau}_2 = -5,4 - (-16,4) = 11.$$

Das hier verwendete Modell ist ein lineares Modell. Lineare Modelle sind häufig gute Annäherungen an die Wirklichkeit. Wegen Ihrer einfachen Handhabbarkeit werden sie oft benutzt.

Ein wichtiger Aspekt der Modellformulierung ist die Annahme über die Verteilung des Fehlers e_{ij} . Wir nehmen hier an, dass die Fehler stochastisch unabhängig sind und dass sie einer Normalverteilung mit Mittelwert Null und konstanter Varianz σ^2 folgen. Dies wird auch wie folgt ausgedrückt:

$$e_{ij} \sim N(0, \sigma^2)$$

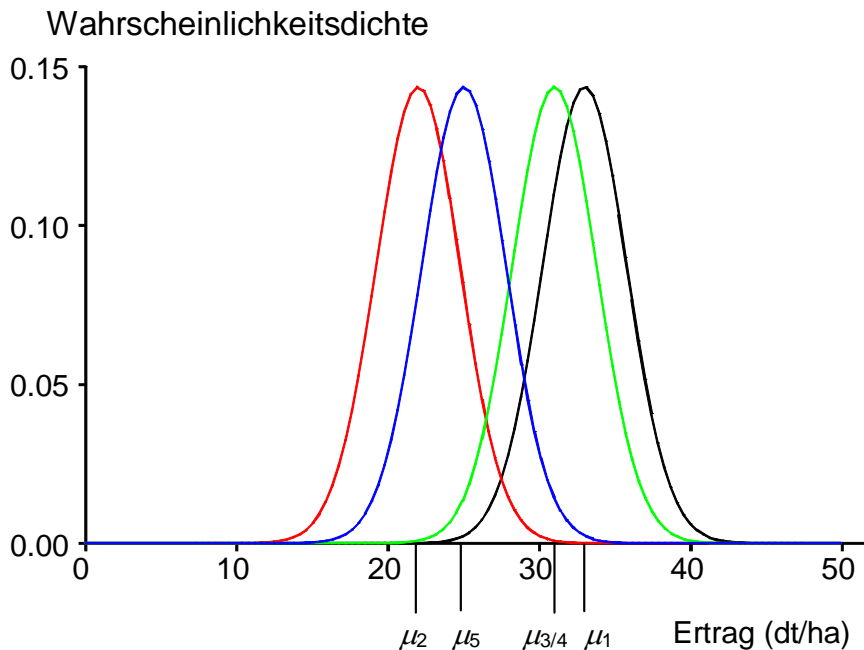


Abb. 4.2.1: Graphische Darstellung des varianzanalytischen Modells: Überlagerung von 5 Normalverteilungen, wobei die Verteilungen der Behandlungen 3 und 4 wegen des gleichen Mittelwertes exakt übereinander liegen. Diese Annahme bedeutet, dass auch die Daten normalverteilt sind, und zwar mit Mittelwert μ_i :

$$y_{ij} \sim N(\mu_i, \sigma^2)$$

Die Annahme der stochastischen Unabhängigkeit ist immer dann gewährleistet, wenn der Versuch richtig randomisiert wurde. Wurde dagegen nicht randomisiert, ist diese Annahme nicht zulässig. Daher ist die Randomisation außerordentlich wichtig.

Das Modell lässt sich graphisch wie in Abb 4.2.1. gezeigt darstellen. Jede Behandlung hat eine eigene Normalverteilung. Die Breite der Dichtefunktion und damit die Varianz ("Streuung") ist für jede Behandlung gleich. Wenn Behandlungsunterschiede im mittleren Ertrag bestehen, so ist die Lage der Verteilungskurven auf der x -Achse verschoben. Die Lage der Kurven ist durch den Mittelwert (Erwartungswert) μ_i festgelegt.

Die Varianzanalyse liefert einen Test der globalen Nullhypothese, dass alle Sorten denselben Ertrag liefern. Die Nullhypothese lässt sich schreiben als

$$H_0: \tau_1 = \tau_2 = \dots = \tau_t$$

Diese wird gegen die Alternative

$$H_A: \{\text{Mindestens zwei der } \tau_i \text{ sind verschieden}\}$$

geprüft.

4.3 Zwei Arten, die Fehlervarianz zu schätzen

Ein wichtiger Aspekt der Varianzanalyse ist die Schätzung der Fehlervarianz σ^2 . Hierfür gibt es in der Varianzanalyse zwei Möglichkeiten.

(1) Wir können für jede Behandlung i die Stichprobenvarianz schätzen:

$$s_i^2 = \frac{\sum_{j=1}^r (y_{ij} - \bar{y}_{i.})^2}{r-1}$$

Da jede dieser Stichprobenvarianzen die Fehlervarianz schätzt, bietet es sich an, die Schätzungen zusammenzufassen (zu "poolen"):

$$s^2 = \frac{\sum_{i=1}^t s_i^2}{t} = \frac{\sum_{i=1}^t \sum_{j=1}^r (y_{ij} - \bar{y}_{i.})^2}{t(r-1)}$$

Der Zähler sind die sog. Fehlerquadratsumme (SQ_{Fehler}):

$$SQ_{Fehler} = \sum_{i=1}^t \sum_{j=1}^r (y_{ij} - \bar{y}_{i.})^2$$

Der Nenner sind die Fehler-Freiheitsgrade. Der Quotient aus SQ_{Fehler} und den Freiheitsgraden wird als Mittelquadrat (MQ) bezeichnet:

$$s^2 = MQ_{Fehler} = \frac{SQ_{Fehler}}{t(r-1)}$$

Das MQ_{Fehler} ist ein Schätzwert für die Fehlervarianz σ^2 . Der Erwartungswert von MQ_{Fehler} entspricht genau der Fehlervarianz, was wir ausdrücken können als:

$$E(MQ_{Fehler}) = \sigma^2$$

Diese Gleichung bedeutet folgendes: Wenn wir uns vorstellen, der Versuch könnte sehr oft unter denselben Bedingungen wiederholt werden, so würde das MQ_{Fehler} im Mittel gleich der Fehlervarianz σ^2 sein.

(2) Betrachten wir nun die Varianz eines Behandlungsmittelwertes. Es gilt:

$$\text{var}(\bar{y}_{i.}) = \frac{\sigma^2}{r}$$

Unter der Nullhypothese $H_0: \tau_1 = \tau_2 = \dots = \tau_t$ haben alle Mittelwerte denselben Erwartungswert:

$$E(\bar{y}_{i.}) = \mu$$

Wir können nun die Stichprobenvarianz der Stichprobenmittelwerte berechnen:

$$s_T^2 = \frac{\sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2}{t-1}$$

Da unter der Nullhypothese alle Behandlungsmittelwerte denselben Erwartungswert haben, gilt:

$$E(s_T^2) = \frac{\sigma^2}{r}$$

Dies aber bedeutet, dass rs_T^2 ein alternativer Schätzer der Fehlervarianz ist, denn es ist

$$E(rs_T^2) = \sigma^2$$

Wir können dies auch schreiben als

$$rs_T^2 = \frac{r \sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2}{t-1} = \frac{SQ_{Beh}}{t-1} = MQ_{Beh}$$

wobei

$$SQ_{Beh} = r \sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2$$

Unter der Nullhypothese gilt:

$$E(MQ_{Beh}) = \sigma^2$$

Entscheidend für die Varianzanalyse ist nun die Tatsache, dass unter der Alternative $H_A: \{\text{Mindestens zwei der } \tau_i \text{ sind verschieden}\}$

$$E(MQ_{Beh}) > \sigma^2$$

Genauer gesagt gilt allgemein:

$$E(MQ_{Beh}) = \sigma^2 + Q(\tau)$$

wobei

$$Q(\tau) = \frac{r \sum_{i=1}^t (\tau_i - \bar{\tau}_\cdot)^2}{t-1}$$

Dies bedeutet, dass der Erwartungswert von MQ_{Beh} umso größer ist, je größer die Behandlungsunterschiede sind, d.h. je größer die Unterschiede der Behandlungseffekte τ_i sind.

Hieraus ergibt sich der Ansatz, den Quotienten

$$F_{Vers} = \frac{MQ_{Beh}}{MQ_{Fehler}}$$

als Teststatistik zur Prüfung der Nullhypothese zu verwenden. Falls die H_0 zutrifft, wird F_{Vers} etwa den Wert 1 haben, natürlich im Rahmen einer Stichprobenstreuung. Trifft dagegen die Alternative zu, so wird F_{Vers} mehr oder weniger deutlich größer als 1 sein. Es kann gezeigt werden, dass F_{Vers} eine F -Verteilung mit $t-1$ und $t(r-1)$ Freiheitsgraden (FG) hat.

4.4 Die Varianzanalyse-Tabelle

SQ , MQ und die Freiheitsgrade (FG) werden üblicherweise in eine Varianzanalyse-Tabelle geschrieben. Zusammenfassend wird die Varianzanalyse im folgenden Kasten beschrieben:

Ursache	FG	SQ	MQ	$E(MQ)$	F_{Vers}
Behandlungen	$t-1$	SQ_{Beh}	$MQ_{Beh} = \frac{SQ_{Beh}}{t-1}$	$\sigma^2 + Q(\tau)$	$F_{Vers} = \frac{MQ_{Beh}}{MQ_{Fehler}}$
Fehler	$t(r-1)$	SQ_{Fehler}	$MQ_{Fehler} = \frac{SQ_{Fehler}}{t(r-1)}$	σ^2	

Quadratsummen:

$$SQ_{Beh} = r \sum_{i=1}^t (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^t y_{i\cdot}^2 / r - y_{\cdot\cdot}^2 / (rt);$$

$$SQ_{Fehler} = \sum_{i=1}^t \sum_{j=1}^r (y_{ij} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \sum_{i=1}^t y_{i\cdot}^2 / r$$

wobei

Beobachtung: y_{ij} = Messwert j -te Wiederholung der i -ten Behandlung
 $(i = 1, \dots, t; j = 1, \dots, r)$

Behandlungssummen:	$y_{i\bullet} = \sum_{j=1}^r y_{ij}$
Gesamtsumme:	$y_{\bullet\bullet} = \sum_{i=1}^t \sum_{j=1}^r y_{ij}$
Verwerfe die Nullhypothese $H_0: \tau_1 = \tau_2 = \dots = \tau_t$ wenn gilt:	
$F_{Vers} = MQ_{Beh}/MQ_{Fehler} > F_{Tab} = F_{(1-\alpha; t-1, t(r-1))}$ (Tab. VI)	

wobei $F_{(1-\alpha; t-1, t(r-1))}$ der kritische Wert zum Niveau α für die F -Verteilung mit $t-1$ und $t(r-1)$ Freiheitsgraden ist. Das Signifikanzniveau α ist die Wahrscheinlichkeit eines Fehlers 1. Art, d.h. die Wahrscheinlichkeit, die Nullhypothese fälschlicherweise zu verwerfen. Diese Wahrscheinlichkeit muss vom Anwender vorgegeben werden. Üblich sind die Werte 0,1%, 1% und 5%. Einige Werte für $\alpha = 5\%$ sind in Tab. VI angegeben. Man beachte, dass hier eine **einseitige** Fragestellung vorliegt, da nur große F -Werte zu einer Verwerfung der Nullhypothese führen.

Beispiel: Für das Beispiel des Sortenversuches wollen wir einen F -Test zum Niveau $\alpha = 5\% = 0,05$ durchführen. Wir berechnen zunächst die SQ und erstellen eine Varianzanalysetabelle. Die "Zutaten" zur Berechnung der SQ , nämlich y_{ij} , $y_{i\bullet}$ und $y_{\bullet\bullet}$, entnehmen wir der Urliste in Abschnitt 4.1. Sodann berechnen wir

$$I = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 = 31^2 + 32^2 + \dots + 26^2 = 16596$$

$$II = \sum_{i=1}^t y_{i\bullet}^2 / r = 132^2/4 + 88^2/4 + 124^2/4 + 124^2/4 + 100^2/4 = 16480$$

$$III = y_{\bullet\bullet}^2 / (rt) = 568^2/20 = 16131,2$$

$$SQ_{Fehler} = I - II = 116$$

$$SQ_{Beh} = II - III = 348,8$$

Ursache	FG	SQ	MQ	F
Behandlungen	4	348,8	87,200	$87,2/7,733 = 11,28$
Fehler	15	116	7,733	

Der Tabellenwert $F_{Tab} = F_{(0,95; 4; 15)} = 3,06$ ist kleiner als $F_{Vers} = 11,28$. Wir schließen daher auf signifikante Sortenunterschiede. □

4.5 Multiple Mittelwertvergleiche

Bei signifikantem F -Test können wir zwar auf signifikante Behandlungsunterschiede schließen, wir wissen dadurch aber noch nicht, welche Behandlungen sich unterscheiden. Hierzu führen wir im Anschluss paarweise Tests durch. Der einfachste Test ist der **t -Test (LSD-Test)**, der zuerst beschrieben wird. Der multiple t -Test birgt das besondere Risiko, viele Fehler 1. Art zu produzieren. Das Risiko wird umso grö-

ßer, je mehr Behandlungen verglichen werden. Ein alternativer Test, der eine bessere Kontrolle des Fehlers 1. Art erlaubt, ist der **Tukey-Test**, der im Anschluss besprochen wird.

4.5.1 LSD-Test

Die Besonderheit des multiplen t -Tests im Rahmen der Varianzanalyse besteht darin, dass die Fehlervarianz aus dem MQ_{Fehler} der Varianzanalyse geschätzt wird, also unter Verwendung der Daten aller Behandlungen.

Die Entscheidungsregel für den multiplen t -Test lautet:

$$\text{Verwerfe } H_0: \tau_s = \tau_u, \text{ wenn } t_{Vers} = \left| \frac{\bar{y}_{s.} - \bar{y}_{u.}}{\sqrt{2MQ_{Fehler}/r}} \right| > t_{Tab} [t(r-1), \alpha]$$

Diese kann man umformen zu folgender Vorschrift:

$$\text{Verwerfe } H_0: \tau_s = \tau_u, \text{ wenn } |\bar{y}_{s.} - \bar{y}_{u.}| > t_{Tab} [t(r-1), \alpha] \sqrt{2MQ_{Fehler}/r} = LSD$$

t_{Tab} ist in der **zweiseitigen** Tabelle II abzulesen.

Eine (absolute) Differenz wird nach dieser Vorschrift signifikant, wenn sie die Grenzdifferenz $LSD = t_{Tab} [t(r-1), \alpha] \sqrt{2MQ_{Fehler}/r}$ überschreitet. Die Grenzdifferenz wird im Englischen als **least significant difference** (LSD) bezeichnet. Die Grenzdifferenz ist für jeden Vergleich dieselbe, wenn die Zahl der Wiederholungen je Behandlung konstant ist.

Beispiel: In unserem Sortenversuch ist $r = 4$. Wir finden mit $MQ_{Fehler} = 7,733$ und $t_{Tab}[t(r-1) = 15, \alpha = 5\%] = 2,131$ die Grenzdifferenz

$$LSD = 2,131 \sqrt{2 \times 7,733 / 4} = 4,19 .$$

Alle Differenzen, die diese Grenzdifferenz überschreiten, sind signifikant. Die Grenzdifferenz eignet sich zur tabellarischen Darstellung wie der folgenden:

Sorte	Mittelwert
A	33
B	22
C	31
D	31
E	25
LSD	4,19

Anhand der tabellierten Mittelwerte kann der Leser Differenzen im Kopf berechnen und mit der *LSD* vergleichen. Beispielsweise sind die Sorten A und B signifikant verschieden, da $33-22 = 11 > \text{LSD}$. □

Mit Hilfe der Grenzdifferenz ist es möglich, signifikante Unterschiede durch Buchstaben kenntlich zu machen. Dies lässt sich am besten direkt am Beispiel erklären.

Beispiel: Für den Sortenversuch sind die Mittelwerte für die Sorten A, B, C, D, und E

$$\bar{y}_1. = 33, \bar{y}_2. = 22, \bar{y}_3. = 31, \bar{y}_4. = 31, \bar{y}_5. = 25$$

Diese ordnen wir der Größe nach und tragen sie in eine Kreuztabelle der folgenden Form ein:

	A (33)	C (31)	D (31)	E (25)	B (22)
A (33)		2	2	8*	11*
C (31)			0	6*	9*
D (31)				6*	9*
E (25)					3
B (22)					

(*Signifikante Differenz: $\text{LSD} = 4,19$)

Zu beachten ist, dass die Behandlungen sowohl in den Zeilen als auch in den Spalten der Größe nach geordnet sind. In der Tabelle sind auch die Werte der paarweisen Differenzen eingetragen, und zwar ausschließlich in die Zellen oberhalb der Diagonalen. Diese Zellen reichen, um alle Vergleiche darzustellen. Sodann wird jede Differenz mit der Grenzdifferenz verglichen. Wir verwenden hier den *LSD*-Wert von 4,19. Die signifikanten Differenzen werden mit einem * gekennzeichnet. Ausgehend von der Tabelle werden nun Mittelwertunterschiede durch Unterstreichen kenntlich gemacht. Hierzu schreiben wir die fünf Behandlungen nach der Größe ihrer Mittelwerte geordnet in eine Reihe:

A C D E B

Dann setzen wir einen Strich unter der ersten Behandlung an und ziehen so lange nach rechts, bis alle diejenigen Behandlungen unterstrichen sind, die nicht von der ersten Behandlung signifikant verschieden sind. Um zu sehen, wie weit nach rechts wir die Linie ziehen müssen, schauen wir einfach in der ersten Zeile der Kreuztabelle nach, wann das erste mal eine Signifikanz auftaucht. Die erste Signifikanz liefert der Vergleich A - E. Also ziehen wir die Linie nur bis Behandlung D:

A C D E B

Dann setzen wir einen Strich bei der zweiten Behandlung an und ziehen wieder nach rechts bis alle Behandlungen unterstrichen sind, die nicht von der zweiten verschieden sind. Hierzu schauen wir in der zweiten Zeile der Kreuztabelle nach, wann die erste Signifikanz auftaucht. Dies wird für alle weiteren Behandlungen ebenfalls gemacht. Das Ergebnis ist:

A	C	D	E	B

Diese Darstellung ist wie folgt zu interpretieren: Behandlungen, die von einer gemeinsamen Linie unterstrichen werden, sind nicht signifikant voneinander verschieden.

Die obige Darstellung enthält noch einige Redundanzen. Die Linien 2 und 3 sind in der ersten Linie "enthalten" und können daher wegfallen. Ebenso ist die 5. Linie in der 4. "enthalten" und kann wegfallen. Die Darstellung sieht dann wie folgt aus:

A	C	D	E	B

Man überzeugt sich leicht, dass durch Weglassen der redundanten Linien keine Information verloren gegangen ist. Die Liniendarstellung kann in eine Tabelle überführt werden. Dazu vergeben wir für jede Linie einen Buchstaben:

A	C	D	E	B
a			b	

Wir schreiben die Behandlungen und ihre Mittelwerte in eine Tabelle. Sodann werden die Buchstaben der Linien, von denen eine Behandlung unterstrichen ist, hinter den Mittelwert der betreffenden Behandlung geschrieben:

Sorte	Mittelwert
A	33 ^a
B	22 ^b
C	31 ^a
D	31 ^a
E	25 ^b
<hr/>	
<i>LSD</i> ($\alpha=5\%$)	4,19

Mittelwerte, die mit einem gemeinsamen Buchstaben versehen sind, sind nicht signifikant voneinander verschieden.

Wir sehen anhand der Darstellung leicht, dass beispielsweise A von B und E verschieden ist, nicht aber von C und D. Das Beispiel zeigt, dass die Buchstabendarstellung die Interpretation der Ergebnisse von multiplen Mittelwertvergleichen erleichtert.

□

Im obigen Beispiel bekommt jede Behandlung nur einen Buchstaben. Es ist allerdings durchaus möglich, dass eine Behandlung mehrere Buchstaben erhält.

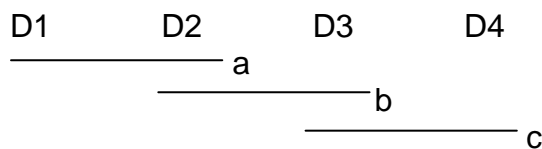
Beispiel: In einem Versuch werden vier Dünger-Behandlungen verglichen. Mittelwerte und Grenzdifferenz sind wie folgt:

Behandlung	Mittelwert
D1	35
D2	41
D3	47
D4	51

LSD($\alpha=5\%$) 7

	D1 (35)	D2 (41)	D3 (47)	D4 (51)
D1 (35)		6	12*	16*
D2 (41)			6	10*
D3 (47)				4
D4 (51)				

(*Signifikante Differenz: LSD = 7)



Behandlung	Mittelwert
D1	35 ^a
D2	41 ^{ab}
D3	47 ^{bc}
D4	51 ^c

LSD($\alpha=5\%$) 7

Mittelwerte, die mit einem gemeinsamen Buchstaben versehen sind, sind nicht signifikant voneinander verschieden.

Man beachte, dass sich beispielsweise die Behandlungen D1 und D2 nicht signifikant unterscheiden, weil sie einen Buchstaben gemeinsam haben. Die Tatsache, dass die Behandlung D2 den Buchstaben b trägt, die Behandlung D1 aber nicht, ist unerheblich. Es reicht die Übereinstimmung in einem einzigen Buchstaben. Man beachte außerdem, dass die Mittelwerte hier aufsteigend und nicht absteigend geordnet wurden. Mit beiden Sortierungen erhält man dieselbe Liniendarstellung. Es ist also egal, ob man von klein nach groß oder von groß nach klein sortiert.

Eine Schnellmethode: Man kann das oben beschriebene Verfahren auch wie folgt vereinfachen. Man berechnet für jede Behandlung die Differenz zwischen Mittelwert und LSD. Die Behandlungen werden dann von groß nach klein sortiert nebeneinander geschrieben und der Mittelwert in Klammern hinter das Symbol für die Behandlung beschrieben. Im letzten Beispiel liefert dies

D4 (51) D3 (47) D2 (41) D1 (35)
 44 40 34 28

Nun zieht man ausgehend von jedem Mittelwert jeweils solange nach rechts unter die benachbarten Mittelwerte, wie die Differenz zwischen Mittelwert und LSD kleiner ist als der gerade zur Unterstreichung anstehende Mittelwert ist. Im Beispiel liefert dies folgendes Ergebnis:

D4 (51)	D3 (47)	D2 (41)	D1 (35)
44	40	34	28

Für D4 ziehen wir die Linie z.B. bis unter D3, da $44 < 47$, aber nicht unter D2, da $44 > 41$ ist. Dies ist offensichtlich dieselbe Liniendarstellung die bereits mit der etwas aufwändigeren Methode erhalten wurde.

4.5.2 Vergleichsbezogener vs. versuchsbezogener Fehler 1. Art

Wenn t Behandlungen geprüft werden, so können $t(t-1)/2$ paarweise Vergleiche durchgeführt werden. Wenn ein einzelner Vergleich bei einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$ durchgeführt wird, so ist die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen (Nullhypothese wird fälschlicherweise verworfen) für diesen Vergleich gleich 5%. Da diese Irrtumswahrscheinlichkeit auf den einzelnen Vergleich bezogen ist, spricht man auch von der **vergleichsbezogenen** Wahrscheinlichkeit eines Fehlers 1. Art. Führen wir jeden der paarweisen Vergleiche mit einem t -Test zum Niveau α durch, so wird die vergleichsbezogene Irrtumswahrscheinlichkeit α eingehalten.

Anstatt den einzelnen Vergleich als Bezugsbasis zu wählen, können wir die Gesamtheit aller $t(t-1)/2$ Vergleiche betrachten. Diese Betrachtung wird auch als versuchsbezogen bezeichnet. Die Wahrscheinlichkeit, dass bei mindestens einem der Vergleiche ein Fehler 1. Art auftritt, heißt **versuchsbezogene** Irrtumswahrscheinlichkeit. Diese Irrtumswahrscheinlichkeit ist in der Regel deutlich höher als die vergleichsbezogene Irrtumswahrscheinlichkeit.

Vergleichsbezogener Fehler 1. Art

Für einen bestimmten Vergleich wird die Nullhypothese fälschlicherweise verworfen (Fehler 1. Art). Dies ist ein vergleichsbezogener Fehler 1. Art.

Versuchsbezogener Fehler 1. Art

Liegt dann vor, wenn unter allen $t(t-1)/2$ paarweisen Vergleichen mindestens ein vergleichsbezogener Fehler 1. Art auftritt. Ein versuchsbezogener Fehler 1. Art liegt nur dann *nicht* vor, wenn bei keinem der Vergleiche ein vergleichsbezogener Fehler 1. Art auftritt.

Abb. 4.4.1: Zwei Irrtumswahrscheinlichkeiten

Wir können die versuchsbezogene und die vergleichsbezogene Betrachtungsweise auch auf Vertrauensintervalle für Differenzen anwenden. Hält ein Verfahren die vergleichsbezogene Irrtumswahrscheinlichkeit α ein, so überdeckt jedes einzelne Intervall die wahre Differenz mit einer Wahrscheinlichkeit von mindestens $(1-\alpha)100\%$. Hält ein Verfahren dagegen die versuchsbezogene Irrtumswahrscheinlichkeit α ein,

so beträgt die Wahrscheinlichkeit, dass **alle** Intervalle die jeweils wahre Differenz überdecken, mindestens $(1-\alpha)100\%$.

Der Unterschied zwischen vergleichs- und versuchsbezogener Irrtumswahrscheinlichkeit wird vielleicht am besten anhand der folgenden Tatsachen deutlich. Halten wir eine vergleichsbezogene Irrtumswahrscheinlichkeit α ein, so müssen wir damit rechnen, dass bis zu $\alpha 100\%$ der Vergleiche zu einem Fehler 1. Art führen (Nullhypothese fälschlicherweise verworfen oder das Vertrauensintervall überdeckt nicht die wahre Differenz). Wir müssen also bereit sein, unter allen Aussagen (Test oder Intervall) $\alpha 100\%$ falsche Aussagen zu treffen. Die Wahrscheinlichkeit, dass bei mindestens einem Vergleich ein Fehler 1. Art auftritt kann sehr groß sein und kann bis zu 100% betragen. Halten wir dagegen eine versuchsbezogene Irrtumswahrscheinlichkeit α ein, so beträgt die Wahrscheinlichkeit, dass bei der Gesamtheit aller Vergleiche überhaupt nur einer oder mehr falsche Aussagen getroffen werden, höchstens $\alpha 100\%$. Die Wahrscheinlichkeit, dass kein einziger Fehler 1. Art auftritt, ist mindestens $(1-\alpha)100\%$.

Es ist eine schwierige Frage, welche der beiden Irrtumswahrscheinlichkeiten eingehalten werden soll. Es gibt hierzu kaum eine einfache Regel. Aussagen, die unter Einhaltung der versuchsbezogenen Irrtumswahrscheinlichkeit getroffen werden, sind stärker abgesichert. Allerdings ist diese Absicherung nicht umsonst: Sie geht bei Tests zu Lasten eines größeren Fehlers 2. Art (Alternativhypothese fälschlicherweise verworfen). Bei Vertrauensintervallen ist der Preis eine größere Breite der Intervalle. Eine denkbare Strategie besteht darin, zu Beginn eines Projektes, wenn noch nach interessanten, weiter zu verfolgenden Zusammenhängen gesucht wird, eher vergleichsbezogen zu testen, während zu Projektende eine stärkere Absicherung durch versuchsbezogene Tests angestrebt wird.

Der multiple t -Test hält die vergleichsbezogene Irrtumswahrscheinlichkeit ein. Es gibt eine ganze Reihe statistischer Verfahren, welche die versuchsbezogene Irrtumswahrscheinlichkeit einhalten (Miller, 1981; Hsu, 1996; Horn und Vollandt, 1995). Hiervon soll hier nur der Tukey-Test angesprochen werden. Ein allgemeines Verfahren zur Einhaltung des versuchsbezogenen Fehlers ist das sog. **Bonferroni-Verfahren** (siehe Anhang).

4.5.3 Tukey-Test

Der Tukey-Test hält die versuchsbezogene Irrtumswahrscheinlichkeit ein. Für den Tukey-Test lässt sich wie für den LSD-Test eine gemeinsame Grenzdifferenz berechnen, sofern der Stichprobenumfang konstant ist. Die Grenzdifferenz für den Tukey-Test wird auch als Honestly Significant Difference (HSD) bezeichnet. Die Entscheidungsvorschrift lautet:

$\text{Verwerfe } H_0: \tau_s = \tau_u, \text{ wenn } \bar{y}_s - \bar{y}_u > q_{Tab} [t, t(r-1), \alpha] \sqrt{MQ_{Fehler}/r} = HSD$

Hierbei ist q_{Tab} die sog. studentisierte Variationsbreite, die in Tab. VII (Anhang) tabelliert ist. Diese hängt neben den Fehlerfreiheitsgraden von der Zahl der Behandlungen (t) ab. Die HSD wird umso größer, je mehr verschiedene Behandlungen

untersucht wurden, und sie ist generell größer als die LSD. Dies ist der Preis für die Einhaltung der versuchsbezogenen Irrtumswahrscheinlichkeit.

Beispiel: Für den Sortenversuch finden wir $q_{Tab}[t=5, t(r-1)=15, \alpha=5\%] = 4,367$ und somit

$$HSD = 4,367 \sqrt{7,733/4} = 6,07$$

Dies ist größer als die *LSD*. Dies ist der Preis, den wir für die Kontrolle der versuchsbezogenen anstelle der vergleichsbezogenen Irrtumswahrscheinlichkeit zahlen müssen. □

Es ist auch für die HSD möglich, signifikante Unterschiede durch Buchstaben kenntlich zu machen.

Beispiel: Für den Sortenversuch sind die Mittelwerte für die Sorten A, B, C, D, und E

$$\bar{y}_1. = 33, \bar{y}_2. = 22, \bar{y}_3. = 31, \bar{y}_4. = 31, \bar{y}_5. = 25$$

Diese ordnen wir der Größe nach und tragen sie in eine Kreuztabelle der folgenden Form ein:

Beispiel: Wir verrechnen die Weizen-Daten mit dem Tukey-Test ($HSD = 6,07$).

	A (33)	C (31)	D (31)	E (25)	B (22)
A (33)		2	2	8*	11*
C (31)			0	6	9*
D (31)				6	9*
E (25)					3
B (22)					

A C D E B

Eine redundante Linie (Nr. 3) kann wegfallen.

A C D E B
 _____ a
 _____ b
 _____ c

Eintragen in die Mittelwert-Tabelle liefert:

Sorte	Mittelwert
A	33 ^a
B	22 ^c
C	31 ^{ab}
D	31 ^{ab}
E	25 ^{bc}
<hr/>	
HSD($\alpha=5\%$)	6,07

Das Beispiel zeigt, dass es nötig sein kann, mehrere Buchstaben an einen Mittelwert zu schreiben, um alle Signifikanzen darzustellen. Wegen der größeren Grenzdifferenz finden wir weniger Signifikanzen: C und B sind nicht signifikant verschieden von E, wenn der Tukey-Test verwendet wird. Diese Vergleiche sind mit dem LSD-Test jedoch signifikant. □

Abschließende Bemerkung: Die Varianzanalyse wurde hier für den Fall balancierter Daten besprochen, also für den Fall, dass jede Behandlung gleich oft wiederholt wurde. Wenn die Daten unbalanciert sind, ist das Vorgehen grundsätzlich analog. Der wesentliche Unterschied besteht darin, dass dann nicht mehr mit einer gemeinsamen Grenzdifferenz gerechnet werden kann. In der Folge ist es nicht immer möglich, mit der beschriebenen Methode eine Liniendarstellung zu finden. Ein alternatives Verfahren, welches es erlaubt, zumindest eine Buchstabendarstellung zu erhalten, wird bei Piepho (2004; *Journal of Graphical and Computational Statistics* **13**, 456-466) beschrieben. Hierzu gibt es ein SAS macro %MULT, welches auf der Homepage des Fachgebietes verfügbar ist (siehe auch Piepho 2012: *Communications in Biometry and Crop Science* **7**, 4-13). Außerdem ist das Verfahren im Statistik-Programm R im Paket **multcomp** implementiert.

4.6 Ausgabe der Varianzanalyse mit einem Statistikpaket

Die Varianzanalyse für den Sortenversuch wird hier mit dem Statistik-Paket SAS durchgeführt. Die ausgegebene Varianzanalysetabelle sieht wie folgt aus:

The GLM Procedure					
Dependent Variable: ertrag					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	348.8000000	87.2000000	11.28	0.0002
Error	15	116.0000000	7.7333333		
Corrected Total	19	464.8000000 [^]			

Die Signifikanz des F-Wertes ($F_{Vers} = 11,28$) kann anhand des p-Wertes ($p = 0.0002$) abgelesen werden. Zur Interpretation von p-Werten, siehe Abschnitt 3.15. Im vorliegenden Fall ist $p < \alpha = 0.05$, also ist der F-Test auf dem 5%-Niveau signifikant.

Der multiple Mittelwertvergleich mit Hilfe des LSD-Test wird wie folgt ausgegeben:

The GLM Procedure

t Tests (LSD) for ertrag

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	15
Error Mean Square	7.733333
Critical Value of t	2.13145
Least Significant Difference	4.1912

Means with the same letter are not significantly different.

t Grouping	Mean	N	sorte
A	33.000	4	sorte_a
A			
A	31.000	4	sorte_d
A			
A	31.000	4	sorte_c
B	25.000	4	sorte_e
B			
B	22.000	4	sorte_b

Diese Ausgabe ist selbsterklärend. Die Buchstabendarstellung mit den Buchstaben A und B wird auf der linken Seite so ausgegeben, dass die ursprüngliche Liniendarstellung sichtbar wird, bei welcher eine Linie die Sorten A, C und D verbindet und die andere die Sorten B und E.

Einige Übersetzungen:

Degrees of freedom (DF)	= Freiheitsgrade
Sum of squares	= Quadratsumme
Mean square	= Mittelquadrat
Dependent variable	= abhängige Variable
Comparison-wise error rate	= vergleichsbezogene Fehlerrate
Experiment-wise error rate	= versuchsbezogene Fehlerrate

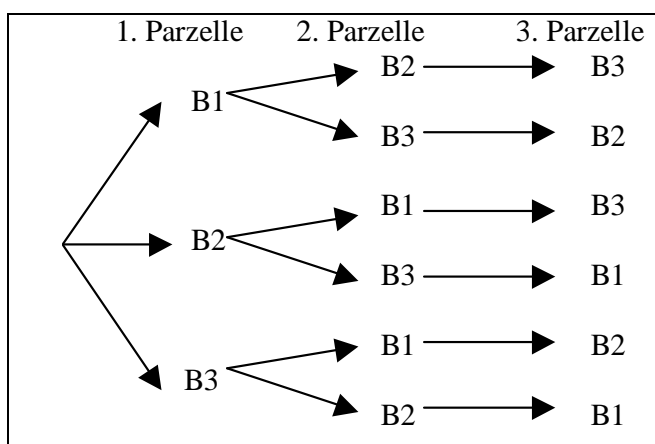
5. Einführung in die schließende Statistik für kategoriale Daten

In den Kapiteln 2, 3 und 4 haben wir uns ausschließlich mit metrischen Daten befasst, wobei im Falle der schließenden Statistik die zusätzliche Annahme getroffen wurde, dass die Daten einer Normalverteilung folgen. In diesem Kapitel kommen wir nun zu den kategorialen Daten, wobei die nominalen Daten im Vordergrund stehen, also Daten, bei welchen keine Rangordnung der Kategorien vorliegt. Bei kategorialen Daten werden in vielen Fällen Häufigkeiten für die verschiedenen Kategorien bestimmt. Für die Modellierung und für Tests solcher Häufigkeiten spielen zwei Verteilungen eine wichtige Rolle, die daher im folgenden detailliert erläutert werden sollen: Die **Binomialverteilung** (5.3) und die **Poissonverteilung** (5.4). Zuvor ist es notwendig, einige Grundzüge der Wahrscheinlichkeitsrechnung darzulegen (5.1 und 5.2). Am Ende des Kapitels werden verschiedene Tests besprochen (5.5 bis 5.7).

5.1 Kombinatorik

Permutationen: In vielen experimentellen Situationen ist eine wichtige Frage, wie viele mögliche Reihenfolgen (Permutationen) es für die Anordnung von k Objekten gibt. Dies ist beispielsweise bei der Randomisation von Versuchen wichtig.

Beispiel: In einem Feldstück, welches in $k = 3$ Parzellen unterteilt ist, sollen $k = 3$ verschiedene Behandlungen (B1, B2, B3) zufällig auf die Parzellen verteilt werden. Wie viele Möglichkeiten gibt es, die Behandlungen auf die Parzellen zu verteilen? Für die erste Behandlung haben wir 3 Wahlmöglichkeiten. Haben wir diese zugeordnet, sind noch 2 Parzellen frei. Für die zweite Behandlung bleiben also noch 2 Wahlmöglichkeiten. Für die letzte Behandlung bleibt dann keine Wahlmöglichkeit. Es ergeben sich daher insgesamt $3 \cdot 2 \cdot 1 = 3!$ (! = Fakultät) Permutationen.



Aus diesem Beispiel ergibt sich folgende Gesetzmäßigkeit:

Es gibt für k Objekte

$$k! = k \cdot (k-1) \cdot (k-2) \dots 3 \cdot 2 \cdot 1$$

mögliche Permutationen (lies: $k!$ = " k Fakultät").

Das Urnenmodell: Viele experimentelle Situationen lassen sich durch ein sog. Urnenmodell beschreiben. Hierbei wird angenommen, dass die Urne n Kugeln, die sich eindeutig unterscheiden lassen, z.B. durch Nummern. Aus der Urne werden k Kugeln gezogen. Wir können nun die Frage stellen, wie viele Möglichkeiten es gibt, aus der Urne k Kugeln zu ziehen. Bei der Antwort auf diese Frage sind zwei Fragen zu beachten:

1. Spielt die Reihenfolge, in der gezogen wird, eine Rolle?
2. Wird eine Kugel wieder zurückgelegt, nachdem sie gezogen wurde?

Wir geben hier zunächst für die vier möglichen Fälle Formeln an, die dann einzeln erläutert werden (Köhler et al., 1995).

Tab. 5.1.1: Zahl der Möglichkeiten, k Objekte aus n auszuwählen.

		Zurücklegen?	
		Ja	Nein
Reihenfolge wichtig?	Ja (Variation)	n^k	$\frac{n!}{(n-k)!}$
	Nein (Kombination)	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Im folgenden wird für jede der vier Möglichkeiten ein Beispiel gegeben.

(1) Variation mit Zurücklegen: Der genetische Code ist durch vier Basen T, C, A und G festgelegt. Ein Triplet von Basen kodiert für eine Aminosäure. Wie viele verschiedene Basentriplets gibt es? Wir können $n = 4$ Kugeln mit den Basenbezeichnungen in eine Urne legen. Ein Triplet ergibt sich durch Ziehen von $k = 3$ Kugeln mit Zurücklegen, wobei die Reihenfolge eine Rolle spielt. Für die erste Position gibt es $n = 4$ Möglichkeiten, für die zweite ebenfalls $n = 4$ Möglichkeiten, und für die dritte auch $n = 4$ Möglichkeiten. Also gibt es $4 \cdot 4 \cdot 4 = 4^3 = 64 = n^k$ Triplets.

(2) Variation ohne Zurücklegen: Bei einer sensorischen Prüfung von Quark sollen Probanden aus 5 Quarkproben die beste, zweitbeste und drittbeste bestimmen. Wie viele mögliche Ergebnisse für die ersten 3 Sorten gibt es? Die Möglichkeiten können mit dem Urnenmodell abgeleitet werden. Wir können uns eine Urne mit $n = 5$ Kugeln vorstellen, die den 5 Quarksorten entsprechen. Wir ziehen der Reihe nach $k = 3$ Kugeln ohne Zurücklegen. Die erste ist die beste, die zweite Kugel die zweitbeste und die dritte die drittbeste. Bei der ersten Kugel haben wir 5 Möglichkeiten, bei der zweiten 4, und bei der dritten 3. Also gibt es insgesamt $5 \cdot 4 \cdot 3$ Möglichkeiten. Dies entspricht $5!/2!$, oder allgemein $n!/(n-k)!$.

(3) Kombination ohne Zurücklegen: Wie viele Möglichkeiten gibt es, beim Lotto 6 aus 49 zu ziehen? Die Urne hat $n = 49$ Kugeln, und wir ziehen $k = 6$ Kugeln. Wir wissen, dass es $n!/(n-k)! = 49!/43!$ mögliche (Variationen) Ergebnisse gibt, wenn die Reihenfolge des Ziehens wichtig wäre. Wenn wir nun aber eine bestimmte Zahlenkombination gezogen haben, ist es unwichtig, in welcher Reihenfolge wir sie gezogen haben. So sind die Ergebnisse 1, 3, 8, 9, 23, 47 und 47, 1, 23, 9, 8, 3 identisch, was die Kombination angeht. Nun gibt es für jede Kombination $6! = k!$ Möglich-

keiten, die Kugeln in eine Reihenfolge zu bringen. Teilen wir daher die Zahl der möglichen Variationen, $n!/(n-k)! = 49!/43!$, durch $k! = 6!$, so erhalten wir die Zahl der möglichen Kombinationen:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{49!}{43!6!} = 13.983.816 \quad (\text{lies: "n über k"})$$

(4) Kombination mit Zurücklegen: Auf einer Grünlandfläche kommen 9 verschiedene Pflanzenarten (a, b, c, d, e, f, g, h, i) vor. Es wird eine Stichprobe von 6 Pflanzen gezogen. Wie viele Möglichkeiten gibt es für die Zusammensetzung der Stichprobe? Zu beachten ist, dass eine Art mehr als einmal in die Stichprobe gelangen kann. Für das Urnenmodell haben wir $n = 9$ verschiedene Kugeln (Arten) und ziehen $k = 6$ Kugeln mit Zurücklegen. Da die Reihenfolge keine Rolle spielt, gibt es gemäß Tab. 5.1.1

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!} = \frac{14!}{6!8!} = 3003$$

Kombinationen. Wie kann diese Formel erklärt werden?

Eine Plausibilitätserklärung für die Gültigkeit der Formel in unserem Beispiel verwendet ein anderes Modell anstelle des Urnenmodells. Es lautet wie folgt: Wir haben einen Kasten, in dem Plätze für 9 verschiedenen Pflanzenarten markiert sind. Diese Plätze werden durch 8 Trennwände voneinander getrennt, so dass 9 verschiedene Abteile entstehen:

a | b | c | d | e | f | g | h | i

Jetzt nehmen wir 6 identischen (!) Murmeln und lassen sie in den Kasten fallen, so, dass die 6 Murmeln sich ganz zufällig auf die 9 markierten Abteile verteilen, z. B. so:

a | b | (c) | (d) | (e) | f | g | (h) | (i)

In unserem Experiment haben die Abteile a, b, f und g keine, die Abteile c, d, e und i eine und das Abteil h zwei Murmeln bekommen. Wir stellen jetzt unsere Pflanzenstichprobe laut der Murmelverteilung zusammen: Wir nehmen je eine Pflanze der Arten c, d, e und i, sowie zwei Pflanzen der Art h und keine Pflanzen der Arten a, b, f und g. Von den Arten a, b, f und g nehmen wir keine Pflanze.

Nun stellen wir die Konstellation der 8 Trennwände und der 6 Murmeln schematisch dar. Für diese insgesamt 14 Objekte zeichnen wir 14 Kreise nebeneinander, die von 1 bis 14 nummeriert sind:

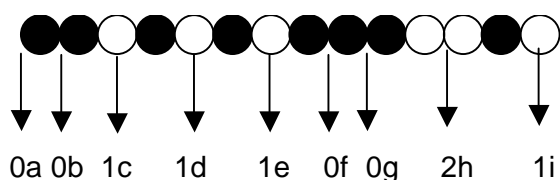


Die Kreise für Trennwände markieren wir schwarz, die Kreise für Murmeln lassen wir weiß.



Die Lage der weißen Kreise (Murmeln) zeigt nun, welche der 9 Arten in welcher Zahl in die Stichprobe gelangt.

Art	Anzahl der Pflanzen in Stichprobe	Beispiel
a	Anzahl weißer Kreise links des 1. schwarzen Kreises	0
b	Anzahl weißer Kreise zwischen 1. und 2. schwarzen Kreis	0
c	Anzahl weißer Kreise zwischen 2. und 3. schwarzen Kreis	1
d	Anzahl weißer Kreise zwischen 3. und 4. schwarzen Kreis	1
e	Anzahl weißer Kreise zwischen 4. und 5. schwarzen Kreis	1
f	Anzahl weißer Kreise zwischen 5. und 6. schwarzen Kreis	0
g	Anzahl weißer Kreise zwischen 6. und 7. schwarzen Kreis	0
h	Anzahl weißer Kreise zwischen 7. und 8. schwarzen Kreis	2
i	Anzahl weißer Kreise rechts des 8. schwarzen Kreis	1



Das Schema beschreibt also den Fall, dass die Stichprobe eine Pflanze der Arten c, d, e, und i sowie zwei Pflanzen der Art h hat ($k = 6$).

Alle möglichen Zusammensetzungen der Pflanzenprobe zu bestimmen bedeutet also alle möglichen Konstellationen der 8 Trennwände und 6 Murmeln zu bestimmen. Das entspricht der Anzahl der Möglichkeiten, 8 Kreise aus 14 vorhandenen zu schwärzen. Zum Lösen dieser Aufgabe können wir zum bekannten Urnenmodell zurückgreifen. Die Urne soll in diesem Fall 14 durchnummerierte Kugeln enthalten – für die 14 möglichen Positionen eines schwarzen Kreises. Wir ziehen dann ohne (!) Zurücklegen 8 Kugeln. Die Konstellation der Schwarzen und weißen Kreise, die auf dem Bild 2 dargestellt ist, entspricht z.B. dem Ziehen der Kugeln mit den Nummern 1, 2, 4, 6, 8, 9, 10, 13.

Insgesamt gibt es $(n+k-1)! = 14!$ Reihenfolgen (Variationen). Unter den schwarzen Kreisen (Trennwände) sowie unter den weißen (Murmeln) spielt die Reihenfolge aber keine Rolle. Die Zahl der möglichen Pflanzenkombinationen erhalten wir daher, indem wir $(n+k-1)! = 14!$ durch $k! = 6!$ und $(n-1)! = 9!$ teilen. So ergibt sich die Formel für Kombination mit Zurücklegen als

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!} = \frac{14!}{6!9!} = 3003$$

5.2 Einige wichtige Grundregeln der Wahrscheinlichkeitsrechnung

Wir benutzen hier folgende Begrifflichkeit und Symbolik:

Experiment: Messen einer Zufallsvariable
Zufallsvariable: X (groß geschrieben)
Realisation der Zufallsvariable (Ereignis): x (klein geschrieben)

Beispiele:

Experiment: Wähle zufällig eine Pflanze aus einem Bestand aus und bestimme den Gesundheitszustand

Zufallsvariable: X = Gesundheitszustand einer Pflanze

Mögliche Realisationen (Ereignis): x = gesund oder x = krank

Experiment: Wähle zufällig eine Pflanze eines Grünlandbestandes mit Arten a bis i aus

Zufallsvariable: X = Pflanzenart der zufällig bestimmten Pflanze

Mögliche Realisationen: $x = a, x = b, \dots, x = i$

Experiment: Wähle zufällig eine Pflanze einer F_2 -Generation einer Kreuzung aus Futtererbse (blaue Blüte) und Speiseerbse (weiße Blüte)

Zufallsvariable: X = Blütenfarbe bei Erbsen

Mögliche Realisationen: x = blau, x = weiss

Ein ganz entscheidender Aspekt bei den hier beschriebene experimentellen Situationen ist, dass das Ergebnis eines einzelnen Experimentes immer nur eines von mehr oder weniger vielen möglichen Ergebnissen (Realisationen einer Zufallsvariablen) ist. Die erste zufällig gewählte Pflanze mag krank sein, die zweite zufällig gewählte Pflanze kann dagegen vielleicht gesund sein. Das Ergebnis des Experimentes ist also nicht sicher voraussehbar.

Wir können allerdings in vielen Fällen Wahrscheinlichkeiten für das Auftreten der verschiedenen möglichen Realisationen einer Zufallsvariablen angeben. Die Auswertung der Daten besteht dann darin, solche Wahrscheinlichkeiten zu schätzen.

Beispiel: In einem Maisfeld mit 100.000 Pflanzen sind 5.000 Pflanzen mit dem Maiszünsler befallen. Wählen wir zufällig mehrere Pflanzen aus, so erwarten wir, dass ca. $5.000/100.000 \cdot 100\% = 5\%$ der Pflanzen befallen sind. Wir können auch sagen, dass die Wahrscheinlichkeit, dass eine einzelne, zufällig bestimmte Pflanze befallen ist, $P(\text{befallen}) = 5\%$ beträgt. Die Wahrscheinlichkeit, dass sie nicht befallen ist, beträgt $P(\text{nicht befallen}) = 95\%$. Wir können die Wahrscheinlichkeit auch als Zahl zwischen 0 und 1 darstellen, also $P(\text{befallen}) = 5\% = 0,05$ und $P(\text{nicht befallen}) = 95\% = 0,95$. Man beachte, dass $P(\text{befallen}) + P(\text{nicht befallen}) = 1$ ist.

Beispiel: Das Landsberger Gemenge ist eine Saatgutmischung aus Winterwicken, Inkarnatklee und Welschem Weidelgras. Angenommen, eine Mischung hat 100.000 Weidelgrassamen, 60.000 Wickensamen und 40.000 Kleesamen. Also sind 50% der Samen vom Weidelgras, 30% von der Wicke und 20% vom Klee. Die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Samenkorn vom Weidelgras (G) stammt, ist daher $P(G) = 0,5$. Die entsprechenden Wahrscheinlichkeiten für Wicke (W) und Klee (K) sind $P(W) = 0,3$ und $P(K) = 0,2$. Auch hier gilt, dass die Summe der Wahrscheinlichkeiten gleich 1 ist.

Beispiel: Ein Gen für die Resistenz der Zuckerrübe gegen Rizomania hat einen dominant-rezessiven Erbgang. Es gibt die Allele A = anfällig und a = resistent. Bei einer Kreuzung zweier Eltern mit Genotypen AA und aa können in der Aufspaltung der F_2 -Generation die Genotypen AA , Aa und aa im Verhältnis $1 : 2 : 1$ auftreten. Wegen der Dominanz von A über a ist AA = anfällig, Aa = anfällig und aa = resistent. Aus dem erwarteten Aufspaltungsverhältnis ergibt sich für die F_2 : $P(\text{anfällig}) = 3/4$ und $P(\text{resistent}) = 1/4$. Es gilt: $P(\text{anfällig}) + P(\text{resistent}) = 1$.

Beispiel (Lorenz, 1996): Das zufällige Ereignis, dass ein Unfallverletzter, der in die Klinik der Stadt XY eingeliefert wird und eine Bluttransfusion benötigt, die Blutgruppe 0 besitzt, beträgt etwa $P(0) = 0,38$. Dieser Wert ergibt sich aus den Unterlagen der Blutzentrale dieser Klinik. Danach wurde bei 38% aller Personen, deren Blutgruppe bestimmt worden war, die Gruppe 0 festgestellt (bei 42% die Blutgruppe A, bei 13% die Blutgruppe B und bei 7% die Blutgruppe AB). Es ist

$$P(0) + P(A) + P(B) + P(AB) = 1.$$

Beispiel: Beim Werfen einer "fairen" Münze ist die Wahrscheinlichkeit für das Ergebnis "Kopf" gleich $P(\text{Kopf}) = 0,5$ und für Zahl gleich $P(\text{Zahl}) = 0,5$. Denn in 50% der Fälle erwarten wir "Kopf" und in 50% erwarten wir "Zahl" als Ergebnis. Es gilt $P(\text{Kopf}) + P(\text{Zahl}) = 1$.

Beispiel: Beim Werfen eines fairen Würfels hat jede Augenzahl die Wahrscheinlichkeit $1/6$. Es gilt $P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$.

Aus den Beispielen ergibt sich die erste Regel:

Die Summe aller Auftretenswahrscheinlichkeiten $P(\cdot)$ eines Zufallsexperiments ist 1.

Eine zweite Regel lautet:

Additionssatz: Sind A , B , usw. zufällige Ereignisse, die sich gegenseitig ausschließen und deren Wahrscheinlichkeiten $P(A)$, $P(B)$, usw. betragen, so gilt:

$$P(A \text{ oder } B \text{ oder } \dots) = P(A) + P(B) + \dots$$

Beispiel: Ein Landsberger Gemeinde hat 50% Weidelgrassamen (G), 30% Wicken-samen (W) und 20% Kleesamen (K). Es ist daher $P(G) = 0,5$, $P(W) = 0,3$ und $P(K) = 0,2$. Wir können nun nach der Wahrscheinlichkeit fragen, dass ein zufällig gezogenes Samenkorn nicht vom Weidelgras stammt. Diese ist gegeben durch

$$P(W \text{ oder } K) = P(W) + P(K) = 0,3 + 0,2 = 0,5.$$

Beispiel: Die Wahrscheinlichkeit, beim Würfeln eine gerade Zahl zu werfen, ist

$$P(\text{gerade Zahl}) = P(2) + P(4) + P(6) = 1/6 + 1/6 + 1/6 = 1/2.$$

Der **Additionssatz** betrifft die Kombination von Ereignissen oder möglichen Realisationen einer Zufallsvariablen, die sich gegenseitig ausschließen. Die möglichen Er-

Ereignisse stehen in einer **Entweder-oder-Beziehung**. Dagegen verknüpft der **Multiplikationssatz** Ereignisse, die in einer **Sowohl-als-auch-Beziehung** stehen. Die Ereignisse müssen voneinander unabhängig sein.

Multiplikationssatz (Produktregel): Sind A, B, \dots usw. voneinander unabhängige zufällige Ereignisse mit den Wahrscheinlichkeiten $P(A), P(B), \dots$ usw., so ist die Wahrscheinlichkeit für das gemeinsame Eintreten der Ereignisse A, B, \dots usw. gleich dem Produkt der Einzelwahrscheinlichkeiten

$$P(A \text{ und } B \text{ und } \dots) = P(A) \cdot P(B) \cdot \dots$$

Beispiel: Ein Supermarkt führt ein Gewinnspiel durch, bei dem jeder Kunde vier mal würfelt. Wenn er vier mal die 6 würfelt, bekommt er seinen Einkauf umsonst. Wie groß ist die Wahrscheinlichkeit, bei diesem Spiel zu gewinnen? Die Antwort ergibt sich durch Anwendung des Multiplikationssatzes. Die einzelnen Würfe des Würfels sind unabhängig. Die Wahrscheinlichkeit, bei einem Wurf eine 6 zu würfeln, ist $1/6$. Diese Wahrscheinlichkeit gilt für alle vier Würfe. Anwendung des Multiplikationssatzes ergibt:

$$\begin{aligned} P(\text{vier mal die 6}) \\ &= P(6 \text{ beim 1. Wurf}) \cdot P(6 \text{ beim 2. Wurf}) \cdot P(6 \text{ beim 3. Wurf}) \cdot P(6 \text{ beim 4. Wurf}) \\ &= 1/6 \cdot 1/6 \cdot 1/6 \cdot 1/6 = 0,00077 = 0,077\% \end{aligned}$$

Bemerkung: Es ergibt sich dasselbe Ergebnis, wenn vier Würfel auf einmal mit einem Knobelbecher geworfen werden. Denn die Augenzahlen der verschiedenen Würfel sind auch in diesem Fall unabhängig, so dass der Multiplikationssatz anwendbar ist.

Beispiel (Lorenz): Wir betrachten zwei verschiedene Blutgruppensysteme. Für das AB0 System gelten die Wahrscheinlichkeiten

$$P(0) = 0,38; P(A) = 0,42; P(B) = 0,13; P(AB) = 0,07$$

Daneben gibt es den sog. Rhesus-Faktor mit Ausprägungen "Rh+" und "Rh-". 85% der (weißen) Bevölkerung sind Rh+ und 15% sind Rh-. Also gilt

$$P(\text{Rh}+) = 0,85 \text{ und } P(\text{Rh}-) = 0,15$$

Es ist bekannt, dass der Rhesus-Faktor unabhängig von der AB0-Blutgruppe vererbt wird, da die verantwortlichen Gene auf verschiedenen Chromosomen liegen. Dies bedeutet, dass die Wahrscheinlichkeit, einen Rh+-Träger anzutreffen, nicht davon abhängt, welche AB0-Blutgruppe er hat. Die Merkmalsausprägung Rh+ ist in allen Gruppen des AB0-Systems gleich häufig vertreten, nämlich zu 85%.

Betrachtet man beide Blutgruppensysteme gleichzeitig, so gliedert sich die Bevölkerung in insgesamt 8 Klassen. Für jede dieser 8 Klassen können wir mit Hilfe der Produktregel berechnen, mit welcher Wahrscheinlichkeit eine Person, die wir zufällig antreffen, in eine bestimmte Klasse fällt. So ist z.B.

$$P(0 \text{ und } \text{Rh}+) = P(0) \cdot P(\text{Rh}+) = 0,38 \cdot 0,85 = 0,3230$$

Zufälliges Ereignis	Wahrscheinlichkeit
0 und Rh+	$0,38 \cdot 0,85 = 0,3230$
A und Rh+	$0,42 \cdot 0,85 = 0,3570$
B und Rh+	$0,13 \cdot 0,85 = 0,1105$
AB und Rh+	$0,07 \cdot 0,85 = 0,0595$
0 und Rh–	$0,38 \cdot 0,15 = 0,0570$
A und Rh–	$0,42 \cdot 0,15 = 0,0630$
B und Rh–	$0,13 \cdot 0,15 = 0,0195$
AB und Rh–	$0,07 \cdot 0,15 = 0,0105$
Summe =	1,0000

Die Summe der Wahrscheinlichkeiten für die acht möglichen Ereignisse hat wiederum den Wert 1. Die Wahrscheinlichkeiten können übersichtlich nach den beiden Blutgruppen-Systemen auch in einer Kreuztabelle angegeben werden. Solche **Kreuztabellen** oder auch **Kontingenztafeln** sind sehr gängig bei der Auswertung kategorialer Daten.

Tab. 5.2.1: Kontingenz-Tafel für Wahrscheinlichkeiten in AB0 × Rhesus System

		Blutgruppe im AB0 System				
		0	A	B	AB	Summe
Blutgruppe im Rhesus-System	Rh+	0,3230	0,3570	0,1105	0,0595	0,8500
	Rh–	0,0570	0,0630	0,0195	0,0105	0,1500
	Summe	0,3800	0,4200	0,1300	0,0700	1,0000

Beispiel: Aus der Mendelschen Genetik ist die Spaltungsregel bekannt, die besagt, dass bei einem durch ein einziges Gen bestimmten Merkmal das Aufspaltungsverhältnis in der F₂ Generation 3:1 beträgt, sofern ein dominant-rezessiver Erbgang vorliegt. Wie es zu dieser Regel kommt, kann mit den elementaren Gesetzen der Wahrscheinlichkeitsrechnung erklärt werden. Ausgangspunkt der Betrachtung ist eine F₁-Erbsen-Pflanze, die aus der Kreuzung reinerbiger Eltern hervorgegangen ist. Der

eine Elter hat den Genotyp AA und den Phänotyp "glatte Samen", der andere den Genotyp aa und den Phänotyp "runzelige Samen". Daher hat eine F₁-Pflanze den Genotyp Aa. Wegen der Dominanz von A über a hat die Aa-Pflanze den Phänotyp "glatte Samen". Werden Samen der F₁-Pflanze ausgesät, so findet man in der F₂-Generation Pflanzen mit glatten Samen und Pflanzen mit runzeligen Samen etwa im Verhältnis 3 : 1. Der Grund ist wie folgt: Die Erbse ist ein Selbstbefruchter. In der Meiose bildet die F₁-Pflanze aus somatischen Zellen (diploid) haploide Gameten. Da die somatischen Zellen den Genotyp Aa haben, werden je zur Hälfte A-Gameten und a-Gameten gebildet. Dies gilt sowohl für die männlichen als auch für die weiblichen Samenanlagen. Formal können wir schreiben:

$$P(\text{weibliche Gamete A}) = 1/2; \quad P(\text{weibliche Gamete a}) = 1/2$$

$$P(\text{männliche Gamete A}) = 1/2; \quad P(\text{männliche Gamete a}) = 1/2$$

Eine F₂-Pflanze hat nun den Genotyp aa, wenn beide Gameten, aus der die Zygote der F₂-Pflanze entstanden ist, den Genotyp a hatten. Die Wahrscheinlichkeit hierfür ist

$$\begin{aligned} P(\text{F}_2\text{-Pflanze ist aa}) &= P(\text{männliche Gamete a und weibliche Gamete a}) \\ &= P(\text{männliche Gamete a}) \cdot P(\text{weibl. Gamete a}) = 1/2 \cdot 1/2 = 1/4 \end{aligned}$$

Analog finden wir

$$\begin{aligned} P(\text{F}_2\text{-Pflanze ist AA}) &= P(\text{männliche Gamete A und weibliche Gamete A}) \\ &= P(\text{männliche Gamete A}) \cdot P(\text{weibl. Gamete A}) = 1/2 \cdot 1/2 = 1/4 \end{aligned}$$

Bei den Aa-Pflanzen bestehen zwei Möglichkeiten, wie es zum Genotyp Aa kommen kann. Entweder, die männliche Gamete ist A und die weibliche ist a, oder aber die männliche ist a und die weibliche A. Um nun zur Wahrscheinlichkeit für eine Aa-Pflanze zu gelangen, müssen wir sowohl die Produktregel als auch die Additionsregel anwenden. Es gilt:

$$P(\text{F}_2\text{-Pflanze ist Aa}) = P[(\text{männliche Gamete A und weibliche Gamete a}) \text{ oder } (\text{männliche Gamete a und weibliche Gamete A})]$$

$$\begin{aligned} &= P(\text{männliche Gamete A und weibliche Gamete a}) \\ &+ P(\text{männliche Gamete a und weibliche Gamete A}) \end{aligned}$$

$$\begin{aligned} &= P(\text{männliche Gamete A}) \cdot P(\text{weibl. Gamete a}) \\ &+ P(\text{männliche Gamete a}) \cdot P(\text{weibl. Gamete A}) \end{aligned}$$

$$= 1/2 \cdot 1/2 + 1/2 \cdot 1/2 = 1/4 + 1/4 = 1/2$$

		Weibliche Gamete	
		a	A
		$P = 1/2$	$P = 1/2$
Männliche Gamete	a	aa $P = 1/4$	aA $P = 1/4$
	A	Aa $P = 1/4$	AA $P = 1/4$
		$P = 1/2$	$P = 1/2$

Nun zum Phänotyp: AA und Aa Pflanzen haben glatte Samen. Also ist

$$P(\text{F}_2\text{-Pflanze hat glatte Samen}) = P(\text{F}_2\text{-Pflanze ist AA}) + P(\text{F}_2\text{-Pflanze ist Aa}) \\ = 1/4 + 1/2 = 3/4$$

Außerdem ist

$$P(\text{F}_2\text{-Pflanze hat runzelige Samen}) = P(\text{F}_2\text{-Pflanze ist aa}) = 1/4$$

Also erwarten wir zu $3/4$ F₂-Pflanzen mit glatte Samen und zu $1/4$ F₂-Pflanzen mit runzeligen Samen, was einem Aufspaltungsverhältnis von 3 : 1 entspricht.

Beispiel: Nach der Mendelschen Unabhängigkeitsregel spalten zwei unabhängig vererbte Merkmale im Verhältnis 9 : 3 : 3 : 1 auf, sofern beide Merkmale dominant-rezessiv sind. Unabhängige Vererbung liegt z.B. dann vor, wenn die betreffenden Gene auf verschiedenen Chromosomen liegen. So sind z.B. bei der Erbse glatte Samen dominant über runzelige Samen und spalten nach dem Verhältnis 3 : 1 auf. Bei der Samenfarbe ist gelb dominant über grün und die Aufspaltung ist ebenfalls 3 : 1. Bei Betrachtung beider Merkmale ergibt sich

glatt +	:	glatt +	:	runzelig +	:	runzelig +
gelb		grün		gelb		grün
9	:	3	:	3	:	1

Die Ableitung ist wie folgt. Es gilt in der F₂-Generation für jedes einzelne Merkmal

$$P(\text{glatt}) = 3/4 \text{ und } P(\text{runzelig}) = 1/4 \text{ sowie}$$

$$P(\text{gelb}) = 3/4 \text{ und } P(\text{grün}) = 1/4$$

Wegen der Unabhängigkeit kann die Produktregel angewendet werden, so dass gilt:

$$\begin{aligned}
P(\text{glatt und gelb}) &= P(\text{glatt}) \cdot P(\text{gelb}) = 3/4 \cdot 3/4 = 9/16 \\
P(\text{glatt und grün}) &= P(\text{glatt}) \cdot P(\text{grün}) = 3/4 \cdot 1/4 = 3/16 \\
P(\text{runzelig und gelb}) &= P(\text{runzelig}) \cdot P(\text{gelb}) = 1/4 \cdot 3/4 = 3/16 \\
P(\text{runzelig und grün}) &= P(\text{runzelig}) \cdot P(\text{grün}) = 1/4 \cdot 1/4 = 1/16
\end{aligned}$$

$$\text{Summe} = 16/16 = 1$$

Aus diesen Wahrscheinlichkeiten ergibt sich das Aufspaltungsverhältnis 9 : 3 : 3 : 1.

Der hier betrachtete Multiplikationssatz gilt ausschließlich bei unabhängigen Ereignissen. Es gibt eine wichtige Verallgemeinerung dieses Satzes für den Fall abhängiger Ergebnisse, die unter der Bezeichnung „Bayes Formel“ oder Bayes Theorem“ firmiert. Da wir diese Verallgemeinerung hier nicht brauchen, wird sie auch nicht näher betrachtet. Sie gehört aber zu den ganz wichtigen elementaren statistischen Resultaten und Regeln, weshalb sie hier zumindest erwähnt werden soll. Näheres ist sehr gut beschrieben in Bosch (2007: Basiswissen Statistik).

5.3 Binomialverteilung

Beispiel: In einem Maisfeld aus 100.000 Maispflanzen befinden sich 5.000 Pflanzen, die mit dem Maiszünsler befallen sind. Die Befallswahrscheinlichkeit beträgt also $P(\text{Befall}) = 0,05$. Wir führen jetzt die folgenden einfachen Bezeichnungen ein:

$$\begin{aligned}
p &= P(\text{Befall}) = 0,05 & \text{und} \\
q &= P(\text{kein Befall}) = 1 - p = 0,95
\end{aligned}$$

Wir überlegen nun, welche möglichen Stichproben wir mit welcher Wahrscheinlichkeit erhalten können in Abhängigkeit von p und vom Stichprobenumfang n , wobei wir uns auf die Fälle $n = 1$ bis $n = 3$ beschränken. Wir beachten zunächst die Reihenfolge, in der die Pflanzen untersucht werden. Die Wahrscheinlichkeiten ergeben sich einfach durch Anwendung des Multiplikationssatzes. So ist die Wahrscheinlichkeit, dass von $n = 2$ Pflanzen beide befallen sind, gleich

$$P(\text{beide Pflanzen befallen}) = P(1. \text{ Pflanze befallen}) \cdot P(2. \text{ Pflanze befallen}) = p \cdot p = p^2$$

Für Stichprobenumfänge von $n = 1, 2$ und 3 sind in der folgenden Tabelle alle möglichen Fälle und Wahrscheinlichkeiten aufgeführt.

Stichproben- umfang (n)	Mögliche Ergebnisse	Wahrscheinlichkeit	Zahl befallener Pflanzen (k)
1	1. befallen	p	1
	1. gesund	q	0
2	1. befallen 2. befallen	p^2	2
	1. befallen 2. gesund	pq	1
	1. gesund 2. befallen	qp	1
	1. gesund 2. gesund	q^2	0
3	1. befallen 2. befallen 3. befallen	p^3	3
	1. befallen 2. befallen 3. gesund	p^2q	2
	1. befallen 2. gesund 3. befallen	pqp	2
	1. gesund 2. befallen 3. befallen	qp^2	2
	1. befallen 2. gesund 3. gesund	pq^2	1
	1. gesund 2. befallen 3. gesund	qpq	1
	1. gesund 2. gesund 3. befallen	q^2p	1
	1. gesund 2. gesund 3. gesund	q^3	0

Für praktische Belange ist es unwichtig, in welcher Reihenfolge die Pflanzen untersucht werden. Wichtig ist nur, wie viele der n untersuchten Pflanzen befallen sind. Daher ist z.B. die Frage von Interesse, mit welcher Wahrscheinlichkeit in einer Stichprobe vom Umfang $n = 2$ keine, eine oder zwei Pflanzen befallen sind. Zu einer befallenen Pflanze kommt es, wenn die erste befallen und die zweite gesund ist, oder die erste gesund und die zweite befallen ist. Die Wahrscheinlichkeit, eine befallene Pflanze in der Stichprobe zu haben, berechnet sich nach der Additionsregel:

$$\begin{aligned} P(1 \text{ befallene Pflanze}) &= P(1. \text{ befallen und } 2. \text{ gesund}) \\ &+ P(1. \text{ gesund und } 2. \text{ befallen}) \\ &= pq + qp = 2pq \end{aligned}$$

Für die Stichprobenumfänge 1, 2 und 3 finden wir:

Stichproben- umfang (n)	Zahl befallener Pflanzen (k)	Wahrscheinlichkeit	---Beispiel ($p = 0,05$)--- P %	
1	1	p	0,05	5%
	0	q	0,95	95%
2	2	p^2	0,0025	0,25%
	1	$2pq$	0,0950	9,50%
	0	q^2	0,9025	90,25%
3	3	p^3	0,000125	0,01%
	2	$3p^2q$	0,007125	0,71%
	1	$3pq^2$	0,135375	13,54%
	0	q^3	0,857375	85,57%

Bei einem Stichprobenumfang von $n = 3$ beträgt die Wahrscheinlichkeit, dass $k = 2$ Pflanzen von dreien befallen ist, nur 0,71%. Wenn sehr viele Stichproben vom Umfang $n = 3$ gezogen werden, erwarten wir in 0,71% der Fälle zwei befallene Pflanzen. Wir können dies formal ausdrücken als:

$$P(X = 2 | n = 3; p = 0,05) = 0,007125 = 0,71\%$$

wobei X die Zufallsvariable "Zahl befallener Pflanzen" bezeichnet. Das geübte Auge erkennt in den oben angegebenen Wahrscheinlichkeiten die binomische Formel wieder. Es ist

$$\begin{aligned} (p + q) &= p + q = 1 \\ (p + q)^2 &= p^2 + 2pq + q^2 = 1 \\ (p + q)^3 &= p^3 + 3p^2q + 3pq^2 + q^3 = 1 \\ (p + q)^4 &= p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4 = 1 \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned}$$

$$(p + q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = 1$$

Hieraus ergeben sich die

Binomialwahrscheinlichkeiten:

$$P(X=k|n;p) = \binom{n}{k} p^k q^{n-k}$$

n = Stichprobenumfang

X = Zahl Stichprobenelemente mit einer bestimmten Eigenschaft (Zufallsvariable)

k = ein möglicher Wert für die Realisation von X

p = Wahrscheinlichkeit, dass ein einzelnes Element die Eigenschaft hat

Zur weiteren Erklärung dieser Formel: Wenn z.B. $n = 4$ Pflanzen untersucht werden, so gibt es insgesamt 6 Möglichkeiten, wie es zu $k = 2$ befallenen Pflanzen kommen kann. Die Zahl 6 ist einfach die Zahl der **Kombinationen** beim Ziehen von $k = 2$ aus $n = 4$ **ohne Zurücklegen**. Und die Zahl der Kombinationen ist, wie wir wissen, gleich

$\binom{n}{k}$. Jede dieser Möglichkeiten hat die Wahrscheinlichkeit $p^2 q^2 = p^k q^{n-k}$. Nach dem

Additionssatz ist dann die Wahrscheinlichkeit, 2 befallene unter insgesamt 4 Pflanzen

zu beobachten, gleich $6 p^2 q^2 = \binom{n}{k} p^k q^{n-k}$.

Man bezeichnet die Werte für $\binom{n}{k}$ ($k = 0, \dots, n$) auch als die Binomialkoeffizienten.

Diese können auch aus dem sog. Paskalschen Dreieck abgeleitet werden:

				1			
			1		1		
		1		2		1	
	1		3		3		1
	1	4		6		4	1
1		5	10		10	5	1
.

Beispiel: Die Keimfähigkeit einer Partie Saatgut für Zucchini beträgt 75%. Ein Saatgutvertrieb packt die Samen in kleinen Tüten à 5 Samen ab. Es soll die Wahrscheinlichkeitsverteilung der Zahl keimfähiger Samen je Beutel berechnet werden. Insbesondere ist von Interesse zu wissen, mit welcher Wahrscheinlichkeit eine Tüte weniger als 3 keimfähigen Samen enthält, also weniger als 50% keimfähige Samen.

Zahl keimfähiger Samen (k); $n = 5$	Wahrscheinlichkeit $P(X = k n; p) = \binom{n}{k} p^k q^{n-k}$
0	0,00098
1	0,01465
2	0,08789
3	0,26367
4	0,39551
5	0,23730

Die Wahrscheinlichkeit, dass eine Tüte weniger als drei Samen enthält, ist

$$P(\text{weniger als 3 keimfähige Samen}) = 0,00098 + 0,01465 + 0,08789 = 0,10352$$

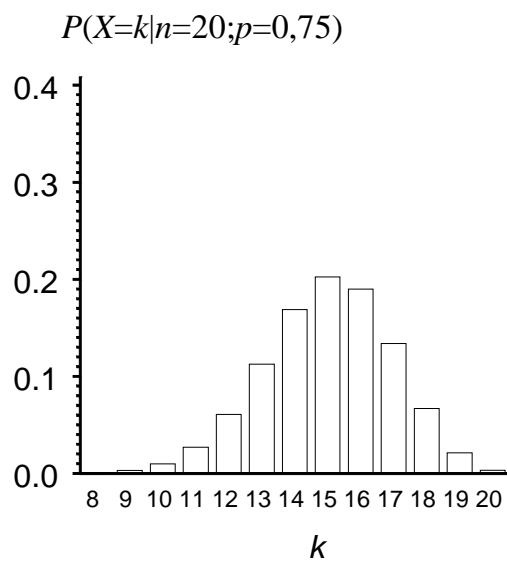
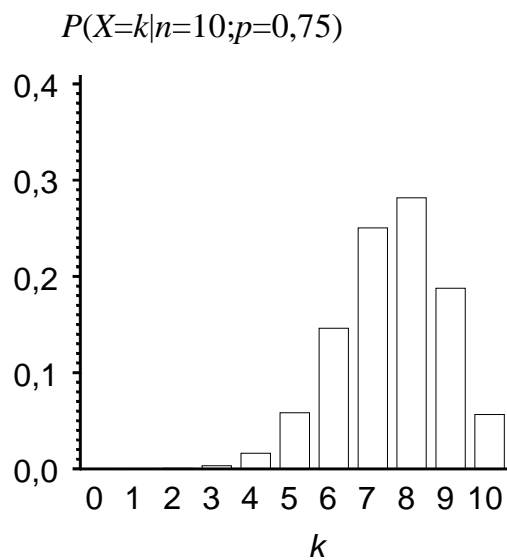
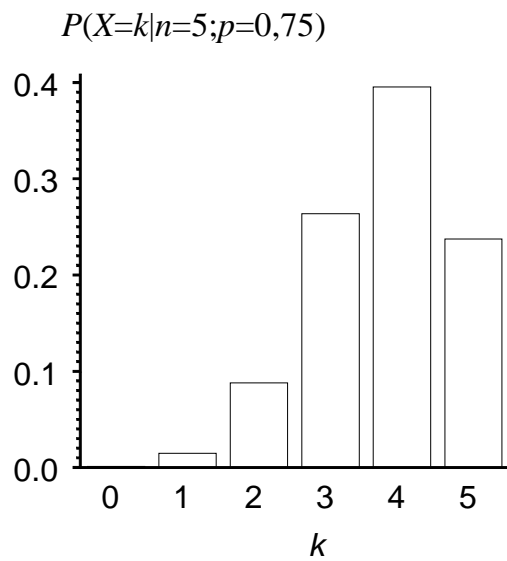
Also ist in etwa 10% der Tüten die Zahl keimfähiger Samen kleiner als 3. Dies wird als zu hoch angesehen. Daher überlegt der Versand, die Zahl der Samen auf 10 zu erhöhen. Hierbei ergeben sich folgende Wahrscheinlichkeiten:

Zahl keimfähiger Samen (k); $n = 10$	Wahrscheinlichkeit $P(X = k n; p) = \binom{n}{k} p^k q^{n-k}$
0	0,00000
1	0,00003
2	0,00039
3	0,00309
4	0,01622
5	0,05840
6	0,14600
7	0,25028
8	0,28157
9	0,18771
10	0,05631

Der Anteil von Tüten mit weniger als 50% keimfähigen Samen (weniger als 5) ist auf unter 2% gesunken.

Beachtenswert ist die Tatsache, dass die Wahrscheinlichkeitsverteilung für wachsendes n der Normalverteilung immer ähnlicher wird. Um dies zu verdeutlichen, ist in Abb. 5.3.1 auch der Fall $n = 20$ abgebildet. Wir sehen auch hier die Wirksamkeit des zentralen Grenzwertsatzes. Für $n \rightarrow \infty$ (n gegen Unendlich) geht die Binomialverteilung in die Normalverteilung über.

Abb. 5.3.1: Verschiedene Binomialverteilungen



Wenn n groß ist und alle Binomialwahrscheinlichkeiten berechnet werden sollen, so

ist für das Rechnen mit dem Taschenrechner die im folgenden beschriebene Regel hilfreich. Hierbei wird zunächst die Wahrscheinlichkeit für $k = 0$ berechnet. Aus dieser wird dann die Wahrscheinlichkeit für $k = 1$ berechnen, welche dann wiederum Grundlage für die Berechnung für $k = 2$ ist, usw. Man kann die Wahrscheinlichkeit $P(X=k+1|p; n)$ so umformen, dass ein Ausdruck entsteht, der von $P(X=k|p; n)$ abhängt:

$$P(X=k+1|n;p) = \binom{n}{k+1} p^{k+1} q^{n-k-1} = \frac{n!}{(n-k-1)!(k+1)!} p^{k+1} q^{n-k-1}$$

$$= \frac{(n-k)n!p}{(n-k)!(k+1)k!q} p^k q^{n-k} = \frac{n-k}{k+1} \frac{p}{1-p} P(X=k|n;p)$$

Wir haben eine Formel gefunden, die es ermöglicht, $P(X=k+1|p; n)$ aus $P(X=k|p; n)$ bzw. $P(X=k|p; n)$ aus $P(X=k-1|p; n)$ zu berechnen. Man nennt dies eine **Rekursionsformel**.

Rekursionsformel:

(1) Berechne

$$P(X=0|n; p) = (1-p)^n$$

(2) Berechne

$$P(X=k|n; p) = \frac{n-k+1}{k} \frac{p}{1-p} P(X=k-1|n; p)$$

für $k = 1, 2, \dots, n$

Beispiel: $n = 5; p = 0,75$

$$P(X=0|n; p) = (1-p)^n = 0,25^5 = 0,0009765625$$

$$P(X=1|n; p) = \frac{5}{1} \frac{0,75}{0,25} 0,0009765625 = 0,014648437$$

$$P(X=2|n; p) = \frac{4}{2} \frac{0,75}{0,25} 0,014648437 = 0,087890625$$

$$P(X=3|n; p) = \frac{3}{3} \frac{0,75}{0,25} 0,087890625 = 0,263671875$$

$$P(X=4|n; p) = \frac{2}{4} \frac{0,75}{0,25} 0,263671875 = 0,395507812$$

$$P(X=5|n; p) = \frac{1}{5} \frac{0,75}{0,25} 0,395507812 = 0,237304687$$

5.3.1 Mittelwert und Varianz der Binomialverteilung

Intuitiv ist zu erwarten, dass bei einer Stichprobe vom Umfang n aus einer Binomialverteilung der Erwartungswert für die Zufallsvariable X gleich np ist. Für die Varianz gilt eine ähnlich einfache Beziehung.

Mittelwert und Varianz der Binomialverteilung sind gegeben durch:

Mittelwert: $E(X) = np$

Varianz: $\text{var}(X) = npq$

wobei $q = 1 - p$.

Beispiel: Wenn im Maisfeld 5% der Pflanzen ($p = 0,05$) mit dem Maiszünsler befallen sind, dann erwarte ich in einer Stichprobe vom Umfang $n = 100$, dass $np = 5$ Pflanzen mit dem Maiszünsler befallen sind.

Diese Resultate lassen sich auch theoretisch zeigen. So ist

$$E(X) = \mu = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np.$$

Die einzelnen Ableitungsschritte werden nicht gezeigt (Tipp: $\sum_{i=0}^n \binom{n}{i} p^i q^{n-i} = 1$ und

$$\sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} p^{(k-1)} q^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{(k-1)} q^{(n-1)-(k-1)}.$$

Ebenso lässt sich zeigen:

$$\text{var}(X) = E[(X - np)^2] = \sigma^2 = \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k q^{n-k} = npq.$$

Eine besonders einfache Herleitung macht Gebrauch von der Tatsache, dass sich die Zufallsvariable X ausdrücken lässt als

$$X = Y_1 + Y_2 + \dots + Y_n,$$

wobei Y_i ($i = 1, \dots, n$) die Werte 0 und 1 annehmen kann. Bezüglich des Urnenmodells gilt, dass $Y_i = 1$ ist wenn eine gezogene Kugel eine bestimmte Eigenschaft aufweist (Kugel schwarz) und dass $Y_i = 0$ ist, wenn die Kugel die Eigenschaft nicht aufweist (Kugel weiß). Y_i folgt einer sog. Bernoulli-Verteilung. Es gilt offensichtlich $E(Y_i) = p \times 1 + q \times 0 = p$ sowie $E(Y_i^2) = E(Y_i) = p$, weil $Y_i^2 = Y_i$. Hieraus folgt zunächst für die Varianz von Y_i :

$$\text{var}(Y_i) = E\{[Y_i - E(Y_i)]^2\} = E\{Y_i^2 - 2Y_i E(Y_i) + [E(Y_i)]^2\} = E(Y_i^2) - [E(Y_i)]^2 = p - p^2 = pq.$$

Da die Y_i stochastisch unabhängig sind, ergibt sich dann für Erwartungswert und Varianz von X :

$$E(X) = E(Y_1 + Y_2 + \dots + Y_n) = E(Y_1) + E(Y_2) + \dots + E(Y_n) = p + p + \dots + p = np \quad \text{sowie}$$

$$\text{var}(X) = \text{var}(Y_1) + \text{var}(Y_2) + \dots + \text{var}(Y_n) = pq + pq + \dots + pq = npq.$$

5.3.2 Schätzen des Parameters p der Binomialverteilung

n = Stichprobenumfang

x = Zahl Stichprobenelemente mit einer bestimmten Eigenschaft

p = Wahrscheinlichkeit, dass ein einzelnes Element die Eigenschaft hat

Parameterschätzung: $\hat{p} = \frac{x}{n}$

Beispiel: Aus einem Maisfeld werden zufällig 100 Pflanzen ausgewählt. Von diesen sind 7 mit dem Maiszünsler befallen. Es soll die Befallswahrscheinlichkeit für das Feld geschätzt werden.

$$n = 100; x = 7; \hat{p} = \frac{7}{100} = 0,07$$

Die geschätzte Befallswahrscheinlichkeit beträgt 7%. Wir schätzen, dass 7% der Maispflanzen mit dem Maiszünsler befallen sind. Man beachte, dass wir auch hier von einer Stichprobe ($n = 100$ Pflanzen) auf eine Grundgesamtheit (Maisfeld) schließen.

Es muss bei der Interpretation des Schätzwertes für den Parameter einer Binomialverteilung berücksichtigt werden, wie groß der Stichprobenfehler ist. Diesen können wir als Varianz oder Standardfehler angeben.

Varianz der Parameterschätzung:

$$\text{var}(\hat{p}) = \frac{p(1-p)}{n}$$

Standardfehler der Parameterschätzung:

$$s_{\hat{p}} = \sqrt{\text{var}(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}$$

Um die Formeln praktisch zu nutzen, muss der unbekannte wahre Wert p durch seinen Schätzwert \hat{p} ersetzt werden.

Beispiel: $n = 100$; $x = 7$; $\hat{p} = 0,07$

$$\text{var}(\hat{p}) = \frac{p(1-p)}{n} \approx \frac{0,07 \cdot 0,93}{100} = 0,000651$$

$$s_{\hat{p}} = \sqrt{\text{var}(\hat{p})} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{0,000651} = 0,0255$$

Wir haben hier das "ungefähr gleich" Zeichen \approx verwendet, weil der wahre Wert p durch den Schätzwert ersetzt wurde.

Stichprobenplanung: Der Standardfehler der Schätzung von p hängt von p selbst ab. Wir können nun z.B. fordern, dass der Standardfehler einem Wert von C (ausgedrückt in %) entspricht, also

$$s_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \frac{C}{100}.$$

Auflösen nach n liefert den notwendigen Stichprobenumfang:

$$n = \frac{p(1-p)10^4}{C^2}.$$

Der ungünstigste Fall tritt dann aus, wenn $p = 0.5$ beträgt. In diesem Fall ist

$$n = \frac{0,5 \cdot 0,5 \cdot 10000}{C^2} = \frac{2500}{C^2}$$

zu wählen.

Beispiel: Bei einer Wahlumfrage soll der Standardfehler der Schätzung des Stimmenanteils einer Partei nicht mehr als 1% ($=0.01$) betragen. Also muss $n = 2500$ gewählt werden.

5.3.3 Tests für den Parameter der Binomialverteilung

Der Schätzer \hat{p} ist für großes n näherungsweise normalverteilt. Hierzu beachte man, dass sich als arithmetischer Mittelwert von Bernoulli Variablen Y_i ($i = 1, \dots, n$) auffassen lässt (siehe 5.3.1), d.h.

$$\hat{p} = \frac{Y_1 + Y_2 + \dots + Y_n}{n},$$

so dass hier der zentrale Grenzwertsatz greift (siehe 3.4). Er besagt im vorliegenden Fall, dass \hat{p} näherungsweise normalverteilt ist, obwohl die Einzelwerte Y_i keine Normalverteilung haben (sondern eine Bernoulli-Verteilung).

Beispiel: Von zehn Maispflanzen sind die erste, die dritte und die siebente mit dem

Maiszünsler befallen. Also gilt:

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
1	0	1	0	0	0	1	0	0	0

und

$$\hat{p} = \frac{1+0+1+0+0+0+1+0+0+0}{10} = 0,3$$

Die asymptotische Normalverteilung von \hat{p} kann zur Konstruktion eines asymptotischen Tests (gültig für großes n) verwendet werden. In kleinen Stichproben muss dagegen "exakt" getestet werden, also die zugrundeliegende Binomialverteilung berücksichtigt werden.

Asymptotischer Test

Test der Nullhypothese

$$H_0: p = p_0$$

Alternativhypothese: $p \neq p_0$

Berechne
$$z_{Vers} = \sqrt{n} \frac{|\hat{p} - p_0|}{\sqrt{p_0(1-p_0)}}$$

wobei

$$\hat{p} = \frac{x}{n}$$

n = Stichprobenumfang

x = Zahl Stichprobenelemente mit einer bestimmten Eigenschaft

Bestimme:

$$z_{Tab} = z_{1-\alpha/2}$$

wobei z_γ das γ -Quantil der Standardnormalverteilung ist (Tab. III) und α die vorgegebene Irrtumswahrscheinlichkeit (Fehler 1. Art).

$$z_{Vers} > z_{Tab} \Rightarrow \text{verwerfe } H_0$$

$$z_{Vers} \leq z_{Tab} \Rightarrow \text{behalte } H_0 \text{ bei}$$

Voraussetzung: $np_0(1-p_0) \geq 9$

Beispiel: In einem Kreuzungsexperiment mit Zuckerrüben soll geklärt werden, ob

eine Resistenz gegen Rizomania von einem einzigen Gen vererbt wird, dass einen dominant-rezessiven Erbgang hat. Falls die Vermutung über den Vererbungsmodus zutrifft, ist in einer F_2 -Generation mit einer Aufspaltung von 3 : 1 für anfällige versus resistente Pflanzen zu rechnen. Mit anderen Worten: Die Wahrscheinlichkeit p , dass eine F_2 -Pflanze resistent ist, beträgt $p = P(\text{resistent}) = 1/4$. Die Pflanzen werden im Gewächshaus einem hohen Infektionsdruck ausgesetzt, so dass anfällige Pflanzen eindeutig zu erkennen sind. Es werden 567 Pflanzen untersucht, von denen 143 resistente Pflanzen gefunden. Um zu prüfen, ob die Daten mit dem vermuteten Erbgang übereinstimmen, testen wir die Nullhypothese

$$H_0: p = 1/4$$

zum Niveau $\alpha = 5\%$

$$n = 567; x = 143; \hat{p} = \frac{143}{567} = 0,252; p_0 = 1/4 = 0,25$$

$$np_0(1-p_0) = 567 \cdot 0,25 \cdot 0,75 = 106 \geq 9 \Rightarrow \text{Test anwendbar}$$

$$z_{\text{Vers}} = \sqrt{n} \frac{|\hat{p} - p_0|}{\sqrt{p_0(1-p_0)}} = \sqrt{567} \frac{|0,252 - 0,25|}{\sqrt{0,25 \cdot 0,75}} = 0,1100 < z_{\text{Tab}} = z_{0,975} = 1,96 \Rightarrow H_0 \text{ beibehalten}$$

Die Nullhypothese eines Aufspaltungsverhältnisses von 3:1 kann nicht verworfen werden. Es gibt somit keinen Hinweis auf eine Abweichung vom vermuteten Vererbungsmodus.

Exakter Test

Falls die Bedingung $np_0(1-p_0) \geq 9$ nicht erfüllt ist, muss ein **exakter Test** durchgeführt werden, der nun besprochen wird. Zur Testkonstruktion betrachten wir die Verteilung der Zufallsvariable X .

Beispiel: In einem Kreuzungsexperiment mit Zuckerrüben werden nur $n = 20$ F_2 -Pflanzen auf Rizomania-Resistenz untersucht (und nicht $n = 567$, wie im vorangegangenen Beispiel). Es werden $X = 7$ resistente Pflanzen gefunden. Bei dem vermuteten Erbgang liegt in der F_2 -Generation eine Aufspaltung von 3 anfälligen : 1 resistenten Pflanzen vor. Es gilt also $H_0: p = p_0 = 1/4$, wobei p der erwartete Anteil resistenter Pflanzen ist. Zum Test der Nullhypothese, dass der vermutete Erbgang tatsächlich vorliegt, betrachten wir die Verteilung von X = Zahl resistenter Pflanzen. Es handelt sich um eine Binomialverteilung mit Parameter $p = 1/4$ und Konstante $n = 20$. Die Wahrscheinlichkeiten sind in Tab. 5.3.1 angegeben. Die Verteilung ist in Abb. 5.3.2 wiedergegeben.

Tab. 5.3.1: Binomialwahrscheinlichkeiten für $n = 20$ und $p = 1/4$.

k	$P(X = k/n; p)$		
0	0,00317	}	$= 0,02431$
1	0,02114		
2	0,06695	}	H_0 annehmen
3	0,13390		
4	0,18969		
5	0,20233		
6	0,16861		
7	0,11241		
8	0,06089		
9	0,02706		
10	0,00992	}	$= 0,01386$
11	0,00301		
12	0,00075		
13	0,00015		
14	0,00003		
15	0,00000		
16	0,00000		
17	0,00000		
18	0,00000		
19	0,00000		
20	0,00000		

$E(X)$ ←
unter H_0

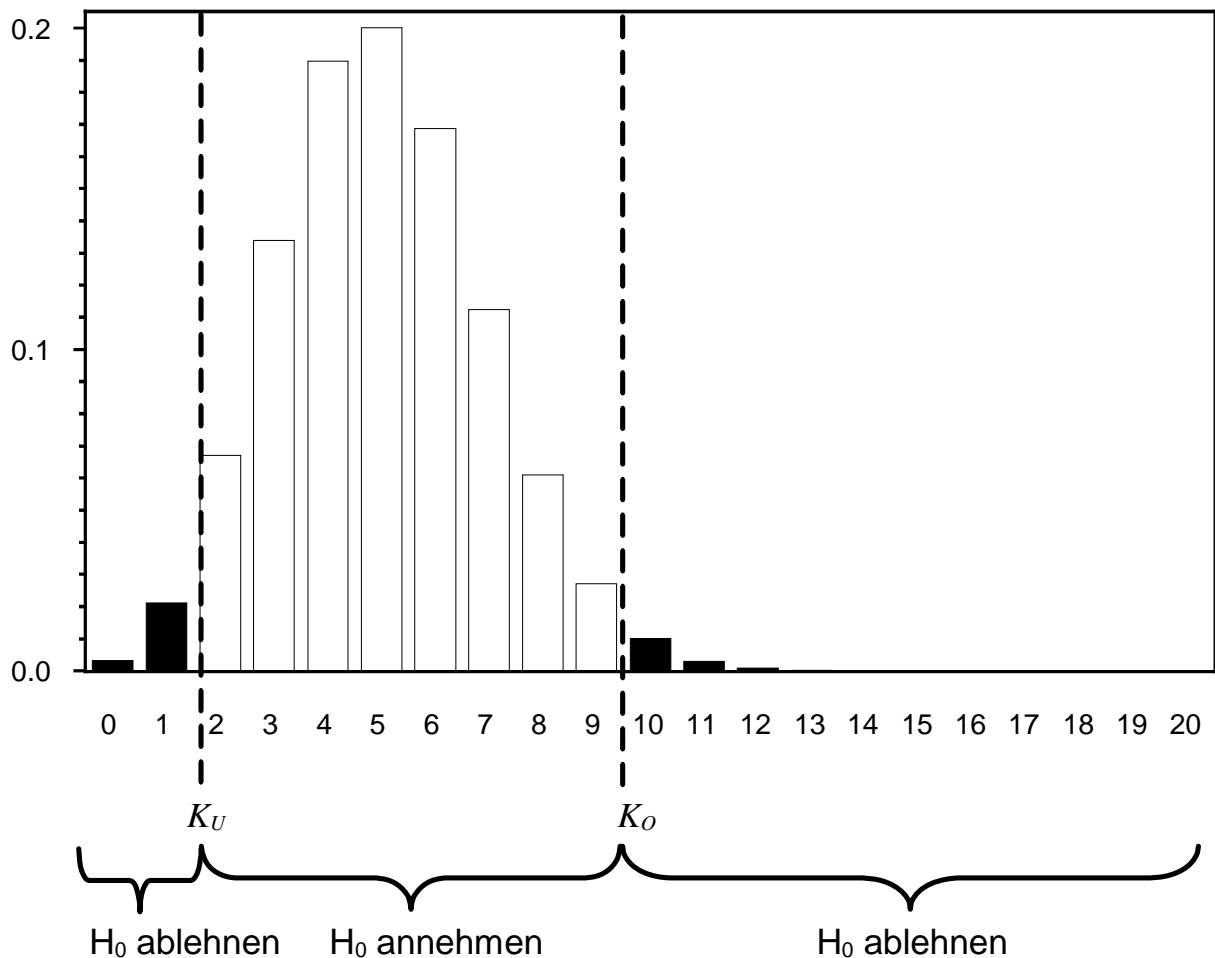


Abb. 5.3.2: Binomialwahrscheinlichkeiten für $n = 20$ und $p = 1/4$ mit kritischen Schwellen K_U und K_O .

Nun zur Konstruktion eines Ablehnungsbereiches für den Test basierend auf der Teststatistik X . Der Erwartungswert unter der Nullhypothese ist

$$E(X) = np = 5 \quad .$$

Wir konstruieren einen Test so, dass die Nullhypothese verworfen wird, wenn X „stark“ vom Erwartungswert abweicht. Hierbei werden die kritische obere und untere Schranke (K_U und K_O) so gelegt, dass die Überschreitungswahrscheinlichkeit eine vorgegebene Irrtumswahrscheinlichkeit (z.B. $\alpha = 5\%$) erreicht bzw. nicht überschreitet. Hierzu addieren wir „von links“ (beginnend bei $X = 0$; siehe Abb. 5.3.2) die Binomialwahrscheinlichkeiten, bis der Wert $\alpha/2$ überschritten wird, wobei α die angestrebte Irrtumswahrscheinlichkeit des Tests ist. Die Wahrscheinlichkeit, dass $X < 2$ wird, beträgt 0,02431 (Tab. 5.3.1). Analog verfahren wir „von rechts“ und finden, dass die Wahrscheinlichkeit, dass $X > 9$ wird, 0,01386 beträgt (Tab. 5.3.2). Betrachten wir nun folgende Testentscheidung:

Verwerfe H_0 , wenn $X < K_U = 2$ oder $X > K_O = 9$

Aus dem Additionssatz der Wahrscheinlichkeitsrechnung (Abschnitt 5.2) folgt, dass dieser Test eine Irrtumswahrscheinlichkeit von $\alpha = 0,02431 + 0,01386 = 0,03817$ hat. Die Nullhypothese wird also mit Wahrscheinlichkeit 3,817% fälschlicherweise verworfen. Das wir in diesem Fall keinen runden Wert für die Irrtumswahrscheinlichkeit erhalten können, liegt daran, dass es sich bei der Binomialverteilung um eine diskrete Verteilung handelt. Wir können in der Regel daher nur Irrtumswahrscheinlichkeiten erhalten, die kleiner als der jeweils angestrebte Wert sind.

[In manchen Fällen kann die Irrtumswahrscheinlichkeit noch etwas näher an den angestrebten Wert herangebracht werden, wenn man zulässt, dass die linke oder die rechte Überschreitungswahrscheinlichkeit $> \alpha/2$ wird, wobei die Summe der beiden Wahrscheinlichkeiten allerdings $< \alpha$ bleiben muss. Im vorliegenden Fall ist dies jedoch nicht möglich.]

Nach dem beschriebenen Test behalten wir bei dem beobachteten Wert von $x = 7$ resistenten Pflanzen die Nullhypothese bei, weil dieser Wert in den Annahmebereich fällt ($K_U < 7 < K_O$). Man beachte, dass

$$\hat{p} = x / n = 7 / 20 = 0,35$$

ist. Dieser Schätzwert liegt relativ weit vom Wert unter der Nullhypothese weg ($p_0 = 0,25$). Dass dies nicht zum Verwerfen der Nullhypothese reicht, liegt hier einfach am geringen Stichprobenumfang.

Man kann den exakten Test auch mit Hilfe eines **exakten p-Wertes** durchführen (siehe Abschnitt 3.15; der p-Wert ist nicht zu verwechseln mit dem Parameter p der Binomialverteilung). Hierzu addiert man alle Binomialwahrscheinlichkeiten, die kleiner oder gleich der für $x = 7$ sind. Dies sind gerade die Wahrscheinlichkeiten für diejenigen X -Werte, die so extrem wie der beobachtete Wert oder noch extremer vom Erwartungswert $E(X) = 5$ abweichen. Die Berechnung ist in Tab. 5.3.2 veranschaulicht. Wir finden $p\text{-Wert} = 0,09126 + 0,21422 = 0,30548$. Dieser Wert ist größer als 5%, also bei $\alpha = 5\%$ nicht signifikant.

Abschließende Bemerkung: Die hier beschriebenen exakten Berechnungen wird man in der Regel mit einem Computerprogramm durchführen, da sie relativ aufwändig sind. Dies gilt nicht nur für den hier vorgestellten exakten Test sondern generell für das Gros der exakten Verfahren für kategoriale Daten.

Tab. 5.3.2: Berechnung des exakten p-Wertes für $X = 7$ aus den Binomialwahrscheinlichkeiten für $n = 20$ und $p = 1/4$.

k	$P(X = k/n; p)$	
0	0,00317	} = 0,09126
1	0,02114	
2	0,06695	
3	0,13390	
4	0,18969	
5	0,20233	
6	0,16861	
7	0,11241	} = 0,21422
8	0,06089	
9	0,02706	
10	0,00992	
11	0,00301	
12	0,00075	
13	0,00015	
14	0,00003	
15	0,00000	
16	0,00000	
17	0,00000	
18	0,00000	
19	0,00000	
20	0,00000	

$E(X)$ ←
unter H_0

p-Wert = 0,30548

5.3.4 Vertrauensintervall für den Parameter p der Binomialverteilung

Asymptotisches Vertrauensintervall

Der Standardfehler ist ein gängiges Maß für die Genauigkeit einer Schätzung. Je kleiner der Standardfehler umso genauer die Schätzung. Aussagekräftiger als der Standardfehler ist ein Vertrauensintervall. Im vorliegenden Fall kann ausgenutzt werden, dass der Schätzer \hat{p} für großes n näherungsweise normalverteilt ist. Daher kann ein asymptotisches Vertrauensintervall (gültig für großes n) wie folgt berechnet werden:

Asymptotisches $100(1-\alpha)\%$ **Vertrauensintervall** für den Parameter p einer Binomialverteilung mit Stichprobenumfang n

Voraussetzung: $n\hat{p}(1-\hat{p}) \geq 9$

Andernfalls muss der Stichprobenumfang erhöht werden, oder es müssen exakte Methoden bzw. bessere Approximationen angewendet werden (siehe unten).

Das asymptotische Intervall ist gegeben durch

$$[\hat{p} - z_{1-\alpha/2} \times s_{\hat{p}}; \hat{p} + z_{1-\alpha/2} \times s_{\hat{p}}]$$

wobei z_γ das γ -Quantil der Standardnormalverteilung ist (Tab. III), α die vorgegebene Irrtumswahrscheinlichkeit (Fehler 1. Art), und $s_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

Bemerkung: Im Gegensatz zum Test wird hier die Entscheidungsregel $n\hat{p}(1-\hat{p}) \geq 9$ mit Hilfe des Schätzwertes von p berechnet. Bei Test kann dagegen statt der Schätzung der Wert p_0 unter der Nullhypothese H_0 verwendet werden, weil für die Gültigkeit des Tests die Verteilung unter H_0 relevant ist.

Beispiel: Für das Maisfeld hatten wir $n = 100$; $x = 7$; $\hat{p} = 0,07$; $s_{\hat{p}} = 0,0255$ berechnet. Nun soll ein 95% Vertrauensintervall für p berechnet werden. Wir stellen fest dass

$$n\hat{p}(1-\hat{p}) = 100 \cdot 0,07 \cdot 0,93 = 6,51 < 9$$

ist. Also ist die Methode nicht anwendbar.

Beispiel: Aus demselben Maisfeld wird eine Stichprobe vom Umfang $n = 250$ gezogen. Es sind $x = 20$ Pflanzen befallen.

$$\hat{p} = 20/250 = 0,08$$

$$n\hat{p}(1-\hat{p}) = 250 \cdot 0,08 \cdot 0,92 = 18,4 > 9$$

Das Verfahren zur Berechnung des Vertrauensintervalls ist anwendbar.

$$s_{\hat{p}}(\hat{p}) = \sqrt{0,08 \cdot 0,92 / 250} = 0,0172$$

$$\alpha = 0,05$$

$$z_{1-\alpha/2} = z_{0,975} = 1,96$$

$$[\hat{p} - z_{1-\alpha/2} s_{\hat{p}}; \hat{p} + z_{1-\alpha/2} s_{\hat{p}}]$$

$$=[0,08 - 1,96 \cdot 0,0172; 0,08 + 1,96 \cdot 0,0172] = [0,08 - 0,034; 0,08 + 0,034]$$

$$=[0,046; 0,112]$$

Das Intervall $[0,046; 0,112]$ überdeckt mit 95%iger Wahrscheinlichkeit den wahren Parameter p . Zwischen 4,6 und 11,2% der Maispflanzen sind mit dem Maiszünsler befallen (Vertrauenswahrscheinlichkeit 95%).

Exaktes Vertrauensintervall

Bei kleinen Stichproben ergeben sich "exakte" Vertrauensintervalle nach Clopper und Pearson durch Auflösen der folgenden Gleichungen nach p_u und p_o :

$$P(X \geq x | p_u) = \sum_{k=x}^n \binom{n}{k} p_u^k (1-p_u)^{n-k} = \frac{\alpha}{2}$$

$$P(X \leq x | p_o) = \sum_{k=0}^x \binom{n}{k} p_o^k (1-p_o)^{n-k} = \frac{\alpha}{2}$$

wobei α die vorgegebene Irrtumswahrscheinlichkeit ist. Dann gilt $P(p_u \leq p \leq p_o) \leq 1 - \alpha$, d.h. p_u und p_o sind untere und obere Vertrauensgrenzen für die Binomialwahrscheinlichkeit, so dass die Überdeckungswahrscheinlichkeit des Intervalls **mindestens** $(1-\alpha)$ beträgt. Das "mindestens" bedeutet, dass Clopper-Pearson Intervalle konservativ sind, was an der Diskretheit der Binomialverteilung liegt (die Zufallsvariable X kann nur eine begrenzte Zahl von Werten annehmen; die Variation der Zufallsvariable X ist daher also nicht kontinuierlich, sondern diskret). In anderen Worten bedeutet "konservativ", dass die Intervalle tendenziell etwas "breiter als nötig" sind.

Erklärung: Am einfachsten lässt sich das Verfahren mit der Äquivalenz von Vertrauensintervall und Hypothesentest erklären. Die obigen beiden Gleichungen definieren gerade den kleinsten und den größten Wert für p_0 , für den die Nullhypothese $H_0: p = p_0$ noch angenommen würde (siehe Abschnitt 5.3.3). Alle Werte von p_0 , für die diese Nullhypothese angenommen würde, definieren ein Vertrauensintervall für p .

Auflösen der obigen Gleichungen nach p_u und p_o führt einigen algebraischen Umformungen zu folgendem Vertrauensintervall:

Exaktes $(1-\alpha)$ -Vertrauensintervall für die Binomialwahrscheinlichkeit p :

n = Stichprobenumfang

x = Zahl Stichprobenelemente mit einer bestimmten Eigenschaft

p = Wahrscheinlichkeit, dass ein einzelnes Element die Eigenschaft hat

α = Irrtumswahrscheinlichkeit

$$p_u = \frac{x}{x + (n - x + 1)F_u}$$

mit

$$F_u = F(1 - \alpha/2; v_1, v_2) \quad (\text{Quantile der F-Verteilung; Tab. V})$$

und

$$v_1 = 2(n - x + 1), v_2 = 2x \quad (\text{Freiheitsgrade})$$

sowie

$$p_o = \frac{(x+1)F_o}{n-x+(x+1)F_o}$$

mit

$$F_o = F(1-\alpha/2; w_1, w_2) \quad (\text{Quantile der } F\text{-Verteilung})$$

und

$$w_1 = 2(x+1), w_2 = 2(n-x) \quad (\text{Freiheitsgrade})$$

In den Grenzfällen $x = 0$ und $x = n$ ergeben sich folgende Lösungen:

$$x = 0: \quad p_u = 0; \quad p_o = 1 - \exp\left[\frac{1}{n} \ln\left(\frac{\alpha}{2}\right)\right]$$

$$x = n: \quad p_u = \exp\left[\frac{1}{n} \ln\left(\frac{\alpha}{2}\right)\right]; \quad p_o = 1$$

Beispiel: Von $n = 10$ zufällig aus einer Saatgutpartie ausgewählten Maiskörnern sind $x = 9$ keimfähig. Berechne ein 95%iges Vertrauensintervall für die Keimfähigkeit (p) der Saatgutpartie.

$$\text{Untere Grenze: } v_1 = 4; v_2 = 18; F(0,975; 4, 18) = 3,61; \quad p_u = \frac{9}{9+(10-9+1) \cdot 3,61} = 0,555$$

$$\text{Obere Grenze: } w_1 = 20; w_2 = 2; F(0,975; 20, 2) = 39,45; \quad p_o = \frac{(9+1) \cdot 39,45}{10-9+(9+1) \cdot 39,45} = 0,997$$

Das 95% Vertrauensintervall hat die Grenzen 0,555 und 0,997. Zum Vergleich nun das asymptotische Vertrauensintervall:

$$\hat{p} = x/n = 0,9; \quad 1,96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,186$$

$$\text{Untere Grenze: } \hat{p} - 1,96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,9 - 0,186 = 0,714$$

$$\text{Obere Grenze: } \hat{p} + 1,96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,9 + 0,186 = 1,086$$

Die obere Grenze (1,086) übersteigt den zulässigen Bereich ($0 \leq p \leq 1$), was ein deutlicher Hinweis ist, dass das asymptotische Intervall die Irrtumswahrscheinlichkeit nicht einhält. Aus praktischen Gründen wird man hier die obere Grenze gleich Eins setzen. Das exakte Vertrauensintervall ist beträchtlich breiter (0,56; 0,98) als das asymptotische (0,71; 1,00).

SAS Anweisungen

```
data;  
input keimfaehig$ zahl;  
datalines;  
nein 1  
ja 9  
;  
proc freq;  
weight zahl;  
tables keimfaehig/binomial;  
run;
```

Ergebnis:

keimfaehig	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ja	9	90.00	9	90.00
nein	1	10.00	10	100.00

Binomial Proportion
for keimfaehig = ja

Proportion	0.9000
ASE	0.0949
95% Lower Conf Limit	0.7141
95% Upper Conf Limit	1.0000
Exact Conf Limits	
95% Lower Conf Limit	0.5550
95% Upper Conf Limit	0.9975

5.3.5 Vergleich von zwei Binomialwahrscheinlichkeiten - unverbundene Stichproben

Für zwei unabhängige Stichproben soll die Binomialwahrscheinlichkeit eines Merkmales verglichen werden.

Asymptotischer Test zum Vergleich von 2 Binomialwahrscheinlichkeiten

Voraussetzung: $n_1 p_1 (1 - p_1) > 9$, $n_2 p_2 (1 - p_2) > 9$

p_1 = Binomialwahrscheinlichkeit der Gruppe 1

p_2 = Binomialwahrscheinlichkeit der Gruppe 2

H_0 : $p_1 = p_2 = p$

Alternativhypothese: $p_1 \neq p_2$

n_1, n_2 = Stichprobenumfang in Gruppe 1 und 2

x_1, x_2 = Zahl Stichprobenelemente mit einer bestimmten Eigenschaft in Gruppe 1 und 2

Berechne:

$$\hat{p}_1 = \frac{x_1}{n_1}; \hat{p}_2 = \frac{x_2}{n_2}; \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$z_{Vers} = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$z_{Tab} = z_{1-\alpha/2}$$

wobei z_γ das γ -Quantil der Standardnormalverteilung ist (Tab. III) und α die vorgegebene Irrtumswahrscheinlichkeit (Fehler 1. Art).

$z_{Vers} > z_{Tab} \Rightarrow$ verwirfe H_0

$z_{Vers} \leq z_{Tab} \Rightarrow$ behalte H_0 bei

Beispiel: Zwei Saatgutpartien sollen bezüglich ihrer Keimfähigkeit verglichen werden. Es soll die Nullhypothese geprüft werden, dass beide Partien die gleiche Keimfähigkeit haben. Von jeder Partie wurden 316 Körner untersucht. Bei der ersten Partie keimten 289 Körner, bei der zweiten 274. Der Test soll zum Niveau $\alpha = 5\%$ durchgeführt werden.

$$n_1 = n_2 = 316; x_1 = 289; x_2 = 274$$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{289}{316} = 0,9146; \hat{p}_2 = \frac{x_2}{n_2} = \frac{274}{316} = 0,8671; \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{289 + 274}{316 + 316} = 0,8908$$

$$z_{Vers} = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{|0,9146 - 0,8671|}{\sqrt{0,8908 * 0,1092 \left(\frac{1}{316} + \frac{1}{316}\right)}} = 1,9188$$

$$z_{Tab} = z_{1-\alpha/2} = 1,96; \quad z_{Vers} < z_{Tab} \Rightarrow \text{behalte } H_0 \text{ bei}$$

Die Keimfähigkeiten der beiden Partien unterscheiden sich nicht signifikant voneinander.

Bemerkung: Wenn die Stichprobenumfänge zu klein sind, sollte man alternativ Fishers exakten Test verwenden (siehe Abschnitt 5.6.3).

Wahl des Stichprobenumfanges für den Test zum Vergleich zweier Binomialverteilungen bei unverbundenen Stichproben:

$$n = \frac{[p_1(1-p_1) + p_2(1-p_2)](z_{1-\alpha/2} + z_{1-\beta})^2}{(p_1 - p_2)^2}$$

Diese Formel ist weitgehend äquivalent zur entsprechenden Formel für den t-Test bei unverbundenen Stichproben (Abschnitt 3.14). Anstelle der Vorinformation über Varianz und kleinste nachzuweisende Differenz tritt hier die Voreinschätzung über die mutmaßlichen Werte der beiden Binomialparameter p_1 und p_2 , aus welchen sich dann Varianz und nachzuweisende Differenz ergeben. Auch für andere in diesem Skript behandelte Tests gibt es Formeln für die Stichprobenplanung, die hier aus Platzgründen aber nicht alle besprochen werden können (Siehe z.B. Rasch, D. et al. 1996 Verfahrensbibliothek. Oldenbourg-Verlag, München).

5.3.6 Vergleich von zwei Binomialwahrscheinlichkeiten - verbundene Stichproben

Beispiel: Es sollen zwei verschiedene Labormethoden A und B zum Nachweis eines Virus geprüft werden. Hierzu wurden insgesamt 243 Blätter von Pflanzen, die eindeutige Befallssymptome zeigen, analysiert. Da deutliche Unterschiede zwischen den Blättern im Virusgehalt erwartet werden, ist eine verbundene Stichprobe sinnvoll. Hierdurch wird die Genauigkeit des Vergleiches erhöht (siehe Abschnitt 3.12). Jedes der Blätter wird in zwei Hälften geteilt. Die eine Hälfte wird mit der Methode A analysiert, die andere mit Methode B.

Nachweis mit Methode		Häufigkeit
A	B	
ja	ja	156
ja	nein	32
nein	ja	18
nein	nein	37

Hier liegt eine verbundene (gepaarte) Stichprobe vor, weil jeweils auf derselben Untersuchungseinheit (Blatt) beide Methoden untersucht werden. Offenbar wird mit keiner der beiden Methoden das Virus sicher nachgewiesen. Es interessiert nun die Frage, ob sich die Methoden in der Wahrscheinlichkeit unterscheiden, mit der das Virus in einer zufälligen Probe nachgewiesen wird. Bezeichnen wir die Nachweiswahrscheinlichkeiten für die Methoden A und B mit p_1 und p_2 , so ist die zu testende Nullhypothese $H_0: p_1 = p_2$.

Beispiel: Zwei verschiedene Nährmedien A und B sollen auf Ihre Eignung zur Anzucht eines Bakteriums überprüft werden, welches im Darm von Kühen vorkommt. Es wird vermutet, dass die Eignung des Mediums auch davon anhängt, aus welcher Umwelt das Bakterium stammt, also von welcher Kuh. Daher wird von insgesamt 266 Kühen eine Darmprobe punktiert und das Bakterium isoliert. Die Probe jeder Kuh wird geteilt, wobei die eine Hälfte auf eine Petrischale mit Nährmedium A und die andere Hälfte auf eine Petrischale mit Nährmedium B ausgebracht wird. Es werden insgesamt $2 \cdot 266 = 532$ Petrischalen benötigt. Auf jeder Petrischale wird nach einer Woche festgestellt, ob das Bakterium überlebt hat oder nicht. Das Ergebnis war wie

folgt:

Wachstum auf Nährmedium		Häufigkeit
A	B	
ja	ja	180
ja	nein	3
nein	ja	73
nein	nein	10

Auch hier liegt eine verbundene (gepaarte) Stichprobe vor, weil jeweils von der selben Kuh eine Bakterienprobe auf Medium A und Medium B ausgebracht wird. Es interessiert die Frage, ob sich die Medien in der Überlebenswahrscheinlichkeit der Bakterien unterscheiden. Bezeichnen wir die Überlebenswahrscheinlichkeiten für die Medien A und B mit p_1 und p_2 , so ist die zu testende Nullhypothese $H_0: p_1 = p_2$.

McNemar-Test (approximativer Test): Es liegen je Versuchseinheit gepaarte Beobachtungen eines dichotomen Merkmals bei zwei Behandlungen A und B vor.

A_1, A_2 = Ausprägungen des Merkmals bei Behandlung A

B_1, B_2 = Ausprägungen des Merkmals bei Behandlung B

a = Häufigkeit der Kombination A_1, B_1

b = Häufigkeit der Kombination A_1, B_2

c = Häufigkeit der Kombination A_2, B_1

d = Häufigkeit der Kombination A_2, B_2

p_1 = Wahrscheinlichkeit des Auftretens von A_1 bei Behandlung A

p_2 = Wahrscheinlichkeit des Auftretens von B_1 bei Behandlung B

		Ausprägung für Behandlung B	
		B_1	B_2
Ausprägung für Behandlung A	A_1	a	b
	A_2	c	d

$H_0: p_1 = p_2$

Voraussetzung: $b, c > 5$ (sonst exakt testen; hier nicht näher besprochen)

Berechne:

$$\hat{p}_1 = \frac{a+b}{a+b+c+d}; \quad \hat{p}_2 = \frac{a+c}{a+b+c+d}$$

$$z_{Vers} = \frac{|b-c|}{\sqrt{b+c}}$$

$$z_{Tab} = z_{1-\alpha/2}$$

wobei z_γ das γ -Quantil der Standardnormalverteilung ist (Tab. III) und α die vorgegebene Irrtumswahrscheinlichkeit (Fehler 1. Art).

$$z_{Vers} > z_{Tab} \Rightarrow \text{verwerfe } H_0$$

$$z_{Vers} \leq z_{Tab} \Rightarrow \text{behalte } H_0 \text{ bei}$$

Beispiel: Wir verrechnen das erste Beispiel zum Vergleich der Labormethoden zum Nachweis eines Virus. Die Nachweiswahrscheinlichkeit der beiden Methoden soll zum Niveau $\alpha = 5\%$ mit dem McNemar-Test verglichen werden.

Nachweis mit Methode B

		+	-
Nachweis mit Methode A	+	$a = 156$	$b = 32$
	-	$c = 18$	$d = 37$

$$\hat{p}_1 = \frac{a+b}{a+b+c+d} = \frac{156+32}{156+32+18+37} = 0,774$$

$$\hat{p}_2 = \frac{a+c}{a+b+c+d} = \frac{156+18}{156+32+18+37} = 0,716$$

$$z_{Vers} = \frac{|b-c|}{\sqrt{b+c}} = \frac{|32-18|}{\sqrt{32+18}} = 1,98$$

$$z_{Tab} = 1,96$$

$$z_{Vers} > z_{Tab} \Rightarrow \text{verwerfe } H_0$$

Die erste Methode (A) hat mit 77,4% eine signifikant höhere Nachweiswahrscheinlichkeit als die zweite Methode (B) mit 71,6%.

*5.3.7 Grenzwerte für GVO in Saatgut – Produzenten- und Konsumentenrisiko

Operationscharakteristik und exakter Test

Im folgenden Beispiel wird ein exaktes Verfahren für den Parameter einer Binomialverteilung betrachtet. Das Beispiel führt die Begriffe des **Produzentenrisikos** und des **Konsumentenrisikos** als alternative Bezeichnungen für α - und β -Fehler ein. Der Bezug des beschriebenen Verfahrens zu einem statistischen Tests wird erst am Ende des Beispiels erläutert.

Beispiel: Auf EU-Ebene werden Grenzwerte für die unvermeidbare Verunreinigung von Saatgut mit GVO (gentechnisch veränderte Organismen) diskutiert (Stand Oktober 2003). Hierbei ist beispielsweise im Gespräch, dass bei Sojabohne und Erbse der Anteil GVO Samen im Saatgut konventioneller Sorten einen Anteil von 0,7% nicht überschreiten darf. Für andere Arten sind Grenzwerte von 0,5% und 0,3% im Gespräch. Für Nahrungsmittel ist der Grenzwert 0,9%. Bei Überschreiten ist das Saatgut bzw. das Nahrungsmittel als GVO-haltig zu kennzeichnen.

Eine wichtige Frage ist, wie die Einhaltung solcher Grenzwerte in der Praxis überprüft werden soll. Ein einfacher Vorschlag besteht darin, eine Stichprobe von Samen molekulargenetisch auf das Vorhandensein von GVO zu untersuchen, wobei dann der Anteil der GVO Samen in der Stichprobe ermittelt wird. Überschreitet dieser den Grenzwert, so muss die Partie gekennzeichnet werden.

Es stellt sich die Frage, ob das beschriebene Entscheidungsverfahren eine gute Strategie ist und mit welchen Risiken es behaftet ist. Zu berücksichtigen ist hierbei, dass von einer Stichprobe auf die Grundgesamtheit einer ganzen Partie zurückgeschlossen werden muss, und dieser Schluss beinhaltet eine gewisse Wahrscheinlichkeit für Fehlentscheidungen. Wir gehen hier zunächst vereinfachend davon aus, dass Partien mit einem Anteil GVO Samen unterhalb des Grenzwertes akzeptabel sind, während solche mit Anteilen über dem Grenzwert nicht akzeptabel sind. Später wird diese Definition modifiziert. Die möglichen Fehler sind:

- (1) Die Partie muss gekennzeichnet werden, obwohl der Grenzwert in der Partie unterschritten wird. Dieser Fehler ist unerwünscht aus Sicht des **Produzenten**.
- (2) Die Partie muss nicht gekennzeichnet werden, obwohl der Grenzwert in der Partie überschritten wird. Dieser Fehler ist unerwünscht aus Sicht des **Konsumenten** (Verbrauchers).

Wir können näherungsweise annehmen, dass die Partie unendlich viele Körner hat und dass eine echte Zufallsstichprobe aus dieser Grundgesamtheit gezogen wird. Wenn p den Anteil GVO Samen in der Partie bezeichnet, n den Stichprobenumfang und x die Anzahl GVO Samen in der Stichprobe, so ist $\hat{p} = x/n$ der entsprechende Anteil in der Stichprobe. Mit anderen Worten, der Anteil p in der Partie ist ein zu schätzender Parameter der Grundgesamtheit, und $\hat{p} = x/n$ ist der entsprechende Stichprobenschätzer. Die beiden Fehler treten nun in den folgenden Fällen ein, wobei G der Grenzwert ist:

- (1) Die Partie wird gekennzeichnet, obwohl Grenzwert in der Partie unterschritten wird:

$\hat{p} > G$ (Partie wird gekennzeichnet), obwohl $p \leq G$ (Partie ist nicht zu kennzeichnen)

(2) Partie wird nicht gekennzeichnet, obwohl Grenzwert in der Partie überschritten wird:

$\hat{p} \leq G$ (Partie wird nicht gekennzeichnet), obwohl $p > G$ (Partie ist zu kennzeichnen)

Interessant ist nun die Frage, wie groß die Fehlerwahrscheinlichkeiten sind, wenn die folgende einfache Entscheidungsregel angewendet wird:

"Kennzeichne Partie, wenn $\hat{p} > G$. Andernfalls nicht.

Diese Wahrscheinlichkeiten lassen sich für verschiedene Werte für den Anteil p in der Partie berechnen unter Nutzung der Tatsache, dass die Zahl GVO Samen (x) einer Binomialverteilung mit Parameter p und Konstante n folgt. Wir finden

$$p(\hat{p} > G) = p(x > G \cdot n) = \sum_{k > gn} \binom{n}{k} p^k (1-p)^{n-k}$$

Angenommen, der Grenzwert ist $G = 0,7\%$. Wir wollen berechnen, mit welcher Wahrscheinlichkeit eine Partie gekennzeichnet wird, deren Anteil genau $p = 0,007$ ist, die also gerade noch akzeptabel ist. Der Stichprobenumfang sei $n = 1000$ Samenkörner. Wir finden $nG = 7$ und somit

$$P(\hat{p} > G) = P(x > 7) = 1 - P(x \leq 7) = 1 - \sum_{k=0}^7 \binom{n}{k} p^k (1-p)^{n-k} = 1 - 0,5987 = 0,4013$$

Diese Wahrscheinlichkeit ist auch in Abb. 5.3.3 veranschaulicht.

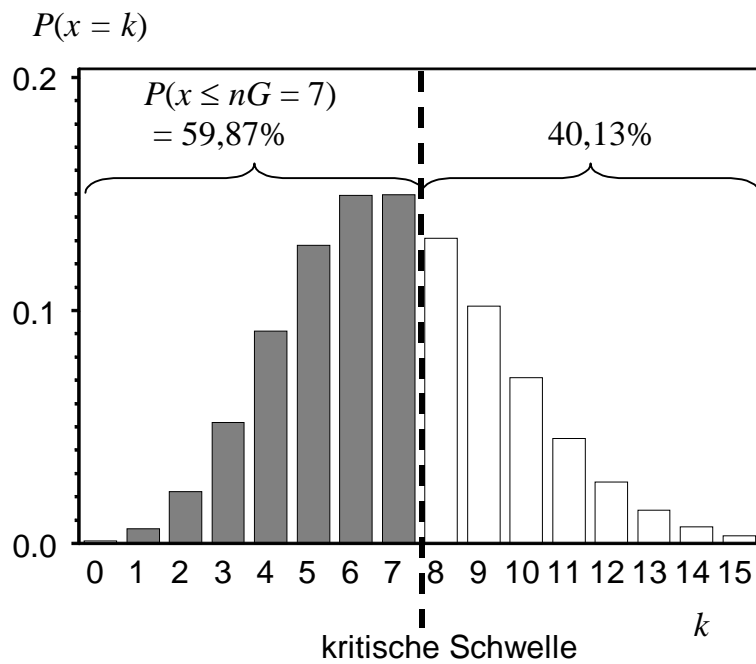


Abb. 5.3.3: Binomialverteilung mit $p = G = 0,007 = 0,7\%$ und $n = 1000$.

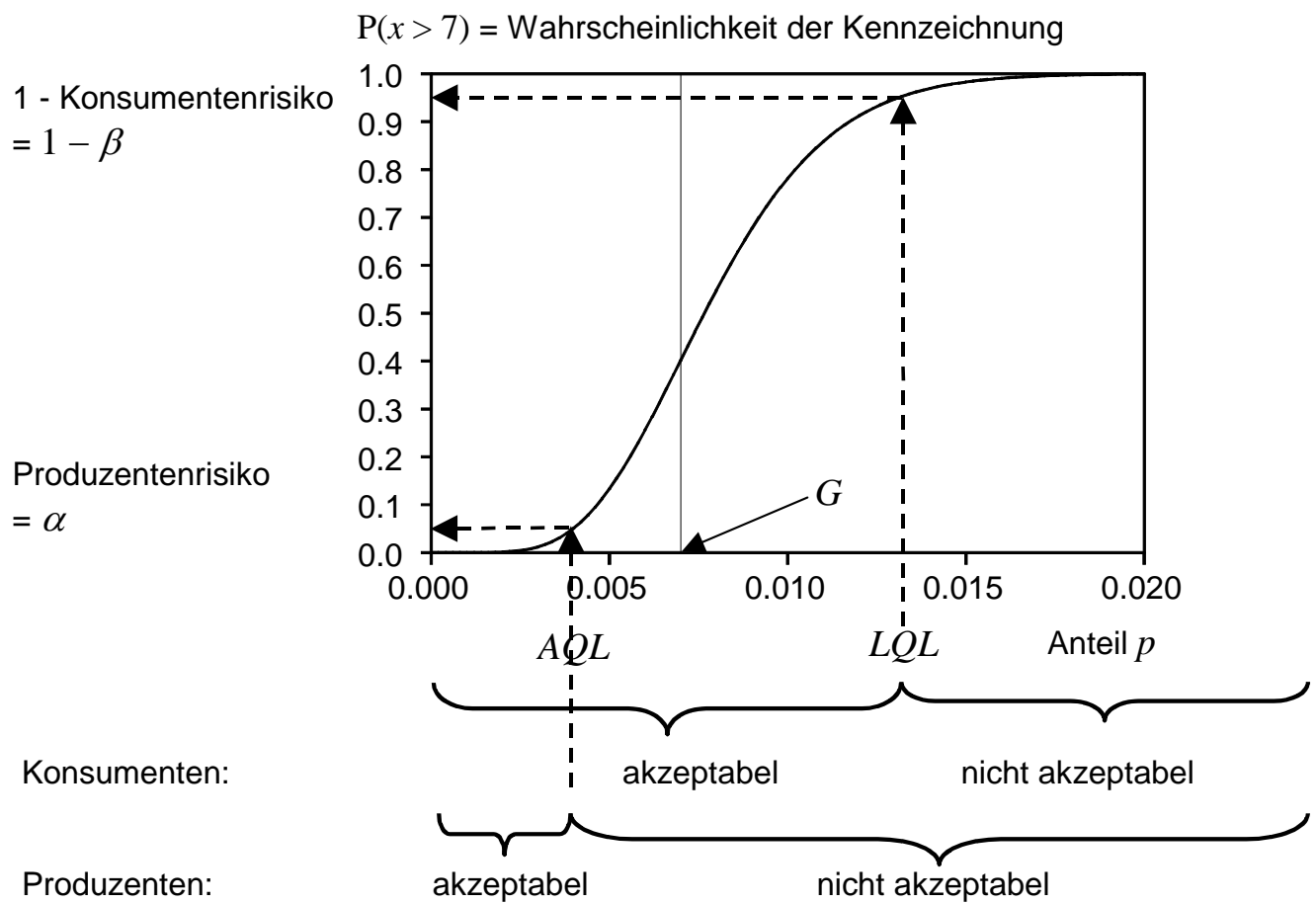


Abb. 5.3.4: Operations-Charakteristik für $n = 1000$. Vertikale Linie: $G = 0,007$.

Mit einer Wahrscheinlichkeit von 40,13% wird also eine Partie, die eigentlich gerade noch akzeptabel ist ($p = 0,007$), als inakzeptabel eingestuft. Diese Wahrscheinlichkeit steigt sogar noch und strebt gegen 50%, wenn n weiter wächst. Je weiter allerdings der tatsächliche Anteil p unter den Grenzwert G fällt, desto kleiner wird die Wahrscheinlichkeit für eine falsche Entscheidung. Der Abfall der Wahrscheinlichkeit ist umso stärker, je größer der Stichprobenumfang n ist.

Man sieht an dieser Betrachtung, dass für Anteile p sehr nahe am Grenzwert die Fehlerwahrscheinlichkeiten beträchtlich sind. Die Fehlerwahrscheinlichkeiten werden erst dann akzeptabel, wenn die wahren Anteile p deutlich vom Grenzwert abweichen. Dies lässt sich am besten anhand der sog. **Operations-Charakteristik** veranschaulichen (Abb. 5.3.4).

Für einen gegebenen Stichprobenumfang kann man die Wahrscheinlichkeit einer Überschreitung des Grenzwertes in der Stichprobe gegen den tatsächlichen Anteil in der Partie (p) abtragen (Abb. 5.3.3). Diese Darstellung wird als **Operations-Charakteristik** bezeichnet. Je größer der Stichprobenumfang ist, desto steiler verläuft die Kurve. Je deutlicher der wahre Anteil p unter dem Grenzwert liegt, umso geringer ist die Wahrscheinlichkeit, dass die betreffende Partie fälschlicherweise gekennzeichnet wird (**Produzentenrisiko**). Ebenso trifft es zu, dass die Wahrscheinlichkeit, eine inakzeptable Partie zu akzeptieren (**Konsumentenrisiko**), um so kleiner wird, je deutlicher der Anteil p über dem Grenzwert liegt.

Hieraus wird deutlich, dass über Partien mit Anteilen von GVO Samen nahe dem Grenzwert keine sehr genaue Entscheidung möglich ist, und zwar weder für Konsumenten noch für Produzenten. Verlässlichere Aussagen sind dagegen möglich für Partien, die deutlich genug über bzw. unter dem Grenzwert liegen. Die EU diskutiert aus diesem Grunde Werte unterhalb und oberhalb des Grenzwertes, die wie folgt bezeichnet werden:

LQL = Low Quality Level (liegt oberhalb des Grenzwertes):

Das **Konsumentenrisiko** (Wahrscheinlichkeit der Nichtkennzeichnung) soll einen vorgegebenen Wert (β) nicht überschreiten, wenn der wahre Anteil p in der Partie größer oder gleich dem *LQL* ist.

AQL = Acceptable Quality Level (liegt unterhalb des Grenzwertes):

Das **Produzentenrisiko** (Wahrscheinlichkeit der Kennzeichnung) soll einen vorgegebenen Wert (α) nicht überschreiten, wenn der wahre Anteil p in der Partie kleiner oder gleich dem *AQL* ist.

In der Praxis sucht man für vorgegebene Werte von *AQL* und *LQL* denjenigen Stichprobenumfang n , für welchen die vorgegebenen Irrtumswahrscheinlichkeiten α und β eingehalten werden.

Hier wollen wir zur Veranschaulichung umgekehrt vorgehen und zunächst bei einem vorgegebenen Stichprobenumfang von $n = 1000$ bleiben. Die Wahl von n wird am Ende dieses Beispiels besprochen. Wir suchen also zunächst die Werte für *AQL* und *LQL*, die sich für gegebenes n ergeben. Es wird also derjenige Wert für *LQL* gesucht, für den

$$P(\hat{p} > G) = \sum_{k > nG} \binom{n}{k} LQL^k (1 - LQL)^{n-k} = 1 - \beta$$

ist, wobei β das Konsumentenrisiko ist. Ebenso sucht man einen Wert für AQL , so dass

$$P(\hat{p} > G) = \sum_{k > nG} \binom{n}{k} AQL^k (1 - AQL)^{n-k} = \alpha$$

ist, wobei α das Produzentenrisiko ist. Ähnlich wie beim Pearson-Clopper-Vertrauensintervall (vgl. Abschnitt 5.3.2) können die Lösungen nach folgenden Formeln gefunden werden:

$$LQL = \frac{x}{x + (n - x + 1)F_{LQL}}$$

mit

$$F_{LQL} = F(\beta; v_1, v_2) = 1 / F(1 - \beta; v_2, v_1)$$

$$v_1 = 2(n - x + 1), v_2 = 2x$$

x = kleinste ganze Zahl so dass $x > nG$

Ebenso berechnet man:

$$AQL = \frac{x}{x + (n - x + 1)F_{AQL}}$$

mit $F_{AQL} = F(1 - \alpha; v_1, v_2)$

Für $n = 1000$, $G = 0,007$ und $\alpha = \beta = 0,05$ findet man $LQL = 0,0131$ und $AQL = 0,0040$. Hiermit lassen sich folgende Aussagen bezüglich eines Entscheidungsverfahrens mit Stichprobenumfang $n = 1000$ und Grenzwert $G = 0,007$ treffen:

Das Konsumentenrisiko, dass eine Partie mit einem Anteil GVO Samen von $LQL = 0,0131 = 1,31\%$ oder mehr in den Handel gelangt, beträgt nicht mehr als $\beta = 0,05 = 5\%$.

Das Produzentenrisiko, dass eine Partie mit einem Anteil kontaminierter Samen von $AQL = 0,0040 = 0,40\%$ oder weniger nicht in den Handel gelangt, beträgt nicht mehr als $\alpha = 0,05 = 5\%$.

Anstelle des Grenzwertes von $G = 0,7\%$ ist für den Verbraucher somit der Wert $LQL = 1,31\%$ relevant. Für den Produzenten ist dagegen bei einem Grenzwert von $G = 0,7\%$ der Wert $AQL = 0,40\%$ relevant. Man sieht an dieser Betrachtung, dass die Aussagefähigkeit des Grenzwertes selbst eingeschränkt ist. Größeren Informationsgehalt bieten LQL und AQL .

Umgekehrt kann man folgendes sagen: Damit bei einem Grenzwert von $G = 0,7\%$ und einem Stichprobenumfang von $n = 1000$ sowohl Produzenten als auch Konsumenten zufriedengestellt werden, muss für Konsumenten ein Anteil von bis zu $1,31\%$ kontaminierter Samen noch akzeptabel sein, während es für Produzenten noch akzeptabel sein muss, dass eine Partie verworfen wird, die einen Anteil von $0,40\%$ oder mehr kontaminierter Samen aufweist.

Erhöht man den Stichprobenumfang n bei gleichbleibendem Grenzwert (G), so sinkt das LQL und steigt das AQL . Beide nähern sich dem Grenzwert G an. Der Grenzwert ist also diejenige Schwelle, über die hinaus sich weder das LQL für den Verbraucher noch das AQL für den Produzenten durch Erhöhung des Stichprobenumfanges verbessern lassen. Dies muss bei der Festlegung des Grenzwertes berücksichtigt werden.

Das beschriebene Entscheidungsverfahren hat einen engen Bezug zu einem Signifikanztest mit einer einseitigen Fragestellung. Hierbei ist G die kritische Schwelle für die Teststatistik x bzw. $\hat{p} = x/n$. Bisher haben wir solche Tests kennen gelernt, bei denen ausgehend von einer Nullhypothese die kritische Schwelle festgelegt wird, die von der Teststatistik überschritten werden muss, damit die Nullhypothese unter Einhaltung einer Irrtumswahrscheinlichkeit α bzw. β verworfen wird. Hier ist es umgekehrt: Die kritische Schwelle, der Grenzwert G , wird zuerst festgelegt. Sodann wird derjenige Wert des Parameters p festgelegt, für den die Irrtumswahrscheinlichkeit einen vorgegebenen Wert (α bzw. β) einhält. Für den **Konsumenten** lautet die Nullhypothese:

$$H_0: p > LQL$$

Die zugehörige Alternative lautet

$$H_A: p \leq LQL$$

Es liegt also eine einseitige Fragestellung vor. Die Nullhypothese wird verworfen, wenn $\hat{p} > G$ ist. Die Wahrscheinlichkeit, dass die Nullhypothese fälschlicherweise verworfen wird, das **Konsumentenrisiko**, beträgt α . Für den **Produzenten** stellt sich die Frage, wie groß die Teststärke ist, wie groß also die Wahrscheinlichkeit ist, dass eine akzeptable Partie auch nicht gekennzeichnet wird. Für $p = AQL$ ist diese Wahrscheinlichkeit gerade gleich $1 - \alpha$. Die Fehlerwahrscheinlichkeit, das **Produzentenrisiko**, beträgt α .

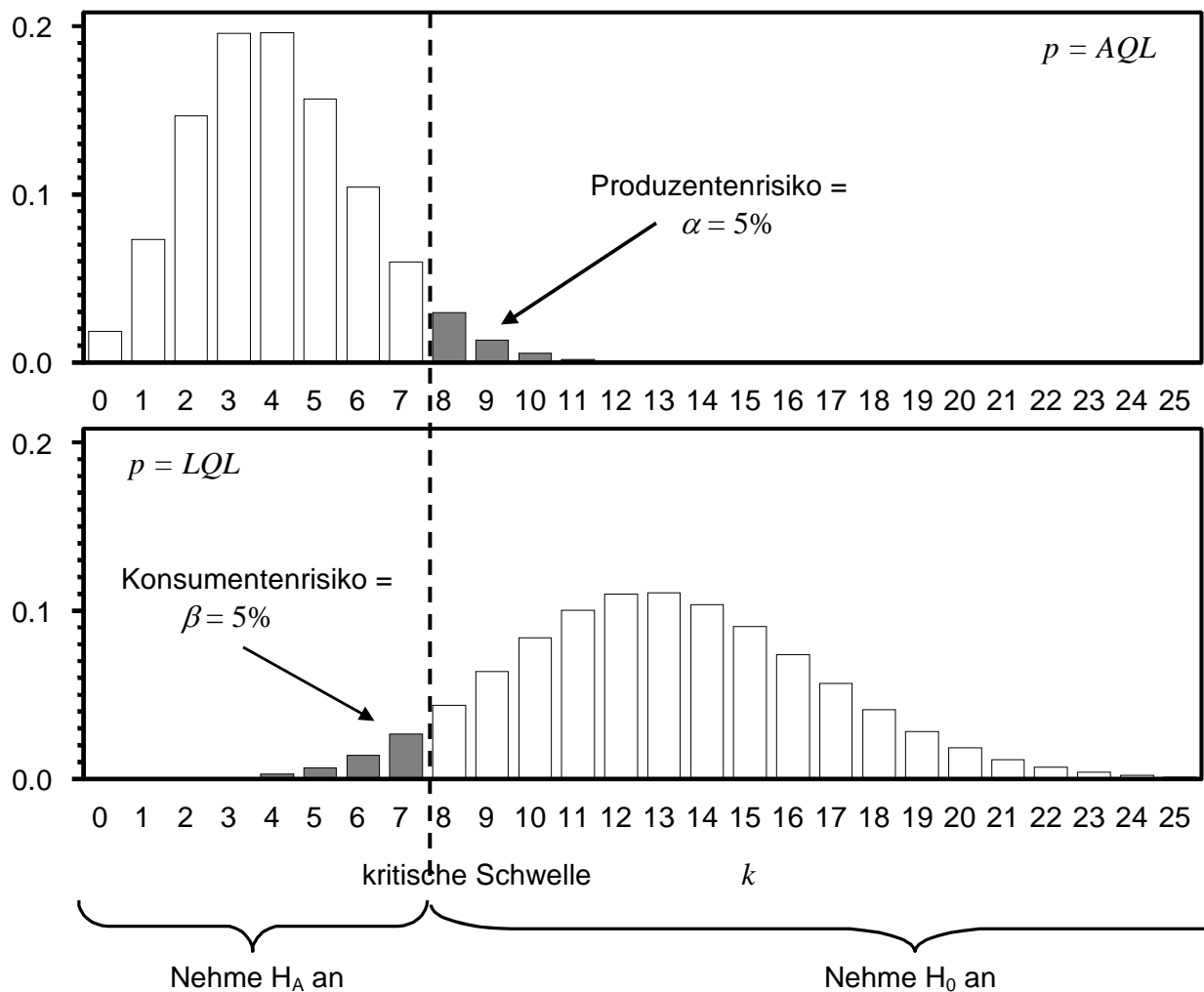


Abb. 5.3.5: Zwei Binomialverteilungen für $n = 1.000$ bei $p = AQL$ und $p = LQL$.

Nun zur **Wahl des Stichprobenumfanges**. Wie findet man beispielsweise für einen vorgegebenen Wert von $LQL = 2G$ und eine vorgegebene Schwelle $G = 0,7\%$ den notwendigen Stichprobenumfang n ? Wir müssen den kleinsten Wert für n finden, der folgende Ungleichung erfüllt:

$$P(\hat{p} > G) = \sum_{k > nG} \binom{n}{k} LQL^k (1 - LQL)^{n-k} \geq 1 - \beta$$

Dies lässt sich einfach mit einem Computer implementieren, wobei man sukzessive n um Eins erhöht, bis die Ungleichung erfüllt ist. Nach diesem Verfahren finden wir für dieses Beispiel einen notwendigen Stichprobenumfang von $n = 552$.

Abschließende Bemerkung: Wir haben hier jeweils exakte Wahrscheinlichkeitsberechnungen durchgeführt. In großen Stichproben kann man die Tatsache nutzen, dass $\hat{p} = x/n$ näherungsweise normalverteilt ist mit Mittelwert (Erwartungswert) p und Varianz $p(1-p)/n$.

*5.3.8 Maximum-Likelihood Schätzung

Der übliche Schätzer des Parameters p der Binomialverteilung lautet $\hat{p} = x/n$. Intuitiv ist dies ein vernünftiger Schätzer, da x/n eine relative Häufigkeit darstellt und somit eine empirische Wahrscheinlichkeit. Es gibt allerdings eine fundiertere Begründung für diesen Schätzer; er folgt aus der Anwendung des sog. **Maximum-Likelihood (ML)** Prinzips. Dieses Prinzip ist neben dem Prinzip der **Kleinsten Quadrate**, welches wir in der linearen Regression kennen lernen werden (Kap. 6), eines von zwei fundamentalen Schätzprinzipien in der Statistik. Es wurde maßgeblich von Sir R.A. Fisher entwickelt, der auch die Varianzanalyse erfunden hat.

Bei der ML Schätzung wird die Wahrscheinlichkeitsfunktion der Daten zugrunde gelegt. Im Fall der Binomialverteilung ist diese gegeben durch

$$P(X=k|n;p) = \binom{n}{k} p^k q^{n-k}.$$

Die Wahrscheinlichkeitsfunktion gibt für jede mögliche Realisation der Zufallsvariablen X die Wahrscheinlichkeit an, wobei die Berechnung dieser Wahrscheinlichkeit die Kenntnis des Parameters p voraussetzt.

Bei der ML-Schätzung wird nun die Betrachtung umgekehrt. Der Parameter p ist unbekannt, während die beobachteten Daten X als feste Größe betrachtet werden. Für die gegebenen Daten wird nun derjenige Wert des Parameters gesucht, der die Wahrscheinlichkeit für die beobachteten Daten maximiert. Der maximierende Wert ist die ML-Schätzung des Parameters p . Bei dieser Betrachtung wird die Wahrscheinlichkeitsfunktion als **Likelihood-Funktion** oder kurz „Likelihood“ bezeichnet:

$$L(p | X = k; n) = L = \binom{n}{k} p^k q^{n-k}.$$

Maximierung der Likelihood L wird vereinfacht, wenn diese zunächst logarithmiert wird:

$$\log L = \log \binom{n}{k} + k \log(p) + (n-k) \log(1-p).$$

Für die Maximierung der log-Likelihood setzen wir die erste Ableitung nach p gleich Null und lösen auf:

$$\frac{\partial \log L}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} = 0 \Leftrightarrow (1-p)k = p(n-k) \Leftrightarrow k = pn \Leftrightarrow p = k/n$$

Dies ist genau der bereits in 5.3.2 verwendete Schätzer des Parameters p ($x = k$).

Die zweite Ableitung der log-Likelihood quantifiziert die Genauigkeit der Parameterschätzung. Die sog. **Information** (oder Fisher Information) ist definiert als der negative Erwartungswert der zweiten Ableitung der log-Likelihood:

$$I(p) = E\left(-\frac{\partial^2 \log L}{\partial p^2}\right).$$

Die zweite Ableitung entspricht der Krümmung oder Kurvatur der Likelihood. Sie zeigt somit an, wie schnell die Likelihood nahe dem Maximum abfällt. Je stärker der Abfall der Likelihood nahe des Maximums, umso höher die Genauigkeit der Parameterschätzung, umso höher also die Information. Die Varianz der Parameterschätzung entspricht dem Kehrwert der Information:

$$\text{var}(\hat{p}) = [I(p)]^{-1}.$$

Wir finden für die Binomialverteilung

$$\frac{\partial^2 \log L}{\partial p^2} = -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} \quad \text{und}$$

$$I(p) = E\left(-\frac{\partial^2 \log L}{\partial p^2}\right) = \frac{np}{p^2} + \frac{n-np}{(1-p)^2} = \frac{n}{p} + \frac{n}{1-p} = \frac{(1-p)n + pn}{p(1-p)} = \frac{n}{p(1-p)},$$

so dass

$$\text{var}(\hat{p}) = [I(p)]^{-1} = \frac{p(1-p)}{n}.$$

Dies ist genau dieselbe Varianzformel, wie sie schon in Abschnitt 5.3.2 verwendet wurde.

Anderes Beispiel (Exkurs): Das ML-Prinzip ist sehr universell. Beispielsweise kann es auch zur Schätzung des Mittelwertes μ einer Normalverteilung mit Varianz σ^2 eingesetzt werden. Gegeben seien Beobachtungen y_1, y_2, \dots, y_n . Für eine Beobachtung y_i ist die Likelihood

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu)^2}{2\sigma^2}}$$

und die log-Likelihood

$$\log L_i = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y_i - \mu)^2}{2\sigma^2}.$$

Bei Unabhängigkeit der Beobachtungen können wir die Produktregel (den Multiplikationssatz) anwenden (Abschnitt 5.2), so dass die Likelihood der Daten gegeben ist durch $L = L_1 L_2 \dots L_n$ und somit

$$\log L = \sum_{i=1}^n \log L_i = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 .$$

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) = 0 \Leftrightarrow \sum_{i=1}^n y_i = \sum_{i=1}^n \mu = n\mu \Leftrightarrow \mu = n^{-1} \sum_{i=1}^n y_i = \bar{y}_{\bullet} .$$

Der resultierende Schätzer ist einfach das arithmetische Mittel der Daten. Weiter finden wir

$$\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{1}{\sigma^2} \sum_{i=1}^n 1 = -\frac{n}{\sigma^2} \quad \text{und} \quad I(\mu) = E\left(-\frac{\partial^2 \log L}{\partial \mu^2}\right) = \frac{n}{\sigma^2}, \text{ so dass}$$

$$\text{var}(\hat{\mu}) = \text{var}(\bar{y}_{\bullet}) = [I(\mu)]^{-1} = \frac{\sigma^2}{n} .$$

Dies ist die aus Abschnitt 3.5 bekannte Formel für die Varianz eines Mittelwertes.

5.4 Poissonverteilung

Ein wichtiges Charakteristikum der **Binomialverteilung** ist, dass in einer Stichprobe die Zahl der Träger einer bestimmten Merkmalsausprägung (z.B. kranke Tiere, rezessive Pflanzen) eine obere Grenze haben: den Stichprobenumfang n . Hierbei ist unerheblich, ob der Stichprobenumfang von vornherein festliegt (Beispiel: Es ist geplant, $n = 100$ Maispflanzen auf Befall mit dem Maiszünsler zu untersuchen), oder ob der Stichprobenumfang selbst eine Zufallsvariable ist (Beispiel: Auf einer Parzelle werden $n = 123$ Vogelmierenpflanzen gefunden; von diesen sind $x = 34$ mit einem Pilz befallen). Wichtig ist, dass die Stichprobe unterteilt werden kann in Individuen /Objekte, die eine bestimmte Merkmalsausprägung haben und solche, die diese Ausprägung nicht haben. Die Zahl der Merkmalsträger x kann natürlicherweise nicht größer als n werden.

In anderen Fällen wird einfach die Zahl der Individuen abgezählt, die Träger eines bestimmten Merkmals sind, **ohne** dass eine bestimmte **obere Grenze** besteht. Es ist dann die Anzahl selbst von Interesse, und nicht ein Anteil oder Prozentsatz. Ein Prozentsatz kann in solchen Fällen oft gar nicht berechnet werden.

Beispiele:

- Anzahl Blattläuse je Pflanze
- Anzahl Nethodenzysten pro cm^3 Boden
- Anzahl Dasseliegen pro Rind
- Anzahl Unkräuter pro m^2
- Keimzahl pro ml Milch
- Zellzahl pro ml Milch
- Anzahl der Kopplungsbrüche auf einem Chromosom
- Anzahl von Samenkörnern pro laufendem Meter bei Drillsaat
- Anzahl Impulse pro Zeiteinheit bei einem Geigerzähler (Zahl radioaktiver Zerfälle)
- Anzahl von Viren pro ml Pflanzensaft
- Anzahl von Rapsglanzkäfern pro Pflanze

In diesen Fällen ist die Binomialverteilung kein geeignetes Modell. Denn es ist hier nicht möglich, eine Stichprobe zu unterteilen in 2 Klassen wie bei der Binomialverteilung. Stattdessen kann die Wahrscheinlichkeitsverteilung der Zählwerte oft durch eine sog. **Poissonverteilung** beschrieben werden.

Binomialverteilung und Poissonverteilung sind Verteilungsmodelle für Zählwerte X . Bei der Binomialverteilung gibt es eine obere Grenze, den Stichprobenumfang, so dass nur Zählwerte $0, 1, \dots, n$ möglich sind, während es bei der Poissonverteilung keine obere Grenze gibt.

Beim **Urnenmodell** für die Binomialverteilung gibt es weiße und schwarze Kugeln. Diese stehen für eine Merkmalsausprägung, die entweder vorhanden ist oder nicht. Anders gesagt, die weißen und schwarzen Kugeln stehen für **zwei Arten von Objekten**. Eine Pflanze kann mit einem Pilz befallen sein ("Kugel schwarz") oder nicht ("Kugel weiß"). Dieses Urnenmodell ist bei der Poissonverteilung so nicht anwendbar. Es gibt dort immer nur eine Art von Objekt. Im Prinzip gibt es nur schwarze Kugeln, und zwar unendlich viele. Die interessierende Frage ist, wie viele schwarze Ku-

geln insgesamt gezogen werden, d.h. wie viele Objekte einer bestimmten Art es je Untersuchungseinheit gibt.

Die Wahrscheinlichkeiten der Anzahl von k Individuen pro Beobachtungseinheit mit einer bestimmten Eigenschaft ist bei Poisson-Verteilung gegeben durch

$$P(X=k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

wobei e die Eulersche Konstante ist ($e \approx 2,7118281828\dots$) und λ = Erwartungswert der Zufallsvariablen X (Zählwert).

Erwartungswert: $E(X) = \lambda$

Varianz: $\text{var}(X) = \lambda$

Bei einer Poissonverteilung stimmen Erwartungswert und Varianz überein.

Zur Berechnung der Wahrscheinlichkeiten kann, ähnlich wie bei der Binomialverteilung, die folgende Rekursionsformel genutzt werden:

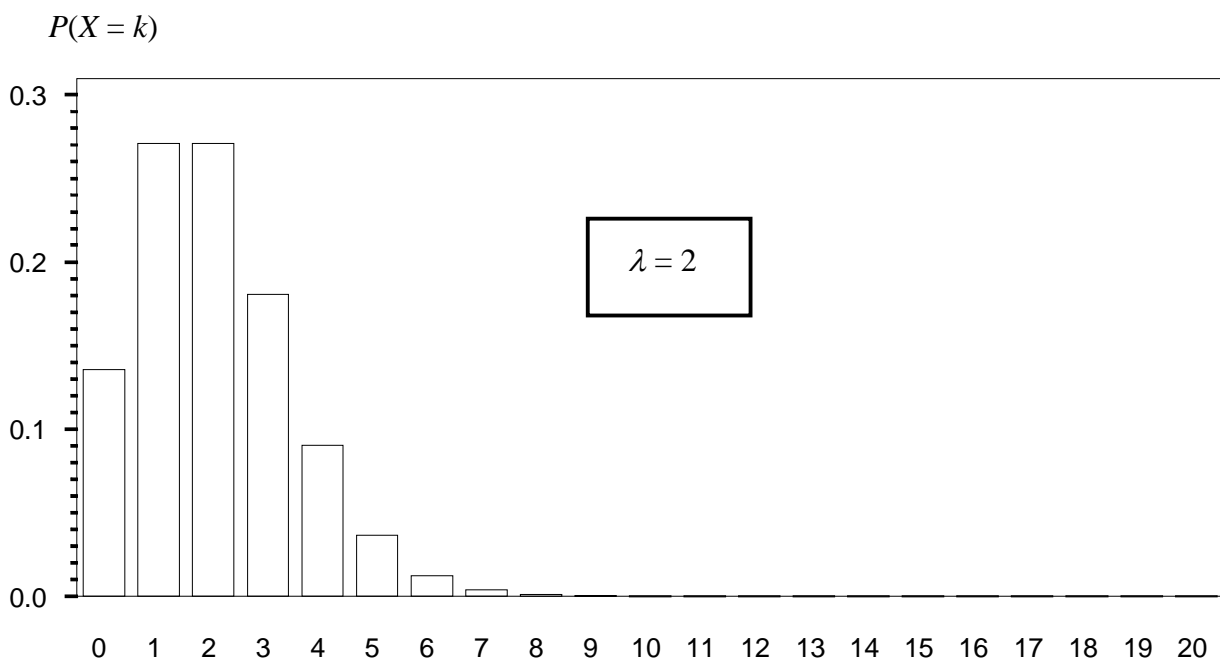
Rekursionsformel:

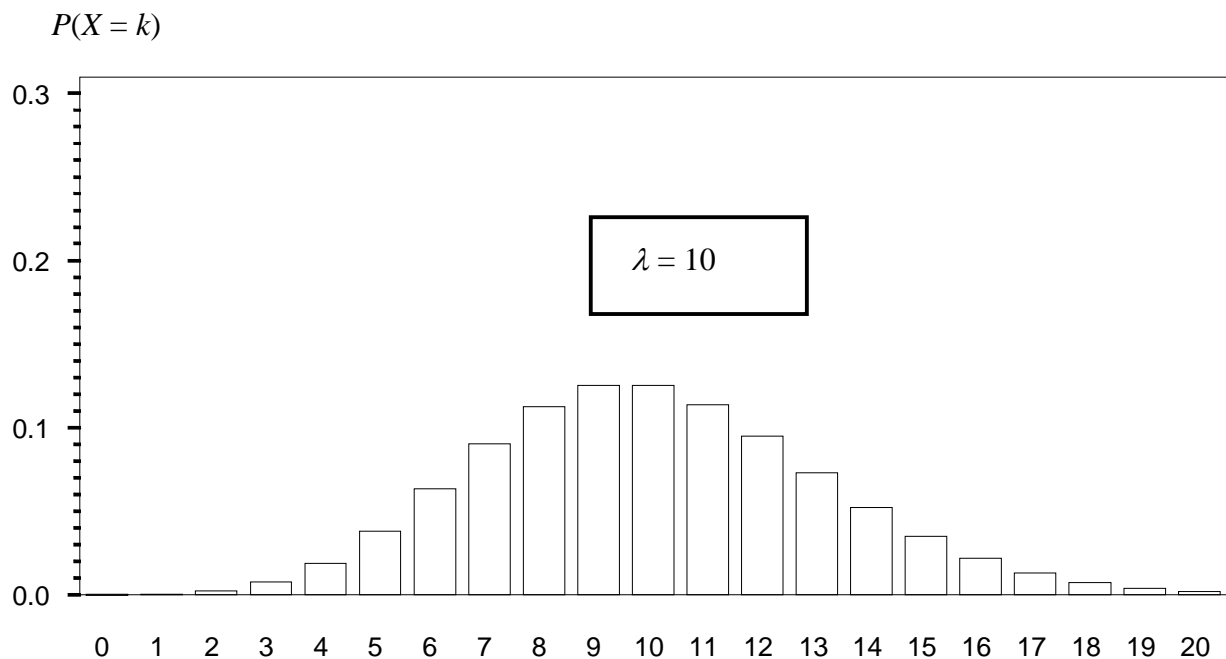
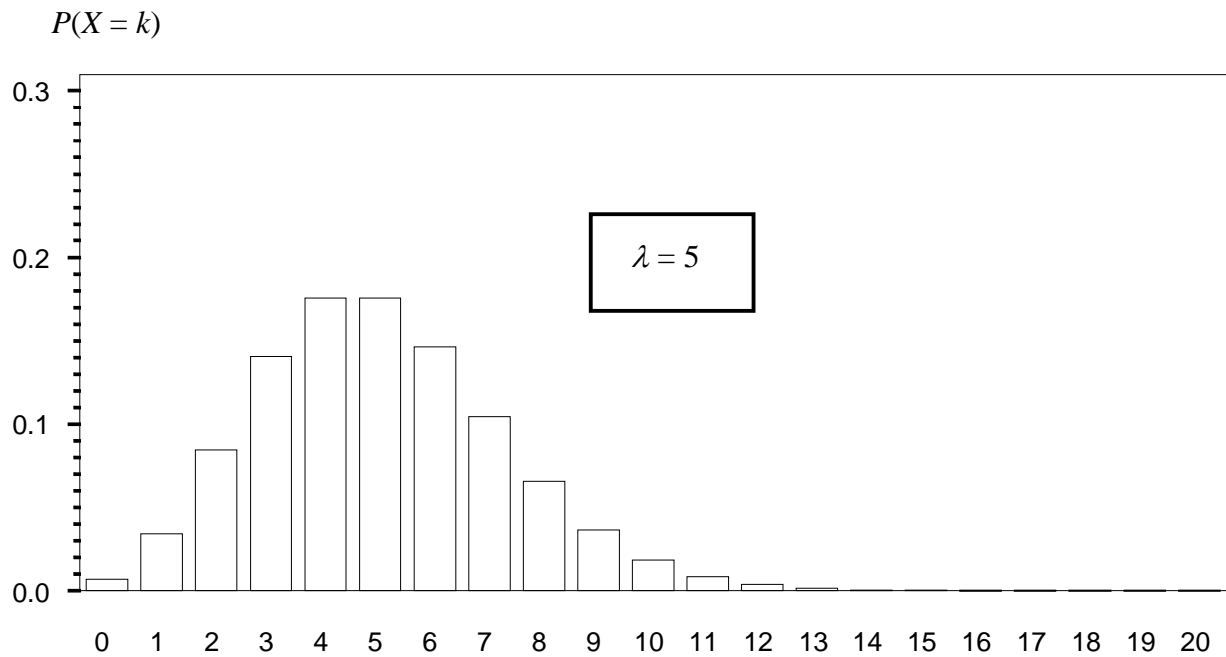
(1) Berechne $P(X=0|\lambda) = e^{-\lambda}$

(2) Berechne

$$P(X=k|\lambda) = \frac{\lambda}{k} P(X=k-1|\lambda)$$

für $k = 1, 2, 3, \dots$





Obenstehend sind einige Beispiele für Poissonverteilungen mit verschiedenen Werten des Parameters λ aufgeführt. Die Verteilung ist für kleine Werte von λ sehr schief und gleicht sich für wachsende Werte von λ immer mehr einer Normalverteilung an.

Beispiel (Lorenz, 1996): Leukozytenzählung. Eine Zählkammer hat eine quadratische Grundfläche von 4mm Seitenlänge, die durch eine Netzteilung in 16 Quadrate unterteilt ist, von denen jedes seinerseits in 16 quadratische Felder unterteilt ist. Die Kammer besteht somit aus 256 Feldern (= Zähleinheiten). Sie ist 0,2 mm hoch; das von ihr eingeschlossene Volumen beträgt mithin $3,2 \text{ mm}^3$. Wird aus einer (in der Regel verdünnten) Leukozytensuspension eine Stichprobe von $3,2 \text{ mm}^3$ eingefüllt, so verteilen sich die Leukozyten regellos über die Netzteilung. Bei Betrachtung durch ein Mikroskop bietet sich ein Bild, das - schematisch - in Abb. 5.4.1 dargestellt ist.

Die Netzteilung dient der Erleichterung beim Auszählen. Durch Hochrechnen der festgestellten Leukozytenzahl auf 1 cm^3 der unverdünnten Suspension bekommt man einen Schätzwert für die Leukozytenkonzentration in der Suspension (= Grundgesamtheit).

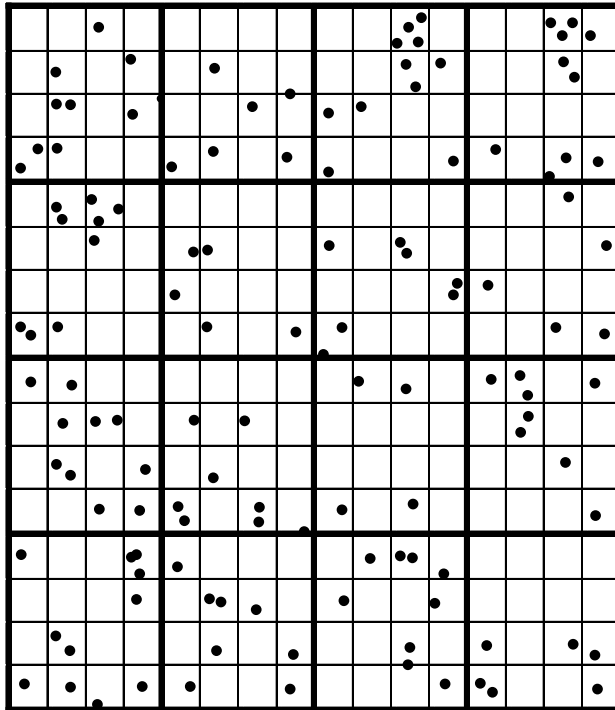


Abb. 5.4.1: Schema einer Zählkammer mit 128 zufällig lokalisierten Leukozyten (nach Lorenz, 1996, S. 96)

Die Verteilung der Leukozyten über das Gitter ist regellos und völlig zufällig. Die Positionen der Leukozyten sind voneinander völlig unabhängig. Dies bedeutet jedoch nicht, dass die Verteilung völlig gleichmäßig wäre. Vielmehr gibt es eine deutliche, rein zufallsbedingte Variation in der Zahl der Leukozyten je Feld:

158 Felder enthalten	keine Leukozyten	($x = 0$)
73 Felder enthalten	je einen Leukozyten	($x = 1$)
21 Felder enthalten	je zwei Leukozyten	($x = 2$)
3 Felder enthalten	je drei Leukozyten	($x = 3$)
1 Feld enthält	je vier Leukozyten	($x = 4$)

256 Felder enthalten zusammen 128 Leukozyten.

Diese Daten stellen eine empirische Häufigkeitsverteilung für die Zufallsvariable X (=Anzahl Leukozyten pro Feld) dar.

5.4.1 Eine interessante Beziehung zur Binomialverteilung

Wir hatten oben festgestellt, dass die Poissonverteilung für Zählwerte gilt, die nach oben keine feste Schranke aufweisen. Dies ist ein fundamentaler Unterschied zur

Binomialverteilung. Interessanterweise kann man allerdings die Poissonverteilung in gewisser Weise als einen Grenzfall der Binomialverteilung auffassen.

Beispiel: Leukozytenzählung. Wir können uns vorstellen, dass die Felder des Zählrasters immer kleiner gemacht werden, so klein, dass in einem Feld höchstens eine Leukozyte Platz hat. Das Feld hat jetzt zwei mögliche Eigenschaften: entweder es ist leer (weiße Kugel) oder es ist von einer Leukozyte besetzt (schwarze Kugel). Bei dieser Betrachtung gibt es somit zwei verschiedene Objekte, so dass die Binomialverteilung als Modell in Frage kommt. Die Zahl der Felder sei n . Wir können nun die Frage stellen, wieviele der n Felder eine Leukozyte aufweisen. Die Zahl der Felder mit Leukozyten (X) ist eine binomialverteilte Zufallsvariable. Der Erwartungswert für die Zahl der Leukozyten je Feld entspricht der Binomialwahrscheinlichkeit p . Er ist hier sehr klein. Umgekehrt ist der Stichprobenumfang (n) sehr groß. Es lässt sich mathematisch zeigen, dass die Binomialverteilung für sehr großes n , bzw. für sehr kleines p , gegen eine Poissonverteilung strebt.

Beispiel: Es wird eine Partie von Saatgut für *Phleum pratense* (Wiesenlieschgras) auf Samen eines Unkrautes untersucht. Von der Partie werden 98 Proben genommen. Jede Probe wiegt 25 g und hat sehr viele Samen, von denen nur sehr wenige von dem gesuchten Unkraut stammen. Der Prozentsatz von Unkrautsamen in einer Probe folgt einer Binomialverteilung. Da die Binomialwahrscheinlichkeit für die Unkrautsamen sehr klein ist bzw. die Zahl aller Samen inklusive der vom Lieschgras sehr groß ist, folgt die Zahl der Unkrautsamen je Probe näherungsweise einer Poissonverteilung. In diesem Beispiel ist diese Tatsache sehr hilfreich, weil hier nur das Gewicht der Probe bekannt ist, nicht aber die Anzahl aller Samen je Probe. Um die Binomialverteilung anwenden zu können, müsste in jeder Probe die Zahl aller Samen gezählt werden, um dann den Prozentsatz von Unkrautsamen zu bestimmen. Wegen des sehr kleinen Prozentsatzes der Unkrautsamen können wir aber die Poissonverteilung annehmen und müssen den Prozentsatz gar nicht bestimmen.

Beispiel (Strickberger, 1988, Genetik, Hanser-Verlag, S. 142): Es wird eine seltene Erbänderung bei *E. coli*-Bakterien untersucht, die zu einer Resistenz gegen Streptomycin führen. Diese Änderungen oder Mutationen kann man dadurch nachweisen, dass man viele Bakterien auf Petrischalen mit antibiotikahaltigem Medium plattiert und die resistenten Kolonien zählt, die von der Mutation eines einzigen Bakteriums herrühren. Es wurden beispielsweise 150 Petrischalen mit Streptomycin-Agar je 1 Millionen Bakterien plattiert und insgesamt 69 resistente Kolonien gefunden. Diese Kolonien verteilten sich wie folgt:

0 Kolonien	98 Petrischalen
1 Kolonie	40 Petrischalen
2 Kolonien	8 Petrischalen
3 Kolonien	3 Petrischalen
4 Kolonien	1 Petrischalen
<hr/>	
150 Petrischalen	

Der Mittelwert der Zahl von Kolonien je Petrischale beträgt $69/150 = 0,46$. Da wir je Petrischale etwa 1 Millionen (10^6) Bakterienzellen plattiert haben, beträgt die geschätzte Mutationswahrscheinlichkeit $p = 0,46 \times 10^{-6}$ Zellen. Für die Zahl der mutier-

ten Bakterienzellen kann eine Binomialverteilung mit Stichprobengröße $n = 10^6$ angenommen werden. Da die Stichprobe sehr groß ist und da die Mutationswahrscheinlichkeit sehr klein ist, kann hier näherungsweise von einer Poissonverteilung der Zahl der Mutationen pro Schale ausgegangen werden.

5.4.2 Parameterschätzung

Schätzen des Parameters der Poisson-Verteilung:

(1) Gegeben sind Zählwerte x_1, x_2, \dots, x_n

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$$

Beachte: Diese Schätzung ist auch möglich, wenn nur ein einziger Zählwert vorliegt.

(2) Gegeben ist eine empirische Häufigkeitsverteilung

Zählwert	Beobachtete Häufigkeit
k	B_k
<hr/>	
0	.
1	.
2	.
.	.
m	.

Berechne:

$$\hat{\lambda} = \frac{\sum_{k=0}^m B_k k}{n}$$

mit

$$n = \sum_{k=0}^m B_k$$

Beispiel: Für die Leukozyten-Daten finden wir:

Zählwert k	Beobachtete Häufigkeit B_k
0	158
1	73
2	21
3	3
4	1

$$n = 256$$

$$\hat{\lambda} = \frac{\sum_{k=1}^m B_k k}{n} = \frac{158 \cdot 0 + 73 \cdot 1 + 21 \cdot 2 + 3 \cdot 3 + 1 \cdot 4}{158 + 73 + 21 + 3 + 1} = \frac{128}{256} = 0,5$$

Der geschätzte Erwartungswert (Mittelwert) der Verteilung beträgt 0,5 Leukozyten pro Feld.

Mit diesen Schätzwert der Verteilung können wir nun die Verteilung der Leukozyten schätzen, die wir erwarten, wenn die Annahme einer Poisson-Verteilung zutrifft.

Erwartete Häufigkeiten einer Poisson-Verteilung:

$$E_k = n \cdot P(X = k | \lambda)$$

$$n = \text{Summe aller beobachteten Häufigkeiten} = \sum_{k=0}^m B_k$$

Rekursionsformel:

(1) Berechne

$$E_0 = n \cdot e^{-\lambda}$$

(2) Berechne

$$E_k = \frac{\lambda}{k} E_{k-1}$$

für $k = 1, 2, 3, \dots$

Praktischer Hinweis:

Für E_{k+1}, E_{k+2}, \dots kann die Berechnung abgebrochen werden, falls E_k klein wird. Man fasst dann die höheren Klassen zu einer Restklasse zusammen. Des weiteren kann es sein, dass die ersten Klassen eine sehr kleine Häufigkeit haben. Auch diese sollten dann zusammengefasst werden.

Beispiel: Leukozytenzählung.

$$\hat{\lambda}=0,5$$

$$P(X=0) = e^{-\lambda} = e^{-0,5} = 0,60653$$

$$E_0 = n \cdot e^{-\lambda} = 256 \cdot 0,60653 = 155,27185$$

$$E_1 = \frac{0,5}{1} 155,27185 = 77,63592$$

$$E_2 = \frac{0,5}{2} 77,63592 = 19,40898$$

$$E_3 = \frac{0,5}{3} 19,40898 = 3,23483$$

$$E_4 = \frac{0,5}{4} 3,23483 = 0,40435 < 1$$

$$E_{Rest} = n - E_0 - E_1 - E_2 = 256 - 155,27 - 77,64 - 19,41 = 3,64$$

Zählwert k	Beobachtete Häufigkeit B_k	Erwartete Häufigkeit E_k
0	158	155,3
1	73	77,6
2	21	19,4
3 oder mehr	4	3,6
Summe	256	255,9 \approx 256

Die erwarteten Häufigkeiten stimmen hier sehr gut mit denen überein, die bei einer Poisson-Verteilung zu erwarten sind. Wir werden in Abschnitt 5.5 einen Test kennenlernen, mit dem man formal die Abweichung von der erwarteten Verteilung auf Signifikanz testen kann, den sog. Chi-Quadrat (χ^2)-Anpassungstest.

Vertrauensintervall für Parameter λ einer Poisson-Verteilung:

$$\lambda_u = \frac{1}{n} \left(\sqrt{n\hat{\lambda}} - \frac{1}{2} z_{Tab} \right)^2; \quad \lambda_o = \frac{1}{n} \left(\sqrt{n\hat{\lambda}+1} + \frac{1}{2} z_{Tab} \right)^2$$

$$z_{Tab} = z_{1-\alpha/2}$$

wobei z_γ das γ -Quantil der Standardnormalverteilung ist (Tab. III) und α die vorgegebene Irrtumswahrscheinlichkeit (Fehler 1. Art). Für große Stichproben (Die Literatur sagt nicht eindeutig wie groß) gilt näherungsweise Normalverteilung von $\hat{\lambda}$ und somit:

$$\lambda_{u,o} = \hat{\lambda} \pm z_{Tab} \sqrt{\hat{\lambda}/n}$$

Beispiel: Die Leukozytenzählung ergab $\hat{\lambda}=0,5$ Leukozyten pro Feld (entspricht $1/256$ von $3,2 \text{ mm}^3$) mit $n = 256$. Es soll ein 95%iges Vertrauensintervall berechnet werden.

$$\lambda_u = \frac{1}{n} \left(\sqrt{n\hat{\lambda}} - \frac{1}{2} z_{Tab} \right)^2 = \frac{1}{256} \left(\sqrt{256 \cdot 0,5} - \frac{1,96}{2} \right)^2 = \frac{1}{256} (-10,33)^2 = 0,41713$$

$$\lambda_o = \frac{1}{n} \left(\sqrt{n\hat{\lambda} + 1} + \frac{1}{2} z_{Tab} \right)^2 = \frac{1}{256} \left(\sqrt{256 \cdot 0,5 + 1} + \frac{1,96}{2} \right)^2 = \frac{1}{256} (12,34)^2 = 0,59462$$

Umrechnung auf 1 mm^3 ergibt die Grenzen $0,41713 \cdot 256 / 3,2 = 33,37$ und $0,59462 \cdot 256 / 3,2 = 47,57$ Leukozyten pro mm^3 . Das Intervall mit den Grenzen von 33,37 und 47,57 überdeckt mit 95%iger Wahrscheinlichkeit die tatsächliche Zahl der Leukozyten je mm^3 .

Beispiel (ungruppierte Daten): In einer Kläranlage wurden 5 Wasserproben untersucht. Diese enthielten 2, 0, 1, 5, und 2 Keime pro ml. Es soll ein 95%-Vertrauensintervall für die Keimzahl je ml berechnet werden.

$$n = 5; \hat{\lambda} = \frac{2+0+1+5+2}{5} = 2$$

$$\lambda_u = \frac{1}{n} \left(\sqrt{n\hat{\lambda}} - \frac{1}{2} z_{Tab} \right)^2 = \frac{1}{5} \left(\sqrt{5 \cdot 2} - \frac{1,96}{2} \right)^2 = \frac{1}{5} (2,18)^2 = 0,952$$

$$\lambda_o = \frac{1}{n} \left(\sqrt{n\hat{\lambda} + 1} + \frac{1}{2} z_{Tab} \right)^2 = \frac{1}{5} \left(\sqrt{5 \cdot 2 + 1} + \frac{1,96}{2} \right)^2 = \frac{1}{5} (4,30)^2 = 3,692$$

Die Keimzahl wird mit 95%iger Wahrscheinlichkeit vom Intervall mit den Grenzen von 0,95 Keime/ml und 3,69 Keime/ml überdeckt.

Beispiel: Es wird eine Partie von Saatgut für *Phleum pratense* (Wiesenlieschgras) auf Samen eines Unkrautes untersucht. Hierzu werden 98 Proben genommen. Jede Probe wiegt 25 g und hat sehr viele Samen, von denen nur sehr wenige von dem gesuchten Unkraut stammen. Wir finden im Durchschnitt 3,02 Samen je Probe. Es soll ein Vertrauensintervall für die durchschnittliche Zahl der Unkrautsamen je Probe berechnet werden. Für das Vertrauensintervall wählen wir $\alpha = 5\%$.

$$n = 98; \hat{\lambda} = 3,02$$

$$\lambda_u = \frac{1}{n} \left(\sqrt{n\hat{\lambda}} - \frac{1}{2} z_{Tab} \right)^2 = \frac{1}{98} \left(\sqrt{98 \cdot 3,02} - \frac{1,96}{2} \right)^2 = \frac{1}{98} (17,203 - 0,98)^2 = 2,69$$

$$\lambda_o = \frac{1}{n} \left(\sqrt{n\hat{\lambda} + 1} + \frac{1}{2} z_{Tab} \right)^2 = \frac{1}{98} \left(\sqrt{98 \cdot 3,02 + 1} + \frac{1,96}{2} \right)^2 = \frac{1}{98} (17,233 + 0,98)^2 = 3,38$$

Die Zahl von Unkrautsamen je Probe (25 g) liegt mit 95%iger Wahrscheinlichkeit zwischen 2,69 und 3,38.

Alternative Formel (große Stichproben):

$$\lambda_{u,o} = \hat{\lambda} \pm z_{Tab} \sqrt{\hat{\lambda}/n} = 3,02 \pm 1,96 \sqrt{3,02/98} = 3,02 \pm 0,34$$

$$\lambda_u = 2,68; \lambda_o = 3,36$$

Die Grenzen mit dieser etwas gröberen Formel unterscheiden sich kaum von denen, die mit der genaueren erhalten wurden. Dies liegt an dem relativ großen Stichprobenumfang.

5.4.3 Test für den Vergleich zweier Parameter λ_1 und λ_2

λ_1 = Poisson-Parameter der Gruppe 1

λ_2 = Poisson-Parameter der Gruppe 2

$$H_0: \lambda_1 = \lambda_2$$

Alternativhypothese: $\lambda_1 \neq \lambda_2$

n_1, n_2 = Stichprobenumfang in Gruppe 1 und 2

Schätzung von λ_1 und λ_2 wie in 5.4.1.

$$z_{Vers} = \frac{|\hat{\lambda}_1 - \hat{\lambda}_2|}{\sqrt{\left(\frac{n_1 \hat{\lambda}_1 + n_2 \hat{\lambda}_2}{n_1 + n_2} \right) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$z_{Tab} = z_{1-\alpha/2}$$

wobei z_γ das γ -Quantil der Standardnormalverteilung ist (Tab. III) und α die vorgegebene Irrtumswahrscheinlichkeit (Fehler 1. Art).

$$z_{Vers} > z_{Tab} \Rightarrow \text{verwerfe } H_0$$

$$z_{Vers} \leq z_{Tab} \Rightarrow \text{behalte } H_0 \text{ bei}$$

Beispiel: Es werden zwei Partien A und B von Saatgut für *Phleum pratense* (Wiesenslieschgras) auf Samen eines Unkrautes untersucht. Von jeder Partie werden 98 Proben genommen. Jede Probe wiegt 25 g und hat sehr viele Samen, von denen nur sehr wenige von dem gesuchten Unkraut stammen. Von der zweiten Partie gehen drei Proben verloren, so dass nur 95 Proben untersucht werden. Die durchschnittliche Zahl von Unkrautsamen je Probe betrug 3,02 für Partie A und 3,75 für Partie B. Es soll zum Niveau $\alpha = 5\%$ geprüft werden, ob sich die beiden Partien im Unkrautbesatz unterscheiden.

$$\hat{\lambda}_1 = 3,02; \hat{\lambda}_2 = 3,75; n_1 = 98; n_2 = 95$$

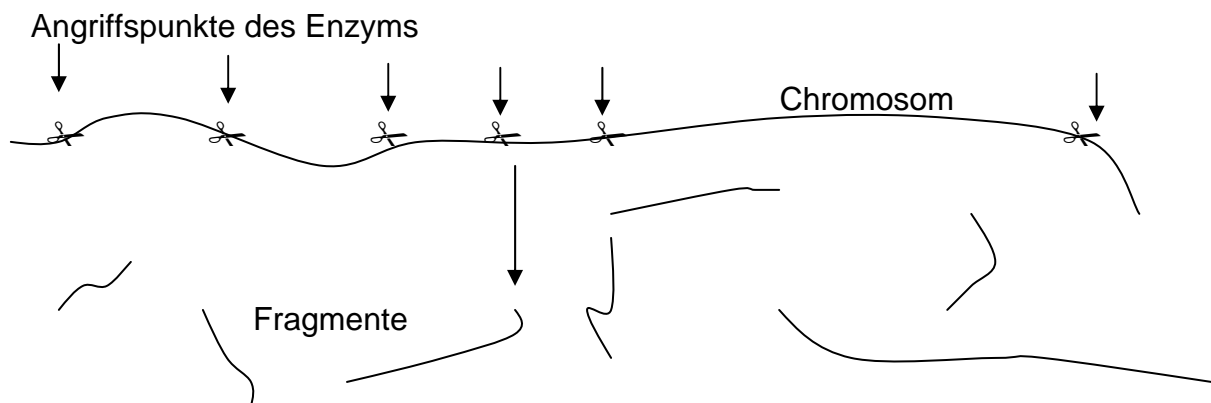
$$\begin{aligned}
z_{Vers} &= \frac{|\hat{\lambda}_1 - \hat{\lambda}_2|}{\sqrt{\left(\frac{n_1 \hat{\lambda}_1 + n_2 \hat{\lambda}_2}{n_1 + n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{|3,02 - 3,75|}{\sqrt{\left(\frac{98 \cdot 3,02 + 95 \cdot 3,75}{98 + 95}\right) \left(\frac{1}{98} + \frac{1}{95}\right)}} \\
&= \frac{0,73}{\sqrt{\left(\frac{295,96 + 356,25}{193}\right) 0,02073}} = \frac{0,73}{\sqrt{0,0701}} = 2,76
\end{aligned}$$

$$z_{Tab} = z_{0,975} = 1,96 < z_{Vers}$$

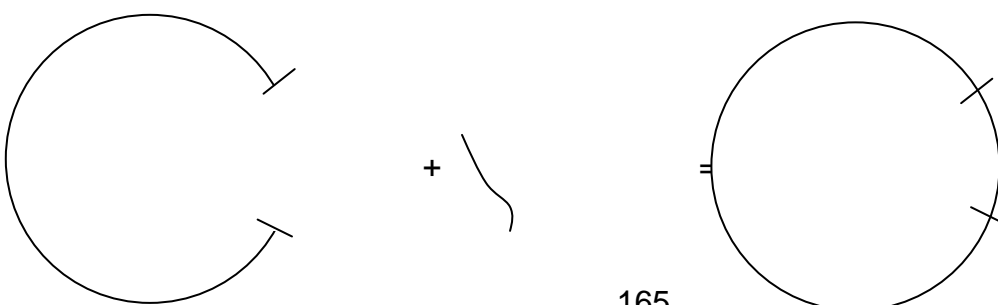
⇒ lehne H_0 ab. Die beiden Partien unterscheiden sich signifikant im Unkrautbesatz. Die Partie B hat mit 3,75 Unkrautsamen je Probe den signifikant höheren Besatz.

*5.4.4 Genbanken - ein einfaches, aber ausführliches Beispiel zu Poisson- und Binomialverteilung

Beispiel (Prof. Dr. G. Weber, Institut 350, FG Pflanzenzüchtung und Biotechnologie): Das Hopfengenom besteht aus ca. 10^9 Basenpaaren (bp). Es soll eine Genbank erstellt werden, welche das gesamte Genom umfasst. Hierzu wird DNA mit Hilfe eines Restriktionsenzym in Fragmente zerschnitten. Das Enzym erkennt eine spezifische Basensequenz als Schnittstelle und schneidet daher nur an solchen Stellen, welche die betreffende Sequenz aufweisen.



Die durchschnittliche Länge eines DNA Fragmentes beträgt 10^5 bp. Die Fragmente werden in sog. BAC (bacterial artificial chromosomes) integriert. Hierzu wird ein Reaktionsansatz mit den Fragmenten und einer Bakterienkolonie gebildet. Während der Reaktion wird jeweils ein Fragment in das kreisförmige Chromosom eines Bakteriums inseriert.



Wichtig bei der Betrachtung dieses Prozesses ist, dass die Auswahl der Fragmente aus dem Reaktionsansatz völlig zufällig verläuft. Eine zentrale Frage beim Aufbau einer Genbibliothek mit der beschriebenen Methode ist, wie viele Bakterien notwendig sind, um eine möglichst vollständige Abdeckung des Genoms zu erreichen. In anderen Worten: Wie viele Bakterien sind notwendig, damit das Hopfengenom z.B. zu 99,9% abgedeckt wird? Hierbei soll die Annahme gemacht werden, dass jedes Fragment mit derselben Wahrscheinlichkeit in ein Bakteriengenom insertiert wird.

Lösung 1 (Binomialverteilung): Zunächst müssen wir die Zahl der verschiedenen Fragmente betrachten. Diese ist über die durchschnittliche Länge der Fragmente zu erhalten. Da die mittlere Fragmentlänge 10^5 bp beträgt und das Gesamtgenom 10^9 bp umfasst, muss es $10^9/10^5 = 10^4$ **verschiedene Fragmente** geben. Diese 10^4 verschiedenen Fragmente bezeichne ich im folgenden mit **Fragmenttypen**. Es gibt also 10^4 verschiedene Fragmenttypen, wobei im Reaktionsansatz jeweils sehr viele Exemplare eines Fragmenttyps vertreten sind.

Man könnte bei naiver Betrachtung den Schluss ziehen, dass genau $10^4 = 10.000$ Bakterien benötigt werden, nämlich genau eines für jeden Fragmenttyp. Dies ist aber deswegen falsch, weil die Insertion von Fragmenten völlig zufällig erfolgt: Es kann nicht gesteuert werden, welches Fragment in ein bestimmtes Bakterium insertiert wird. Vielmehr erfolgt die Auswahl der Fragmente zufällig. Es kann daher leicht passieren, dass ein Fragmenttyp gar nicht repräsentiert wird, obwohl insgesamt sehr viele Fragmente insertiert werden. Aus diesem Grund kann das Problem nur mit Hilfe statistischer Methoden gelöst werden!

Betrachten wir nun einen bestimmten Fragmenttyp, also denjenigen Bereich des Hopfengenoms, der Fragmente von diesem Typ liefert. Der Bereich wird genau dann in der Bibliothek repräsentiert sein, wenn ein Fragment dieses Typs mindestens einmal in ein Bakterium insertiert wird. Umgekehrt wird der Bereich genau dann nicht repräsentiert sein, wenn bei keinem der Bakterien der betreffende Fragmenttyp insertiert wird. Zur Beantwortung der Frage, wie viele Bakterien notwendig sind, um das Genom zu 99,9% abzudecken, ist es am einfachsten, zunächst die Wahrscheinlichkeit zu berechnen, dass ein bestimmter Fragmenttyp bei keiner der Bakterien insertiert wird. Da es 10^4 verschiedene Fragmenttypen gibt, ist die Wahrscheinlichkeit, dass ein bestimmter Fragmenttyp bei einem Bakterium insertiert wird, gleich $p = 1/10^4 = 10^{-4}$. Die Gegenwahrscheinlichkeit, dass es bei einem einzelnen Bakterium nicht insertiert wird, ist $q = 1 - p = 1 - 10^{-4}$. Wenn es n Bakterien gibt, so beträgt die Wahrscheinlichkeit, dass ein Fragmenttyp bei keinem der Bakterien insertiert wird, nach dem Multiplikationssatz der Wahrscheinlichkeitsrechnung (beachte: Insertionen bei verschiedenen Bakterien sind statistisch unabhängig) gleich

$$W = P(\text{Fragmenttyp bei keiner der } n \text{ Bakterien inseriert}) =$$

$$\begin{aligned} & P(\text{Fragmenttyp beim 1. Bakterium nicht inseriert}) \\ & \times P(\text{Fragmenttyp beim 2. Bakterium nicht inseriert}) \\ & \times \dots \\ & \times P(\text{Fragmenttyp beim } n\text{-ten Bakterium nicht inseriert}) \end{aligned}$$

$$= (1 - p) \times (1 - p) \times \dots \times (1 - p) \quad ("n \text{ mal} ")$$

$$= (1 - p)^n$$

Man beachte, dass sich diese Formel auch einfach ergibt aus der Annahme einer **Binomialverteilung** für die Anzahl X von Bakterien, die den betreffenden Fragmenttyp tragen, ergibt, wobei p die Binomialwahrscheinlichkeit ist. Es ist

$$P(X = 0|n; p) = (1 - p)^n$$

(siehe Abschnitt 5.3). Nun soll $W = 1 - 0,999 = 0,001$ werden, und zwar für jede Position, also jeden Fragmenttyp des Genoms. Wenn die Wahrscheinlichkeit, nicht repräsentiert zu sein, für jeden Fragmenttyp $W = 0,001$ beträgt, so ist diese Wahrscheinlichkeit gleich dem erwarteten Anteil des Genoms, der nicht repräsentiert ist. Der erwartete repräsentierte Anteil ist demnach 99,9%.

Um die notwendige Zahl von Bakterien zu errechnen, so dass $W = 0,001$ ist, müssen wir die Werte für p und W in die Gleichung einsetzen und nach n auflösen. Der Einfachheit halber können wir auch erst die Gleichung nach n auflösen und dann den Wert für W und p einsetzen.

$$W = (1 - p)^n$$

$$\Leftrightarrow \log(W) = n \log(1 - p)$$

$$\Leftrightarrow n = \log(W)/\log(1 - p)$$

$$\text{Hier: } p = 10^{-4}, W = 0,001 \Rightarrow n = \log(0,001)/\log(1 - 10^{-4}) \approx 69.074$$

Es werden also rund 70.000 Bakterien benötigt, um das Genom zu 99,9% zu repräsentieren. Dies ist das siebenfache von 10.000, der naiven Abschätzung für n basierend auf der falschen Annahme, dass es technisch möglich ist, jeden Fragmenttyp genau einmal repräsentiert zu haben. Wir benötigen hier somit eine **siebenfache Redundanz** in der Bibliothek, um 99,9% des Genoms zu repräsentieren.

Wir können an dieser Stelle auch sagen, welcher Teil des Genoms nicht repräsentiert wird, wenn wir nur $n = 10.000$ Bakterien verwenden. In diesem Fall ist

$$W = (1 - p)^n = (1 - 10^{-4})^{10.000} = 0,36786 \approx 37\%$$

Es wäre damit zu rechnen, dass 37% des Genoms nicht in der Genbibliothek repräsentiert ist!

Lösung 2 (Poissonverteilung): Die Wahrscheinlichkeit, das ein zufällig ausgewähltes Bakterium einen bestimmten Fragmenttyp trägt, ist $p = 10^{-4}$. Unter n Bakterien erwarten wir daher, dass

$$\lambda = np$$

Bakterien diesen Fragmenttyp tragen. Dieser Erwartungswert gilt für jeden Fragmenttyp. Wenn wir nun n Insertionen in Bakteriengenome betrachten, so können wir

fragen, wie viele der Fragmenttypen sind keinmal, wie viele einmal, wie viele zweimal, etc. in ein Bakterium insertiert worden sind. Diese Anzahlen ergeben eine Häufigkeitsverteilung, die vom Erwartungswert λ abhängt. Die Verteilung ist eine Poissonverteilung mit Erwartungswert

$$\lambda = np.$$

Um den Typ der Verteilung zu erkennen, mag auch folgende Analogie bemüht werden: Wir stellen uns eine Platte mit 10^4 gleich großen Quadraten vor. Jedes Quadrat repräsentiert einen Fragmenttyp. Auf diese Platte fallen n Tropfen. Auf manche Quadrate fällt kein Tropfen, auf manche ein Tropfen, auf manche zwei Tropfen, etc. Die Zahl der Tropfen auf einem Quadrat entspricht der Zahl der Bakterien, die den betreffenden Fragmenttyp insertiert haben. Die Häufigkeiten der Quadrate mit keinem, einem, zwei, ... Tropfen folgen näherungsweise einer Poissonverteilung. Man beachte auch die Ähnlichkeit zum Leukozytenbeispiel in Abschnitt 5.4.

Wir können nun also annehmen, dass die beobachtete Anzahlen, X , von Fragmenttypen, die, keinmal, einmal, zweimal, repräsentiert sind, einer Poissonverteilung folgen. Insbesondere ist die Wahrscheinlichkeit, dass es keinen Fragmenttyp gibt, der nicht repräsentiert ist, nach Poissonverteilung gleich

$$P(X = 0) = e^{-\lambda}$$

Diese Wahrscheinlichkeit soll nun $W = 0,001$ betragen, also

$$P(X = 0) = e^{-\lambda} = W$$

$$\Leftrightarrow \log(W) = -\lambda = -np$$

$$\Leftrightarrow n = -\log(W)/p = 69.078 \quad (\text{mit } p = 10^{-4})$$

Dies ist fast dieselbe Zahl, die wir bei *Lösung 1* unter Annahme einer Binomialverteilung gefunden haben.

Abschließende Bemerkung: Das Beispiel zeigt die enge Beziehung zwischen Binomialverteilung und Poissonverteilung bei sehr kleiner Binomialwahrscheinlichkeit p . Man kann hier auch ein mathematisches Argument (siehe Mathematik-Vorlesung) verwenden. Es gilt die folgende Taylor-Reihen-Entwicklung:

$$\log(1+x) = \sum_{v=1}^{\infty} \frac{(-1)^{v-1}}{v} x^v = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \dots \quad ; \quad x \in (-1, 1]$$

Für $|x| \ll 1$ kann man die Taylorreihe in guter Näherung nach dem ersten Glied abbrechen:

$$\log(1+x) \approx x$$

Setzen wir nun $x = -p$, so finden wir

$$\log(1-p) \approx -p \Leftrightarrow -\log(1-p) \approx p$$

Beispiele:

p	$-\log(1-p)$
10^{-1}	$1,05361 \times 10^{-1}$
10^{-2}	$1,00503 \times 10^{-2}$
10^{-3}	$1,00050 \times 10^{-3}$
10^{-4}	$1,00005 \times 10^{-4}$
10^{-5}	$1,00001 \times 10^{-5}$
10^{-6}	$1,00000 \times 10^{-6}$
10^{-7}	$1,00000 \times 10^{-7}$
10^{-8}	$1,00000 \times 10^{-8}$

Bei Annahme einer **Binomialverteilung** hatten wir gefunden:

$$n = \log(W)/\log(1-p)$$

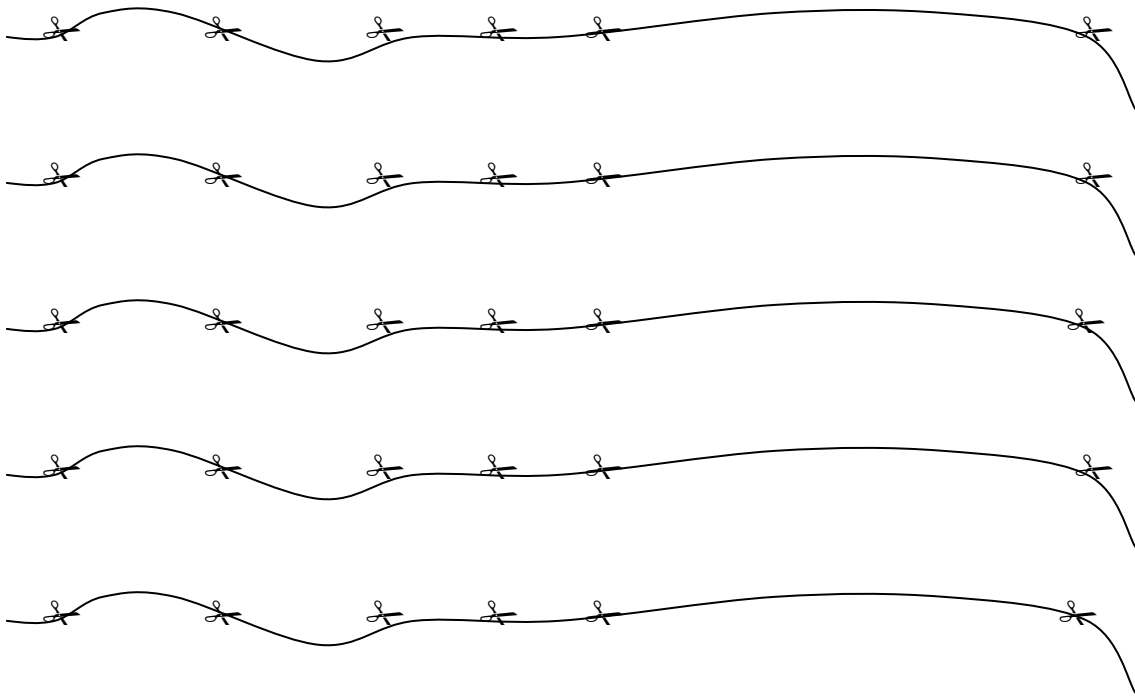
Einsetzen der Approximation liefert

$$n = -\log(W)/p$$

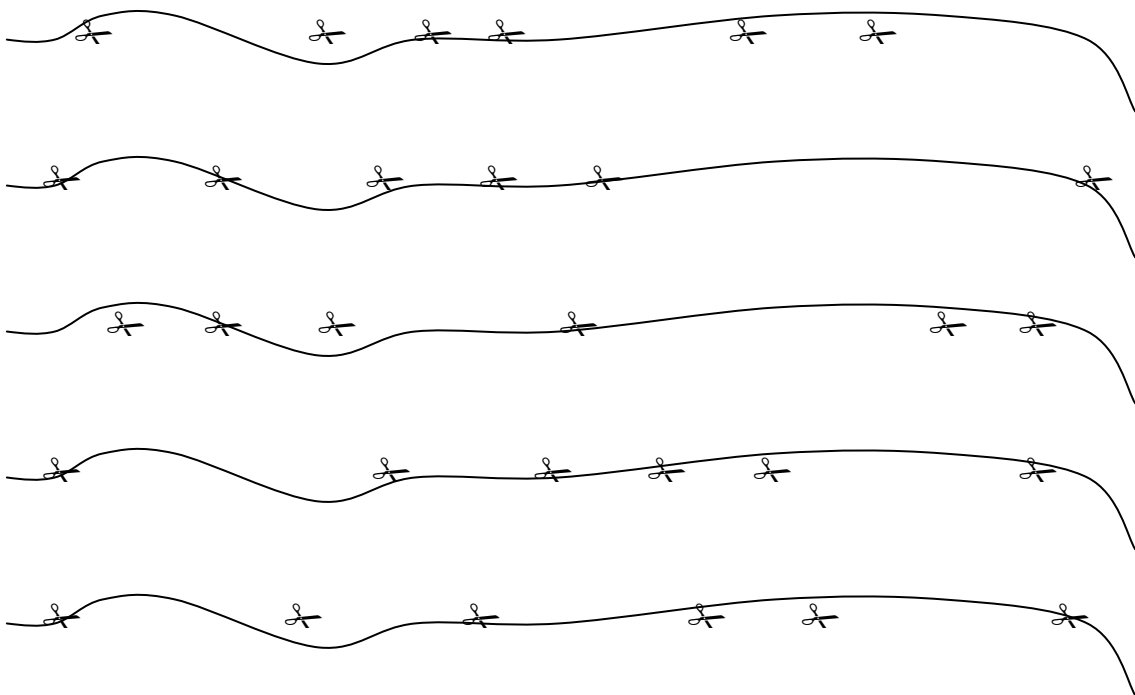
was genau der Gleichung nach **Poissonverteilung** entspricht. Dies erklärt die gute Übereinstimmung beider Ergebnisse.

Beispiel (Prof. Dr. G. Weber, Institut 350, FG Pflanzenzüchtung und Biotechnologie): Eine alternative Methode zur Generierung einer Genbibliothek beruht auf der sog. **DNA-Scherung**. Hierbei wird die DNA physikalisch, z.B. durch Ultraschall, in Fragmente zerteilt. Ein Vorteil dieses Verfahrens besteht darin, dass sehr kleine Fragmente nicht zu Problemen führen. Sehr kleine Fragmente sind grundsätzlich problematisch, weil diese aus bestimmten Gründen nicht in das Bakteriengenom insertiert werden können und somit verloren gehen. Wenn die Fragmente mit einem bestimmten Restriktionsenzym erzeugt werden, muss man damit rechnen, dass ein Teil des Genoms nicht in der Genbibliothek repräsentiert ist, weil es ausschließlich durch zu kleine Fragmente repräsentiert wird. Bei der DNA-Scherung sind dagegen die Bruchstellen völlig zufällig über das Genom verteilt. Die Bruchstellen sind somit nicht auf bestimmte Nukleotidsequenzen beschränkt. Zwar kann es auch hier geschehen, dass sehr kleine Fragmente auftreten. Aber diese sind, und das ist wesentlich, nicht ausschließlich auf bestimmte Bereiche des Genoms beschränkt. In anderen Worten: es ist nicht wie bei der Fragmentierung mit Restriktionsenzymen, bei der bestimmte Regionen ausschließlich durch zu kleine Fragmente repräsentiert werden, die nie in ein Bakteriengenom insertiert werden. Bei der Scherung wird jeder Ort im Genom mit großer Wahrscheinlichkeit auch durch Fragmente repräsentiert, die groß genug sind für eine Insertion in ein Bakteriengenom.

Mit Restriktionsenzymen:



Mit DNA-Scherung:



Die Fragmente werden mittels Pulsfeldelektrophorese nach ihrer Größe aufgetrennt. Hierdurch ist es möglich, Fragmente des für die Insertion optimalen Größenbereiches von 10^5 bp abzutrennen. Wegen der Art der Erzeugung der Fragmente kann prinzipiell jeder Punkt des Genoms durch ein Fragment dieses Größenbereiches repräsentiert werden. Die Fragmente werden dann mit den Bakterien in einen Reaktionsansatz gebracht, bei dem die Fragmente in Genome der Bakterien inseriert werden.

Auch bei der DNA-Scherung stellt sich die Frage, wie viele Fragmente inseriert werden müssen, um dass Genom z.B. zu 99,9% zu repräsentieren.

Lösung (Binomialverteilung): Wir können davon ausgehen, dass die Länge bei allen Fragmenten näherungsweise 10^5 bp beträgt. Die Gesamtlänge des Hopfengenoms beträgt etwa 10^9 bp. Betrachten wir nun einen bestimmten Punkt im Genom. Wird zufällig ein Fragment aus dem Reaktionsansatz ausgewählt, so ist die Wahrscheinlichkeit, dass dieses den betrachteten Punkt im Genom enthält, proportional zur Länge des Fragmentes. Die Wahrscheinlichkeit beträgt

$$p = 10^5/10^9 = 10^{-4}.$$

Die Wahrscheinlichkeit, dass von keinem von n Bakterien ein Fragment insertiert wird, welches den Punkt des Genoms enthält, beträgt

$$W = (1 - p)^n$$

Da diese Wahrscheinlichkeit für jeden Punkt im Genom gilt (wenn man Randeffekte ignoriert), wird ein Anteil von W des Genoms nicht repräsentiert sein, wenn Fragmente in n Bakterien insertiert werden. Hiermit ergibt sich, genau wie im Beispiel mit Restriktionsenzymen,

$$n = \log(W)/\log(1 - p)$$

$$\text{Hier: } p = 10^{-4}, W = 0,001 \Rightarrow n = \log(0,001)/\log(1 - 10^{-4}) \approx 69.074$$

Also müssen, wie bei Verwendung von Restriktionsenzymen, ca. 70.000 Bakterien verwendet werden, um eine Abdeckung des Genoms von 99,9% zu erreichen.

Eine alternative Lösung basierend auf der Poissonverteilung, wie bei Restriktionsenzymen, ist hier nicht möglich, weil die bei der Scherung entstehenden Fragmente nicht für jedes Individuum (Zelle einer Hopfenpflanze) dieselben sind.

[Siehe auch: Gardner et al. (1999) Journal of Agricultural, Biological and Environmental Statistics 4, 1-8.]

5.5 Chi-Quadrat-Anpassungs-Test

Beispiel: Für die Leukozyten-Daten aus Abschnitt 5.4 hatten wir festgestellt, dass die beobachteten Häufigkeiten (B_k) gut mit den unter der Annahme einer Poisson-Verteilung erwarteten Häufigkeiten (E_k) übereinstimmten.

Zählwert k	Beobachtete Häufigkeit B_k	Erwartete Häufigkeit E_k
0	158	155,3
1	73	77,6
2	21	19,4
3 oder mehr	4	3,6
Summe	256	255,9 \approx 256

Die Güte der Übereinstimmung kann auch formal mittels eines Tests geprüft werden. Der sog. Anpassungs-Test prüft, ob die beobachteten Häufigkeiten signifikant von einer theoretischen Verteilung abweichen oder nicht.

Zielstellung: Es soll eine beobachtete Häufigkeitsverteilung mit einer theoretischen verglichen werden.

H_0 : Die Beobachtete Verteilung entstammt der vermuteten theoretischen Verteilung

Alternative: Die Beobachtete Verteilung entstammt einer anderen als der vermuteten theoretischen Verteilung

B_k = beobachtete Häufigkeit in der k -ten Klasse

E_k = erwartete Häufigkeiten in der k -ten Klasse

c = Zahl der Klassen

α = vorgegebene Irrtumswahrscheinlichkeit

Voraussetzung: $E_k > 1$.

Berechne:

$$X_{Vers}^2 = \sum \frac{(B_k - E_k)^2}{E_k}$$

Bestimme:

$$X_{Tab}^2 = \chi_{1-\alpha; c-1-q}^2$$

wobei q die Zahl der geschätzten Parameter zur Berechnung von E_k ist.

$X_{Tab}^2 = \chi_{1-\alpha; c-1-q}^2$ ist der kritische Wert einer χ^2 -Verteilung mit $c-1-q$ Freiheitsgraden (Tab. VIII).

$$X_{Vers}^2 > X_{Tab}^2 \Rightarrow \text{verwerfe } H_0$$

$$X_{Vers}^2 \leq X_{Tab}^2 \Rightarrow \text{behalte } H_0 \text{ bei}$$

Beispiel: Leukozyten-Daten. Es soll zum Niveau $\alpha = 5\%$ geprüft werden, ob die Poisson-Verteilung ein geeignetes Modell für die Leukozyten-Daten darstellt.

Zählwert k	Beobachtete Häufigkeit B_k	Erwartete Häufigkeit E_k	$\frac{(B_k - E_k)^2}{E_k}$
0	158	155,3	0,0469
1	73	77,6	0,2727
2	21	19,4	0,1320
3 oder mehr	4	3,6	0,0444
Summe	256	255,9 \approx 256	$X_{Vers}^2 = 0,4960$

Wir haben $c = 4$ Klassen. Zur Berechnung von E_k mussten wir den Parameter λ schätzen (siehe oben), so dass $q = 1$ ist. Da $\alpha = 5\%$ ist, lesen wir den kritischen Wert $X_{Tab}^2 = \chi_{1-\alpha; c-1-q}^2 = \chi_{0,95;2}^2$ in Tab. VIII ab und finden:

$$X_{Tab}^2 = 5,991 > X_{Vers}^2 = 0,4960$$

\Rightarrow Behalte H_0 bei; beobachtete und erwartete Verteilung weichen nicht signifikant voneinander ab. Die beobachtete Verteilung spricht nicht gegen eine Poissonverteilung.

Beispiel: Nach der Mendelschen Unabhängigkeitsregel spalten zwei unabhängig vererbte Merkmale im Verhältnis 9 : 3 : 3 : 1 auf, sofern beide Merkmale dominant-rezessiv sind. Unabhängige Vererbung liegt z.B. dann vor, wenn die betreffenden Gene auf verschiedenen Chromosomen liegen. So sind z.B. bei der Erbse glatte Samen dominant über runzelige Samen und spalten nach dem Verhältnis 3 : 1 auf. Bei der Samenfarbe ist gelb dominant über grün und die Aufspaltung ist ebenfalls 3 : 1. Bei Betrachtung beider Merkmale ergibt bei Gültigkeit der Unabhängigkeitsannahme:

glatt + gelb	:	glatt + grün	:	runzelig + gelb	:	runzelig + grün
9	:	3	:	3	:	1

In einem Kreuzungsexperiment werden unter 1600 Pflanzen folgende Häufigkeiten beobachtet:

glatt + gelb	:	glatt + grün	:	runzelig + gelb	:	runzelig + grün
909	:	310	:	288	:	93

Es soll geprüft werden, ob die beobachtete Verteilung mit dem erwarteten Verhältnis 9 : 3 : 3 : 1 übereinstimmt.

Phänotyp	Beobachtete Häufigkeit B_k	Erwartete Häufigkeit E_k	$\frac{(B_k - E_k)^2}{E_k}$
glatt+gelb	909	$9/16 \cdot 1600 = 900$	0,09
glatt+grün	310	$3/16 \cdot 1600 = 300$	0,33
runzelig+gelb	288	$3/16 \cdot 1600 = 300$	0,48
runzelig+grün	93	$1/16 \cdot 1600 = 100$	0,49
Summe	1600	1600	$X_{Vers}^2 = 1,39$

Wir haben $c = 4$ Klassen. Zur Berechnung von E_k mussten wir keinen Parameter schätzen, weil wir hier vom Aufspaltungsverhältnis $9 : 3 : 3 : 1$ ausgehen; also ist $q = 0$. Da $\alpha = 5\%$ ist, lesen wir den kritischen Wert $X_{Tab}^2 = \chi_{1-\alpha; c-1-q}^2 = \chi_{0,95;3}^2$ in Tab. VIII ab und finden:

$$X_{Tab}^2 = 7,815 > X_{Vers}^2 = 1,39$$

\Rightarrow Behalte H_0 bei; beobachtete und erwartete Verteilung weichen nicht signifikant voneinander ab. Das Aufspaltungsverhältnis weicht nicht signifikant von $9 : 3 : 3 : 1$ ab. Die Unabhängigkeitsannahme kann aufrechterhalten werden.

Beispiel: Für ein Erdnussfeld wurde der Inokulumbesatz mit Mikrosklerotien des Pilzes *Cylindrosporium crotolariae* bestimmt. Hierzu wurde an 96 Stellen im Feld von je 531 cm^2 Boden eine Bodenprobe gezogen. In den Bodenproben wurde die Anzahl von Mikrosklerotien bestimmt. Die Ergebnisse der Auszählung sind in Tab. 5.5.1 wiedergegeben.

Tab. 5.5.1: Inokulumverteilung für *Cylindrocladium crotolariae* in einem Erdnussfeld (Hau FC, Campbell CL, Beute MK 1982 Plant Disease 66, 568-571)

k	B_k	k	B_k	k	B_k
0	2	10	5	20	2
1	5	11	9	21	1
2	5	12	6	22	2
3	13	13	2	23	0
4	5	14	2	24	0
5	6	15	0	25	0
6	9	16	1	26	1
7	6	17	1		
8	7	18	2		
9	4	19	0		

Die mittlere Zahl der Sklerotien beträgt $\hat{\lambda} = 7,99$. Dieser Wert bezieht sich auf die Fläche von 531 cm^2 , das sind $7,99 \text{ Mikrosklerotien}/531 \text{ cm}^2 = 0,01504 \cdot 10.000/\text{m}^2 \approx 150 \text{ Mikrosklerotien}/\text{m}^2$.

Es soll nun geprüft werden, ob die Verteilung dieser Daten einer Poissonverteilung folgt. Mit dem Schätzwert $\hat{\lambda} = 7,99$ berechnen wir die erwarteten Häufigkeiten E_k und stellen sie den beobachteten Häufigkeiten B_k gegenüber. Die Übereinstimmung wird mittels Anpassungstest geprüft. Die Häufigkeiten sind in den beiden folgenden Abbildungen dargestellt.

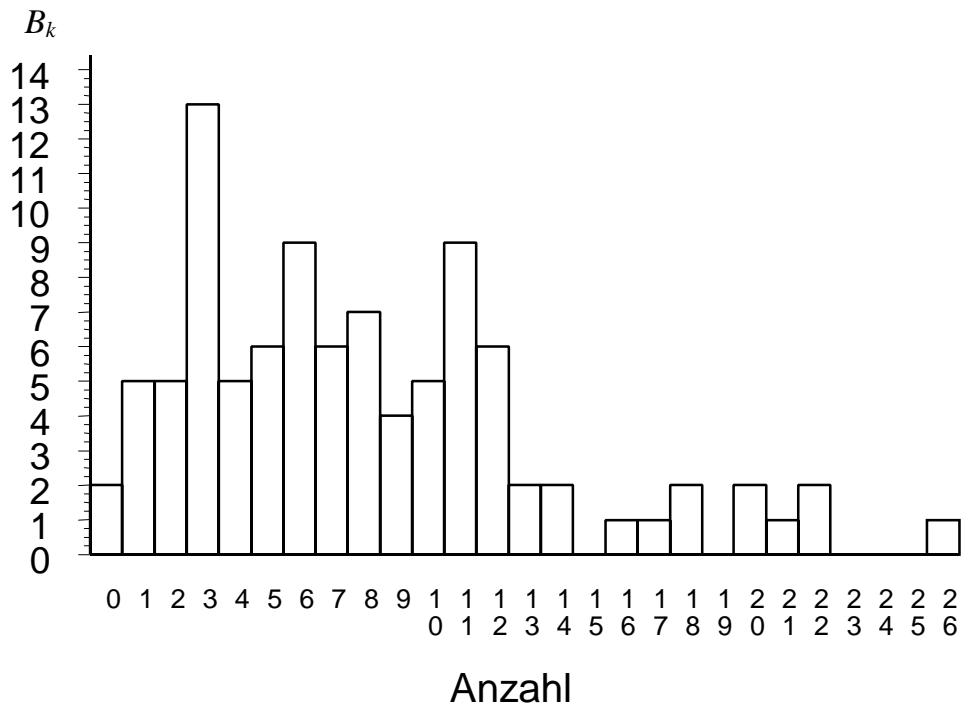


Abb. 5.5.1: Beobachtete Häufigkeiten (B_k) für Inokulumverteilung.

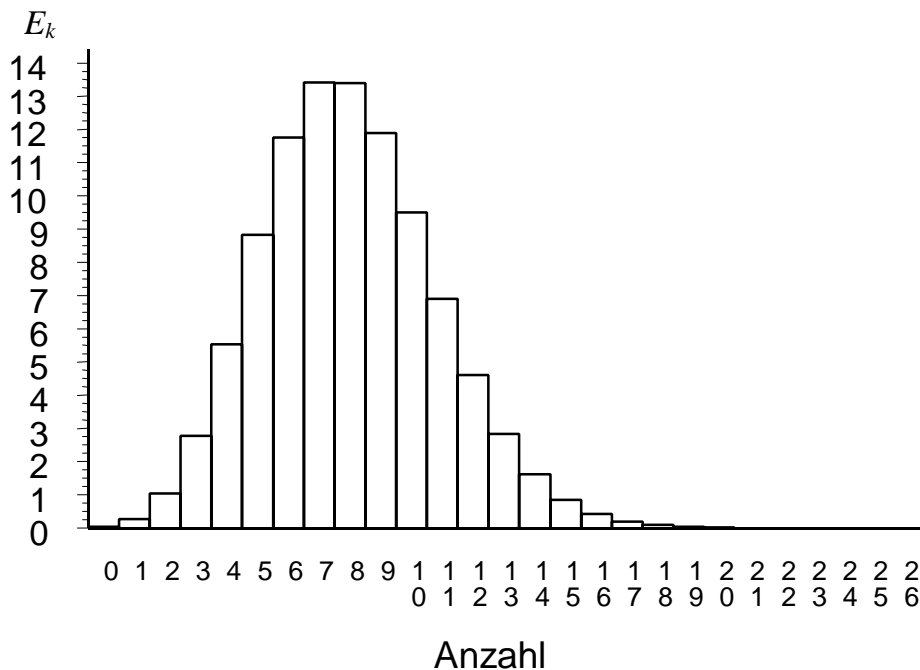


Abb. 5.5.2: Erwartete Häufigkeiten (E_k) für Inokulumverteilung, falls Poissonverteilung gilt ($\hat{\lambda}=7,99$).

Wir sehen, dass die erwartete Verteilung enger begrenzt ist, als die beobachtete. Im Vergleich zur Poissonverteilung liegt eine deutlich höhere Streuung vor. Man spricht auch von **Überdispersion**. Nun zur Berechnung des Anpassungstests. Die erwarteten Häufigkeiten in den Klassen $k < 3$ sowie in den Klassen $k > 11$ sind klein. Aus diesem Grund fassen wir die erwarteten und beobachteten Häufigkeiten dieser Klassen zusammen, so dass insgesamt 11 Klassen verbleiben ($c=11$). Es ist $q=1$, da ein Parameter (λ) geschätzt wurde. Die Details der Berechnung werden hier nicht ange-

geben. Wir finden $X_{Vers}^2 = 148,3$ und lesen $X_{Tab}^2 = \chi_{1-\alpha; c-1-q}^2 = \chi_{0,95;9}^2$ in Tab. VIII ab:

$$X_{Tab}^2 = 16,919 < X_{Vers}^2 = 148,3$$

Die Nullhypothese einer Poissonverteilung muss verworfen werden. Wir haben hier eine signifikante Überdispersion im Verhältnis zur Poissonverteilung.

Der biologische Grund für die Abweichung von der Poissonverteilung ist, dass das Inokulum nesterweise auftritt. Daher ist der Erwartungswert nicht an jeder Stelle im Feld gleich. In den Nestern ist der Erwartungswert größer als außerhalb der Nester. Es können hier andere Verteilungsmodelle angewendet werden, die das nesterweise Auftreten berücksichtigen, z.B. die sog. negative Binomialverteilung (Lorenz, 1996, siehe unter "Klumpenverteilungen").

5.5.1 Berechnung der erwarteten Häufigkeiten für eine Binomialverteilung

Im folgenden soll der Anpassungstest im Falle einer Binomialverteilung angewendet werden. Hierfür müssen Messwiederholungen mit demselben Stichprobenumfang n vorliegen, damit erwartete Häufigkeiten berechnet werden können.

Schätzen des Parameters p der Binomialverteilung bei Messwiederholungen

Es liegen Messwiederholungen bei gleichen Stichprobenumfang n vor.

(1) Ungruppierte Daten

n = Stichprobenumfang der Proben

h = Zahl der Proben

x_i = Zahl Stichprobenelemente mit einer bestimmten Eigenschaft in der i -ten Probe

p = Wahrscheinlichkeit, dass ein Element eine vorgegebene Eigenschaft hat

Parameterschätzung:

$$\hat{p} = \frac{x_1 + x_2 + \dots + x_h}{hn} = \frac{\sum_{i=1}^h x_i}{hn}$$

(2) Gegeben ist eine empirische Häufigkeitsverteilung

Zählwert k	Beobachtete Häufigkeit B_k
0	·
1	·
2	·
·	·
n	·
Berechne:	
$\hat{p} = \frac{\sum_{k=0}^n B_k k}{nh}$	
mit	
$h = \sum_{k=0}^n B_k$	

Beispiel: In einem Kartoffelfeld soll der Befall mit Kartoffenkäferlarven bestimmt werden. Hierzu werden 347 Kontrollpunkte ausgewählt. An jedem Punkt werden 20 Pflanzen auf Befall untersucht und die Zahl der befallenen Pflanzen bestimmt.

Tab. 5.5.1: Beobachtete Häufigkeitsverteilung (B_k) der Kontrollpunkte nach Anzahl von Pflanzen (k), die Befall durch Kartoffelkäferlarven aufweisen ($h = 347$, $n = 20$).

k	B_k	k	B_k	k	B_k
0	104	7	8	14	11
1	33	8	7	15	8
2	20	9	9	16	5
3	16	10	11	17	4
4	20	11	9	18	8
5	21	12	10	19	13
6	9	13	6	20	15

Die beobachtete Verteilung ist in Abb. 5.5.3 wiedergegeben. Es wird nun der mittlere Befall berechnet.

Berechne

$$\hat{p} = \frac{\sum B_k * k}{hn} = \frac{104*0 + 33*1 + 20*2 + \dots + 15*20}{20*347} = \frac{2073}{6940} = 0,2987 = 29,87\%$$

29,87% der untersuchten Pflanzen zeigten also einen Befall mit Kartoffelkäferlarven.

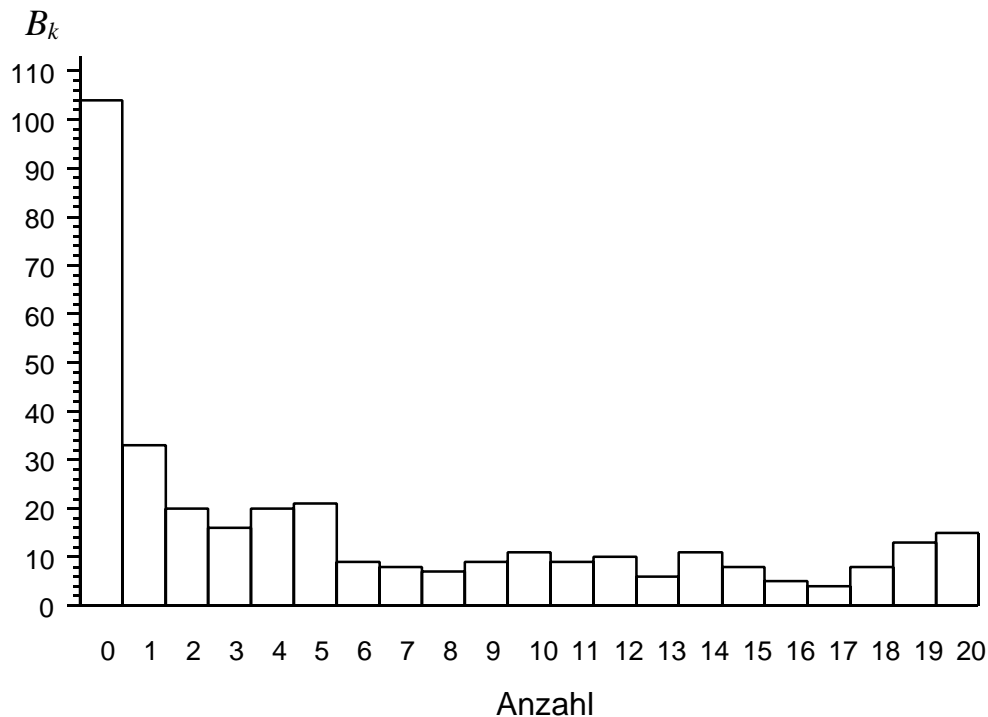


Abb.5.5.3: Beobachtete Häufigkeiten (B_k) der Anzahl von mit Kartoffelkäferlarven befallener Pflanzen in $h = 347$ Stichproben von je $n = 20$ Pflanzen.

Wir wollen nun die erwartete Verteilung berechnen, falls die Binomialverteilungsannahme zutrifft.

Erwartete Häufigkeiten einer Binomial-Verteilung:

$$E_k = h \cdot P(X = k|n; p)$$

$$h = \text{Summe aller beobachteten Häufigkeiten} = \sum_{k=0}^n B_k$$

$P(X = k|n; p)$ = Binomialwahrscheinlichkeiten (siehe Abschnitt 5.3)

Rekursionsformel:

(1) Berechne

$$E_0 = h \cdot P(X = 0|n; p) = h \cdot (1 - p)^n$$

(2) Berechne

$$E_k = \frac{n - k + 1}{k} \times \frac{p}{1 - p} E_{k-1} \quad \text{für } k = 1, 2, \dots, n$$

Praktischer Hinweis: Wenn E_k klein wird, sollten die betreffenden Klassen zusammengefasst werden (Linder und Berchtold, 1979). Falls dieser Fall eintritt, sind dies die niedrigen Klassen ab $k = 0$ und die hohen Klassen bis $k = n$.

Beispiel: Wir berechnen die erwarteten Häufigkeiten für die Kartoffelkäferlarven-Daten.

$$h = 347$$

$$n = 20$$

$$\hat{p} = 0,2987$$

$$E_0 = h \cdot P(X=0|n; p) = h \cdot (1 - p)^n = 347 \cdot (1 - 0,2987)^{20} = 0,2873$$

$$E_1 = \frac{20}{1} \times \frac{0,2987}{0,7013} \cdot 0,2873 = 2,4476$$

$$E_2 = \frac{19}{2} \times \frac{0,2987}{0,7013} \cdot 2,4476 = 9,9037$$

$$E_3 = \frac{18}{3} \times \frac{0,2987}{0,7013} \cdot 9,9037 = 25,3096$$

usw. bis E_{20} .

k	B_k		E_k		$(B_k - E_k)^2/E_k$
0	104	137	0,2873	2,7349	6591,51
1	33		2,4476		
2	20		9,9037		
3	16		25,3096		10,29
4	20		45,8155		3,42
5	21		62,4454		14,55
6	9		66,4934		27,51
7	8		56,6430		49,71
8	7		39,2047		41,77
9	9		22,2646		26,45
10	11		10,4315		7,90
11	9		4,0392		0,03
12	10	80	1,2903	1,7146	3574,36
13	6		0,3382		
14	11		0,0720		
15	8		0,0123		
16	5		0,0016		
17	4		0,0002		
18	8		0,0000		
19	13		0,0000		
20	15		0,0000		

$$\text{Summe: } X_{Vers}^2 = 10353,61$$

Die erwarteten Häufigkeiten sind in Abb. 5.5.4 dargestellt.

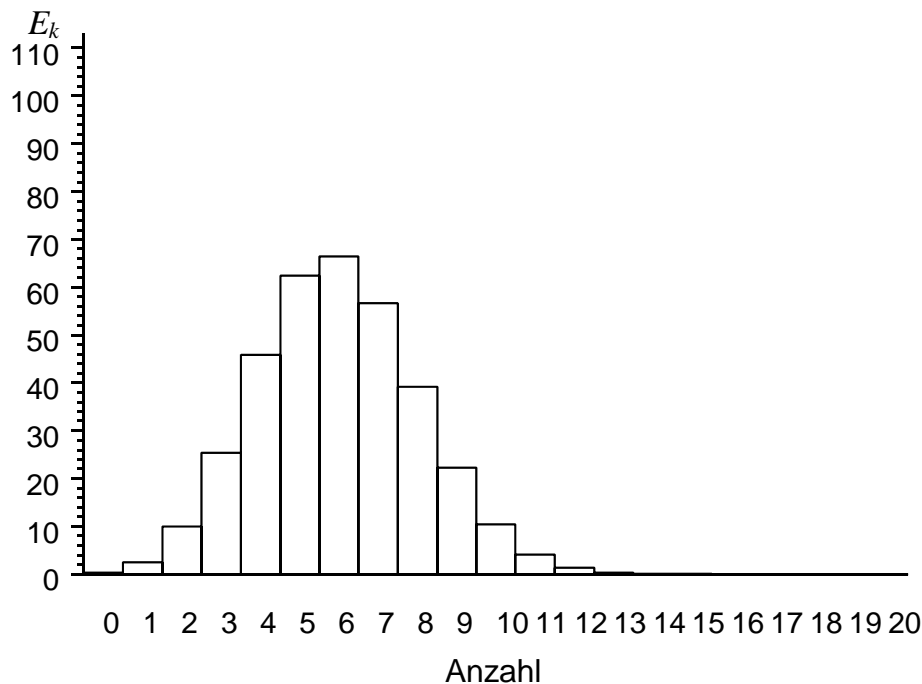


Abb. 5.5.4: Erwartete Häufigkeitsverteilung (E_k), wenn Kartoffenkäfer-Daten binomialverteilt wären mit Erwartungswert 0,2987.

Wir sehen auch hier, dass die erwartete Verteilung enger ist als die beobachtete. Wir sprechen auch hier von **Überdispersion**. Der Grund ist, ähnlich wie bei den Inokulum-Daten, dass die Käferlarven nesterweise auftreten. Aus diesem Grund ist die Binomialwahrscheinlichkeit bzw. die Befallswahrscheinlichkeit nicht konstant über das ganze Feld. In Befallsnestern ist die Wahrscheinlichkeit überdurchschnittlich, zwischen den Nestern dagegen ist sie unterdurchschnittlich.

Wir führen nun noch den Anpassungs-Test durch. Da einige E_k Werte < 1 sind, fassen wir die Klassen wie in der Tabelle gezeigt zusammen und berechnen

$$X_{Vers}^2 = 10353,61$$

Es liegen $c=12$ Klassen vor. Ein Parameter (p) musste geschätzt werden, so dass $q=1$. Wir lesen $X_{Tab}^2 = \chi_{1-\alpha; c-1-q}^2 = \chi_{0,95;10}^2$ in Tab. VIII ab und finden:

$$X_{Tab}^2 = 18,307 < X_{Vers}^2 = 10353,61$$

Die Nullhypothese einer Binomialverteilung muss verworfen werden.

5.5.2 Was tun bei Überdispersion?

Wir haben zwei Beispiele gesehen, bei denen Überdispersion durch nesterweises Auftreten einer Krankheit bzw. eines Schädlings auftrat. Dieses Phänomen ist sehr typisch für sehr viele Daten, die in Form von Zählwerten erhoben werden. Dies trifft in besonderem Maße für Feld- und Gewächshausversuche sowie für Untersuchungen mit Tieren zu. Das eine untersuchte Beispiel betraf Zählwerte, die nach oben unbe-

grenzt sind (Inokulum bei Erdnuss), so dass zunächst eine Poissonverteilung als sinnvolles Verteilungsmodell in Frage kam. Das andere betraf Zählwerte, die nach oben begrenzt sind, weil je Probefläche eine bestimmte Zahl von Pflanzen untersucht wurde. Hier war zunächst an eine Binomialverteilung zu denken. In beiden Fällen stellten wir aber eine deutliche Überdispersion fest. Jedes Ergebnis eines Tests oder eines Vertrauensintervalls, dass nun diese spezifischen Verteilungsannahmen macht (Poisson- bzw. Binomialverteilung), ist fragwürdig, da die Verteilungsannahmen verletzt sind. Dies ist vor allem deswegen der Fall, weil viele der in Kap. 5 besprochenen Verfahren relativ empfindlich auf Abweichungen vom unterstellten Verteilungsmodell reagieren.

Zur Auswertung von Daten, die Überdispersion aufweisen, gibt es verschiedene Ansätze. Hier soll nur der einfachste genannt werden. Er besteht darin, die Zählwerte so auszuwerten, als wären sie normalverteilt. Zur Anwendung können alle Verfahren kommen, die in den Kapiteln 3 und 4 besprochen wurden. Diese Verfahren haben sich als relativ robust gegenüber Abweichungen von der Annahme der Normalverteilung erwiesen. Dies ist ein großer Vorteil gegenüber den in Kapitel 5 besprochenen Verfahren. Voraussetzung ist, dass der Stichprobenumfang nicht zu klein ist. Um die Voraussetzungen der Normalverteilung und Varianzhomogenität (jede Behandlung hat dieselbe Varianz) besser zu erfüllen, müssen die Daten oft transformiert werden. Hierbei sind z.B. folgende Transformationen gebräuchlich:

Eine Transformation für Prozentzahlen:

$$y = \sin^{-1}[\sqrt{(x/100)}] \quad (\text{Arcus-Sinus-Wurzel Transformation; Winkel-Transformation})$$

Beispiele:

$$\begin{aligned} x = 40\%; & y = \sin^{-1}[\sqrt{(40/100)}] = \sin^{-1}[\sqrt{(0,4)}] = 0,685 \\ x = 5\%; & y = \sin^{-1}[\sqrt{(0,05)}] = 0,332 \end{aligned}$$

Eine Transformation für Zählwerte:

$$y = \log_{10}(x)$$

Beispiele:

$$\begin{aligned} x = 100; & y = 2 \\ x = 10; & y = 1 \\ x = 1; & y = 0 \end{aligned}$$

Alternativ kann auch der natürliche Logarithmus berechnet werden. Hierdurch ergeben sich zwar andere Werte, aber die Verhältnisse bleiben gleich.

$$\begin{aligned} x = 100; & y = \log_e(100) = 4,6052 \\ x = 10; & y = \log_e(10) = 2,3026 \\ x = 1; & y = \log_e(1) = 0 \end{aligned}$$

Man transformiert die Daten und analysiert die transformierten Werte nach Standard-

verfahren, z.B. mit dem t -Test oder einer Varianzanalyse. Ergeben sich für die transformierten Daten signifikante Unterschiede, kann auf Unterschiede für die untransformierten Daten zurückgeschlossen werden. Dies ist zulässig, wenn man eine sog. monotone Transformation verwendet, und die oben genannten Beispiele haben diese Eigenschaft. Für monotone Transformationen gilt, dass der größte transformierte Wert zum größten untransformierten Wert korrespondiert, der zweitkleinste transformierte Wert korrespondiert zum zweitkleinsten, etc. Mathematisch ausgedrückt gilt für eine monotone Transformation $y = f(x)$:

$$x_1 > x_2 \Rightarrow f(x_1) > f(x_2)$$

Beispiel:

$$x_1 = 10 > x_2 = 1 \Rightarrow \log_{10}(x_1) = 1 > \log_{10}(x_2) = 0$$

5.6 Test auf Unabhängigkeit in der 2×2 Feldertafel (4-Feldertafel)

5.6.1 Test bei großen Stichproben

Wenn in einer Untersuchung an jeder Untersuchungseinheit zwei verschiedene kategoriale Merkmale erhoben werden, die jeweils zwei Ausprägungen haben (dichotome Merkmale), kann die Frage gestellt werden, ob zwischen den Merkmalen ein Zusammenhang besteht. Dabei ist es hilfreich, die Zählwerte in einer **2×2 Feldertafel** anzuordnen, die folgende Struktur hat:

Es liegen gepaarte Beobachtungen für zwei dichotome Merkmale auf jeder Versuchseinheit vor.

A_1, A_2 = Ausprägungen des ersten Merkmals

B_1, B_2 = Ausprägungen des zweiten Merkmals

a = Häufigkeit der Kombination A_1, B_1

b = Häufigkeit der Kombination A_1, B_2

c = Häufigkeit der Kombination A_2, B_1

d = Häufigkeit der Kombination A_2, B_2

$n = a + b + c + d$ = Stichprobenumfang

p_1 = Wahrscheinlichkeit des Auftretens von A_1

p_2 = Wahrscheinlichkeit des Auftretens von B_1

Die Häufigkeiten können in eine 2×2 Kreuztabelle eingetragen werden:

		Ausprägung Merkmal B		
		B ₁	B ₂	
Ausprägung Merkmal A	A ₁	<i>a</i>	<i>b</i>	<i>a+b</i>
	A ₂	<i>c</i>	<i>d</i>	<i>c+d</i>
		<i>a+c</i>	<i>b+d</i>	<i>n</i>

Wir hatten diese Struktur bereits im Zusammenhang mit dem Vergleich von zwei Binomialwahrscheinlichkeiten in verbundenen Stichproben mittels des McNemar-Test kennengelernt (5.3.5). Dort ging es darum zu prüfen, ob $P(A_1) = P(B_1)$ ist oder nicht. Hier geht es dagegen um die Frage, ob die Merkmale A und B unabhängig sind.

Beispiel (Linder & Berchtold, 1979): Anzahl der von Typhus befallenen Personen ohne und mit Impfung.

		Gesundheitszustand		
		erkrankt	gesund	
Impfung?	Geimpft	56	6759	6815
	Nicht geimpft	272	11396	11668
		328	18155	18483

Eine sinnvolle Frage ist hier, ob ein Zusammenhang zwischen Gesundheitszustand und Impfung besteht. (Es macht dagegen wenig Sinn, zu fragen, ob der Anteil geimpfter Personen gleich dem Anteil gesunder Personen ist, eine Frage, die mit dem McNemar-Test zu beantworten wäre, und die sich auf die **Randverteilungen** bezieht. Dies ist keine sinnvolle Nullhypothese, es sei denn, man ginge z.B. davon aus, dass alle geimpften Personen nicht erkranken, und alle nicht geimpften erkranken. Dies ist aber offensichtlich nicht der Fall, wie ein Blick auf die Kreuztabelle zeigt).

Bei der obigen Datenstruktur ist einzig der Stichprobenumfang n vorgegeben. Wir können nun zunächst jedes Merkmal für sich betrachten. Im Beispiel können wir fragen, mit welcher Wahrscheinlichkeit eine Person in der Stichprobe geimpft ist. Diese Wahrscheinlichkeit ist

$$p_1 = P(\text{geimpft}) \text{ .}$$

Ebenso können wir nach der Wahrscheinlichkeit fragen, mit der eine zufällig ausgewählte Person an Typhus erkrankt. Diese können wir schreiben als

$$p_2 = P(\text{erkrankt}) \quad .$$

Nun können wir nach der gemeinsamen Wahrscheinlichkeit fragen, dass eine zufällig ausgewählte Person geimpft ist und trotzdem erkrankt:

$$P(\text{geimpft und erkrankt}) \quad .$$

Wenn nun Impfen und Gesundheitszustand voneinander unabhängig sind, das eine also das andere nicht beeinflusst, so können wir diese Wahrscheinlichkeit aus denen für die beiden einzelnen Merkmale berechnen, indem wir den Multiplikationssatz der Wahrscheinlichkeitsrechnung (Produktregel) anwenden (Abschnitt 5.2):

$$P(\text{geimpft und erkrankt}) = P(\text{geimpft}) \cdot P(\text{erkrankt}) = p_1 \cdot p_2 \quad .$$

Die erwartete Zahl von Personen, die geimpft und krank sind, ist dann

$$n \cdot P(\text{geimpft und erkrankt}) \quad .$$

Um diese Formel anzuwenden, müssen wir die Wahrscheinlichkeiten p_1 und p_2 schätzen. Dies tun wir einfach anhand der Randhäufigkeiten.

Schätzung der Randwahrscheinlichkeiten:

$$\hat{p}_1 = \frac{a+b}{n} \quad ; \quad \hat{p}_2 = \frac{a+c}{n}$$

Beispiel:

$$\hat{p}_1 = \hat{P}(\text{geimpft}) = \frac{a+b}{n} = \frac{6815}{18483} = 0,3687$$

$$\hat{p}_2 = \hat{P}(\text{krank}) = \frac{a+c}{n} = \frac{328}{18483} = 0,01775$$

Falls nun die Nullhypothese der Unabhängigkeit zutrifft, so ist

$$\hat{P}(\text{geimpft und krank}) = \hat{P}(\text{geimpft}) \cdot \hat{P}(\text{krank}) = 0,0006543$$

und

$$n \cdot \hat{P}(\text{geimpft und krank}) = 18483 \cdot 0,0006543 = 120,93$$

Bei Unabhängigkeit würden wir also etwa 121 Personen erwarten, die geimpft und krank sind. Beobachtet haben wir aber nur 56, was eine merkliche Diskrepanz darstellt. Dies ist ein Hinweis darauf, dass die Impfung die Erkrankungswahrscheinlichkeit reduziert.

Ob die Abweichung so groß ist, dass die Nullhypothese der Unabhängigkeit verworfen werden muss, können wir hier wieder mit einem Anpassungstest wie in Abschnitt

5.5 prüfen. Hierzu müssen wir für jede Zelle die erwarteten Häufigkeiten unter der Nullhypothese der Unabhängigkeit schätzen.

Klasse	$B(\text{eobachtet})$	$E(\text{rwartet})$	Schätzung von E
A_1 und B_1	a	$n \cdot p_1 \cdot p_2$	$\frac{(a+b)(a+c)}{n}$
A_1 und B_2	b	$n \cdot p_1 \cdot (1-p_2)$	$\frac{(a+b)(b+d)}{n}$
A_2 und B_1	c	$n \cdot (1-p_1) \cdot p_2$	$\frac{(c+d)(a+c)}{n}$
A_2 und B_2	d	$n \cdot (1-p_1) \cdot (1-p_2)$	$\frac{(c+d)(b+d)}{n}$

Es lässt sich zeigen, dass die Statistik des Anpassungstests in diesem speziellen Fall die folgende Form hat:

$$X_{\text{Vers}}^2 = \sum \frac{(B-E)^2}{E} = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

Bleibt noch die Frage der Freiheitsgrade. Um die erwarteten Häufigkeiten (E) zu berechnen, müssen hier zwei Parameter geschätzt werden, nämlich p_1 und p_2 . Da wir 4 Klassen haben, hat die χ^2 -Statistik $4 - 2 - 1 = 1$ Freiheitsgrad. Nun können wir den Test durchführen.

H_0 : Die dichotomen Merkmale A und B sind unabhängig

Alternative: Es besteht eine Abhängigkeit

Voraussetzung: $n > 200$ und alle Erwartungswerte > 1

Berechne:

$$X_{\text{Vers}}^2 = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

Bestimme:

$$X_{\text{Tab}}^2 = \chi_{1-\alpha;1}^2 \quad (\text{Tab. VIII})$$

$$X_{\text{Vers}}^2 > X_{\text{Tab}}^2 \Rightarrow \text{verwerfe } H_0$$

$$X_{\text{Vers}}^2 \leq X_{\text{Tab}}^2 \Rightarrow \text{behalte } H_0 \text{ bei}$$

Beispiel: Impfung gegen Typhus. Wir wollen zum Niveau $\alpha = 5\%$ einen Test auf Unabhängigkeit durchführen. Für die angegebenen Daten finden wir

$$X_{Vers}^2 = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)} = \frac{(56 \cdot 11396 - 6759 \cdot 272)^2 18483}{(6815)(11668)(328)(18155)} = 56,2341$$

$$X_{Tab}^2 = \chi_{0,95;1}^2 = 3,841 < X_{Vers}^2 \Rightarrow \text{verwerfe } H_0$$

Es besteht eine Abhängigkeit von Impfung und Erkrankung. Die Impfung reduziert das Erkrankungsrisiko.

Der Test ist anwendbar, weil der Stichprobenumfang sehr groß ist ($n = 18483 > 200$) und die Erwartungswerte alle > 1 sind:

Klasse	B(eobachtet)	Schätzung von E
geimpft, krank	$a = 56$	$\frac{(a+b)(a+c)}{n} = \frac{6815 \cdot 328}{18483} = 120,9$
geimpft, gesund	$b = 6759$	$\frac{(a+b)(b+d)}{n} = \frac{6815 \cdot 18155}{18483} = 6694,1$
nicht geimpft, krank	$c = 272$	$\frac{(c+d)(a+c)}{n} = \frac{11668 \cdot 328}{18483} = 207,1$
nicht geimpft, gesund	$d = 11396$	$\frac{(c+d)(b+d)}{n} = \frac{11668 \cdot 18155}{18483} = 11460,9$

Bemerkung: Wir könnten den Test auf Unabhängigkeit auch umformulieren in einen Test zum Vergleich von zwei Binomialwahrscheinlichkeiten. Wir würden hierbei die Nullhypothese testen, dass die Erkrankungswahrscheinlichkeit mit Impfung genau so groß ist wie ohne Impfung. Diese Nullhypothese können wir mit dem Test in Abschnitt 5.3.5 prüfen. Es kommt dabei exakt dasselbe Ergebnis heraus, wie mit dem χ^2 -Test. Es kann sogar gezeigt werden, dass $X_{Vers}^2 = z_{Vers}^2$. Hierzu setzen wir $x_1 = a$, $n_1 = a + b$ und $b = n_1 - x_1$ sowie $x_2 = c$, $n_2 = c + d$ und $d = n_2 - x_2$. Hiermit lässt sich die Chi-Quadrat-Statistik des Tests auf Unabhängigkeit in der 2 x 2 Tafel umformen zu

$$X_{Vers}^2 = \frac{(ad-bc)^2 n}{(a+c)(b+d)(a+b)(c+d)} = \frac{[x_1(n_2 - x_2) - x_2(n_1 - x_1)]^2 (n_1 + n_2)}{(x_1 + x_2)(n_1 + n_2 - x_1 - x_2)n_1 n_2} =$$

$$\frac{\left(\frac{x_1 n_2}{n_1 n_2} - \frac{x_2 n_1}{n_1 n_2} \right)^2 (n_1 n_2)^2 (n_1 + n_2)}{\frac{x_1 + x_2}{n_1 + n_2} \frac{n_1 + n_2 - x_1 - x_2}{n_1 + n_2} n_1 n_2 (n_1 + n_2)^2} = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = z_{Vers}^2,$$

wobei $\hat{p}_1 = x_1 / n_1$, $\hat{p}_2 = x_2 / n_2$, $\hat{p} = (x_1 + x_2) / (n_1 + n_2)$ und z_{Vers} die Test-Statistik zum Vergleich zweier Binomialverteilungen in Abschnitt 5.3.4 ist.

5.6.2 Die Yates Korrektur

In kleinen Stichproben kann die Test-Statistik nur ganz bestimmte Werte annehmen. Daher ist die Approximation der Prüfverteilung durch eine χ^2 -Verteilung schlecht, wenn die Stichprobe klein ist. Eine verbesserte Approximation ergibt sich durch die sog. Kontinuitäts-Korrektur von Yates. Diese sollte berücksichtigt werden, falls $20 < n < 200$ ist.

Kontinuitätskorrektur von Yates:

Voraussetzungen: $200 \geq n > 20$; Erwartungswerte > 1

$$\text{Berechne: } X_{Vers}^2 = \frac{(|ad - bc| - n/2)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

Hypothesen und weiteres Vorgehen wie in 5.6.1.

Beispiel (Weber, 1980, S. 205): Ein Versuch zur Ermittlung des Zusammenhanges zwischen Wasserart und Keimung bei *Primula sinensis* wurde mit 32 Pflanzen durchgeführt. Es ergab sich folgendes Ergebnis:

Wasser	Keimverhalten		Alle Fälle
	nicht gekeimt	gekeimt	
Lehmwasser	5	9	14
Regenwasser	15	3	18
Insgesamt	20	12	32

$$X_{Vers}^2 = \frac{\left(|ad - bc| - \frac{n}{2}\right)^2 n}{(a+b)(c+d)(a+c)(b+d)} = \frac{\left(|5 \cdot 3 - 15 \cdot 9| - \frac{32}{2}\right)^2 \cdot 32}{14 \cdot 18 \cdot 20 \cdot 12} = 5,7228$$

$X_{Tab}^2 = 3,841 < X_{Vers}^2 = 5,7228 \Rightarrow$ Verwerfe H_0 . Es gibt einen signifikanten Zusammenhang zwischen Keimverhalten und Wasserart. Die Keimung ist in Lehmwasser signifikant höher als in Regenwasser.

5.6.3 Fishers exakter Test

Bei sehr kleinen Stichproben ($n \leq 20$) und/oder wenn eine erwartete Häufigkeit < 1 ist, sollte Fishers exakter Test angewendet werden.

Die Idee des Tests ist es, die Randverteilung, also die Randhäufigkeiten, als gegeben zu betrachten. Man spricht von einem bedingten (konditionalen) Test. Unter der Voraussetzung gegebener Randhäufigkeiten können wir die Frage stellen, wie wahrscheinlich es ist, dass a einen bestimmten Wert annimmt. Es reicht, hierbei eine

der vier Zellen zu betrachten. Denn wenn die Randhäufigkeiten bekannt sind, können mit einem vorgegebenen Wert für a alle anderen Häufigkeiten b , c , und d berechnet werden.

Es kann gezeigt werden, dass die Verteilung von a bei gegebenen Randhäufigkeiten **unter der Nullhypothese** der Unabhängigkeit einer **hypergeometrischen Verteilung** folgt. Die Wahrscheinlichkeit für einen bestimmten Wert a ist gegeben durch

$$P(a) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

Bei Fishers exaktem Test wird von der gegebenen Konfiguration der 2×2 Tafel ausgegangen und dann nach extremeren, also unwahrscheinlicheren Konfigurationen gesucht, welche dieselben Randsummen aufweisen.

Beispiel: Heilungserfolg von zwei Behandlungen in einem klinische Versuch (Lienert et al.).

Behandlung	Ergebnis der Behandlung		Alle Fälle
	Erfolg	Kein Erfolg	
X	4	1	5
Y	3	7	10
Insgesamt	7	8	15

Wir wollen prüfen, ob der Behandlungserfolg von der Behandlung abhängt. Mit $a = 4$, $b = 1$, $c = 3$ und $d = 7$ können wir zunächst die Wahrscheinlichkeit für die vorliegende Tabelle unter der Nullhypothese nach der hypergeometrischen Verteilung berechnen.

$$P(\text{Daten}) = P(a=4) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} = \frac{5!10!7!8!}{15!4!1!3!7!} = 0,09324$$

Dann schauen wir uns alle anderen möglichen Konfigurationen an, die bei den gegebenen Randsummen möglich sind. **Um zu überlegen, welche Möglichkeiten es gibt, suchen wir die kleinste Randsumme.** Diese ist $(a+b) = 5$. Da bei allen Konfigurationen die Randsumme dieselbe sein soll, ist klar, dass a bei gegebenen Randsummen nur zwischen 0 und 5 liegen kann. Für diese 6 Möglichkeiten berechnen wir $P(a)$. **(Wenn die kleinste Randsumme a nicht enthält, vertauschen wir einfach die Spalten der Tabelle und wenden dann dasselbe Verfahren an!).**

Konfiguration		$P(a)$
5	0	0,00699
2	8	
4	1	0,09324 = $P(\text{Daten})$
3	7	
3	2	0,32634
4	6	
2	3	0,39161
5	5	
1	4	0,16317
6	4	
0	5	0,01865
7	3	

$P(a)$ kursiv: $P(a) \leq P(\text{Daten})$

Danach addieren wir die Wahrscheinlichkeiten aller Konfigurationen, für die $P(a) \leq P(\text{Daten})$ ist. Dies ist dann die Wahrscheinlichkeit, unter der Nullhypothese eine so extreme Konfiguration zu beobachten, wie die vorliegende, oder eine noch extremere. Wir finden

$$p_{Vers} = \sum_{P(a) \leq P(\text{Daten})} P(a) = 0,00699 + 0,09324 + 0,01865 = 0,11888$$

Die vorliegenden Daten oder extremere sind unter der Nullhypothese nicht unwahrscheinlich. Daher ist die Nullhypothese plausibel, angesichts der beobachteten Daten. Es ist daher sinnvoll, die Nullhypothese beizubehalten. Testen wir nun bei einem Signifikanzniveau von $\alpha = 5\% = 0,05$, so ist

$$p_{Vers} > \alpha$$

Daher behalten wir die Nullhypothese bei.

Fishers exakter Test

H_0 : Es besteht kein Zusammenhang zwischen den Merkmalen A und B.

Voraussetzung: Die Daten entstammen einer einfachen Zufallsstichprobe. Der Test sollte anstelle der χ^2 -Test angewendet werden, wenn $n \leq 20$ ist. Eine Anwendung für größere Stichproben ist auch möglich, aber der Rechenaufwand wird größer.

α = vorgegebene Irrtumswahrscheinlichkeit

(1) Prüfe, ob die kleinste Randsumme entweder in der ersten Zeile oder in der ersten Spalte auftritt. Ist dies nicht der Fall, werden die beiden Zeilen bzw. die beiden Spalten der 2 x 2 Tafel vertauscht, so dass die kleinste Randsumme in der ersten Zeile bzw. Spalte auftritt.

(2) Bestimme alle möglichen Konfigurationen der 2 x 2 Tabelle mit denselben Randsummen wie bei den vorliegenden Daten. Die Häufigkeit a in der ersten Zeile und Spalte nimmt hierbei Werte zwischen 0 und der kleinsten Randsumme an.

(3) Berechne für jede Konfiguration:

$$P(a) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

Für die vorliegenden Daten bezeichne man diese Wahrscheinlichkeit mit $P(\text{Daten})$.

(4) Berechne:

$$p_{Vers} = \sum_{P(a) \leq P(\text{Daten})} P(a)$$

$p_{Vers} \geq \alpha \Rightarrow$ behalte H_0 bei

$p_{Vers} < \alpha \Rightarrow$ verwirfe H_0

Beispiel (Weber, 1980, S. 205): Ein Versuch zur Ermittlung des Zusammenhanges zwischen Wasserart und Keimung bei *Primula sinensis* wurde mit 32 Pflanzen durchgeführt. Dieses Beispiel hatten wir schon zur Illustration der Yates-Korrektur betrachtet, die hier anwendbar ist und natürlich weniger Rechenaufwand bedeuten würde. Wir führen hier zur Erläuterung trotzdem den exakten Test durch.

Wasser	Keimverhalten		Alle Fälle
	nicht gekeimt	gekeimt	
Lehmwasser	5	9	14
Regenwasser	15	3	18
Insgesamt	20	12	32

Die kleinste Randsumme ist 12. Diese enthält aber nicht a . Daher tauschen wir die Spalten, damit unser Vorgehen von oben anwendbar ist.

Wasser	Keimverhalten		Alle Fälle
	gekeimt	nicht gekeimt	
Lehmwasser	9	5	14
Regenwasser	3	15	18
Insgesamt	12	20	32

Jetzt umfasst die kleinste Randsumme a . Wir sehen, dass a bei festen Randsummen die Werte von 0 bis 12 annehmen kann. Also gibt es insgesamt 13 verschiedene Konfigurationen. Wir bestimmen diese und berechnen $P(a)$.

Konfiguration		$P(a)$
0	14	<i>0,00008</i> ←
12	6	
1	13	<i>0,00197</i> ←
11	7	
2	12	0,01764
10	8	
3	11	0,07838
9	9	
4	10	0,19399
8	10	
5	9	0,28217
7	11	
6	8	0,24690
6	12	
7	7	0,13023
5	13	
8	6	0,04070
4	14	
9	5	0,00724 = P(Daten)
3	15	
10	4	<i>0,00068</i> ←
2	16	
11	3	<i>0,00003</i> ←
1	17	
12	2	<i>0,00000</i> ←
0	18	

$P(a)$ kursiv: $P(a) \leq P(\text{Daten})$

$$p_{Vers} = \sum_{P(a) \leq P(\text{Daten})} P(a) = 0,00000 + 0,00003 + 0,00068 + 0,00724 \\ + 0,00197 + 0,00008 = 0,0100$$

Man beachte, dass alle addierten Wahrscheinlichkeiten kleiner sind als $P(\text{Daten})$.
 $p_{Vers} < \alpha = 0,05 \Rightarrow H_0$ wird verworfen. Es gibt einen signifikanten Zusammenhang zwischen Keimverhalten und Wasserart. Die Keimung ist in Lehmwasser signifikant höher als in Regenwasser.

Zum Vergleich: Mit der Yates-Korrektur hatten wir $X_{Vers}^2 = 5,7228$ gefunden. Hierzu kann auch ein p-Wert berechnet werden. Mit dem Paket SAS finden wir: $P(\chi^2 > X_{Vers}^2 = 5,7228) = 0,01675$, was etwas vom exakten Wert mit Fisher's Test abweicht. Allerdings sind beide Tests signifikant.

Bemerkung: Die Durchführung des Tests von Hand ist sehr mühsam. In der Praxis wird man die Rechnung mit einem PC machen. Viele Statistik-Pakete bieten den Fisher-Exakt-Test an.

Der oben beschriebene Test ist zweiseitig. Manche Statistik-Pakete berechnen auch einen sog. einseitigen Test (siehe Abschnitt 3.19). Eine einseitige Fragestellung liegt nur in sehr wenigen Fällen vor. Daher sollte in der Regel der zweiseitige Test verwendet werden. Auf den einseitigen Fall gehen wir nicht näher ein. Der Output der Berechnung mit dem Statistik-Paket SAS (PROC FREQ) ist wie folgt:

STATISTICS FOR TABLE OF WASSER BY KEIMUNG

Statistic	DF	Value	Prob

Fisher's Exact Test (Left)			0.999
(Right)			7.94E-03
(2-Tail)			1.00E-02
Sample Size = 32			

Unter "2-Tail" finden wir "Prob=1.00E-02", den p-Wert für unseren zweiseitigen Test. Dies ist die wissenschaftliche Notation für kleine Zahlen. "1,00E-2" bedeutet so viel wie $1,00 \cdot 10^{-2} = 0,0100$. Also finden wir $p_{Vers} = 0,0100$.

Zur hypergeometrischen Verteilung

Abschließend soll hier noch erklärt werden, wie es zur Formel für $P(a)$ kommt, d.h., warum a unter der Nullhypothese einer hypergeometrischen Verteilung folgt. Hierzu werden zwei alternative Erklärungen angeboten.

Erklärung 1: Wir betrachten das Beispiel über den Zusammenhang zwischen Medikament (X oder Y) und Heilungserfolg (ja oder nein) anhand eines Urnenmodells, welches die festen Randsummen berücksichtigt. In der Urne seien insgesamt n Kugeln = Patienten. Hiervon sind

m_1 Zahl der Patienten (Kugeln) mit Heilungserfolg
 $m_2 = n - m_1$ Zahl der Patienten (Kugeln) ohne Heilungserfolg

Diese vorgegebenen Zahlen entsprechen den fixierten Randsummen für den Heilungserfolg. Nun werden k_1 Kugeln zufällig aus der Urne gezogen. Wir legen fest, dass

k_1 = Zahl der Patienten (Kugeln), die Medikament X bekommen
 $k_2 = n - k_1$ = Zahl der Patienten (Kugeln), die Medikament Y bekommen

Unter den k_1 Patienten, die Medikament X erhalten, seien a gesund. Wenn Medikamentenwahl und Heilungserfolg unabhängig sind, hängt die Wahrscheinlichkeit, dass a der gezogenen k_1 Patienten gesund sind, nur von der Zahl m_1 der Patienten in der Urne ab, die gesund sind. Wir müssen hierzu fragen, wie viele Möglichkeiten (Kombinationen!) es gibt, k_1 aus n Patienten auszuwählen. Jede dieser Möglichkeiten ist gleich wahrscheinlich. Von diesen führt nur ein Teil der Möglichkeiten zu einer Stichprobe, in der gerade a Personen gesund sind. Der Quotient der beiden Anzahlen ergibt genau die Wahrscheinlichkeit, a Patienten in der Stichprobe zu haben, die Medikament X bekommen haben und geheilt sind. Bei der Berechnung der beiden Anzahlen von möglichen Kombinationen benutzen wir Resultate der Kombinatorik und berücksichtigen, dass es sich um einen Fall von **Ziehen ohne Zurücklegen** handelt. Die beiden Anzahlen sind wie folgt:

(1) Die Zahl der Möglichkeiten, k_1 aus n Patienten zu wählen. Diese ist gleich

$$\binom{n}{k_1}$$

(2) Die Zahl der Möglichkeiten a aus m_1 gesunden **und** $k_1 - a$ aus m_2 kranken Patienten auszuwählen. Diese ist gegeben durch das Produkt

$$\binom{m_1}{a} \binom{n-m_1}{k_1-a}$$

Der Quotient der beiden Anzahlen ergibt die Wahrscheinlichkeit für a :

$$P(a) = \frac{\binom{m_1}{a} \binom{n-m_1}{k_1-a}}{\binom{n}{k_1}}$$

Setzt man nun noch

$$\begin{aligned}
 k_1 &= a + c \\
 k_2 &= b + d \\
 m_1 &= a + b
 \end{aligned}$$

$$m_2 = c + d$$

so ergibt sich

$$P(a) = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}} = \frac{\frac{(a+b)!}{a!b!} \times \frac{(c+d)!}{c!d!}}{\frac{n!}{(a+c)!(b+d)!}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!},$$

die Wahrscheinlichkeit nach der hypergeometrischen Verteilung.

Erklärung 2: Die Idee von Fishers exaktem Tests besteht darin, die Randsummen als gegeben zu betrachten. Ausgehend vom Stichprobenumfang n kann man nun die Zahl der Möglichkeiten (Kombinationen) bestimmen, die **beobachtete Randverteilung** zu erhalten ($a+b, c+d, a+c, b+d$). Von diesen liefert nur ein Teil die **beobachteten Zellhäufigkeiten** (a, b, c, d). Auch für diesen Teil kann man die Zahl der möglichen Kombinationen berechnen. Da nun jede der Kombinationen unter der Nullhypothese dieselbe Wahrscheinlichkeit hat, berechnet sich die Wahrscheinlichkeit für die gegebenen Zellhäufigkeiten und damit $P(a)$ einfach als

$$P(a) = \frac{\text{Zahl der Kombinationen, die zu den beobachteten Zellhäufigkeiten führen}}{\text{Zahl der Kombinationen, die zu der beobachteten Randverteilung führen}}$$

Die Zahl der Kombinationen für die Randverteilung von Merkmal A ist $\binom{n}{a+b}$.

Die Zahl der Kombinationen für die Randverteilung von Merkmal B ist $\binom{n}{a+c}$.

Die Zahl der Kombinationen für die gemeinsame Randverteilung der Merkmale A und B ist gleich dem Produkt dieser beiden Anzahlen, also gleich $\binom{n}{a+b} \binom{n}{a+c}$.

Die Zahl der Kombinationen für die beobachteten Zellhäufigkeiten kann wie folgt bestimmt werden. Wir stellen uns ein Urnenmodell vor, bei dem aus n Kugeln (z.B. Patienten) zuerst die a Kugeln für Zeile 1 und Spalte 1 der Kreuzklassifikation

gezogen werden. Hierfür gibt es $\binom{n}{a}$ Kombinationen. Bleiben noch $n-a$ Kugeln, aus

denen dann b Kugeln für die Zelle in Zeile 1 und Spalte 2 gezogen werden. Hierfür gibt es $\binom{n-a}{b}$ Kombinationen. Schließlich bleiben noch $n-a-b$ Kugeln, die auf die

beiden Zellen in Zeile 2 zu verteilen sind. Hierfür gibt es $\binom{n-a-b}{c} = \binom{n-a-b}{d}$

Kombinationen. Die Zahl der Kombinationen für die gesamte Klassifikation ergibt sich aus dem Produkt der Kombinationen für die drei Ziehungen:

$$\binom{n}{a} \binom{n-a}{b} \binom{n-a-b}{c} = \frac{n!}{a!b!c!d!}. \text{ Hiermit finden wir schließlich}$$

$$P(a) = \frac{\frac{n!}{a!b!c!d!}}{\binom{n}{a+b} \binom{n}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}.$$

5.6.4 Unterschied zwischen Test auf Unabhängigkeit in 2 x 2 Felder-Tafel und McNemar-Test

McNemar-Test (siehe 5.3.6) und der Test auf Unabhängigkeit in einer 2 x 2 Tafel verwenden dieselbe Art von Häufigkeitstabelle. In beiden Fällen werden jeweils zwei verschiedene Messungen am selben Untersuchungsobjekt zugrunde gelegt. Worin besteht der Unterschied?

McNemar-Test: Es wird jeweils dasselbe Merkmal gemessen, aber für zwei verschiedene Behandlungen bzw. Bedingungen. Der Hypothesentest bezieht sich auf den Vergleich von Randwahrscheinlichkeiten, die sich auf die Randverteilung der 2 x 2 Tafel beziehen.

Unabhängigkeitstest: Es werden zwei ganz verschiedene Merkmale gemessen. Die zu prüfenden Wahrscheinlichkeiten sind bedingte Wahrscheinlichkeiten, die sich auf die vier Zellen der 2 x 2 Tafel beziehen.

5.7 Test auf Unabhängigkeit in einer $r \times c$ Tafel

Wenn in einer Stichprobe aus einer Grundgesamtheit jedes Element nach 2 Kriterien klassifiziert ist, können die Häufigkeiten in den verschiedenen Kategorien als 2-Weg-Tabelle dargestellt werden. Man kann den für die 2 x 2 Tafel beschriebenen Test auch auf den Fall verallgemeinern, dass eines oder beide Merkmale mehr als zwei Kategorien haben. Hat das erste Merkmal r Klassen, das zweite c Klassen, so erhalten wir eine $r \times c$ Tabelle (Kontingenztafel).

Beispiel: Snedecor und Cochran (1971: 250) berichten über eine Erhebung in Audubon County (Iowa) (siehe Tab. 5.7.1). Eine Zufallsstichprobe von Betrieben wurde nach Besitzverhältnissen (Eigenbesitz, Pacht, Mischung aus Eigenbesitz und Pacht) sowie nach Bodenfruchtbarkeit (Klassen I, II, III) klassifiziert. Nun kann gefragt werden, ob die Verteilung von Bodentypen bei jeder Besitzform gleich ist. Ebenso kann gefragt werden, ob die Verteilung in Besitzklassen bei jedem Bodentyp gleich ist. Beide Fragestellungen sind äquivalent. Sie können allgemeiner auch so formuliert werden: Sind Bodenfruchtbarkeit und Besitzverhältnis unabhängig?

Tab. 5.7.1: Anzahl von Betrieben klassifiziert nach Besitzform und Bodentyp; Erhebung aus Audubon County (Iowa); Daten aus Snedecor und Cochran (1971: 250).

Bodentyp	Eigenbesitz	Pacht	Gemischt	Randsumme
I	36	67	49	152
II	31	60	49	140
III	58	87	80	225
Randsumme	125	214	178	517

Allgemein:

Bodentyp	Eigenbesitz	Pacht	Gemischt	Randsumme
I	B_{11}	B_{12}	B_{13}	R_1
II	B_{21}	B_{22}	B_{23}	R_2
III	B_{31}	B_{32}	B_{33}	R_3
Randsumme	C_1	C_2	C_3	n

Für die Beantwortung dieser Frage ist eine Betrachtung von Wahrscheinlichkeiten sinnvoll. Betrachten wir z.B. die Wahrscheinlichkeit, dass ein zufällig gezogener Betrieb in die erste Zeile (Bodentyp I) und die zweite Spalte (Pacht) fällt. Dies entspricht dem relativen Anteil der Kombination von Bodentyp I und der Besitzform Pacht in der Grundgesamtheit (Alle Betriebe in Audubon County). Diese Wahrscheinlichkeit können wir anhand der beobachteten Häufigkeiten schätzen. Mit B_{ij} sei die beobachtete Häufigkeit in Zeile i und Spalte j bezeichnet, mit n der Gesamtstichprobenumfang (siehe unterer Teil von Tab. 5.7.1). Für Bodentyp I und Pacht finden wir

$$\hat{P}(\text{Pacht} + \text{Boden I}) = B_{12} / n = 67 / 517 = 0,1296$$

Von Interesse sind außerdem die sog. Randwahrscheinlichkeiten: Die Wahrscheinlichkeit, dass ein Betrieb den Bodentyp I aufweist, ist gegeben durch die Summe der Zellwahrscheinlichkeiten in Zeile 1 (Bodentyp I):

$$\begin{aligned} P(\text{Boden I}) &= P(\text{Boden I und Eigenbesitz}) \\ &+ P(\text{Boden I und Pacht}) \\ &+ P(\text{Boden I und Gemischt}) \end{aligned}$$

(siehe Additionssatz in Abschnitt 5.2!). Dies können wir anhand der Randsumme (Zeilensumme) für Bodentyp I (siehe Tab. 5.7.1) schätzen. Die Randsummen werden mit R_i bezeichnet. Für Bodentyp I finden wir $R_1 = 152$ und

$$\hat{P}(\text{Boden I}) = B_{11} / n + B_{12} / n + B_{13} / n = R_1 / n = 152 / 517 = 0,2940$$

Genauso können wir die Randwahrscheinlichkeit für den Besitztyp Pacht (Spalte 2) schätzen (C_j sind die Spaltensummen):

$$\hat{P}(\text{Pacht}) = B_{12}/n + B_{22}/n + B_{32}/n = C_2/n = 214/517 = 0,4139$$

Wenn nun Bodentyp und Besitzform unabhängig sind, so kann der Multiplikationsatz (Abschnitt 5.2!) angewendet werden: Die Wahrscheinlichkeit, dass ein Betrieb in die Besitzklasse Pacht und die Bodenklasse I fällt, ergibt sich aus dem Produkt der Randwahrscheinlichkeiten für diese Kategorien, also

$$P(\text{Boden I und Pacht}) = P(\text{Boden I}) * P(\text{Pacht})$$

Setzen wir die geschätzten Randwahrscheinlichkeiten ein, finden wir

$$\hat{P}(\text{Boden I}) * \hat{P}(\text{Pacht}) = \frac{R_1}{n} * \frac{C_2}{n} = 0,2940 * 0,4139 = 0,1217$$

Bei einem Stichprobenumfang von n erwarten wir bei Unabhängigkeit

$$E_{12} = n * P(\text{Boden I}) * P(\text{Pacht})$$

Betriebe mit Bodentyp I und Pacht. Wir finden hierfür die Schätzung

$$\hat{E}_{12} = n * \hat{P}(\text{Boden I}) * \hat{P}(\text{Pacht}) = n * \frac{R_1}{n} * \frac{C_2}{n} = 517 * 0,2940 * 0,4139 = 62,91$$

Dieser Wert stimmt ziemlich gut mit der beobachteten Häufigkeit

$$B_{12} = 67$$

überein, was darauf hindeutet, dass die Unabhängigkeitsregel hier angewendet werden kann, dass also Betriebstyp und Bodentyp unabhängig sind. Ein Vergleich der beobachteten und der bei Unabhängigkeit erwarteten Häufigkeiten führt wieder zu einem χ^2 -Anpassungs-Test:

$$X_{Vers}^2 = \sum \frac{(B_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} = \sum \frac{\left(B_{ij} - \frac{R_i C_j}{n}\right)^2}{\frac{R_i C_j}{n}} = n \left[\left(\sum \frac{B_{ij}^2}{R_i C_j} \right) - 1 \right]$$

X_{Vers}^2 ist verteilt wie χ^2 mit $(r-1)(c-1)$ Freiheitsgraden, wobei r und c die Zahlen der Zeilen (rows) und Spalten (columns) der Kontingenztafel sind.

Test auf Unabhängigkeit in der $r \times c$ Tafel

Voraussetzung: $E_{ij} > 1$. Andernfalls muss exakt getestet werden (siehe 5.7.1).

Stelle Kontingenz-Tafel auf:

		Merkmal Y					
		Y ₁	Y ₂	Y _c			
Merkmal X	X ₁	B ₁₁	B ₁₂	.	.	.	B _{1c} R ₁
	X ₂	B ₂₁					B _{2c} R ₂
		.					.
		.			B _{ij}		B _{ic} R _i
		.					.
	X _r	B _{r1}	.	.	B _{rj}	.	B _{rc} R _r
		C ₁			C _j		C _c n

H₀: Die Merkmale X und Y sind unabhängig.

Berechne: $X_{Vers}^2 = n \left[\left(\sum \frac{B_{ij}^2}{R_i C_j} \right) - 1 \right]$

Bestimme: $X_{Tab}^2 = \chi_{1-\alpha; (r-1)(c-1)}^2$

r = Zahl der Zeilen; c = Zahl der Spalten

$X_{Vers}^2 > X_{Tab}^2 \Rightarrow$ verwirfe H₀

$X_{Vers}^2 \leq X_{Tab}^2 \Rightarrow$ behalte H₀ bei

Beispiel: Für die Erhebung aus Audubon County (Iowa) finden wir:

$$\sum \frac{B_{ij}^2}{R_i C_j} = \frac{36^2}{152 \cdot 125} + \frac{67^2}{152 \cdot 214} + \frac{49^2}{152 \cdot 178} + \frac{31^2}{140 \cdot 125} + \frac{60^2}{140 \cdot 214} + \frac{49^2}{140 \cdot 178} + \frac{58^2}{225 \cdot 125} + \frac{87^2}{225 \cdot 214} + \frac{80^2}{225 \cdot 178} = 1,0029847902$$

$$X_{Vers}^2 = n \left[\left(\sum \frac{B_{ij}^2}{R_i C_j} \right) - 1 \right] = 517 \cdot 0,0029847902 = 1,54$$

Mit $\alpha = 5\%$ und $r = c = 3$ finden wir

$$X_{Tab}^2 = \chi_{0,95;4}^2 = 9,49$$

$$X_{Vers}^2 \leq X_{Tab}^2 \Rightarrow \text{behalte } H_0 \text{ bei}$$

Die Nullhypothese der Unabhängigkeit zwischen Betriebstyp und Bodentyp wird nicht verworfen.

5.7.1 Exakter Test

Es gibt auch einen exakten Test für $r \times c$ Tafeln, der bei kleineren Stichproben anzuwenden ist, insbesondere dann, wenn einige Zellen der Tafel eine Null aufweisen. Der exakte Test erfordert die Anwendung eines sog. Netzwerkalgorithmus und ist sehr rechenaufwendig. Er kann nur mit Hilfe eines Computerprogramms durchgeführt werden (z.B. SAS PROC FREQ).

*5.8 Bemerkung zur Chi-Quadrat-Verteilung

Eine einfache mathematisch-statistische Charakterisierung der Chi-Quadrat-Verteilung ist wie folgt: Wenn y_i ($i=1, \dots, c$) normalverteilt ist mit Mittelwert μ_i und Varianz σ_i^2 , so hat $z_i = (y_i - \mu_i) / \sigma_i$ eine Standardnormalverteilung. Sind die y_i außerdem stochastisch unabhängig, so hat

$$X^2 = \sum_{i=1}^c z_i^2 = \sum_{i=1}^c \frac{(y_i - \mu_i)^2}{\sigma_i^2}$$

eine Chi-Quadrat-Verteilung mit c Freiheitsgraden (FG). Wenn die z_i stochastisch abhängig sind, müssen die FG reduziert werden. Eine wichtige Eigenschaft der Chi-Quadrat-Verteilung ist, dass $E(X^2) = \text{FG}$ ist.

Anwendung 1: Wenn alle y_i aus derselben Normalverteilung mit Mittelwert μ und Varianz σ^2 stammen, dann hat

$$X^2 = \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2}$$

eine Chi-Quadrat-Verteilung mit n FG. Ersetzt man den Parameter μ durch den Schätzwert \bar{y}_\bullet , so erhält man

$$X^2 = \sum_{i=1}^n \frac{(y_i - \bar{y}_\bullet)^2}{\sigma^2} = \frac{SQ}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2},$$

wobei s^2 die Stichprobenvarianz ist. Dies ist die Basis der Verfahren für s^2 in 3.19 und 3.20. Die Zahl der FG reduziert sich um 1, weil die Ersetzung von μ durch \bar{y}_\bullet zu Abhängigkeiten zwischen den Beiträgen $m_i = (y_i - \bar{y}_\bullet) / \sigma$ führt (es ist immer $\sum_i m_i = 0$, so dass die m_i abhängig sind); es bleiben $n-1$ FG.

Anwendung 2: In einer Kontingenztafel folgen die beobachteten Häufigkeiten näherungsweise einer Poissonverteilung. Für diese gilt Gleichheit von Varianz und Mittelwert. Außerdem geht die Poissonverteilung für große Erwartungswerte in eine Normalverteilung über. Bezeichnet man nun wie für kategoriale Daten üblich den Erwartungswert und die beobachtete Häufigkeit in der i -ten Kategorie mit E_i und B_i , so erhält man mit $y_i = B_i$ und $\mu_i = \sigma_i^2 = E_i$ den Ausdruck

$$X^2 = \sum_{i=1}^c \frac{(B_i - E_i)^2}{E_i}.$$

Beim Chi-Quadrat-Anpassungstest ist die Zahl der FG im Fall bekannter Erwartungswerte nicht gleich c , sondern gleich $c-1$, weil die beobachteten Häufigkeiten B_i sich zum Stichprobenumfang addieren müssen: $n = \sum_i B_i$. Dies induziert eine Korrelation zwischen den beobachteten Häufigkeiten, die zu einer Reduktion der FG führt. Außerdem muss für jeden geschätzten Parameter ein FG abgezogen werden.

Freiheitsgrade: Dieser Begriff wurde von R. A. Fisher, dem Begründer der Varianzanalyse, eingeführt, und zwar in Analogie zum physikalischen Begriff der Freiheitsgrade in dynamischen Systemen, also der Zahl unabhängiger Koordinaten, die notwendig sind, um ein System eindeutig zu bestimmen (Kendall & Buckland: A dictionary of statistical terms. Longman, London). Im Falle der Bewegung von Molekülen in idealen Gasen sind dies beispielsweise die Geschwindigkeiten in Richtung der drei Raumkoordinaten (siehe Maxwell-Boltzmann-Gesetz - Physik-Vorlesung Prof. Wulfmeyer). In analogem Sinne sind die Freiheitsgrade eines Datensatzes die Zahl der frei wählbaren Werte im Rahmen der Spezifikation des Systems. Wird beispielsweise ein Datensatz vom Umfang n in c Klassen aufgeteilt, so können bei festem Stichprobenumfang n nur für $c-1$ der c Klassen die Häufigkeiten frei gewählt werden; die Häufigkeit in der c -ten Klasse liegt eindeutig fest. Ebenso können in einer $r \times c$ -Tafel bei fester Randverteilung nur $(r-1)(c-1)$ der Klassenhäufigkeiten frei gewählt werden. Generell hat ein Datensatz vom Umfang n auch n Freiheitsgrade. Werden aber k Funktionen der Daten fixiert, so werden die Freiheitsgrade um k reduziert. So hat die Fehler-Quadratsumme $SQ = \sum_{i=1}^n (y_i - \bar{y})^2$ nur $n-1$ Freiheitsgrade, falls der Mittelwert bei \bar{y} fixiert wird, obwohl ihr n Einzelwerte zugrunde liegen. All diese Erklärungen für den Begriff der Freiheitsgrade haben allerdings stark heuristischen Charakter. Eine genaue Definition ergibt sich jeweils erst im Zusammenhang mit der mathematisch-statistischen Betrachtung der Verteilung der Daten und von aus ihnen berechneten Größen.

Exkurs: An dieser Stelle ergibt sich ein interessanter Bezug zur Maxwell-Boltzmann-Verteilung der kinetischen Energie von Molekülen in idealen Gasen (siehe Physik-Vorlesung Prof. Wulfmeyer). Die kinetische Energie eines Moleküls ist $E = \frac{1}{2}mv^2$, wobei m die Masse und v die Geschwindigkeit ist. Da die Masse je Molekül bei idealen Gasen konstant ist, hängt die Verteilung der Energie nur von der Verteilung der Geschwindigkeiten v ab. Um diese abzuleiten, betrachtet man zunächst die Verteilung der Geschwindigkeiten v_1 , v_2 und v_3 eines Moleküls entlang der drei Koordinaten des kartesischen Koordinatensystems und nutzt dann aus, dass die quadrierte Geschwindigkeit

$$v^2 = v_1^2 + v_2^2 + v_3^2$$

beträgt (siehe Vektorrechnung, Mathe-Vorlesung Prof. Jetter). Die Verteilung der Geschwindigkeit v_i ($i = 1, 2, 3$) ist jeweils eine Normalverteilung mit Mittelwert 0 und Varianz $\sigma^2 = kT/m$, wobei k die Boltzmann-Konstante und T die absolute Temperatur in Grad Kelvin ist (Moore & Hummel, 1983: Physikalische Chemie. 3. Auflage. De Gruyter, Berlin). Diese hat die Form

$$f(v_i) = \frac{1}{\sqrt{2\pi kT/m}} e^{-\frac{v_i^2}{2kT/m}}.$$

Da die Geschwindigkeitsverteilungen in Richtung der drei kartesischen Koordinaten stochastisch unabhängig sind (die Bewegung eines Moleküls hat 3 Freiheitsgrade im physikalischen Sinne!!!), so hat

$$X^2 = \frac{v_1^2}{kT/m} + \frac{v_2^2}{kT/m} + \frac{v_3^2}{kT/m} = \frac{mv^2}{kT}$$

eine Chi-Quadrat-Verteilung mit 3 Freiheitsgraden. Da $E = \frac{1}{2}mv^2 = (\frac{1}{2}kT)X^2$ ist und der Erwartungswert von X^2 gleich FG = 3 (siehe oben) beträgt, ist die mittlere kinetische Energie gleich $\bar{E} = (1/2)kT * 3 = (3/2)kT$. Außerdem hat die kinetische Energie, abgesehen von einem Skalierungsfaktor, ebenfalls eine Chi-Quadrat-Verteilung mit 3 FG! Explizite Ableitung basierend auf der Chi-Quadrat-Verteilung von X^2 liefert die sog. **Maxwell-Boltzmann-Verteilung der kinetischen Energie**:

$$f(E) = \frac{2}{\sqrt{\pi}} (kT)^{-3/2} \sqrt{E} e^{-E/kT}.$$

Die Verteilung des Geschwindigkeitsvektors $\vec{v} = (v_1, v_2, v_3)'$ ist eindeutig bestimmt durch die gemeinsame Verteilung seiner Komponenten. Da die Komponenten unabhängig verteilt sind, gilt aufgrund des Multiplikationssatzes der Wahrscheinlichkeitsrechnung (Abschnitt 5.2)

$$f(\vec{v}) = f(v_1)f(v_2)f(v_3) = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-m\vec{v}^2/2kT}.$$

Dies ist die **Maxwellsche Geschwindigkeitsverteilung** (siehe Vorlesung Prof. Wulfmeyer).

6. Korrelation und Regression

Bisher (vorangegangene Vorlesung Statistik) wurden bereits verschiedene Verfahren zur Erfassung des Zusammenhangs zweier Variablen behandelt, ohne dass diese entsprechend bezeichnet wurden. Der Zusammenhang zweier Variablen x und y kann symbolisch wie folgt dargestellt werden:

$x \longrightarrow y$ (x bewirkt y)

$x \longleftrightarrow y$ (x und y hängen wechselseitig voneinander ab)

$\begin{array}{c} \nearrow x \\ z \\ \searrow y \end{array}$ (x und y hängen direkt von z ab)

Die Wahl des Verfahrens zur Ermittlung des Zusammenhanges hängt vom Skalenniveau der Variablen ab. Sind beide Variablen metrisch skaliert, kommen die Korrelation und die Regression zur Auswertung in Frage. Diese Verfahren stehen im Zentrum dieses Kapitels. Im folgenden wird eine kleine Übersicht gegeben über bisher behandelte Verfahren und deren Abhängigkeit vom Skalenniveau der Variablen.

Tab. 6.1: Abhängigkeit der Wahl des Verfahrens zur Ermittlung eines Zusammenhangs.

Skalenniveau		Verfahren (Beispiele)	Kapitel
x	y		
metrisch	metrisch	Korrelation, Regression	6
kategorial	metrisch	paarweiser t-Test, Varianzanalyse	3, 4
kategorial	kategorial	χ^2 -Test, Binomialtest, Anpassungstest	5

Beispiel: Es soll der Zusammenhang ermittelt werden zwischen Sortenwahl (kategorial) und Ertrag (metrisch). Hierzu wird ein Sortenversuch mit fünf Sorten durchgeführt. Auswertung: Varianzanalyse mit anschließendem multiplen t-Test (Kap. 3 und 4).

Beispiel: Es soll der Zusammenhang ermittelt werden zwischen vorherrschendem Bodentyp und Besitzform landwirtschaftlicher Betriebe in einer Region. Hierzu wird eine Zufallsstichprobe von Betrieben befragt. Die Betriebe werden bezüglich Besitzform und Bodentyp in Klassen eingeteilt. Auswertung: χ^2 -Test (Kap. 5).

Beispiel: In einer Erhebung wurden für einen Zeitraum von 26 Jahren die erzielten Ernten (y) den Regenmengen von April bis Juni (x) gegenübergestellt (Jacob, A. und Rüter H. 1961 Der Vegetationsversuch. VDLUFA, Berlin; Daten aus: Scheinert, R.: Pflanzenbau 5, 236-239, 1928/29). Beide Merkmale sind metrisch.

Jahr	x	y	x^2	y^2	xy
1900	177	26,2	31329	686,44	4637,4
1901	96	25,0	9216	625,00	2400,0
1902	144	32,6	20736	1062,76	4694,4
1903	105	26,6	11025	707,56	2793,0
1904	111	19,6	12321	384,16	2175,6
1905	135	20,4	18225	416,16	2754,0
1906	209	29,2	43681	852,64	6102,8
1907	161	33,8	25921	1142,44	5441,8
1908	246	26,6	60516	707,56	6543,6
1909	108	22,6	11664	510,76	2440,8
1910	137	24,2	18769	585,64	3315,4
1911	71	16,6	5041	275,56	1178,6
1912	119	29,8	14161	888,04	3546,2
1913	108	19,4	11664	376,36	2095,2
1914	132	30,6	17424	936,36	4039,2
1915	89	16,4	7921	268,96	1459,6
1916	147	30,4	21609	924,16	4468,8
1917	98	19,2	9604	368,64	1881,6
1918	106	20,2	11236	408,04	2141,2
1919	123	25,6	15129	655,36	3148,8
1920	156	31,0	24336	961,00	4836,0
1921	191	35,8	36481	1281,64	6837,8
1922	162	31,6	26244	998,56	5119,2
1923	235	33,6	55225	1128,96	7896,0
1924	147	30,4	21609	924,16	4468,8
1925	110	30,2	12100	912,04	3322,0
Summe	3623	687,6	553187	18988,96	99737,8

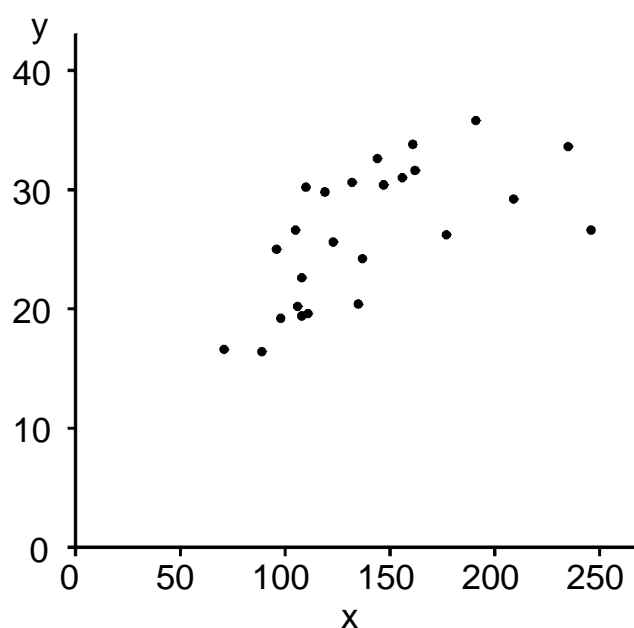


Abb. 6.1: Plot der Erträge (y ; dt/ha) gegen die in den Monaten April bis Juni gefallene Regenmenge (x ; mm).

Diese Daten können graphisch dargestellt werden, indem die Messwerte (x_i = Regenmenge; y_i = Ertrag) als Koordinaten von Punkten in einem kartesischen Koordinatensystem aufgefaßt werden. Hierdurch wird jedes Jahr bzw. jedes Messwertpaar (x_i, y_i) als Punkt repräsentiert (siehe Abb. 6.1). Offenbar besteht ein gewisser **Zusammenhang** zwischen Niederschlag und Ertrag. Je höher der Niederschlag, desto höher der Ertrag. Man spricht auch von einer positiven **Korrelation**. Die Korrelation ist allerdings nicht perfekt, da der Ertrag nicht ausschließlich durch den Niederschlag bestimmt wird. Ziel der weiteren statistischen Auswertung kann es nun sein, den Zusammenhang zwischen den beiden Variablen zu quantifizieren. Hierzu können die Verfahren der Korrelation und Regression verwendet werden, die in diesem Kapitel besprochen werden.

6.1 Die Pearsonsche Produkt-Moment Korrelation

Die Korrelation ist wie folgt definiert:

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}$$

wobei σ_{xy} die sog. **Kovarianz** von x_i und y_i ist, während σ_x^2 und σ_y^2 die Varianzen von x_i und y_i sind. Gibt es keinen Zusammenhang zwischen x_i und y_i , ist die Kovarianz gleich Null. Gibt es dagegen einen Zusammenhang, weicht die Kovarianz um so stärker von Null ab, je stärker der Zusammenhang ist. Die Division durch die Wurzel aus dem Produkt der Varianzen ist eine **Standardisierung**, die bewirkt, dass die **Korrelation** eine dimensionslose Maßzahl zwischen -1 (negativer Zusammenhang) und 1 (positiver Zusammenhang) ist. Die Korrelation kann auch als eine **standardisierte Kovarianz** bezeichnet werden. Wir kommen auf den beschränkten Wertebereich der Korrelation zurück, wenn wir deren Schätzung besprochen haben.

Nun zur Schätzung der Korrelation. Die Varianz wird bekanntermaßen wie folgt geschätzt:

$$s_x^2 = \frac{SQ_x}{n-1} \quad \text{und} \quad s_y^2 = \frac{SQ_y}{n-1}$$

mit

$$SQ_x = \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{und} \quad SQ_y = \sum_{i=1}^n (y_i - \bar{y})^2$$

wobei (x_i, y_i) das i -te Messwertpaar ist. Die Kovarianz wird geschätzt durch

$$s_{xy} = \frac{SP_{xy}}{n-1}$$

$$SP_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Die Korrelation wird geschätzt durch

$$r = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}}$$

SP_{xy} ist die Summe der Kreuzprodukte. Man entnimmt der Formel für SP_{xy} leicht folgendes: Wenn sowohl x_i als auch y_i größer ist als der Mittelwert, ist das Kreuzprodukt $(x_i - \bar{x})(y_i - \bar{y})$ größer als Null. Ebenso ist das Kreuzprodukt größer als Null, wenn beide Abweichungen negativ sind. Hieraus folgt, dass die Summe der Kreuzprodukte und somit auch die Korrelation im Falle eines positiven Zusammenhanges positiv werden muss. Haben die Abweichungen dagegen umgekehrte Vorzeichen, so wird der Beitrag zur Summe der Kreuzprodukte negativ. Ist schließlich das Vorzeichen der Abweichung für x_i unabhängig vom Vorzeichen der Abweichung für y_i , dann wird die Summe der Kreuzprodukte mehr oder weniger nahe bei Null liegen.

Nun zur Erklärung, warum die Korrelation zwischen -1 und 1 liegen muss. Man überzeugt sich leicht, dass die Korrelation auch ausgedrückt werden kann als Kovarianz der z -transformierten Variablen x und y :

$$r = \frac{\sum_{i=1}^n z_{xi} z_{yi}}{(n-1)}$$

wobei

$$z_{xi} = \frac{x_i - \bar{x}}{s_x} \quad \text{und} \quad z_{yi} = \frac{y_i - \bar{y}}{s_y}$$

Es ist klar, dass

$$\sum_{i=1}^n (z_{xi} - z_{yi})^2 \geq 0$$

ist, da hier quadrierte Differenzen summiert werden, und das Quadrat einer beliebigen Zahl größer oder gleich Null sein muss. Nun ist aber

$$\sum_{i=1}^n (z_{xi} - z_{yi})^2 = \sum_{i=1}^n z_{xi}^2 - 2 \sum_{i=1}^n z_{xi} z_{yi} + \sum_{i=1}^n z_{yi}^2 = (n-1) - 2(n-1)r + (n-1) \geq 0$$

woraus durch Umformung folgt, dass $r \leq 1$ ist (Achtung: Bei Division beider Seiten einer Ungleichung durch eine negative Zahl muss "größer gleich" in "kleiner gleich" umgewandelt werden und umgekehrt). Analog folgt aus

$$\sum_{i=1}^n (z_{xi} + z_{yi})^2 = \sum_{i=1}^n z_{xi}^2 + 2 \sum_{i=1}^n z_{xi} z_{yi} + \sum_{i=1}^n z_{yi}^2 = (n-1) + 2(n-1)r + (n-1) \geq 0,$$

dass $r \geq -1$ sein muss. Somit haben wir gezeigt, dass

$$-1 \leq r \leq 1$$

Man kann auch folgende einfache Betrachtung anstellen: Bei perfektem linearen Zusammenhang ist entweder $z_{xi} = z_{yi}$, so dass $r = s_z^2 = 1$ ist, oder es ist $z_{xi} = -z_{yi}$, so dass $r = -s_z^2 = -1$ ist. Somit muss r zwischen -1 und 1 liegen. In Abb. 6.2 werden einige Beispiele für Werte der Korrelation bei verschiedenen Ausprägungen und Vorzeichen des Zusammenhanges zwischen X und Y gegeben.

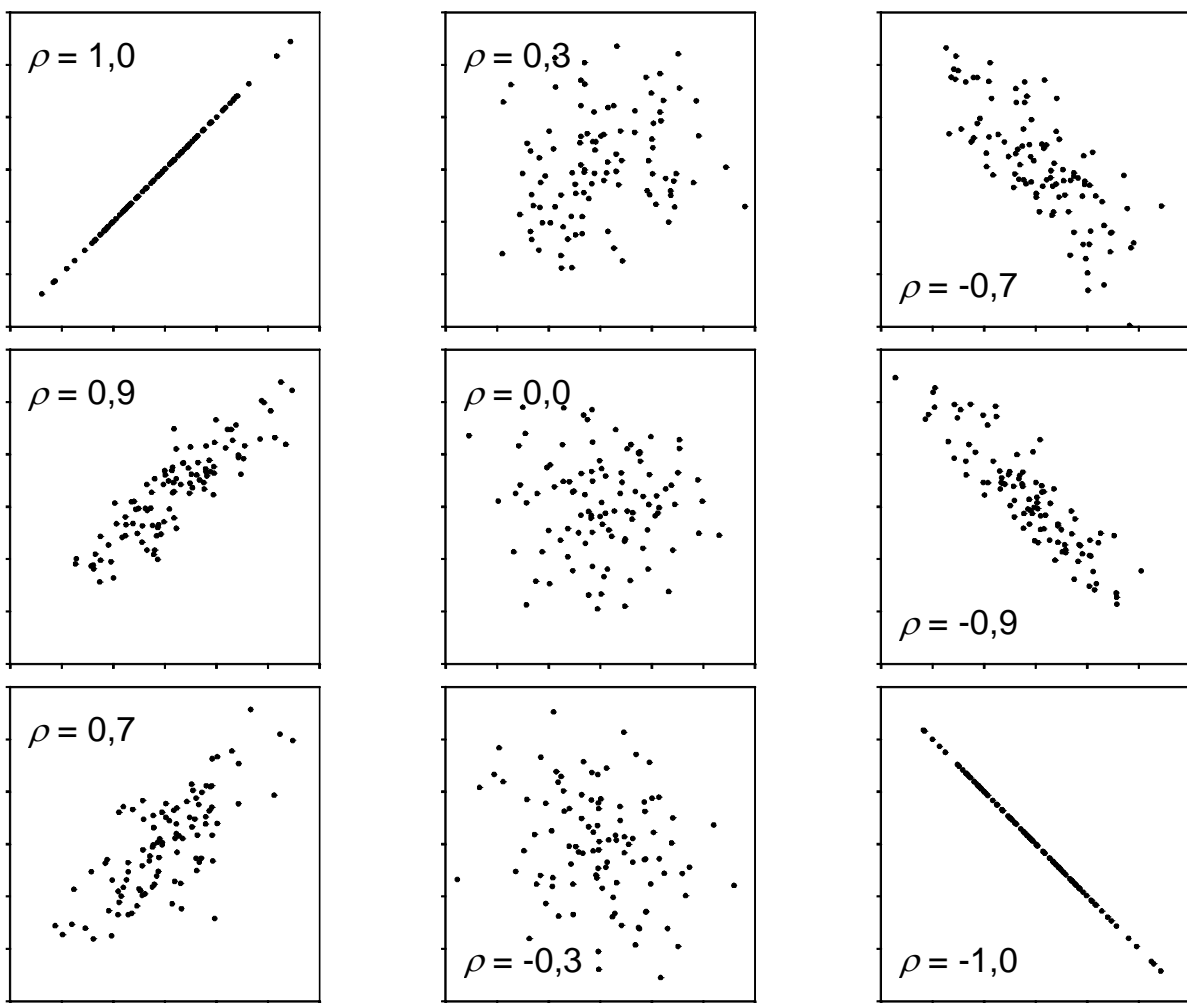


Abb. 6.2: Einige Beispiele (simulierte Daten) für verschiedene Werte der Korrelation.

Es ist wichtig, zu beachten, dass die Korrelation ein Maß für den linearen Zusammenhang zweier Variablen ist. Bei nichtlinearen Zusammenhängen sollte ein anderes Korrelationsmaß verwendet werden, z.B. die Rangkorrelation von Spearman.

Für die praktische Anwendung ist folgende Rechenformel für die Korrelation hilfreich:

Schätzen des Korrelationskoeffizienten

$$r = \frac{SP_{xy}}{\sqrt{SQ_x SQ_y}}$$

$$SP_{xy} = \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right) \left(\sum_{i=1}^n y_i\right)}{n} \quad (\text{Summe der Kreuzprodukte})$$

$$SQ_x = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} \quad (\text{Summe der Quadrate für X})$$

$$SQ_y = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \quad (\text{Summe der Quadrate für Y})$$

$(x_i, y_i) = i\text{-tes Messwertpaar.}$

Beispiel: Für die Regendaten (y = Erträge und x = Regenmenge zwischen April und Juni) finden wir

$$SP_{xy} = 99737,8 - \frac{3623 \cdot 687,6}{26} = 3923,38$$

$$SQ_x = 553187 - \frac{3623^2}{26} = 48335,88$$

$$SQ_y = 18988,96 - \frac{687,6^2}{26} = 804,58$$

$$r = \frac{3923,38}{\sqrt{48335,88 \cdot 804,88}} = 0,63$$

Die Korrelation ist positiv, da der Zusammenhang zwischen Ertrag und Niederschlag positiv ist. Da $r < 1$ ist, ist der Zusammenhang aber nicht perfekt. Im weiteren ist die Frage zu klären, ob der Stichprobenkorrelationskoeffizient r signifikant vom Wert $\rho = 0$ abweicht, ob es also überhaupt einen signifikanten Zusammenhang gibt.

Test der Korrelation ρ

Frage: Gibt es einen echten Zusammenhang zwischen zwei Variablen X und Y ?

Voraussetzung: Die Daten sind bivariat normalverteilt

$$H_0: \rho = 0$$

$$H_A: \rho \neq 0$$

Rechenweg:

(1) Berechne $t_{Vers} = \frac{|r|}{\sqrt{1-r^2}} \sqrt{n-2}$

(2) Lies in der t -Tabelle $t_{Tab}(FG; \alpha)$ ab (Tab. II, zweiseitig) wobei α das Signifikanzniveau und $FG = n - 2$ die Freiheitsgrade sind.

(3) Vergleiche t_{Vers} mit t_{Tab} :

Falls $t_{Vers} \leq t_{Tab} \Rightarrow H_0 (\rho = 0)$ (Kein Zusammenhang)

Falls $t_{Vers} > t_{Tab} \Rightarrow H_1 (\rho \neq 0)$ (Signifikanter Zusammenhang)

Man beachte, dass der Test die Annahme macht, dass die Daten einer bivariaten Normalverteilung folgen. Außerdem erfasst die Korrelation nur lineare Zusammenhänge, und es muss angenommen werden, dass kein nichtlinearer Zusammenhang besteht. Die (geschätzte) bivariate Normalverteilung für die Regendaten ist in der folgenden Graphik wiedergegeben.

Wahrscheinlichkeitsdichte

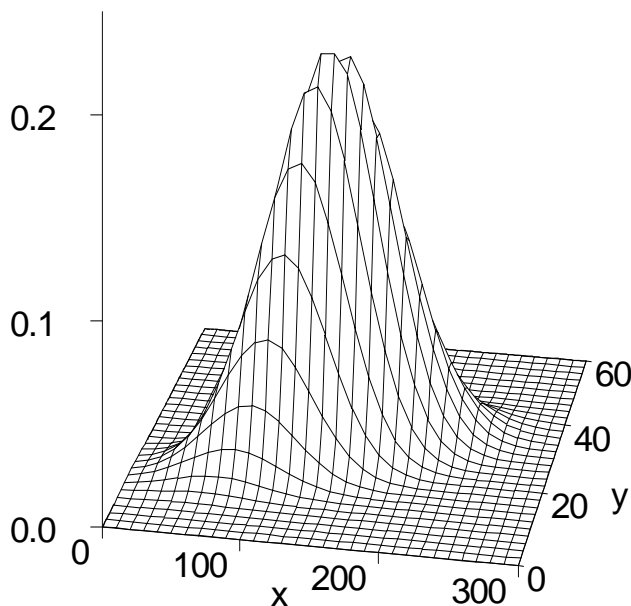


Abb.6.2(b): Geschätzte bivariate Normalverteilung für die Regendaten (F = Wahrscheinlichkeitsdichte).

Beispiel: Wir wollen die Korrelation zwischen Regenmenge und Ertrag zum 5%-Niveau testen.

$r = 0,63, n = 26, \alpha = 5\%$

$$t_{Vers} = \frac{|0,63|}{\sqrt{1-0,63^2}} \sqrt{26-2} = 3,97$$

$t_{Tab} = 2,064 < t_{Vers} \Rightarrow$ Die Korrelation ist signifikant

Eine signifikante Korrelation sagt zunächst nichts über die **Kausalität** des Zusammenhangs aus. Sie besagt lediglich, dass ein Zusammenhang besteht, ohne dass damit geklärt wäre, wie es zu diesem Zusammenhang kommt. Insbesondere ist damit nichts gesagt über eine **Ursache-Wirkungs-Beziehung** oder **Kausalbeziehung**.

Beispiel: In Industrieländern ist oft eine signifikante Korrelation zwischen der Zahl der Geburten und der Zahl der Störche pro Flächeneinheit festgestellt worden. Eine Ursache-Wirkungs-Beziehung besteht hier nicht. Vielmehr ist zu vermuten, dass sowohl Geburten- als auch Storchzahlen beispielsweise vom Grad der Industrialisierung beeinflusst werden. Andere Erklärungen sind möglich.

Beispiel: Es kann oft eine hohe Korrelation zwischen der Größe des Schadens eines Feuers und der Zahl der an der Brandlöschung beteiligten Feuerwehrmänner festgestellt werden.

Man spricht in Fällen wie den vorangegangenen auch von einer **Scheinkorrelation**, obwohl diese Bezeichnung den Punkt nicht ganz trifft, denn die Korrelation ist ja real. Problematisch ist lediglich die fälschliche Interpretation im Sinne einer Ursache-Wirkungs-Beziehung. In beiden Fällen liegt eine Abhängigkeit von einer dritten Größe Z vor, welche zu der Scheinkorrelation von X und Y führt.

Beispiel: Bei den Regendaten liegt von der Sachlage her eine Ursache-Wirkungs-Beziehung auf der Hand: Die Höhe des Niederschlages bedingt den Ertrag, und nicht umgekehrt. Es liegt hier eine **einseitige Abhängigkeit** vor. Allerdings liefern die Regendaten, die durch eine Erhebung ermittelt wurden, keinen strengen Beweis, dass die Regenmenge wirklich ursächlich für den Ertrag verantwortlich ist, weil in dieser Erhebung viele Umweltfaktoren zufällig variieren, die den Ertrag beeinflussen können (Temperatur, Aussaattermin, Wärmesumme, Bodenzustand zum Zeitpunkt der Aussaat, Keimfähigkeit des Saatgutes, etc.). Um den Einfluss der Regenmenge zweifelsfrei nachweisen zu können, ist ein Experiment erforderlich. Es ist generell festzuhalten, dass **Erhebungen** im allgemeinen nicht für den Nachweis von Kausalbeziehungen geeignet sind, auch wenn eine solche noch so plausibel erscheint. Erhebungen können lediglich Hinweise für die Bildung von Hypothesen über Kausalbeziehungen geben. Diese Hypothesen müssen dann in einem **Experiment** überprüft werden, in welchem der ursächliche Faktor gezielt variiert wird, während alles andere möglichst konstant gehalten wird (**ceteris paribus Klausel**). Erst wenn sich unter diesen Bedingungen der Zusammenhang reproduzieren lässt, ist die Kausalbeziehung nachgewiesen. Das für den Nachweis notwendige Konstanthalten aller anderen Einflussfaktoren ist in einer Erhebung generell nicht oder nur sehr eingeschränkt möglich. Im Fall der Regendaten ist es beispielsweise unmöglich, die Temperaturbedingungen konstant zu halten. Wir werden auf diesen Punkt in Abschnitt 6.2 zurückkommen.

Beispiel: Es kann oft eine Korrelation zwischen Pflanzenlänge und Blattflächenindex (BFI) (Blattfläche pro Standfläche) bei Raps festgestellt werden. Diese beiden Merkmale bedingen sich gegenseitig, es besteht also eine **zweiseitige Abhängigkeit**. Je länger die Pflanze, desto mehr Verzweigungen bildet die Pflanze und umso mehr Blätter werden gebildet. Umgekehrt ermöglicht ein höherer Blattflächenindex eine erhöhte Assimilation und somit eine bessere Bedingung für das Längenwachstum der Pflanze.

Neben einem Test der Korrelation können wir auch ein Vertrauensintervall berechnen. R. A. Fisher hat durch eine komplizierte Approximation (A. Stuart & K. Ord: Kendall's advanced theory of statistics. 6th edition. Volume 1, § 16.33) gezeigt, dass

$$q = \frac{1}{2} \ln \left[\frac{1+r}{1-r} \right]$$

für nicht zu kleines n näherungsweise einer Normalverteilung mit Mittelwert

$$\theta = \frac{1}{2} \ln \left[\frac{1+\rho}{1-\rho} \right]$$

und Varianz

$$\sigma_q^2 = \frac{1}{n-3}$$

folgt (der Term $n-3$ im Zähler der Varianz hat nichts mit Freiheitsgraden zu tun). Daher ist

$$z = \frac{q - \theta}{\sigma_q}$$

näherungsweise standardnormalverteilt. Somit sind

$$\theta_u = q - z_{1-\alpha/2} \sigma_q \text{ und } \theta_o = q + z_{1-\alpha/2} \sigma_q$$

die $(1-\alpha)100\%$ -Vertrauensgrenzen für θ , wobei $z_{1-\alpha/2}$ das $(1-\alpha/2)100\%$ -Quantil der Standardnormalverteilung ist. Die entsprechenden Grenzen für ρ erhält man durch Rücktransformation der Grenzen θ_u und θ_o . Die praktisch durchzuführenden Berechnungen enthält der folgende Kasten.

Vertrauensintervall für die Korrelation ρ

Berechne:

$$q = \frac{1}{2} \ln \left[\frac{1+r}{1-r} \right]$$

wobei $\ln()$ der natürliche Logarithmus ist (Basis e)

$$\theta_u = q - \frac{z_{1-\alpha/2}}{\sqrt{n-3}} \text{ und } \theta_o = q + \frac{z_{1-\alpha/2}}{\sqrt{n-3}}$$

wobei

$z_{1-\alpha/2} = (1-\alpha/2)100\%$ -Quantil der Standardnormalverteilung (Tab. III)

Die $(1-\alpha)100\%$ -Vertrauensgrenzen sind gegeben durch

$$\rho_u = \frac{e^{2\theta_u} - 1}{e^{2\theta_u} + 1} \quad \text{und} \quad \rho_o = \frac{e^{2\theta_o} - 1}{e^{2\theta_o} + 1}$$

Voraussetzung: Die Daten sind bivariat normalverteilt.

Beispiel: $r = 0,63$, $n = 26$, $\alpha = 5\%$

$$q = \frac{1}{2} \ln \left[\frac{1+0,63}{1-0,63} \right] = 0,741$$

$z_{1-\alpha/2} = 1,96$ (siehe Tab. III: $\alpha = 5\% = 0,05 \Rightarrow \gamma = 1-\alpha/2 = 1-0,05/2 = 0,975$

$$\Rightarrow z_{1-\alpha/2} = z_{\gamma} = z_{0,975} = 1,95996 \approx 1,96)$$

$$\theta_u = 0,741 - \frac{1,96}{\sqrt{23}} = 0,741 - 0,409 = 0,332 \quad \text{und} \quad \theta_o = 0,741 + 0,409 = 1,150$$

$$\rho_u = \frac{e^{2 \cdot 0,332} - 1}{e^{2 \cdot 0,332} + 1} = \frac{1,943 - 1}{1,943 + 1} = 0,32 \quad \text{und} \quad \rho_o = \frac{e^{2 \cdot 1,150} - 1}{e^{2 \cdot 1,150} + 1} = \frac{9,974 - 1}{9,974 + 1} = 0,82$$

Mit 95%-iger Wahrscheinlichkeit überdeckt das Intervall von 0,32 bis 0,82 die Korrelation ρ . Wir sehen, dass die Schätzung der Korrelation relativ ungenau ist, u.a. aufgrund des geringen Stichprobenumfanges.

6.2 Regression

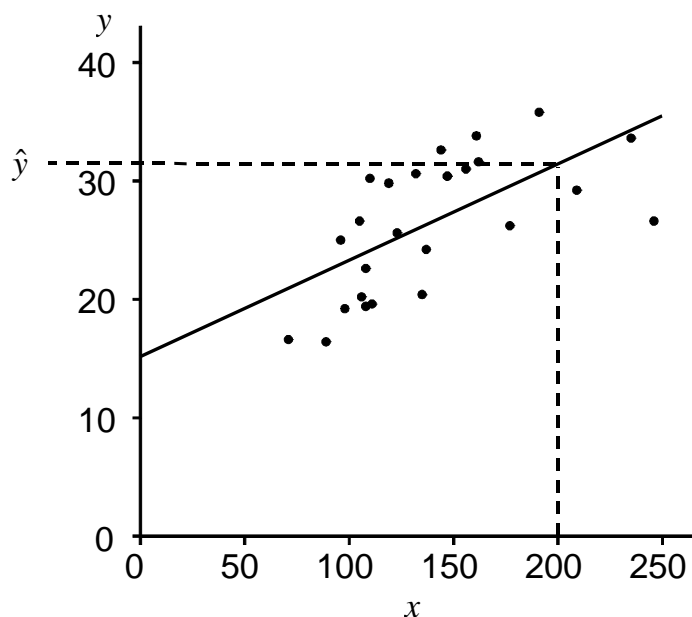


Abb. 6.3: Regressionsgerade für die Regendaten, mit Prognose des Ertrages (\hat{y}) bei Niederschlag $x = 200$ mm.

Zur besseren Beschreibung eines linearen Zusammenhanges ist es oft hilfreich, eine Gerade durch die Punktwolke zu legen. Man spricht in diesem Zusammenhang von einer **Regressionsgeraden**. Neben einer Beschreibung des Zusammenhanges kann eine Regression aber auch zur Vorhersage (Prognose) dienen.

Beispiel: Für die Regendaten ist eine Regressionsgerade in der Abb. 6.3 dargestellt.

Die Regressionsgerade hilft zum einen bei der Interpretation der Punktwolke, indem sie den linearen Trend veranschaulicht. Sie kann hier außerdem für eine Ertragsprognose in einem neuen Jahr genutzt werden, vorausgesetzt, wir haben für dieses Jahr die Niederschlagsmenge zwischen April und Juni (x) ermittelt. Bei einem Niederschlag vom $x = 200$ mm erwarten wir einen Ertrag von etwa $\hat{y} = 32$ dt/ha, wie ein Blick auf Abb. 6.3 zeigt. Die Vorhersage wird mit einem "Dach" auf dem y kenntlich gemacht. Eine Regressionsgerade ist also in verschiedener Hinsicht sehr hilfreich.

Die nun zu klärende Frage ist, wie man am besten eine Gerade durch die Punktwolke legt. Dazu betrachten wir zunächst die Geradengleichung. Sie ist gegeben durch

$$\eta = E(y) = \alpha + \beta x$$

wobei

$E(y)$ = Erwartungswert von y

α = Achsenabschnitt

β = Steigung

Die wesentliche Interpretation der Steigung β ist folgende: Steigt x um eine Einheit, so steigt der erwartete Wert für y um β Einheiten. Dies veranschaulicht das

Steigungsdreieck in Abb. 6.4(a). Der Achsenabschnitt α ist der erwartete Wert für y wenn $x = 0$ ist.

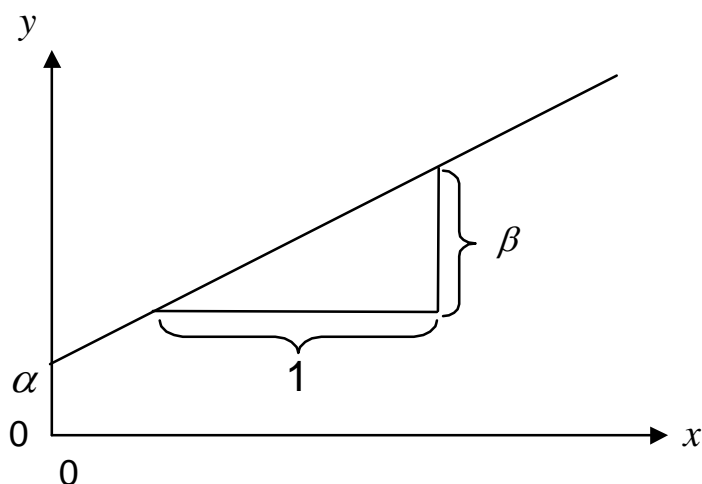


Abb. 6.4(a): Veranschaulichung der Interpretation der Regressionsgleichung $\eta = \alpha + \beta x$ mit Steigungsdreieck.

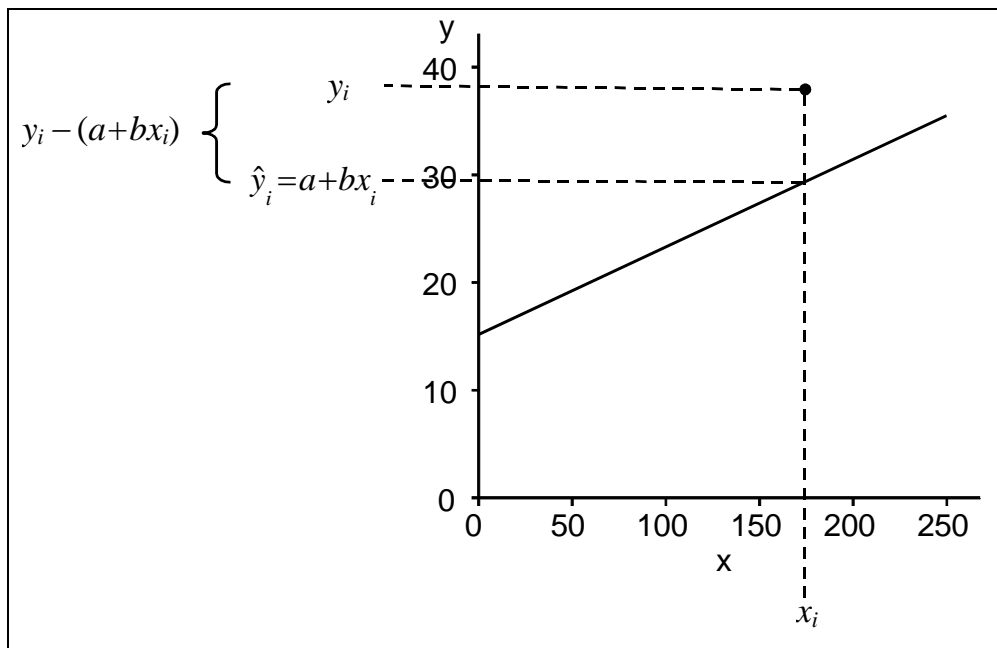


Abb. 6.4(b): Veranschaulichung der Abweichung eines Datenpunktes von der geschätzten der Regressionsgerade.

Eine Gerade legen wir nun so durch die Punktwolke, dass die Abweichung zwischen Gerade und den einzelnen Punkten möglichst klein wird. Die vertikale Abweichung des Punktes für die i -te Beobachtung von der zu schätzenden Regressionsgeraden ist gegeben durch

$$y_i - (a + bx_i) \quad ,$$

wobei a und b Schätzwerte für die Parameter α und β sind [siehe Abb. 6.4(b)]. Diese Abweichungen sollen minimiert werden. Bei der Regression wird die vertikale Abweichung betrachtet, weil es bei der Regression um die möglichst genaue Schätzung von y_i mittels der angepassten Regressionsgerade geht.

Um die Minimierung der vertikalen Abweichungen praktikabel zu gestalten, suchen wir nach einem Kriterium für die Güte der Anpassung der Geraden an die Punkte, welches dann optimiert wird. Zu denken wäre zunächst an die Summe der Abweichungen $y_i - (a + bx_i)$. Da die Abweichungen aber für gut passende Geraden mal positiv und mal negativ werden können, ist dies kein geeignetes Maß. Um das Problem der wechselnden Vorzeichen auszuschließen, können stattdessen die Abweichungen quadriert und dann summiert werden. Dies führt zur bereits von der Varianz her bekannten **Summe der Abweichungsquadrate** oder auch **Summe der Fehlerquadrate**:

$$SQ_{Fehler} = \sum_{i=1}^n [y_i - (a + bx_i)]^2$$

Dieses ist nun das Zielkriterium, welches minimiert werden soll. Wir bestimmen die Werte für a und b so, dass SQ_{Fehler} minimal wird. Hierzu können wir SQ_{Fehler} als Funktion der zwei Variablen a und b betrachten. Das Optimum einer Funktion

mehrerer Variablen bestimmt man, indem die partiellen Ableitungen der Funktion nach den Variablen gleich Null gesetzt werden (bei der Berechnung der partiellen Ableitung nach einer Variablen werden alle anderen Variablen als Konstanten betrachtet und dann die gewöhnliche Ableitung berechnet). In unserem Fall finden wir:

$$\frac{\partial S Q_{\text{Fehler}}}{\partial b} = 2 \sum_{i=1}^n [y_i - (a + bx_i)](-1)x_i = 0$$

$$\frac{\partial S Q_{\text{Fehler}}}{\partial a} = 2 \sum_{i=1}^n [y_i - (a + bx_i)](-1) = 0$$

Diese Gleichungen sind die sog. **Normalengleichungen**. Auflösen der Normalengleichungen nach den Unbekannten a und b liefert die Kleinstquadratschätzungen für die Parameter α und β . Lösen von $\partial S Q_{\text{Fehler}} / \partial a = 0$ nach a liefert:

$$\sum_{i=1}^n [y_i - (a + bx_i)] = n\bar{y} - n(a + b\bar{x}) = 0 \Leftrightarrow a = \bar{y} - b\bar{x}$$

Einsetzen dieser Gleichung in $\partial S Q_{\text{Fehler}} / \partial b = 0$ liefert:

$$\begin{aligned} \frac{\partial S Q_{\text{Fehler}}}{\partial b} &= \sum_{i=1}^n [y_i - (\bar{y} - b\bar{x} + bx_i)]x_i \\ &= \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - b \left[\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right] \\ &= \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n} - b \left[\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right] \\ &= SP_{xy} - b SQ_x = 0 \\ \Leftrightarrow b &= \frac{SP_{xy}}{SQ_x} \end{aligned}$$

Die hier beschriebene Methode heißt **Methode der kleinsten Quadrate**. Die Kleinstquadratschätzungen nach Auflösen der Normalengleichungen sind noch einmal in dem folgenden Kasten zusammengefaßt:

Schätzung der Regressionsgerade (Methode der kleinsten Quadrate):

$$\begin{aligned} b &= \frac{SP_{xy}}{SQ_x} \\ a &= \bar{y} - b\bar{x} \end{aligned}$$

wobei

$$SP_{xy} = \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n}$$

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

Geschätzte Regressionsgerade:

$$\hat{y} = a + bx$$

Man beachte, dass SQ_x und SP_{xy} dieselben Größen sind, die auch zur Berechnung der Korrelation benötigt wurden.

Beispiel: Für die Regendaten finden wir:

$$SP_{xy} = 3923,38$$

$$SQ_x = 48335,88$$

$$\bar{x} = 139,35$$

$$\bar{y} = 26,45$$

$$b = \frac{3923,38}{48335,88} = 0,0812$$

$$a = 26,45 - 0,0812 \cdot 139,35 = 15,14$$

$$\hat{y} = 15,14 + 0,0812x$$

Dies ist die Gleichung, welche für das Zeichnen der Gerade in Abb. 6.3 verwendet wurde. Die Steigung kann wie folgt interpretiert werden: steigt die Regenmenge um einen mm, so steigt der Ertrag um 0,0812 dt/ha = 8,12 kg/ha. Außerdem können wir die Regressionsgleichung zur Prognose nutzen. Wenn in einem neuen Jahr von April bis Juni 200 mm Regen gefallen sind, so schätzen wir den erwarteten Ertrag nach

$$\hat{y} = 15,14 + 0,0812 \cdot 200 = 31,37$$

Wir erwarten also einen Ertrag von 31,37 dt/ha, was nicht heißt, dass dies der tatsächlich realisierte Ertrag sein wird. Die Schätzung besagt lediglich, dass wir bei einem Niederschlag von 200 mm im Mittel 31,37 dt/ha erwarten. Da der Ertrag von vielen Faktoren beeinflusst wird, kann in verschiedenen Jahren mit derselben Niederschlagsmenge von 200 mm mal ein niedrigerer, mal ein höherer Ertrag als 31,37 dt/ha realisiert werden. 31,37 dt/ha ist einfach die beste Schätzung, die wir im Voraus aufgrund der historischen Daten machen können (Die Daten sind hier sehr alt und müssten durch neuere Daten aktualisiert werden, weil das Ertragsniveau im Laufe der Jahre aufgrund von Fortschritten in der Züchtung und im Pflanzenbau gestiegen ist).

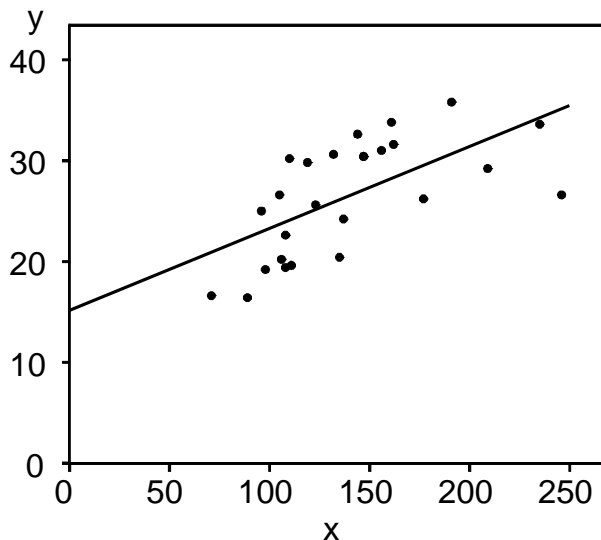


Abb. 6.4(c): Lineare Regression für Regendaten mit Streudiagramm [x = Regenmenge zwischen April und Juni in (mm); y = Ertrag (dt/ha)].

Im Zusammenhang mit der Prognose ist folgende Bezeichnungsweise üblich:

x = Prädiktorvariable, Einflussvariable (unabhängige Variable)

y = Zielvariable (abhängige Variable)

Das Wortpaar "abhängig/unabhängig" impliziert eine Ursache-Wirkungs-Beziehung der Form

$$x \rightarrow y$$

Eine solche muss aber nicht notwendigerweise gegeben sein, wie bereits im Zusammenhang mit der Korrelation in Abschnitt 6.1 diskutiert wurde. Ich bevorzuge daher im allgemeinen das Begriffspaar "Prädiktorvariable/Zielvariable" bzw. Einflussvariable/Zielvariable). Um eine Ursache-Wirkungs-Beziehung eindeutig nachzuweisen, muss die Prädiktorvariable in einem Versuch gezielt variiert werden, während alle anderen Umweltfaktoren konstant gehalten oder durch Randomisation ausgeschaltet werden. Bei der Erhebung zum Zusammenhang zwischen Ertrag und Regenmenge ist dies beispielsweise nicht gegeben. Es ist daher nicht auszuschließen, dass nicht die variierende Regenmenge, sondern ausschließlich andere Umweltfaktoren, die auch von Jahr zu Jahr schwankten, für die Ertragsbildung verantwortlich sind. Dies mag in diesem Beispiel als spitzfindig erscheinen, weil unsere praktische Erfahrung eindeutig zu beweisen scheint, dass der Regen ursächlich mit dem Ertrag zusammenhängt. Und trotzdem liefern die Erhebungsdaten nur einen Hinweis, aber keinen Beweis, dass diese Annahme zutrifft. Um den Einfluss der Wasserzufuhr experimentell einwandfrei nachzuweisen, kann beispielsweise ein Experiment durchgeführt werden, bei dem die Versuchsfläche überdacht wird und die Wasserzufuhr künstlich in mehreren Abstufungen variiert wird. Wir wiederholen hiermit die wichtige Feststellung, dass der strenge Nachweis von Kausalbeziehungen nur in einem Experiment möglich ist, nicht aber in einer Erhebung. Allerdings können in einer Erhebung weitere theoretische Überlegungen neben der Signifikanz eines Zusammenhanges sehr wohl weitere Indizien für eine Kausalbeziehung liefern.

Bei der linearen Regression ist weiterhin zu beachten, dass die Zuordnung der Variablen als x - oder y -Variable wichtig ist. Die Lage der angepassten Geraden ändert sich, wenn man x und y vertauscht. Die vorherzusagende Variable ist immer als y -Variable zu wählen. Außerdem ist es in bei geplanten Versuchen oft so, dass die beobachteten Stufen von x gezielt ausgewählt werden können, z.B. die Aufwandmengen für einen Dünger. Die x -Variable ist in solchen Fällen keine Zufallsvariable, wohl aber der Ertrag als "abhängige" Variable. Die Düngermenge als fixe Größe muss daher als x -Variable verrechnet werden. Als y -Variable ist in solchen Fällen immer diejenige Variable zu verwenden, die eine Zufallsgröße darstellt, hier der Ertrag.

6.2.1 Streuungszerlegung

Wenden wir uns nun der **Streuungszerlegung** für die lineare Regression zu. Die Streuungszerlegung ist die Grundlage der Varianzanalyse für die lineare Regression, die im folgenden hergeleitet wird.

Wir betrachten zunächst die Abweichung einer Beobachtung vom Gesamtmittel. Diese kann wie folgt aufgespalten werden:

$$y_i - \bar{y} = y_i - \hat{y}_i + \hat{y}_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

wobei $\hat{y}_i = a + bx_i$ der durch die Regression an der Stelle x_i vorhergesagte Erwartungswert ist und $\bar{y} = \sum_i y_i / n$. Die Gesamtabweichung lässt sich also aufspalten in eine Abweichung der Beobachtung von der Regression $y_i - \hat{y}_i$ (Fehler) und eine Abweichung des durch die Regression vorhergesagten Wertes vom Gesamtmittel $\hat{y}_i - \bar{y}$ (durch Regression erklärter Teil). Die Aufspaltung der Gesamtstreuung in zwei Komponenten kann auch der Abb. 6.5 entnommen werden.

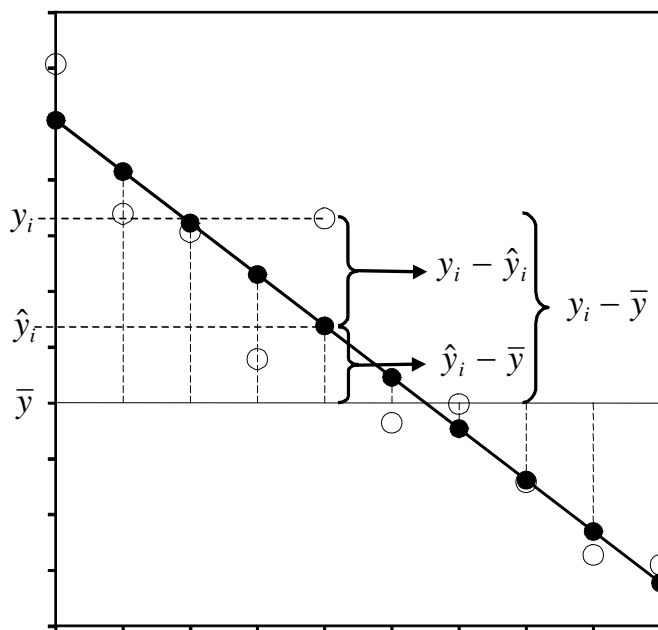


Abb. 6.5: Zerlegung der Gesamtabweichung $y_i - \bar{y}$ in einen durch die Regression erklärten Teil $\hat{y}_i - \bar{y}$ und eine Abweichung von der Regression oder Fehler $y_i - \hat{y}_i$.

Analog zu der einzelnen Abweichung vom Gesamtmittel lässt sich SQ_y aufspalten (**Quadratsummenzerlegung**). Es gilt

$$\begin{aligned} SQ_y &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \end{aligned} \quad (6.2.1)$$

wobei

$$\hat{y}_i = a + bx_i = \bar{y} - b\bar{x} + bx_i = \bar{y} + b(x_i - \bar{x})$$

Die Summe der Kreuzprodukte in (6.2.1) ist gleich Null, wie wir im folgenden sehen. Es ist

$$y_i - \hat{y}_i = y_i - \bar{y} - b(x_i - \bar{x}) \quad \text{und} \quad (6.2.2)$$

$$\hat{y}_i - \bar{y} = b(x_i - \bar{x}) \quad (6.2.3)$$

so dass

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^n [y_i - \bar{y} - b(x_i - \bar{x})]b(x_i - \bar{x}) \\ &= b \sum_{i=1}^n [(y_i - \bar{y})(x_i - \bar{x}) - b(x_i - \bar{x})^2] = b SP_{xy} - b^2 SQ_x \end{aligned}$$

Da $b SP_{xy} = (SP_{xy})^2 / SQ_x = (SP_{xy} / SQ_x)^2 SQ_x = b^2 SQ_x$ ist, folgt, dass $b SP_{xy} = b^2 SQ_x$, so dass die Summe der Kreuzprodukte in (6.2.1) gleich Null ist. Somit ist

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ SQ_y &= SQ_{Fehler} + SQ_{Regression} \end{aligned}$$

Die Gesamtstreuung zerfällt also in einen durch die Regression erklärten Teil ($SQ_{Regression}$) und die Streuung um die Regression, also den unerklärten Teil (SQ_{Fehler}). Zur einfachen Berechnung nutzen wir, dass aus (6.2.3) folgt:

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SQ_{Regression} = b^2 SQ_x$$

Hiermit finden wir SQ_{Fehler} einfach durch Differenzbildung:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = SQ_{Fehler} = SQ_y - b^2 SQ_x$$

$SQ_{Regression}$ ist der durch die Regression erklärte Anteil von SQ_y . Diesen Anteil können wir auch relativ ausdrücken und als **Bestimmtheitsmaß** bezeichnen. Es kann gezeigt werden, dass

$$B = \frac{SQ_{Regression}}{SQ_y} = r^2$$

Das **Bestimmtheitsmaß** für die einfache lineare Regression ist definiert durch

$$B = \frac{SQ_{Regression}}{SQ_y} = r^2$$

wobei

$$SQ_y = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{und} \quad SQ_{Regression} = b^2 SQ_x$$

Es entspricht dem quadrierten Korrelationskoeffizienten (r^2) und gibt den Anteil der gesamten Streuung der Zielvariablen (y) an, der durch die Prädiktorvariable (x) erklärt werden kann.

Die Streuungszerlegung kann in einer Varianzanalyse-Tabelle zusammengefaßt werden:

Varianzanalyse-Tabelle:					
Ursache	FG	SQ	MQ	E(MQ)	F
Regression	1	$SQ_{Regression} = b^2 SQ_x$	$SQ_{Regression}$	$\sigma^2 + \beta^2 SQ_x$	$F_{Vers} = \frac{b^2 SQ_x}{s^2}$
Fehler	$n - 2$	$SQ_{Fehler} = SQ_y - b^2 SQ_x$	s^2	σ^2	

wobei

$$s^2 = MQ_{Fehler} = SQ_{Fehler} / (n - 2)$$

Wenn die Steigung $\beta = 0$ ist, so folgt F_{Vers} einer F -Verteilung mit 1 und $(n - 2)$ Freiheitsgraden. Dies führt zu folgender Entscheidungsregel:

Fragestellung: Ist die Steigung der Regression signifikant von Null verschieden?

$H_0: \beta = 0$

$H_A: \beta \neq 0$

Berechne:

$$F_{Vers} = \frac{b^2 SQ_x}{s^2}$$

mit

$$s^2 = MQ_{Fehler} = SQ_{Fehler} / (n - 2)$$

$$SQ_{Fehler} = SQ_y - b^2 SQ_x$$

Bestimme:

α = Signifikanzniveau

$$F_{Tab} = F_{(1 - \alpha; 1; n - 2)} \quad (\text{Tab. VI})$$

$$F_{Vers} > F_{Tab} \Rightarrow \text{verwerfe } H_0$$

$$F_{Vers} \leq F_{Tab} \Rightarrow \text{behalte } H_0 \text{ bei}$$

Voraussetzung: Abweichungen von der Geraden sind normalverteilt.

Man beachte, dass die zu testende Nullhypothese bezüglich β eine zweiseitige Alternative hat (Abweichung von H_0 , wenn $\beta < 0$ oder $\beta > 0$). Allerdings ist eine einseitige F-Verteilung für den Test zu wählen: Grund: F_{Vers} wird groß wenn β^2 groß wird [siehe $E(MQ_{Regression})$], und dies ist immer dann der Fall, wenn β stark von Null abweicht, egal ob diese Abweichung positiv oder negativ ist. Daher ist der Ablehnungsbereich für F_{Vers} einseitig.

Beispiel: Für die Regendaten finden wir

$$SP_{xy} = 3923,38$$

$$SQ_x = 48335,88$$

$$SQ_y = 804,58$$

$$b = 0,0812$$

$$SQ_{Regression} = 0,0812^2 \cdot 48335,88 = 318,70$$

$$SQ_{Fehler} = 804,58 - 318,70 = 485,88$$

Varianzanalyse-Tabelle:

Ursache	FG	SQ	MQ	F_{Vers}
Regression	1	318,70	318,70	15,74
Fehler	24	485,88	20,25	

$$F_{Tab} = F_{(\alpha = 5\%; 1; 24)} = 4,26 < F_{Vers} = 15,74$$

⇒ Die Steigung ist signifikant von Null verschieden

Das Bestimmtheitsmaß beträgt $B = 318,70/804,58 = 0,396 = 39,6\%$. Alternativ berechnen wir dies auch mittels des in Abschnitt 6.1 ermittelten Korrelationskoeffizienten: $r = 0,629 \Rightarrow B = r^2 = 0,629^2 = 0,396$. Somit kann rund 40% der Ertragschwankungen durch die Regenmenge zwischen April und Juni erklärt werden.

Dem F-Test der obigen Varianzanalyse liegt das folgende **lineare Modell** zugrunde:

$$y_i = \alpha + \beta x_i + e_i$$

wobei

y_i = i -ter Messwert der Zielvariable

x_i = i -ter Messwert der Einflussvariable

e_i = Abweichung des i -ten Messwertes der Zielvariable von der Regressionsgeraden; normalverteilt mit Mittelwert 0 und Varianz σ^2

Die Geradengleichung besagt folgendes: Wenn die Einflussvariable den Wert x_i hat, erwarten wir für die Zielvariable im Mittel den Wert $\alpha + \beta x_i$. Diesen Erwartungswert kann man direkt anhand der graphischen Darstellung der Regressionsgeraden ablesen, wie in Abb. 6.6 gezeigt. Der im Einzelfall beobachtete Wert y_i entspricht aber nicht exakt dem Erwartungswert, sondern er weicht mehr oder weniger stark davon ab. Die Abweichung e_i folgt einer Normalverteilung mit Mittelwert 0 und Varianz σ^2 .

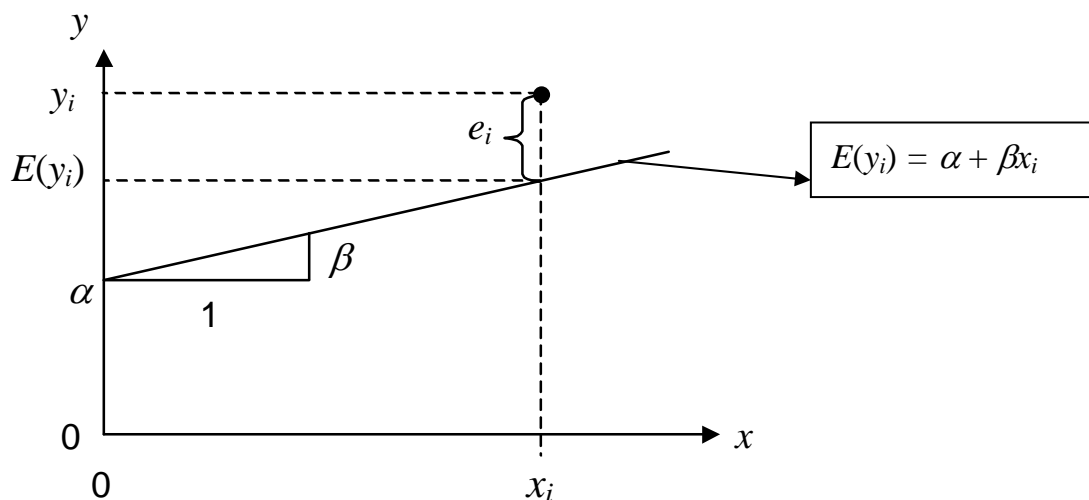


Abb. 6.6: Schematische Darstellung des Modells der linearen Regression.

6.2.2 t-Tests und Vertrauensintervalle

Anstelle des F-Tests der Varianzanalyse kann der Regressionskoeffizient auch mit einem t-Test geprüft werden. Dies liegt daran, dass

$$t = \sqrt{F}$$

einer t-Verteilung folgt, falls die F-Statistik nur einen Freiheitsgrad im Zähler hat, wie in unserem Fall. Der t-Test kann wie folgt durchgeführt werden:

Fragestellung: Ist die Steigung der Regression signifikant von Null verschieden?

$$H_0: \beta = 0$$

$$H_A: \beta \neq 0$$

Berechne:

$$t_{Vers} = \frac{|b| \sqrt{SQ_x}}{s}$$

mit

$$s^2 = SQ_{Fehler} / (n - 2)$$

$$SQ_{Fehler} = SQ_y - b^2 SQ_x$$

Bestimme:

$$t_{Tab}(FG = n - 2; \alpha) \quad (\text{Tab. II, zweiseitig})$$

α = Signifikanzniveau (nicht zu verwechseln mit dem Achsenabschnitt)

$$t_{Vers} > t_{Tab} \Rightarrow \text{verwerfe } H_0$$

$$t_{Vers} \leq t_{Tab} \Rightarrow \text{behalte } H_0 \text{ bei}$$

Voraussetzung: y ist normalverteilt.

Beispiel: Für die Regendaten finden wir

$$SQ_x = 48335,88$$

$$SQ_y = 804,58$$

$$b = 0,0812$$

$$SQ_{Fehler} = 804,58 - 0,0812^2 \cdot 48335,88 = 485,88$$

$$s^2 = 485,88 / 24 = 20,25$$

$$n = 26$$

$$t_{Vers} = \frac{0,0812 \sqrt{48335,88}}{\sqrt{20,25}} = 3,97$$

$$t_{Tab}(FG = 24; \alpha = 5\%) = 2,064 < t_{Vers} \Rightarrow \text{Steigung ist signifikant von Null verschieden}$$

Für dieselben Daten hatten wir in Abschnitt 6.1 bereits $H_0: \rho = 0$ geprüft und denselben t-Wert erhalten ($t_{Vers} = 3,97$)! Dies ist kein Zufall. Wenn die Korrelation Null ist, muss dies auch für den Regressionskoeffizienten gelten. Außerdem kann gezeigt werden, dass

$$t_{Vers} = \frac{|b|\sqrt{SQ_x}}{s} = \frac{|SP_{xy}|\sqrt{SQ_x}}{SQ_x\sqrt{SQ_y - SP_{xy}^2 / SQ_x}}\sqrt{n-2} = \frac{|SP_{xy}|}{\sqrt{SQ_x SQ_y} \sqrt{1 - SP_{xy}^2 / (SQ_x SQ_y)}}\sqrt{n-2}$$

$$= \frac{|r|}{\sqrt{1-r^2}}\sqrt{n-2}$$

ist, was der t-Statistik für den Test des Korrelationskoeffizienten entspricht. Beide Tests liefern also immer exakt dasselbe Ergebnis.

Neben einem Test können für verschiedene Größen Vertrauensintervalle berechnet werden, wie der folgende Kasten zeigt:

Vertrauensintervalle

Steigung: $b \pm t_{Tab} \frac{s}{\sqrt{SQ_x}}$

Regressionslinie an Stelle x_0 : $a + bx_0 \pm t_{Tab} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SQ_x}}$

Vorhersage eines neuen Einzelwertes an Stelle x_0 (Prognoseintervall):

$$a + bx_0 \pm t_{Tab} s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SQ_x}}$$

$t_{Tab}(FG = n - 2; \alpha)$ (Tab. II, zweiseitig)

α = Signifikanzniveau (nicht zu verwechseln mit dem Achsenabschnitt)

Beispiel: Für die Regendaten berechnen wir alle oben genannten Vertrauensintervalle ($\alpha = 5\%$).

Regressionskoeffizient:

$$0,0812 \pm 2,064 \frac{\sqrt{20,25}}{\sqrt{48335,88}} = 0,0812 \pm 0,0422 = (0,0390; 0,1234)$$

Die 95% Vertrauensgrenzen für den wahren Regressionskoeffizienten β sind 0,0390 und 0,1234.

Für die Regressionsgerade sowie für die Vorhersage von y können wir ein Vertrauensband berechnen, indem für verschiedene Werte x_0 die Vertrauensgrenzen berechnet und dann zu Kurven verbunden werden. Dies ist von Hand sehr mühsam, so dass wir hier nur einen Punkt $x_0 = 200$ betrachten.

Regressionsgerade:

$$SQ_x = 48335,88$$

$$b = 0,0812$$

$$a = 15,14$$

$$a + bx_0 = 15,14 + 0,0812 \cdot 200 = 31,37$$

$$t_{Tab} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SQ_x}} = 2,064 \cdot \sqrt{20,25} \sqrt{\frac{1}{26} + \frac{(200 - 139,35)^2}{48335,88}} = 9,288 \sqrt{0,1178} = 3,15$$

Grenzen: $31,37 - 3,15 = 28,22$ und $31,37 + 3,15 = 34,52$.

Vorhersage eines neuen Wertes:

$$t_{Tab} s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SQ_x}} = 2,064 \cdot \sqrt{20,25} \sqrt{1 + \frac{1}{26} + \frac{(200 - 139,35)^2}{48335,88}} = 9,288 \sqrt{1,1178} = 9,81$$

Grenzen: $31,37 - 9,81 = 21,56$ und $31,37 + 9,81 = 41,18$.

Die Grenzen für die Vorhersage sind viel weiter als die Grenzen für die Regression. Der Grund ist, dass die Regressionslinie nur den Erwartungswert an einer Stelle x_0 wiedergibt, während bei der Vorhersage eines neuen Wertes noch die Streuung der Einzelwerte um diesen Erwartungswert in Rechnung zu stellen ist. In den Abbildungen 6.7 und 6.8 sind für beide Fälle Vertrauensbänder für beliebige Werte x_0 gezeigt.

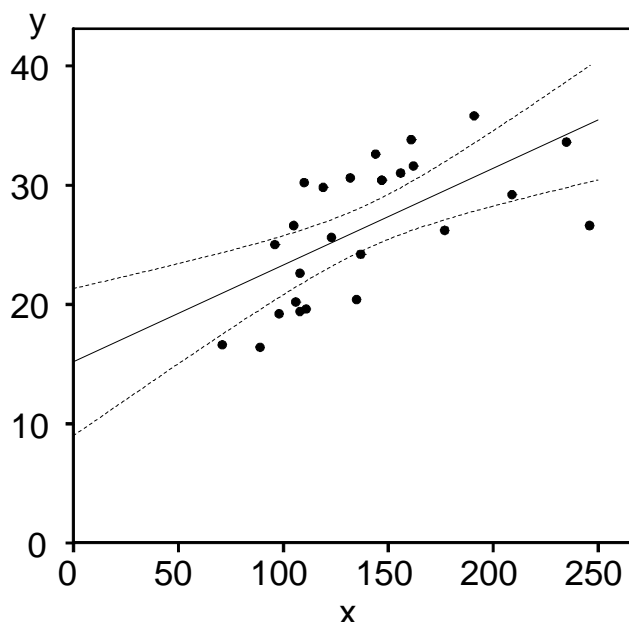


Abb. 6.7: Regressionsgerade für Regendaten mit 95%-Vertrauensintervall für den Verlauf der Regressionsgerade.

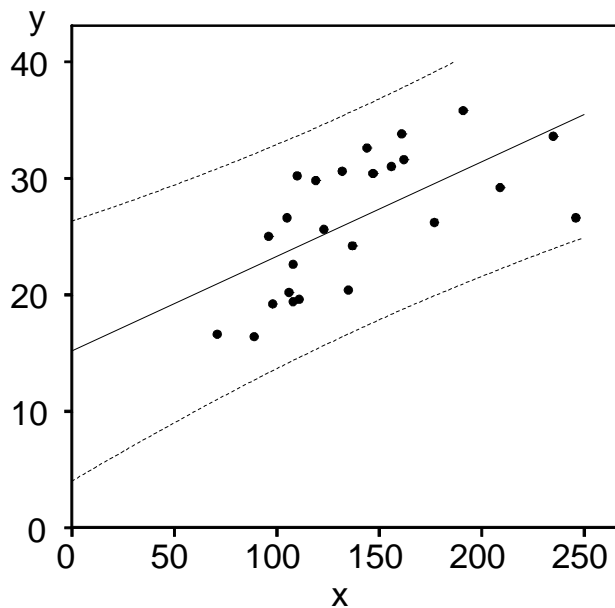


Abb. 6.8: Regressionsgerade für Regendaten mit 95%-Vertrauensintervall für Vorhersage eines neuen y -Wertes bei gegebenem x -Wert.

Wichtig ist hier der Hinweis, dass das Vertrauensband (Konfidenzband) in den Abbildungen 6.7 und 6.8 kein simultanes Konfidenzband ist, welches simultan, also versuchsbezogen, an allen x -Stellen den Fehler 1. Art einhalten wurde. Hierzu wären ähnliche Modifikationen notwendig, wie die des Tukey-Test gegenüber des LSD-Tests (siehe Kap. 4.5).

Extrapolation: Im obigen Beispiel haben wir eine Vorhersage des Ertrages mit der Regenmenge betrachtet. Eine Vorhersage ist nur für solche x -Werte sinnvoll und zulässig, die sich im Bereich der beobachteten x -Werte befinden. Man spricht hierbei von **Interpolation**. Es ist dagegen nicht zulässig, eine Vorhersage für x -Werte außerhalb des beobachteten Bereichs zu machen (**Extrapolation**). Es ist beispielsweise nicht zulässig, eine Vorhersage für $x = 0$ mm Regen zu machen, weil dieser x -Wert deutlich unterhalb der beobachteten x -Werte liegt. Ob der Ertragsverlauf bei niedrigen Niederschlägen tatsächlich linear verläuft, was für diese Voraussage (Extrapolation) angenommen werden müßte, können wir nicht nachweisen, da in diesem Bereich keine x -Werte beobachtet wurden (Dies gilt ungeachtet der Tatsache, dass wir hier aus biologischen Gründen natürlich $y = 0$ bei $x = 0$ erwarten würden).

Regression zum Mittel: Man beachte, dass eine angepasste Regressionsgerade immer durch den Punkt (\bar{x}, \bar{y}) verlaufen muss. Dies folgt aus der Tatsache, dass der für die Stelle \bar{x} vorhergesagte Wert gleich \bar{y} ist, denn es gilt:

$$a + b\bar{x} = \bar{y} - b\bar{x} + b\bar{x} = \bar{y}$$

Schauen wir uns nun den Mittelwert und die Varianz der für die beobachteten Datenpunkte vorhergesagten Werte

$$\hat{y}_i = a + bx_i$$

an. Wir finden

$$\bar{\hat{y}} = \frac{\sum_{i=1}^n \hat{y}_i}{n} = \frac{\sum_{i=1}^n (a + bx_i)}{n} = a + b\bar{x} = \bar{y} - b\bar{x} + b\bar{x} = \bar{y}$$

Die vorhergesagten y -Werte haben also denselben Mittelwert wie die beobachteten y -Werte! Außerdem finden wir

$$s_{\hat{y}}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}{n-1} = \frac{\sum_{i=1}^n (a + bx_i - a - b\bar{x})^2}{n-1} = \frac{b^2 SQ_x}{n-1} = \frac{(SP_{xy})^2}{(n-1)SQ_x} = \frac{(SP_{xy})^2}{(n-1)SQ_x SQ_y} SQ_y = r^2 s_y^2$$

Da $r^2 \leq 1$ ist, gilt $s_{\hat{y}}^2 \leq s_y^2$. Die vorhergesagten y -Werte streuen also weniger als die beobachteten y -Werte! Da beide denselben Mittelwert \bar{y} haben, müssen die vorhergesagten Werte tendenziell näher beim Mittel liegen als die Originalwerte. Man spricht in diesem Zusammenhang von der "**Regression zum Mittel**" (Rückfall oder Rückschritt zum Mittel).

Beispiel: Die "Regression zum Mittel" wurde erstmals von Francis Galton in der Genetik beobachtet. Er untersuchte den Zusammenhang zwischen der Körpergröße von Eltern und ihren Kindern (im Erwachsenenalter). Er fand einen linearen Zusammenhang. Die Steigung der Regression war jedoch nicht genau Eins, wie man hätte erwarten können, sondern kleiner als Eins. Die Kinder wichen im Schnitt weniger stark vom Mittel ab als ihre Eltern. Es hatte also offenbar eine "Regression" zurück zum Mittel stattgefunden (Lynch, M. und Walsh, B. 1998, Genetics and the analysis of quantitative traits, Sinauer, Sunderland).

Die oben abgeleitete Beziehung $s_{\hat{y}}^2 = r^2 s_y^2$ hat noch eine weitere Interpretation: Von der Gesamtstreuung der beobachteten Werte (s_y^2) wird immer nur ein Teil durch die Regression erklärt. Dieser Anteil ist gegeben durch die Varianz der vorhergesagten Werte ($s_{\hat{y}}^2$). Der relative Anteil der erklärten Streuung ist $B = r^2$ (Bestimmtheitsmaß), wie wir bereits weiter oben festgestellt hatten.

6.3 Vergleich von Korrelation und Regression

Im folgenden werden zusammenfassend einige Eigenschaften von Regression und Korrelation gegenübergestellt.

(1) Sowohl Korrelation als auch Regression eignen sich nur zur Erfassung **linearer Zusammenhänge**. Sie sind nicht geeignet, nichtlineare Zusammenhänge aufzudecken. Beide machen die Voraussetzung der Normalverteilung zumindest einer der beiden Variablen.

(2) Die Korrelation beschreibt den Zusammenhang zwischen zwei Variablen mit einer **dimensionslosen Maßzahl**. Die Daten müssen einer bivariaten Normalverteilung folgen. Sowohl x als auch y müssen als zufallsverteilt betrachtet werden können.

(3) Die Regression beschreibt, wie sich y ändert, falls x um eine Einheit erhöht wird. Die Zielvariable y muss hier einer Normalverteilung folgen und als zufallsverteilt betrachtet werden können. Die Einflussvariable x kann dagegen (muss aber nicht; siehe Regendaten) sogar in einem Experiment systematisch variiert werden. In diesem Fall ist nur die Regression, nicht aber die Korrelation ein geeignetes Auswertungsverfahren.

Beispiel: In einem Feldversuch wird die Wirkung einer Mistdüngung (x) auf den Ertrag von Weizen (y) geprüft. Hierzu werden systematisch die Gaben 0, 2, 4, 6, 8 und 10 kg pro Parzelle geprüft. Dieser Versuch kann mittels Regression, nicht aber mittels Korrelation ausgewertet werden.

(4) Die Güte der Anpassung einer Regressionsgeraden kann durch das Bestimmtheitsmaß $B = r^2$ beschrieben werden. Das Bestimmtheitsmaß gibt an, welcher Teil der Gesamtstreuung der y -Variable durch die Regression auf x erklärt werden kann. Die Korrelation r sollte dagegen nicht zur Beschreibung der Güte der Anpassung einer Regression verwendet werden, insbesondere dann nicht, wenn die Einflussvariable x systematisch variiert wurde.

(5) Die Regression kann für Prognosen herangezogen werden, die Korrelation nicht.

(6) Der Regressionskoeffizient b sagt nichts über die Stärke des Zusammenhanges aus, wohl aber das Bestimmtheitsmaß.

(7) In vielen Lehrbüchern findet man die Aussage, eine Regression solle nur durchgeführt werden, wenn eine einseitige Ursache-Wirkungs-Beziehung der Form $x \rightarrow y$ bestehe. Diese Aussage ist m.E. nicht zutreffend. Wichtig ist aber, dass die vorherzusagende Variable als y gewählt wird.

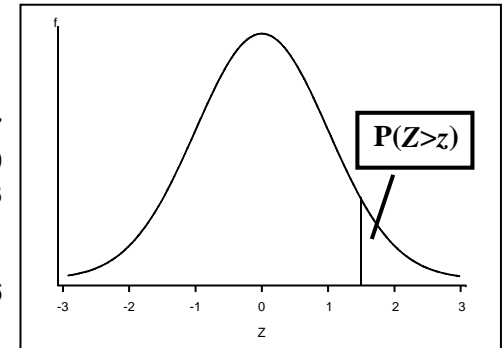
Beispiel: Bei Futter- und Nahrungsgetreide wird ein möglichst geringer Besatz mit Pilztoxinen gefordert. Manche Toxine, wie das Ergosterol, sind sehr leicht zu messen, während andere, wie das Nivalenol, nur sehr schwer messbar sind. Es bietet sich daher an, für eine Zahl von Proben beide Toxine zu bestimmen und dann eine Regression des Ergosterolgehaltes (x) auf den Nivalenolgehalt (y) durchzuführen. Die Regressionsgerade kann dann genutzt werden, um in Routineuntersuchungen, bei denen aus Kostengründen nur das Ergosterol (x) gemessen wird, den Gehalt an Nivalenol (y) vorherzusagen. Obwohl hier die Regression sehr sinnvoll eingesetzt werden kann, besteht keine eindeutige Kausalbeziehung zwischen den beiden Variablen. Beide sind vielmehr kausal beeinflusst von der Stärke des Pilzbefalls.

Beispiel: Bei Bäumen findet man oft eine sehr enge Beziehung zwischen Brusthöhenumfang des Stammes und seines Volumens. Das Stammvolumen ist von unmittelbarem ökonomischen Interesse, lässt sich aber nur mit großem Aufwand vor dem Fällen des Baumes bestimmen. Daher wird die enge Beziehung zum Brusthöhenumfang zur indirekten Messung des Volumens herangezogen. Hierbei wird eine Regression des Volumens (y) auf den Brusthöhenumfang (x) durchgeführt

und die Regressionsgleichung anschließend für eine Volumenprognose herangezogen. Hiermit ist in keiner Weise eine Ursache-Wirkungs-Beziehung zwischen Brusthöhenumfang und Volumen impliziert, und eine solche Beziehung gibt es auch nicht.

Tab. I: Überschreitungswahrscheinlichkeiten $P(Z > z)$ für die Standardnormalverteilung, Beispiel: $P(Z > 1,96) = 0,025$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014



Tab. II: Kritische Werte der t-Verteilung (zweiseitig)

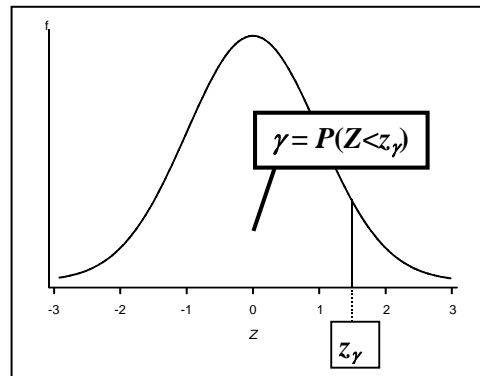
Irrtumswahrscheinlichkeit α				Irrtumswahrscheinlichkeit α			
α FG	0,05	0,01	0,001	α FG	0,05	0,01	0,001
1	12,706	63,657	636,619	51	2,008	2,676	3,492
2	4,303	9,925	31,599	52	2,007	2,674	3,488
3	3,182	5,841	12,924	53	2,006	2,672	3,484
4	2,776	4,604	8,610	54	2,005	2,670	3,480
5	2,571	4,032	6,869	55	2,004	2,668	3,476
6	2,447	3,707	5,959	56	2,003	2,667	3,473
7	2,365	3,499	5,408	57	2,002	2,665	3,470
8	2,306	3,355	5,041	58	2,002	2,663	3,466
9	2,262	3,250	4,781	59	2,001	2,662	3,463
10	2,228	3,169	4,587	60	2,000	2,660	3,460
11	2,201	3,106	4,437	61	2,000	2,659	3,457
12	2,179	3,055	4,318	62	1,999	2,657	3,454
13	2,160	3,012	4,221	63	1,998	2,656	3,452
14	2,145	2,977	4,140	64	1,998	2,655	3,449
15	2,131	2,947	4,073	65	1,997	2,654	3,447
16	2,120	2,921	4,015	66	1,997	2,652	3,444
17	2,110	2,898	3,965	67	1,996	2,651	3,442
18	2,101	2,878	3,922	68	1,995	2,650	3,439
19	2,093	2,861	3,883	69	1,995	2,649	3,437
20	2,086	2,845	3,850	70	1,994	2,648	3,435
21	2,080	2,831	3,819	71	1,994	2,647	3,433
22	2,074	2,819	3,792	72	1,993	2,646	3,431
23	2,069	2,807	3,768	73	1,993	2,645	3,429
24	2,064	2,797	3,745	74	1,993	2,644	3,427
25	2,060	2,787	3,725	75	1,992	2,643	3,425
26	2,056	2,779	3,707	76	1,992	2,642	3,423
27	2,052	2,771	3,690	77	1,991	2,641	3,421
28	2,048	2,763	3,674	78	1,991	2,640	3,420
29	2,045	2,756	3,659	79	1,990	2,640	3,418
30	2,042	2,750	3,646	80	1,990	2,639	3,416
31	2,040	2,744	3,633	81	1,990	2,638	3,415
32	2,037	2,738	3,622	82	1,989	2,637	3,413
33	2,035	2,733	3,611	83	1,989	2,636	3,412
34	2,032	2,728	3,601	84	1,989	2,636	3,410
35	2,030	2,724	3,591	85	1,988	2,635	3,409
36	2,028	2,719	3,582	86	1,988	2,634	3,407
37	2,026	2,715	3,574	87	1,988	2,634	3,406
38	2,024	2,712	3,566	88	1,987	2,633	3,405
39	2,023	2,708	3,558	89	1,987	2,632	3,403
40	2,021	2,704	3,551	90	1,987	2,632	3,402
41	2,020	2,701	3,544	91	1,986	2,631	3,401
42	2,018	2,698	3,538	92	1,986	2,630	3,399
43	2,017	2,695	3,532	93	1,986	2,630	3,398
44	2,015	2,692	3,526	94	1,986	2,629	3,397
45	2,014	2,690	3,520	95	1,985	2,629	3,396
46	2,013	2,687	3,515	96	1,985	2,628	3,395
47	2,012	2,685	3,510	97	1,985	2,627	3,394
48	2,011	2,682	3,505	98	1,984	2,627	3,393
49	2,010	2,680	3,500	99	1,984	2,626	3,392
50	2,009	2,678	3,496	∞	1,960	2,576	3,291

Tab. II(b): Kritische Werte der t-Verteilung (einseitig)

Irrtumswahrscheinlichkeit α				Irrtumswahrscheinlichkeit α			
α FG	0,05	0,01	0,001	α FG	0,05	0,01	0,001
1	6,314	31,821	318,309	51	1,675	2,402	3,258
2	2,920	6,965	22,327	52	1,675	2,400	3,255
3	2,353	4,541	10,215	53	1,674	2,399	3,251
4	2,132	3,747	7,173	54	1,674	2,397	3,248
5	2,015	3,365	5,893	55	1,673	2,396	3,245
6	1,943	3,143	5,208	56	1,673	2,395	3,242
7	1,895	2,998	4,785	57	1,672	2,394	3,239
8	1,860	2,896	4,501	58	1,672	2,392	3,237
9	1,833	2,821	4,297	59	1,671	2,391	3,234
10	1,812	2,764	4,144	60	1,671	2,390	3,232
11	1,796	2,718	4,025	61	1,670	2,389	3,229
12	1,782	2,681	3,930	62	1,670	2,388	3,227
13	1,771	2,650	3,852	63	1,669	2,387	3,225
14	1,761	2,624	3,787	64	1,669	2,386	3,223
15	1,753	2,602	3,733	65	1,669	2,385	3,220
16	1,746	2,583	3,686	66	1,668	2,384	3,218
17	1,740	2,567	3,646	67	1,668	2,383	3,216
18	1,734	2,552	3,610	68	1,668	2,382	3,214
19	1,729	2,539	3,579	69	1,667	2,382	3,213
20	1,725	2,528	3,552	70	1,667	2,381	3,211
21	1,721	2,518	3,527	71	1,667	2,380	3,209
22	1,717	2,508	3,505	72	1,666	2,379	3,207
23	1,714	2,500	3,485	73	1,666	2,379	3,206
24	1,711	2,492	3,467	74	1,666	2,378	3,204
25	1,708	2,485	3,450	75	1,665	2,377	3,202
26	1,706	2,479	3,435	76	1,665	2,376	3,201
27	1,703	2,473	3,421	77	1,665	2,376	3,199
28	1,701	2,467	3,408	78	1,665	2,375	3,198
29	1,699	2,462	3,396	79	1,664	2,374	3,197
30	1,697	2,457	3,385	80	1,664	2,374	3,195
31	1,696	2,453	3,375	81	1,664	2,373	3,194
32	1,694	2,449	3,365	82	1,664	2,373	3,193
33	1,692	2,445	3,356	83	1,663	2,372	3,191
34	1,691	2,441	3,348	84	1,663	2,372	3,190
35	1,690	2,438	3,340	85	1,663	2,371	3,189
36	1,688	2,434	3,333	86	1,663	2,370	3,188
37	1,687	2,431	3,326	87	1,663	2,370	3,187
38	1,686	2,429	3,319	88	1,662	2,369	3,185
39	1,685	2,426	3,313	89	1,662	2,369	3,184
40	1,684	2,423	3,307	90	1,662	2,368	3,183
41	1,683	2,421	3,301	91	1,662	2,368	3,182
42	1,682	2,418	3,296	92	1,662	2,368	3,181
43	1,681	2,416	3,291	93	1,661	2,367	3,180
44	1,680	2,414	3,286	94	1,661	2,367	3,179
45	1,679	2,412	3,281	95	1,661	2,366	3,178
46	1,679	2,410	3,277	96	1,661	2,366	3,177
47	1,678	2,408	3,273	97	1,661	2,365	3,176
48	1,677	2,407	3,269	98	1,661	2,365	3,175
49	1,677	2,405	3,265	99	1,660	2,365	3,175
50	1,676	2,403	3,261	∞	1,645	2,327	3,091

Tab. III: Quantile der Standardnormalverteilung. Werte für z_γ , so dass $P(Z < z_\gamma) = \gamma$, wobei Z einer Standardnormalverteilung folgt.

γ	z_γ
0,600	0,25335
0,700	0,52440
0,800	0,84162
0,900	1,28155
0,950	1,64485
0,975	1,95996
0,990	2,32635
0,995	2,57583



Ablese-Beispiel:

Suche $z_{1-\alpha/2}$ für $\alpha = 5\% = 0,05$

$$\Rightarrow \gamma = 1 - \alpha/2 = 1 - 0,05/2 = 0,975$$

$$\Rightarrow z_{1-\alpha/2} = z_\gamma = z_{0,975} = 1,95996 \approx 1,96$$

Tab. IV: Kritische Werte $\chi^2_{\gamma;v}$ der χ^2 -Verteilung mit v Freiheitsgraden, so dass $P(\chi^2 < \chi^2_{\gamma;v}) = \gamma$.

$\gamma \backslash v$	0,0005	0,005	0,025	0,975	0,995	0,9995
1	0,000	0,000	0,001	5,024	7,879	12,116
2	0,001	0,010	0,051	7,378	10,597	15,202
3	0,015	0,072	0,216	9,348	12,838	17,730
4	0,064	0,207	0,484	11,143	14,860	19,997
5	0,158	0,412	0,831	12,833	16,750	22,105
6	0,299	0,676	1,237	14,449	18,548	24,103
7	0,485	0,989	1,690	16,013	20,278	26,018
8	0,710	1,344	2,180	17,535	21,955	27,868
9	0,972	1,735	2,700	19,023	23,589	29,666
10	1,265	2,156	3,247	20,483	25,188	31,420
11	1,587	2,603	3,816	21,920	26,757	33,137
12	1,934	3,074	4,404	23,337	28,300	34,821
13	2,305	3,565	5,009	24,736	29,819	36,478
14	2,697	4,075	5,629	26,119	31,319	38,109
15	3,108	4,601	6,262	27,488	32,801	39,719
16	3,536	5,142	6,908	28,845	34,267	41,308
17	3,980	5,697	7,564	30,191	35,718	42,879
18	4,439	6,265	8,231	31,526	37,156	44,434
19	4,912	6,844	8,907	32,852	38,582	45,973
20	5,398	7,434	9,591	34,170	39,997	47,498
21	5,896	8,034	10,283	35,479	41,401	49,011
22	6,404	8,643	10,982	36,781	42,796	50,511
23	6,924	9,260	11,689	38,076	44,181	52,000
24	7,453	9,886	12,401	39,364	45,559	53,479
25	7,991	10,520	13,120	40,646	46,928	54,947
26	8,538	11,160	13,844	41,923	48,290	56,407
27	9,093	11,808	14,573	43,195	49,645	57,858
28	9,656	12,461	15,308	44,461	50,993	59,300
29	10,227	13,121	16,047	45,722	52,336	60,735
30	10,804	13,787	16,791	46,979	53,672	62,162
31	11,389	14,458	17,539	48,232	55,003	63,582
32	11,979	15,134	18,291	49,480	56,328	64,995
33	12,576	15,815	19,047	50,725	57,648	66,403
34	13,179	16,501	19,806	51,966	58,964	67,803
35	13,787	17,192	20,569	53,203	60,275	69,199
36	14,401	17,887	21,336	54,437	61,581	70,588
37	15,020	18,586	22,106	55,668	62,883	71,972
38	15,644	19,289	22,878	56,896	64,181	73,351
39	16,273	19,996	23,654	58,120	65,476	74,725
40	16,906	20,707	24,433	59,342	66,766	76,095
41	17,544	21,421	25,215	60,561	68,053	77,459
42	18,186	22,138	25,999	61,777	69,336	78,820
43	18,832	22,859	26,785	62,990	70,616	80,176
44	19,483	23,584	27,575	64,201	71,893	81,528
45	20,137	24,311	28,366	65,410	73,166	82,876
46	20,794	25,041	29,160	66,617	74,437	84,220
47	21,456	25,775	29,956	67,821	75,704	85,560
48	22,121	26,511	30,755	69,023	76,969	86,897
49	22,789	27,249	31,555	70,222	78,231	88,231
50	23,461	27,991	32,357	71,420	79,490	89,561

Tab. V: Kritische Werte $F_{(0,975; \nu_1, \nu_2)}$ der F -Verteilung mit ν_1 und ν_2 Freiheitsgraden, so dass $P(F < F_{(0,975; \nu_1, \nu_2)}) = 0,975$.

ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	50	∞
ν_2																
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3	968,6	976,7	984,9	993,1	998,1	1008	1018
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45	39,46	39,48	39,50
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17	14,12	14,01	13,90
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,75	8,66	8,56	8,50	8,38	8,26
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,27	6,14	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,37	5,27	5,17	5,11	4,98	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,40	4,28	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,94	3,81	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,60	3,47	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,35	3,22	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,23	3,16	3,03	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,07	3,01	2,87	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	2,95	2,88	2,74	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,05	2,95	2,84	2,78	2,64	2,49
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,86	2,76	2,69	2,55	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,89	2,79	2,68	2,61	2,47	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,82	2,72	2,62	2,55	2,41	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,77	2,67	2,56	2,49	2,35	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,72	2,62	2,51	2,44	2,30	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,40	2,25	2,09
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,64	2,53	2,42	2,36	2,21	2,04
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,60	2,50	2,39	2,32	2,17	2,00
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,57	2,47	2,36	2,29	2,14	1,97
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,26	2,11	1,94
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,30	2,23	2,08	1,91
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,49	2,39	2,28	2,21	2,05	1,88
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,47	2,36	2,25	2,18	2,03	1,85
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,45	2,34	2,23	2,16	2,01	1,83
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,43	2,32	2,21	2,14	1,99	1,81
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,12	1,97	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	1,99	1,83	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,87	1,70	1,48
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,75	1,56	1,31
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,63	1,43	1,00

Tab. VI: Kritische Werte $F_{(0,95; \nu_1, \nu_2)}$ der F -Verteilung mit ν_1 und ν_2 Freiheitsgraden, so dass $P(F < F_{(0,95; \nu_1, \nu_2)}) = 0,95$.

ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	25	50	∞
ν_2																
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,3	251,8	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,46	19,48	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,63	8,58	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,70	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,52	4,44	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,83	3,75	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,40	3,32	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,11	3,02	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,89	2,80	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,73	2,64	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,60	2,51	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,50	2,40	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,41	2,31	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,34	2,24	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,28	2,18	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,23	2,12	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,18	2,08	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,14	2,04	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,00	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,07	1,97	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	1,94	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,02	1,91	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,00	1,88	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,97	1,86	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,84	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,94	1,82	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,92	1,81	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,79	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,89	1,77	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,88	1,76	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,78	1,66	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,69	1,56	1,39
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,60	1,46	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,51	1,35	1,00

Tab, VII: Studentisierte Variationsbreiten $q_{Tab}(t, FG, \alpha = 5\%)$,

FG/t	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	17,97	26,98	32,82	37,08	40,41	43,12	45,40	47,36	49,07	50,59	51,96	53,19	54,32	55,36	56,32	57,21	58,04	58,82	59,55
2	6,08	8,33	9,80	10,88	11,73	12,43	13,03	13,54	13,99	14,39	14,75	15,08	15,37	15,65	15,91	16,14	16,36	16,57	16,77
3	4,50	5,91	6,82	7,50	8,04	8,48	8,85	9,18	9,46	19,72	9,95	10,15	10,35	10,52	10,69	10,84	10,98	11,11	11,24
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,03	8,21	8,37	8,52	8,66	8,79	8,91	9,03	9,13	9,23
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99	7,17	7,32	7,47	7,60	7,72	7,83	7,93	8,03	8,12	8,21
6	3,46	4,34	4,90	5,30	5,63	5,90	6,12	6,32	6,49	6,65	6,79	6,92	7,03	7,14	7,24	7,34	7,43	7,51	7,59
7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,30	6,43	6,55	6,66	6,76	6,85	6,94	7,02	7,10	7,17
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05	6,18	6,29	6,39	6,48	6,57	6,65	6,73	6,80	6,87
9	3,20	3,95	4,41	4,76	5,02	5,24	5,43	5,59	5,74	5,87	5,98	6,09	6,19	6,28	6,36	6,44	6,51	6,58	6,64
10	3,15	3,88	4,33	4,65	4,91	5,12	5,30	5,46	5,60	5,72	5,83	5,93	6,03	6,11	6,19	6,27	6,34	6,40	6,47
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49	5,61	5,71	5,81	5,90	5,98	6,06	6,13	6,20	6,27	6,33
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,39	5,51	5,61	5,71	5,80	5,88	5,95	6,02	6,09	6,15	6,21
13	3,06	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32	5,43	5,53	5,63	5,71	5,79	5,86	5,93	5,99	6,05	6,11
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36	5,46	5,55	5,64	5,71	5,79	5,85	5,91	5,97	6,03
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,31	5,40	5,49	5,57	5,65	5,72	5,78	5,85	5,90	5,96
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,26	5,35	5,44	5,52	5,59	5,66	5,73	5,79	5,84	5,90
17	2,98	3,63	4,02	4,30	4,52	4,70	4,86	4,99	5,11	5,21	5,31	5,39	5,47	5,54	5,61	5,67	5,73	5,79	5,84
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,17	5,27	5,35	5,43	5,50	5,57	5,63	5,69	5,74	5,79
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23	5,31	5,39	5,46	5,53	5,59	5,65	5,70	5,75
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,11	5,20	5,28	5,36	5,43	5,49	5,55	5,61	5,66	5,71
21	2,94	3,56	3,94	4,21	4,42	4,60	4,74	4,87	4,98	5,08	5,17	5,25	5,33	5,40	5,46	5,52	5,58	5,63	5,68
22	2,93	3,55	3,93	4,20	4,41	4,58	4,72	4,85	4,96	5,06	5,14	5,23	5,30	5,37	5,43	5,49	5,55	5,60	5,65
23	2,93	3,54	3,91	4,18	4,39	4,56	4,70	4,83	4,94	5,03	5,12	5,20	5,27	5,34	5,41	5,46	5,52	5,57	5,62
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,01	5,10	5,18	5,25	5,32	5,38	5,44	5,49	5,55	5,59
25	2,91	3,52	3,89	4,15	4,36	4,53	4,67	4,79	4,90	4,99	5,08	5,16	5,23	5,30	5,36	5,42	5,47	5,52	5,57
26	2,91	3,51	3,88	4,14	4,35	4,51	4,65	4,77	4,88	4,98	5,06	5,14	5,21	5,28	5,34	5,40	5,45	5,50	5,55
27	2,90	3,51	3,87	4,13	4,33	4,50	4,64	4,76	4,86	4,96	5,04	5,12	5,19	5,26	5,32	5,38	5,43	5,48	5,53
28	2,90	3,50	3,86	4,12	4,32	4,49	4,62	4,74	4,85	4,94	5,03	5,11	5,18	5,24	5,30	5,36	5,41	5,46	5,51
29	2,89	3,49	3,85	4,11	4,31	4,47	4,61	4,73	4,84	4,93	5,01	5,09	5,16	5,23	5,29	5,34	5,40	5,44	5,49
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	4,92	5,00	5,08	5,15	5,21	5,27	5,33	5,38	5,43	5,47
31	2,88	3,48	3,84	4,09	4,29	4,45	4,59	4,71	4,81	4,90	4,99	5,06	5,13	5,20	5,26	5,31	5,36	5,41	5,46
32	2,88	3,48	3,83	4,09	4,28	4,45	4,58	4,70	4,80	4,89	4,98	5,05	5,12	5,18	5,24	5,30	5,35	5,40	5,45
33	2,88	3,47	3,83	4,08	4,28	4,44	4,57	4,69	4,79	4,88	4,97	5,04	5,11	5,17	5,23	5,29	5,34	5,39	5,43
34	2,87	3,47	3,82	4,07	4,27	4,43	4,56	4,68	4,78	4,87	4,96	5,03	5,10	5,16	5,22	5,27	5,33	5,37	5,42
35	2,87	3,46	3,81	4,07	4,26	4,42	4,56	4,67	4,77	4,86	4,95	5,02	5,09	5,15	5,21	5,26	5,31	5,36	5,41
36	2,87	3,46	3,81	4,06	4,25	4,41	4,55	4,66	4,76	4,85	4,94	5,01	5,08	5,14	5,20	5,25	5,30	5,35	5,40
37	2,87	3,45	3,80	4,05	4,25	4,41	4,54	4,66	4,76	4,85	4,93	5,00	5,07	5,13	5,19	5,24	5,29	5,34	5,39
38	2,86	3,45	3,80	4,05	4,24	4,40	4,53	4,65	4,75	4,84	4,92	4,99	5,06	5,12	5,18	5,23	5,28	5,33	5,38
39	2,86	3,45	3,79	4,04	4,24	4,39	4,53	4,64	4,74	4,83	4,91	4,98	5,05	5,11	5,17	5,22	5,27	5,32	5,37
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,73	4,82	4,90	4,98	5,04	5,11	5,16	5,22	5,27	5,31	5,36
50	2,84	3,42	3,76	4,00	4,19	4,34	4,47	4,58	4,68	4,77	4,85	4,92	4,98	5,04	5,10	5,15	5,20	5,24	5,29
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81	4,88	4,94	5,00	5,06	5,11	5,15	5,20	5,24
120	2,80	3,36	3,68	3,92	4,10	4,24	4,36	4,47	4,56	4,64	4,71	4,78	4,84	4,90	4,95	5,00	5,04	5,09	5,13
∞	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62	4,68	4,74	4,80	4,85	4,89	4,93	4,97	5,01

Tab, VIII: Kritische Werte $\chi^2_{\gamma;v}$ der χ^2 -Verteilung mit v Freiheitsgraden, so dass $P(\chi^2 < \chi^2_{\gamma;v}) = \gamma$,

Siehe auch
Tab. IV, wo
andere Werte
für γ tabelliert
sind.

$\gamma \backslash v$	0,90	0,95	0,99
1	2,706	3,841	6,635
2	4,605	5,991	9,210
3	6,251	7,815	11,345
4	7,779	9,488	13,277
5	9,236	11,070	15,086
6	10,645	12,592	16,812
7	12,017	14,067	18,475
8	13,362	15,507	20,090
9	14,684	16,919	21,666
10	15,987	18,307	23,209
11	17,275	19,675	24,725
12	18,549	21,026	26,217
13	19,812	22,362	27,688
14	21,064	23,685	29,141
15	22,307	24,996	30,578
16	23,542	26,296	32,000
17	24,769	27,587	33,409
18	25,989	28,869	34,805
19	27,204	30,144	36,191
20	28,412	31,410	37,566
21	29,615	32,671	38,932
22	30,813	33,924	40,289
23	32,007	35,172	41,638
24	33,196	36,415	42,980
25	34,382	37,652	44,314
26	35,563	38,885	45,642
27	36,741	40,113	46,963
28	37,916	41,337	48,278
29	39,087	42,557	49,588
30	40,256	43,773	50,892
31	41,422	44,985	52,191
32	42,585	46,194	53,486
33	43,745	47,400	54,776
34	44,903	48,602	56,061
35	46,059	49,802	57,342
36	47,212	50,998	58,619
37	48,363	52,192	59,893
38	49,513	53,384	61,162
39	50,660	54,572	62,428
40	51,805	55,758	63,691
41	52,949	56,942	64,950
42	54,090	58,124	66,206
43	55,230	59,304	67,459
44	56,369	60,481	68,710
45	57,505	61,656	69,957
46	58,641	62,830	71,201
47	59,774	64,001	72,443
48	60,907	65,171	73,683
49	62,038	66,339	74,919
50	63,167	67,505	76,154

Anhang: Die Bonferroni Methode

In Abschnitt 4.5.2 hatten wir das Problem des multiplen Testens in der Varianzanalyse behandelt. Um den versuchsbezogenen Fehler 1. Art zu kontrollieren, reicht es nicht, einen t-Test (LSD-Test) durchzuführen, sondern man muss spezielle Verfahren anwenden wie den Tukey-Test (siehe Abschnitt 4.5.3). Dieses Verfahren ist allerdings auf die Varianzanalyse beschränkt. Eine ganz allgemein anwendbare Methode zur Kontrolle des versuchsbezogenen Fehlers 1. Art ist das Bonferroni-Verfahren. Um dies zu erläutern, betrachten wir folgendes Beispiel.

Beispiel: Artikel aus der Stuttgarter Zeitung vom 30.12.1010 (Seite 18). Titel „Besonders begabt oder nur gut im Raten?“ Untertitel: „Parapsychologie: Versuchspersonen haben zukünftige Ereignisse vorhergesagt. Die Wiederholung scheiterte aber. Von *Jochen Paulus*.“

„Auf einem Computerbildschirm sind zwei geschlossene Vorhänge zu sehen, einer links, einer rechts. Hinter einem erscheint beim Klicken darauf ein erotisches Bild, hinter dem anderen eine nackte Wand. Es gilt natürlich, auf den Vorhang zu klicken, der das interessantere Bild verbirgt. Wer bei diesem Experiment einfach rät – und etwas anderes bleibt ja offenbar nicht übrig – hat eine Chance von 50%, per Zufall richtig zu tippen. Doch als hundert Versuchspersonen je 36-mal raten mussten, kamen sie auf eine Trefferquote von 53 Prozent. Durch pures Glück ist das kaum zu erklären. [...]“

Der Signifikanztest für die Nullhypothese, dass die Erfolgswahrscheinlichkeit 50% beträgt, ergibt einen p-Wert von $p = 0,011$. Der Test ist also signifikant auf dem 5% Niveau.

Es stellte sich beim Durchlesen der dem Artikel zugrunde liegenden Studie allerdings heraus, dass neben den erotischen Bildern weitere Experimente gemacht wurden, und zwar mit insgesamt fünf weiteren Bilderkategorien (<http://dbem.ws/FeelingFuture.pdf>). Für keinen dieser fünf weiteren Kategorien war der einfache Signifikanztest signifikant. Mit dieser Zusatzinformation erscheint die Signifikanz für einen Bildertyp in einem anderen Licht.

Gehen wir von insgesamt $m = 6$ unabhängigen Tests aus, dann können wir die Frage stellen, mit welcher Wahrscheinlichkeit mindestens einer der Tests einen Fehler 1. Art produziert, wenn in allen Fällen die Nullhypothese gilt. Dies beantwortet man am einfachsten durch Betrachtung der Gegenwahrscheinlichkeit. Für einen einzelnen Test beträgt die Wahrscheinlichkeit, keinen Fehler zu machen, $1 - \alpha$. Für m unabhängige Tests ist nach dem Multiplikationssatz (siehe Abschnitt 5.2) die Wahrscheinlichkeit gleich $(1 - \alpha)^m$. Somit gilt:

$$P(\text{Mindestens ein Fehler 1. Art}) = \alpha_v = 1 - (1 - \alpha)^m$$

Hierbei bezeichnet α_v das versuchsbezogene α . Wenn nun sechs mal bei $\alpha = 5\%$ getestet wird, so ist

$$\alpha_v = 1 - (1 - \alpha)^m = 1 - 0,95^6 = 0,265.$$

Das ist so hoch, dass die Signifikanz des einen Tests bei einem vergleichsbezogenen Niveau von 5% als wenig aussagefähig erscheint. Generell besteht die Gefahr, dass man einfach viele Tests durchführt und sich dann die raussucht, die signifikant sind. Geht man so vor, dann halten die so selektierten Tests keinesfalls mehr das nominelle Signifikanzniveau ein. Um sich gegen solche Fehler zu schützen, muss man das versuchsbezogene α_v einhalten. Die Frage ist nun, welchen Wert für α der einzelnen Tests man wählen muss, damit α_v eingehalten wird. Die Antwort liefert eine einfache Umformung der Gleichung für α_v :

$$\alpha = 1 - (1 - \alpha_v)^{1/m} .$$

Mit $\alpha_v = 5\%$ finden wir

$$\alpha = 1 - (1 - 0,05)^{1/6} = 0,0085 .$$

Wir müssen also mit einem deutlich kleineren vergleichsbezogenen Signifikanzniveau den einzelnen Test durchführen, um das versuchsbezogene Niveau einzuhalten.

Die **Bonferroni-Ungleichung** liefert nun eine sehr einfache allgemeine Methode zur Bestimmung von α , die sogar dann gültig ist, wenn die m Tests nicht voneinander unabhängig sind, wie zum Beispiel bei multiplen Mittelwertvergleichen.

Bonferroni-Ungleichung

Seien E_1, E_2, \dots, E_m irgendwelche zufälligen Ereignisse. Dann besagt die Bonferroni-Ungleichung, daß die Wahrscheinlichkeit für das Eintreten von mindestens einem der m Ereignisse nicht größer als die Summe der Wahrscheinlichkeiten der m Ereignisse ist. Es gilt also

$$P(E_1 \text{ oder } E_2 \cdots \text{oder } E_m) \leq \sum_{i=1}^m P(E_i)$$

[Im Fall unabhängiger Ereignisse wird aus der Ungleichung übrigens eine Gleichung, was dann dem **Additionssatz** (Abschnitt 5.2) entspricht.] Aus dieser Ungleichung ergibt sich ein sehr einfaches Vorgehen, um das versuchsbezogene Signifikanzniveau einzuhalten.

Bonferroni-Adjustierung

Will bei man m Tests die Wahrscheinlichkeit für einen versuchsbezogenen Fehler 1. Art bei einem Wert von α_v kontrollieren, so wählt man für den einzelnen Test (vergleichsbezogen)

$$\alpha = \alpha_v / m .$$

Einen vergleichsbezogenen p -Wert muss man dementsprechend mit m multiplizieren, um einen versuchsbezogenen p -Wert p_v zu erhalten:

$$p_v = p \cdot m$$

Um den p -Wert von 0,011 für das Auffinden erotischer Bilder für das multiple Testen ($m = 6$ Tests) zu adjustieren, berechnen wir $p_v = p \cdot m = 0,011 \cdot 6 = 0,066$. Dies ist nicht signifikant für dem 5% Niveau. Ein parapsychologischer Effekt kann somit nach Bonferroni-Adjustierung nicht nachgewiesen werden.

Für den Einzel-Test müßten wir nach Bonferroni mit einem α von $0,05/6 = 0,0083$ prüfen. Da der p -Wert von 0,011 im vorliegenden Fall größer als 0,0083, liegt nach Adjustierung keine Signifikanz vor. Der Bonferroni-adjustierte Wert $\alpha = 0,0083$ ist nur unwesentlich kleiner als der exakte Wert bei unabhängigen Tests ($\alpha = 0,0085$), den wir oben errechnet hatten. Die Bonferroni-Ungleichung liefert also bei unabhängigen Tests eine sehr gute Approximation der nominalen Irrtumswahrscheinlichkeit. Ihr großer Vorteil ist aber die Anwendbarkeit auch bei abhängigen Tests.

Statistik Übungsaufgaben

Die hier abgedruckten Übungen werden zum Großteil in den Übungen zur Statistik bearbeitet, eventuell werden aber noch weitere Aufgaben in den Übungen selbst verteilt. Die Übungsaufgaben dienen der Vorlesungsnachbereitung, sie ersetzen **nicht** den Besuch der Vorlesung, noch decken Sie den gesamten Stoff der Vorlesung ab. Die Übungsaufgaben sind allein mit Skript und Taschenrechner lösbar. Uns ist es wichtig, dass Sie selbständig auf den Lösungsweg kommen. Wenn Sie hierbei Probleme haben, holen Sie sich in den Übungen Hilfe. Löffern Sie die Tutoren, bis Sie verstanden haben, wie Sie zur Lösung kommen! Im Nachgang zu den hier abgedruckten Übungsaufgaben finden sich einige leere Seiten, auf denen Sie ihre Musterlösungen sowie sonstige Aufschriebe aller Art, die Sie zur Klausur mitnehmen wollen, niederschreiben können.

1. Welchen Datentyp haben die folgenden Beispiele (nominal, ordinal oder metrisch)

- Rasse beim Rind (Schwarzbunt, Fleckvieh, ...)
- Schulnoten (1, 2, 3, 4, 5, 6)
- Befall von Weizenpflanzen mit Mehltau (keiner, gering, mittel, stark)
- Anzahl Unkräuter pro m²
- Unkrautdeckungsgrad (%)
- Geschlecht (männlich, weiblich)
- Geschmack für Quark (sehr gut, gut, neutral, schlecht, sehr schlecht)
- Zustimmung zu der Aussage "Ich würde Biomilch kaufen, wenn es im Umkreis von 5 km einen Bioladen geben würde, der Biomilch anbietet" (auf jeden Fall, wahrscheinlich, wahrscheinlich nicht, auf keinen Fall)
- Temperatursumme (Summe aller positiven Tagesmitteltemperaturen in einem bestimmten Zeitraum; Anhaltspunkt für Wärmebedarf einer Kulturart)
- Ackerzahl (0-100; Indexzahl für Ertragsfähigkeit eines Ackerschläges)
- Kornfraktion eines Partikels im Boden (Blöcke, Sand, Schluff, Ton):

Durchmesser (mm)	Bezeichnung
>2	Blöcke
2 - 0,063	Sand
0,063 - 0,002	Schluff
<0,002	Ton

- Bodenart (Ton = T, toniger Sand = tS, lehmiger Ton = IT, sand. Schluff = sU, etc.)
- Genotyp für Blütenfarbe (AA = weiss, Aa = rosa, aa = rot)
- Resistenz gegen *Phoma lingam* bei Raps (ja, nein)
- Resistenz gegen *Phoma lingam* bei Raps (nein, schwache Resistenz, starke Resistenz, völlige Resistenz)
- Resistenz gegen *Phoma lingam* bei Raps (Anteil befallener Sämlinge in %)
- EC-Stadien bei Getreide (00 = Trockenes Saatkorn, 10 = Auflaufen, 21 = Bestockungsbeginn, ..., 92 = Totreife)

2. Überlegen Sie sich je zwei Beispiele für metrische, kategoriale und nominale Daten. Die Beispiele sollten weder in bisherigen Übungsaufgaben noch im Skript/Vorlesung vorgekommen sein. Fragen Sie anschließend Ihren Nachbarn, welchen Datentyp dieser ihren Beispielen zuordnen würde.

3. Welche dieser Aussagen treffen zu?

- Binäre Merkmale haben zwei Ausprägungen
- Für metrische Merkmale ist die Mittelwertbildung nicht sinnvoll
- Ordinale Daten sind geordnete kategoriale Daten, bei denen die Kategorien in einer Rangordnung stehen
- Dichotome Daten sind metrische Daten
- Polytome Daten sind kategoriale Daten
- Bei nominalen Daten liegen keine geordneten Kategorien vor
- Nominale Daten lassen sich auf einem Zahlenstrahl abbilden
- Für geordnete kategoriale Daten sind Differenzen sinnvoll interpretierbar
- Metrische Daten sind immer auch stetig

4. Wofür ist es wichtig, den Datentyp der Daten zu bestimmen/ zu kennen?

5. Eine im SS 1983 durchgeführte Befragung ergab folgende Urliste der Körpergröße von Studentinnen (R.Schlittgen, 1996)

170	177	168	161	170	174	168	162	175	170
170	172	158	170	170	165	166	158	164	173
172	174	169	165	175	178	167	168	160	165
168	174	165	162	168	168	172	162	158	170
170	165	166	168	158	167	170	172	156	166

- Welche Darstellungen für Häufigkeitsverteilungen kennen Sie?
- Erstellen Sie diese.

6. In einer kleinen Rinderherde ($n = 13$ Tiere) wurden die Lebendgewichte der Tiere gemessen. Es ergaben sich folgende Messwerte:

295 248 260 **223 306** 234 263 248 235 251 232 244 267

- Berechnen Sie die Quantile Q_0 , Q_{25} , Q_{50} , Q_{75} , Q_{100} .
- Berechnen Sie die Quantile Q_0 , Q_{25} , Q_{50} , Q_{75} , Q_{100} ohne den letzten Messwert.
- Berechnen Sie für alle gegebenen Daten den Median und den Mittelwert.
- Berechnen Sie Varianz, Standardabweichung und Variationskoeffizient.
- Berechnen Sie die Variationsbreite und den Interquartilabstand.
- Zeichnen Sie einen Box-Plot.

7. Sie wollen eine Erdmischung aus je einem Kilogramm Hochmoortorf (Dichte 0,9kg/l) und Bims (Dichte 2,4 kg/l) herstellen. Wie groß ist die mittlere Dichte (gehen Sie davon aus, dass sich das Volumen durch den Akt des Mischens nicht verändert)?

8. Vier Jahre lang wurden die Produktionssteigerungen eines Betriebs pro Jahr gemessen. Es ergaben sich die folgenden Werte (J.Hartung, B.Elzelt, K.-H.Klösener, 1995):

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

Berechnen Sie die durchschnittliche Produktionssteigerung. Begründen Sie die Wahl ihres Lagemaßes.

Achtung: ‚log‘ im Skript und in den Lösungen meint den natürlichen Logarithmus – das ist im englischsprachigen Raum und in der Statistik gebräuchlich.

9. Die mittlere Alkoholkonzentration im Blut wurde bei je acht Frauen eine halbe Stunde nach Verabreichung einer Menge unterschiedlich starker alkoholischer Getränke ermittelt (R.Schlittgen, 1996):

Alkohol-Dosis	Alkoholkonzentration*100							
200 mg/kg	2,4	1,3	2,5	1,7	2,9	1,4	2,3	1,5
400 mg/kg	5,2	3,9	4,9	5,9	4,9	6,1	4,7	5,2

Welche Streuungsmaße kennen Sie? Mit welcher Maßzahl würden Sie die Streuung in den beiden Gruppen vergleichen? Begründen Sie Ihre Entscheidung kurz. Berechnen Sie diese Maßzahl.

10. Erstellen Sie eine Liste statistischen Fachbegriffen, die Sie bisher kennen gelernt haben, und erläutern Sie die Fachbegriffe mit wenigen Stichworten.

11. In einer Untersuchung zur Beurteilung der Sortierung einer großen Partie Kartoffeln für die Produktion von Chips soll geklärt werden, welcher Anteil der Knollen einen Durchmesser zwischen 5 und 6 cm hat, da diese Sortierung besonders für die Chipsproduktion geeignet ist. Zur Klärung dieser Frage wurde eine Stichprobe von 1000 Kartoffeln untersucht. Der Mittelwert betrug 5,3 cm, die Standardabweichung 0,5 cm. Nehmen Sie an, dies seien wegen der großen Stichprobe der wahre Mittelwert μ und die wahre Standardabweichung σ_x der Partie. Welcher Anteil der Kartoffeln in der Partie hat einen Durchmesser zwischen 5 und 6 cm, falls von einer Normalverteilung des Durchmessers ausgegangen werden kann?

12. Sie wollen die Milchleistung (pro Laktation) Schwarzbunter Kühe bestimmen und erheben daher die Milchleistung von 10 zufällig ausgewählten Kühen. Erläutern Sie anhand des Beispiels die Begriffe Stichprobe und Grundgesamtheit (ohne den Text im Skript abzuschreiben).

13. Gegeben sei eine Herdbuchpopulation von 200.000 Schwarzbunt-Kühen. Das mittlere Körpergewicht der Tiere beträgt $\mu = 650$ kg. Die Standardabweichung in der Population beträgt $\sigma_x = 25$ kg. Die Körpergewichte sind annähernd normalverteilt. Angenommen, eine Person untersucht $n = 100$ Kühe aus dieser Herdbuchpopulation.

- Welchen Anteil der Einzelwerte der Stichprobe erwarten Sie in den Grenzen zwischen 645 und 655 kg?
- Aus den $n = 100$ Einzelmessungen soll ein Stichprobenmittelwert \bar{x} berechnet werden. Mit welcher Wahrscheinlichkeit wird dieser Stichprobenmittelwert in den Grenzen zwischen 645 und 655 liegen?
- Wie würden sich die Antworten zu a) und b) ändern, wenn 130 Kühe untersucht worden wären?

14. In der Mathematik- und Statistiklausur werden 24 der 50 möglichen Punkte für die Note 4 (D) benötigt. Zwei Studierende des letzten Jahres haben mit den angegebenen Punkten folgende Noten erreicht (gekürzte Liste)

Studierender	Punkte	Note
A	48	1 (A)
B	0	5 (F)

Den Noten liegt folgende Skala zu Grunde:

Punkte	Note
0,0-23,5	5 F
24,0-26,5	4 D
27,0-29,0	4+ D+
29,5-31,5	3- C-
32,0-34,0	3 C
34,5-36,5	3+ C+
37,0-39,5	2- B-
40,0-42,0	2 B
42,5-44,5	2+ B+
45,0-47,0	1- A-
47,5-50,0	1 A

Berechnen Sie deren Durchschnittsnote. (Für Interessierte: Vergleichen Sie Ihr Ergebnis mit der Berechnung der Durchschnittsnote bei Studium online.)

15. Für eine große Bodenprobe (ca. 20 kg), welche nach der Entnahme aus den oberen 30 cm des Bodenkörpers gut durchmischt wurde, soll in einer Laboruntersuchung der Gehalt an Fe_2O_3 in % bestimmt werden. Von der Bodenprobe wurden $n = 9$ Teilproben à 10 g eingewogen und chemisch untersucht. Der Mittelwert des normalverteilten Fe_2O_3 -Gehaltes in der Stichprobe betrug 5,7%. Die Standardabweichung der Einzelmessungen betrug 0,6%.

- Geben Sie ein Intervall an, welches mit 95%iger Wahrscheinlichkeit den wahren Gehalt des Bodens an Fe_2O_3 enthält. Wie breit ist dieses Intervall?
- Für zukünftige Untersuchungen wird festgelegt, dass das 95%-Vertrauensintervall nicht breiter als 0,4% sein darf. Welcher Stichprobenumfang ist mindestens nötig, um diese Genauigkeit zu erzielen?
- Für eine aufwendigere Methode zur Bestimmung des Fe_2O_3 -Gehaltes wurde eine Standardabweichung der Messwerte von 0,15% ermittelt. Welche Breite erwarten Sie für ein 95%-Vertrauensintervall des Mittelwertes, wenn mit der neuen Methode ebenfalls $n = 9$ Teilproben untersucht werden?

16. Was sind Freiheitsgrade?

17. Steel und Torrie (1980) berichten über eine Untersuchung, in der die Wirkung zweier unterschiedlicher Umweltbedingungen auf Luzerneblüten untersucht wurde. Sie wählten 10 Pflanzen, die jeweils einige Blüten am oberen Teil der Pflanze und

einige Blüten in Bodennähe hatten. Sie bestimmten bei jeder Pflanze die mittlere Zahl der Samen pro Hülse an den oberen und an den unteren Blüten.

Pflanze	1	2	3	4	5	6	7	8	9	10
Obere Blüten	4,0	5,2	5,7	4,2	4,8	3,9	4,1	3,0	4,6	6,8
Untere Blüten	4,4	3,7	4,7	2,8	4,2	4,3	3,5	3,7	3,1	1,9

a. Bestimmen Sie ein 95%iges Vertrauensintervall für die Differenz der Zahl der Samen an oberen und unteren Blüten. Enthält dieses Intervall die Null?

b. Bestehen signifikante Unterschiede zwischen den beiden Blütenpositionen? Führen Sie zur Beantwortung dieser Frage einen t-Test zum Signifikanzniveau $\alpha = 5\%$ durch. Hinweis: $s_d^2 \doteq 2,556$; $\bar{d} = 1$.

c. Deckt sich das Ergebnis des Tests mit der Information des Vertrauensintervalls (siehe Lösungsblatt 6)?

c. Statt zu testen, ob die wahre Differenz von Mittelwerten zweier Grundgesamtheiten Null ist, können Sie auch testen, ob die wahre Differenz von Mittelwerten zweier Grundgesamtheiten δ_0 ist. Hierbei wird die Formel für den verbundenen Test wie folgt

verallgemeinert:
$$t_{Vers} = \frac{|\bar{d} - \delta_0|}{s_d} \sqrt{n}.$$

Als Testentscheidung ergibt sich: Falls $t_{Vers} \leq t_{Tab} \Rightarrow H_0 (\delta = \delta_0)$ (Differenz nicht signifikant von δ_0 verschieden), andernfalls verwerfe H_0 . Alle weiteren Rechnungen sind identisch zum Test für verbundene Stichproben. Für welche Werte von δ_0 behalten Sie die Nullhypothese bei? Versuchen Sie zunächst eine Formel abzuleiten. Wenden Sie diese Formel dann auf den Fall zweier verbundene Stichproben der Größe $n=10$, mit Standardabweichung der Differenz $s_d=1,599$ und einer geschätzten Differenz von 1 an. Vergleichen Sie diesen Wertebereich mit dem Vertrauensintervall (Lösungsblatt 6).

18. In einem Fütterungsversuch (siehe Steel und Torrie, 1980) wurde die Verdaulichkeit (%) von Maissilage bei Schafen und bei Mastbullen untersucht.

x_1 (Schafe)	x_2 (Bullen)
57,8	64,2
56,2	58,7
61,9	63,1
54,4	62,5
53,6	59,8
56,4	59,2
53,2	

a. Führen Sie einen t-Test der Nullhypothese durch, dass beide Tierarten die gleiche Verdaulichkeit haben ($\alpha = 5\%$).

b. Berechnen sie ein 95%-Vertrauensintervall für die Differenz der Verdaulichkeiten bei beiden Tierarten.

19. Sie ziehen aus der Grundgesamtheit aller Höckerschwäne, die im Jahr 2014 in Deutschland gebrütet haben, eine zufällig gezogene Stichprobe von drei Tieren. Bei jedem Tier der Stichprobe messen Sie das Gewicht des Tieres. Das Merkmal Gewicht ist normalverteilt. Aus der Stichprobe berechnen Sie das arithmetische Mittel der Gewichte sowie deren Varianz. Welche der folgenden Aussagen sind richtig?

- a. Da die Maßeinheit Schwan im Zähler steht, muss das harmonische Mittel verwendet werden.
- b. Eine Stichprobengröße von 3 ist zu klein, um Aussagen über die Grundgesamtheit treffen zu können.
- c. Der aus der Stichprobe errechnete Mittelwert der drei Tiere ist identisch mit dem Mittelwert aller Höckerschwäne, die im Jahr 2014 in Deutschland gebrütet haben.
- d. Die mit Hilfe der Stichprobe geschätzte Varianz der Höckerschwangewichte ist wegen der kleinen Stichprobe immer größer als die Varianz in der Grundgesamtheit.
- e. Das berechnete arithmetische Mittel schätzt den wahren Mittelwert der Grundgesamtheit.
- f. Sowohl das berechnete arithmetische Mittel, als auch die errechnete Varianz schätzen Parameter der Grundgesamtheit.
- g. Das arithmetische Mittel der Stichprobe ist fast nie identisch mit dem wahren arithmetischen Mittelwert der Grundgesamtheit.

20. Entscheiden Sie, ob in den folgenden Fällen verbundene (**v**) oder unverbundene (**u**) Stichproben vorliegen. Geben Sie an von welchen Grundgesamtheiten Stichproben genommen wurden und begründen Sie ihre zuvor getroffene Entscheidung.

- a. In einer Herde wird das Verhalten von zwei Kühen beobachtet, um festzustellen, welches von beiden hierarchisch höher steht. Die Tiere werden in einem Fischgrätenmelkstand gemolken. Die hierarchisch höher stehenden Tiere betreten jeweils tendenziell eher den Melkstand, als untergeordnete Tiere. An einer Reihe von aufeinander folgenden Tagen wird die relative Position der beiden Tiere (also ob sie als erstes oder als zweites in den Fischgrätenstand kommen) erhoben.
- b. Es soll untersucht werden, ob Stengel oder Wurzel von Reis einen höheren Silikatgehalt aufweist. Hierzu wird eine Reihe von Reispflanzen in Wurzel und Stengel geerntet und jede Pflanze geteilt. Anschließend wird für beide Organe der Silikatgehalt bestimmt.
- c. In einem Laborversuch wird die Wirksamkeit von zwei verschiedenen Fungiziden A und B gegen Hefepilze untersucht. Der Pilz wird auf einem Nährmedium in insgesamt 20 Petrischalen angezogen. Auf 10 der Schalen werden die gewachsenen Pilzkolonien mit Fungizid A ausgezählt, auf den anderen 10 Schalen wird Fungizid B angewendet und ebenfalls die Kolonien ausgezählt.
- d. In einem sensorischen Test wird der Geschmack von zwei Quarksorten geprüft. Hierzu werden beide Quarksorten von einer Gruppe von Testpersonen probiert. Jede Person beurteilt den Geschmack beider Quarksorten auf einer Boniturskala von 1 bis 9.

- e. Es soll untersucht werden, ob Stengel oder Wurzel von Reis einen höheren Silikatgehalt aufweist. Hierzu wird eine Reihe von Reispflanzen in Wurzel und Stengel geerntet und jede Pflanze geteilt. Anschließend wird für beide Organe der Silikatgehalt bestimmt.
- f. In einem Laborversuch wird die Wirksamkeit von zwei verschiedenen Fungiziden A und B gegen Hefepilze untersucht. Der Pilz wird auf einem Nährmedium in insgesamt 20 Petrischalen angezogen. Auf 10 der Schalen werden die gewachsenen Pilzkolonien mit Fungizid A ausgezählt, auf den anderen 10 Schalen wird Fungizid B angewendet und ebenfalls die Kolonien ausgezählt.
- g. In einem sensorischen Test wird der Geschmack von zwei Quarksorten geprüft. Hierzu werden beide Quarksorten von einer Gruppe von Testpersonen probiert. Jede Person beurteilt den Geschmack beider Quarksorten auf einer Boniturskala von 1 bis 9.

21. Überlegen Sie sich je ein Beispiel, bei dem die Stichproben verbunden und unverbunden sind. Die Beispiele sollten weder in bisherigen Übungsaufgaben, noch im Skript oder der Vorlesung vorkommen. Erzählen Sie Ihrem Nachbar eins der beiden Beispiele und fragen Sie anschließend, ob es sich in dem Beispiel um eine verbundene oder unverbundene Stichprobe handelt

22. Bei Kindern wurde bis 2005 häufig ein Tuberkulintest zum Test auf Tuberkulose durchgeführt. Hierbei wird das Tuberkulin mit einem Stempel auf die Unterarminnenseite aufgetragen. Bei einer fühlbaren Papel galt der Test als positiv. Erklären Sie an dem Beispiel den α - und β -Fehler.

23. Welche Informationen brauchen Sie für eine Stichprobenplanung?

24. In einem Sortenversuch mit Raps sollen 10 Sorten miteinander verglichen werden. Aus Vorversuchen ist die Fehlervarianz beim Zielmerkmal Ertrag bekannt: $\sigma^2 = 4 \text{ (dt/ha)}^2$. Der Versuch soll vollständig randomisiert werden. Es soll eine vergleichsbezogene Irrtumswahrscheinlichkeit (Fehler 1. Art) von 5% eingehalten werden. Die kleinste nachzuweisende Differenz ist 5 dt/ha. Unterschiede sollen mit einer Wahrscheinlichkeit von mindestens 80% nachgewiesen werden. Wie groß ist der Stichprobenumfang zu wählen?

25. Nach der Durchführung von zweiseitigen t-Tests mit drei verschiedenen Datensätzen wurden folgende p -Werte aus dem Computer-Output abgelesen:

- a) **Datensatz A:** $p=0,0073$
- b) **Datensatz B:** $p=0,3762$
- c) **Datensatz C:** $p=0,0586$

Wie lauten die Testsergebnisse, wenn $\alpha = 5\%$ gewählt wurde?

26. Erklären Sie mit maximal 2 Sätzen, was ein p -Wert ist.

27. Die Erfolgsquote einer Standardtherapie wurde mit der Erfolgsquote einer neuen Therapie verglichen. Pro Therapie wurden 15 Patienten behandelt. Die Standardabweichung betrug 3,91%. Wie ist die Teststärke für eine nachzuweisende Differenz von $\delta = 4\%$ ($\alpha = 5\%$)?

28. Von einer Brotfirma werden Brötchen produziert. Das Brötchengewicht x wird als normalverteilt vorausgesetzt. Die Standardabweichung ist bekannt: $s = 1,2$ Gramm. Die Brötchen werden an die Lebensmittelhändler mit dem Hinweis ausgeliefert, dass das durchschnittliche Gewicht 50 Gramm betrage. Sie wiegen eine zufällig gezogene Stichprobe von 25 Brötchen und finden $\bar{x} = 49,0$. Prüfen Sie, ob das gefundene Gewicht signifikant von 50 g abweicht. Formulieren Sie hierzu zunächst die Nullhypothese.

29. Sie vergleichen die Größe von weiblichen und männlichen Kälbern. Stellen Sie Nullhypothesen für eine einseitige und zweiseitige Fragestellung auf.

30. Eine Maschine soll Zucker in Tüten zu je 500 g verpacken. Man kann dabei annehmen, dass die tatsächliche Füllmenge normalverteilt ist. Es wurde die Füllmenge von 9 Tüten ermittelt, die von dieser Maschine verpackt wurden. Die Werte sind in der Tabelle wiedergegeben.

Lfd. Nr. der Messung	1	2	3	4	5	6	7	8	9
Füllmenge, g	506	502	500	505	499	505	501	504	498

Prüfen sie mit dem passenden Testverfahren, ob die von dieser Maschine verpackte Füllmenge von der voreingestellten Füllmenge $\mu_0 = 500$ g abweicht ($\alpha = 0,05$).

Prüfen Sie ferner, ob die Maschine weniger als 500 g verpackt.

31. In einer Studie zur Therapie der essentiellen arteriellen Hypertonie sollte nachgewiesen werden, dass ein neu entwickeltes Medikament A mit erheblich verbesserter Verträglichkeit bezüglich der blutdrucksenkenden Wirkung äquivalent ist mit einem Referenzpräparat B. Die Behandlung erfolgte in zwei Patientengruppen über einen Zeitraum von 8 Wochen. Als Zielkriterium wurde die prozentuale Verringerung des diastolischen Blutdrucks herangezogen. Die Patientenzahl betrug in beiden Gruppen je 20. Die Mittelwerte und Standardabweichungen in den Gruppen waren wie folgt:

Patientengruppe	\bar{x}	s
Medikament A	10,75	5,380
Medikament B	10,39	3,023

Äquivalenz beider Medikamente wird angenommen, wenn die wahre Differenz der prozentualen Blutdruckverringerung weniger als $\delta = 0,50$ beträgt. Testen Sie die Medikamente auf Äquivalenz ($\alpha = 0,05$).

32. In einem Fütterungsversuch mit Ratten wurden sechs verschiedene Fütterungsarten untersucht (Linder, 1969, S.26). Hierzu wurden 30 Ratten zufällig auf die Futterarten verteilt. Nach 56 Tagen wurde das Gewicht der Tiere (in g) ermittelt. Die Ergebnisse waren wie folgt:

Futterart					
A	B	C	D	E	F
119	123	130	144	159	139
90	121	163	172	172	146

102	159	159	165	210	161
85	138	140	143	171	149
113	178	121	179	232	124

- a. Geben Sie für den Versuch ein Modell an.
- b. Führen Sie eine Varianzanalyse sowie einen anschließenden multiplen Mittelwertvergleich durch. Bei den Tests soll eine Irrtumswahrscheinlichkeit von $\alpha = 5\%$ eingehalten werden.
- c. Wie lautet die Nullhypothese, die Sie bei der Varianzanalyse in b. getestet haben?

33. Erklären Sie die Begriffe vergleichs- und versuchsbezogene Irrtumswahrscheinlichkeit.

34. Welches Testergebnis (bei versuchsbezogenem Fehler 1. Art) erwarten Sie, wenn es in dem Versuch nur einen Faktor mit exakt zwei Stufen gibt und bei vergleichsbezogener Irrtumswahrscheinlichkeit signifikante Unterschiede nachgewiesen wurden?

35. Zwei miteinander verwandte Tomatenhybridsorten A und B unterscheiden sich in der Frühzeitigkeit der Ernte. Beide Sorten werden in einem Feldversuch auf ihren Gesamtertrag bis Ende Juli geprüft. Alle 20 Versuchspartzellen (von den 20 Versuchspartzellen werden je 10 Partzellen völlig zufällig ausgewählt und mit einer der beiden Sorten bepflanzt, siehe Plan unten) werden zum selben Zeitpunkt gesät und mit derselben Menge Dünger gedüngt. Ende Juli wird von jeder Partzelle die Masse der bis dahin geernteten Tomaten bestimmt. Sie wollen wissen, ob es signifikante Unterschiede zwischen den Sorten gibt.

Versuchsplan:

A	B	A	B	A	A	A	A	B	B
A	B	B	A	A	B	B	B	A	B

- a) Stellen Sie die Nullhypothese auf.
- b) Stellen Sie die Alternativhypothese auf.
- c) Geben Sie die Formel für den t-Test der Nullhypothese an ($\alpha=0,05$). Den Test selbst können und müssen Sie nicht durchführen.

36. In einem Rapsfeld sind 20% der Pflanzen mit dem Pilz *Phoma lingam* befallen. Es werden zufällig 5 Pflanzen ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass alle fünf Pflanzen befallen sind?

37. Wie unterscheiden sich Variation und Kombination?

38. Ein Feldversuch mit 7 Sorten wird als Blockanlage mit 4 Blöcken angelegt. Ein Block besteht aus 7 Partzellen = Feldstücken. Die 7 Sorten werden zufällig auf die 7 Partzellen eines Blocks verteilt. Die zufällige Verteilung der Behandlungen auf die Versuchseinheiten (hier: Partzellen) wird als Randomisation bezeichnet. In jedem der 4 Blöcke wird eine neue Randomisation durchgeführt. Wie viele mögliche Randomisationen gibt es für einen Block und für die ganze Anlage?

39. Das Bundessortenamt erfasst Krankheiten auf einer Boniturskala mit Noten von 1 (kein Befall) bis 9 (sehr starker Befall).

a. Wenn für eine Sorte 4 Bonituren einer Krankheit erhoben werden, wie viele mögliche Ergebnisse gibt es?

b. Wenn für eine Sorte 4 Bonituren von vier verschiedenen Krankheiten erhoben werden, wie viele mögliche Ergebnisse gibt es?

40. Ein Tierzüchter hat 10 Milchkühe zur Auswahl und selektiert anhand der Milchleistungsdaten das beste, das zweitbeste und das drittbeste Tier. Wie viele mögliche Selektionsergebnisse gibt es?

41.* Knobelaufgabe zur Kombinatorik: In einem Raum befinden sich 23 Personen. Wie groß ist die Wahrscheinlichkeit, dass mindestens 2 Personen am selben Tag Geburtstag haben? (Ein Jahr hat 365 Tage).

42. In einer Umfrage wurde eine Zufallsstichprobe von 100 Haushalten einer Großstadt befragt. Auf die Frage, ob der Haushalt Biomilch für einen Preis von 1,50 € pro Liter kaufen würde, antworteten 24 Haushalte mit ja. Berechnen Sie ein 95%-iges Vertrauensintervall für den Anteil von Haushalten in der Stadt (Grundgesamtheit), die Biomilch zum Preis von 1,50 € kaufen würden.

43. In einer Befragung von 3243 Personen, welche Partei sie wählen würden, wenn am Sonntag Bundestagswahl wäre, antwortet 42% der Befragten mit CDU (Ernid-Umfrage vom 17.01.2015). Prüfen Sie, ob dieser Wert signifikant vom Bundestagswahlergebnis von 41,5% abweicht.

44. Ein Produktionssystem A mit Sorghum wurde im Sahel auf einer großen Zahl von Farmen untersucht. Hierbei wurde auf jeder Farm auf je einem Feld der Ertrag (dt/ha) ermittelt. Es ergaben sich ein Mittelwert von 25 dt/ha und eine Standardabweichung von 3,5 dt/ha. Ein Histogramm zeigte, dass die Daten annähernd normalverteilt sind. Wie groß ist die Wahrscheinlichkeit, dass eine Subsistenzschwelle von 20 dt/ha unterschritten wird?

Wie ist im Vergleich dazu eine Produktionssystems B bezüglich des Risikos die Subsistenzschwelle zu unterschreiten zu beurteilen, wenn das Produktionssystems B einen höheren Ertrag von 30 dt/ha, aber auch eine höhere Standardabweichung von 4,2 dt/ha hat? Wird die höhere Streuung durch den höheren Erwartungswert aufgewogen oder nicht?

45. Entscheiden Sie, ob in den folgenden Fällen eine verbundene (v) oder unverbundene (u) Stichprobe vorliegt:

- a. In der Ferkelfütterung soll die Wirkung eines Futterzusatzpräparates auf Silikatbasis auf die Verdaulichkeit von Getreide untersucht werden. Da genetische Unterschiede in der Verdaulichkeit zu erwarten sind, werden je Wurf zwei Ferkel ausgesucht, von denen je ein Tier den Futterzusatz erhält, während das andere normal gefüttert wird.
- b. In einer Erhebung im Stadtgebiet Kassels soll ermittelt werden, ob in Stadtkern eine andere Kaufbereitschaft für Biomilch besteht als im Stadtrand. Das Stadtgebiet wird in Kern- und Randbereich unterteilt. Aus jedem der beiden Bereiche wird eine Zufallsstichprobe von Haushalten ermittelt und nach ihrer Kaufbereitschaft befragt.
- c. Zwei Rinderrassen sollen in ihrer Milchleistung verglichen werden. Hierzu werden jeweils 25 Tiere einer Rasse ausgewählt. Beide Rassen werden zusammen gehalten. An jedem der insgesamt 50 Tiere wird die Milchleistung erfasst.

46. In der Hühnerhaltung ist das Federpicken ein großes Problem. Dieses wird durch einzelne Hühner verursacht, die bei den anderen Tieren im Gefieder herum-picken und dadurch Verletzungen verursachen. Es wurde festgestellt, dass es in vielen Populationen neben normalen „Pickern“ sogenannte Primärpicker gibt, die sich durch eine besonders aggressive Picketätigkeit auszeichnen. Untersuchungen in einer deutschen Hühnerpopulation zeigen, dass der Anteil dieser Primärpicker bei 5% liegt (Prof. Bessei, Institut 470, Uni Hohenheim).

Beim Federpicken ist festzustellen, dass die Präsenz eines einzigen Primärpickers in der Gruppe zu erheblichen Schädigungen der Tiere führt, weil andere Tiere, die nicht Primärpicker sind, ebenfalls zum verstärkten Picken animiert werden. Eine wichtige Frage für die Hühnerhaltung betrifft die Gruppengröße. Je größer eine Gruppe ist, umso größer ist die Wahrscheinlichkeit, dass Primärpicker dabei sind.

- a. Nehmen Sie an, die Gruppengröße beträgt fünf Tiere. Wie groß ist die Wahrscheinlichkeit, dass mindestens ein Primärpicker in der Gruppe ist? Nehmen sie hierzu an, dass die Gruppe aus Tieren besteht, die eine zufällige Stichprobe aus der Gesamtpopulation darstellen.
- b. Wie groß ist die Wahrscheinlichkeit, wenn Sie die Gruppengröße auf 10 erhöhen?

47. In einer Umfrage wurde zu zwei verschiedenen Zeitpunkten eine Zufallsstichprobe von jeweils 500 Haushalten einer Großstadt befragt. Auf die Frage, ob der Haushalt Biomilch für einen Preis von 1,50 € pro Liter kaufen würde, antworteten beim ersten Mal 150 Haushalte mit ja, beim darauf folgenden zweiten Mal waren es 180. Ist dieser Anstieg signifikant, d.h. kann damit gerechnet werden, dass der Anteil von Haushalten in der Stadt (Grundgesamtheit), die Biomilch zum Preis von 1,50 € kaufen würden, gestiegen ist.

48. In einer sensorischen Prüfung testeten 200 Probanden den Geschmack zweier Brotaufstriche A und B. Jeder Proband musste entscheiden, ob der getestete Brotaufstrich gut schmeckt oder nicht. 56 Probanden gaben an, dass Aufstrich A schmeckt, Aufstrich B aber nicht. 41 Probanden schmeckten dagegen nur Aufstrich B gut. 70 Probanden schmeckten beide Aufstriche gut. Den verbleibenden Probanden schmeckte keine der beiden Aufstriche. Entscheiden Sie, ob es signifikante Unterschiede in der Akzeptanz der beiden Aufstriche gibt.

49. In einer BUND Studie zur Keimbelastung von konventionellem Putenfleisch wurden die Fleischproben auf zwei Arten von Keimen getestet, sogenannte MRSA-

Keime, das sind Bakterien, die gegen viele Antibiotika resistent sind und ESBL-Keime. Letzteres sind Bakterien, die Antibiotikaresistenzen an andere Bakterien weitergeben können. In 50 der 57 Proben wurden Keime gefunden, in 42 Proben wurden MRSA-Keime nachgewiesen, in 30 Proben wurden ESBL-Keime gefunden.

a. Unterscheiden sich die Keime in ihrer Wahrscheinlichkeit in einer Probe nachgewiesen zu werden?

50. In einer Untersuchung wurde an 360 Tabakpflanzen die Anzahl von *Heliothis* Eigelegen bestimmt. Überprüfen Sie, ob die Annahme einer Poisson-Verteilung für die Anzahl je Pflanze gerechtfertigt ist. Wenn ja, berechnen Sie ein 95% Vertrauensintervall für den Parameter der Verteilung.

Anzahl *Heliothis* Eigelege je Pflanze (Richter und Sönderath, 1990: 85)

Anzahl	Häufigkeit
0	211
1	110
2	29
3	7
4	2
5	0
6	0
7	1

51. In einem Zuchtprogramm für Weizen wurden zwei Resistenzgene, Javardo und H5, gegen die Hessenfliege untersucht (In Srivastava JP, Damania AB 1990 Wheat genetic resources: meeting diverse needs. Wiley, New York, pp. 303-309). Es wurde eine Kreuzung zweier Eltern durchgeführt. Die resultierenden 89 F_2 -Pflanzen wurden geselbstet. Die Resistenzprüfung erfolgte anhand der F_3 -Nachkommenschaften. In der F_3 traten drei verschiedene Nachkommenschaften von F_2 -Pflanzen auf: nur resistent, aufspaltend in resistent und anfällig, sowie nur anfällig. Unter der Annahme einer Dominanz beider Resistenzgene und unabhängiger Vererbung würde hierfür ein Verhältnis von 7:8:1 erwartet. Im Zuchtprogramm waren 33 Nachkommenschaften nur resistent, 42 aufspaltend und 14 nur anfällig. Prüfen Sie zum Niveau $\alpha = 5\%$ die Nullhypothese, dass beide Resistenzgene dominant sind.

52. Welches Auswertungsverfahren wählen Sie, wenn (i) die abhängige Variable (Y) metrisch und unabhängigen (X) Variablen kategorial ist, (ii) beide Variablen metrisch oder (iii) beide Variablen kategorial sind?

53. Die Pollenmutterzellen in den Staubgefäßen der Tomate tragen zwei Merkmale: Die Länge des Chromosoms Nr. 11 und den Schenkelquotienten des Chromosoms (Verhältnis des langen Schenkels zum kurzen Schenkel des Chromosoms). Bei einer zytologischen Untersuchung wurden Pollenmutterzellen von zwölf Staubfäden einer Tomate mit folgendem Ergebnis untersucht:

Staubfaden	Länge	Schenkelquotient
1	50	2,8
2	52	3,0
3	62	2,9
4	47	3,1
5	55	2,7

6	64	3,6
7	55	3,5
8	65	4,4
9	67	4,1
10	58	3,8
11	57	3,5
12	64	4,1

a. Plotten Sie die Daten. Erwarten Sie einen signifikanten Zusammenhang? Berechnen Sie die Pearsonsche Korrelation.

b. Prüfen Sie die Nullhypothese, dass keine Korrelation besteht.

c. Berechnen Sie ein 95% Vertrauensintervall für den Korrelationskoeffizienten.

54. In einer Rinderherde ($n = 8$ Tiere) wurden die Lebendgewichte der Tiere (kg) gemessen.

a. Wie viele mögliche Reihenfolgen gibt es die 8 Tiere je einmal zu wiegen?

b. Es ergaben sich folgende Messwerte:

295 248 260 **223 306** 234 263 248

Erstellen Sie ein sinnvolles Histogramm.

55. Berechnen Sie für die Daten aus Aufgabe **53** eine Regressionsgerade, wobei die Länge als Einflussvariable und der Schenkelquotient als Zielvariable verwendet wird (Diese kann zur Vorhersage des Quotienten aus der Länge genutzt werden). Berechnen Sie das Bestimmtheitsmaß der Regression. Führen Sie einen Test der Nullhypothese durch, dass die Steigung der Geraden Null ist.

56. Am 10.01.2013 gab es in der Frankfurter Rundschau einen Bericht über die Keimbelastung von Mettbrötchen. In 10 deutschen Großstädten (Berlin, Erfurt, Essen, Frankfurt a.M., Hamburg, Hannover, Köln, Leipzig, München und Osnabrück) wurden jeweils fünf Proben genommen (je eine bei ReWe, Penny, Lidl, Tengelmann sowie einer Bäckerei). In 8 Proben fanden sich resistente Keime.

a. Repräsentativ bedeutet, dass jedes Element einer Grundgesamtheit eine bekannte Wahrscheinlichkeit größer Null hat, in die Stichprobe zu kommen. Wie sehen hier die Grundgesamtheit und die Stichprobe aus? Lassen sich mit den gegebenen Daten die folgenden Fragen beantworten: Ist die hier erhobene Stichprobe repräsentativ? Ist die Stichprobe zufällig aus der Grundgesamtheit gezogen worden? Welche Konsequenzen hätte dies für die Aussagekraft der oben vorgestellten Stichprobe? Welche Aussagen trifft die Studie über ein von Ihnen verzehrtes Mettbrötchen?

b. Nehmen sie an, die 50 Mettbrötchen seien eine Zufallsstichprobe aller verzehrten Mettbrötchen in Deutschland, wie groß wäre dann das 95%ige Vertrauensintervall für die Wahrscheinlichkeit, dass ein Mettbrötchen resistente Keime enthält? Hinweis:

$F_{0,975;86;16}=2,41$; $F_{0,975;18;84}=1,92$

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)

Leere Seiten für Ihre handgeschriebenen Musterlösungen (und sonstige Aufschriebe)