Aufgaben und Lösungen zum Buch

Statistik für Wirtschaftswissenschaftler

Grundlagen und praktische Anwendungen

Ira Frost

1 Grundbegriffe

1. Die *Nationale Verzehrsstudie* ist eine Studie zur Erforschung der Ernährungsgewohnheiten der deutschsprachigen Bevölkerung zwischen 14 und 80 Jahren. Im Rahmen dieser Studie werden u. a. folgende Merkmale erfasst:

   a) Geschlecht
   b) Geburtsjahr
   c) Religion
   d) Zuständigkeit für Einkauf
   e) Raucher
   f) Gesundheitszustand
   g) Familienstand
   h) Anzahl der Personen im Haushalt
   i) Nettoeinkommen des Haushalts
   j) Ausgaben für Lebensmittel und Getränke

   Stellen Sie für jedes Merkmal fest, ob es qualitativ oder quantitativ und ob es diskret oder stetig ist. Geben Sie dazu die jeweilige Skala an.

2. Zur Identifikation von Büchern dient die ISBN. In welche Kategorie würden Sie die ISBN einordnen: Quantitativ oder qualitativ? Welche Skala ist dafür geeignet?

3. Laut einer Umfrage sind in Deutschland Preis, Lage, Bauqualität, Größe, nachbarschaftliche Umgebung und Infrastruktur die wichtigsten Kriterien für einen Immobilienerwerb. Bestimmen Sie für jedes der Kriterien das Skalenniveau.

4. Der zur Klassifizierung des Körpergewichts verwendete Body Mass Index BMI ist definiert als der Quotient zwischen dem Körpergewicht in Kilogramm und dem Quadrat der Körpergröße in Meter:

	BMI (kg/m^2)
Untergewicht	$< 18,5$
Normalgewicht	$18,5 - 24,9$
Übergewicht	$25,0 - 29,9$
Adipositas	$\geq 30,0$

   Welche Skala ist für das Merkmal BMI geeignet? Dem BMI nach sind George Clooney, Brad Pitt und Johnny Depp übergewichtig;

der junge Arnold Schwarzenegger war sogar fettleibig[1]. Diskutieren Sie angesichts dieser Tatsache die Problematik des BMI.

5. Um Kunden regelmäßig zu informieren und zu binden, gibt die Marketing-Abteilung eines großen Auto-Konzerns ein Kunden-Magazin heraus. Die Inhalte des Magazins sollen den Wünschen der Leser angepasst werden. Deshalb wurde eine Leserumfrage durchgeführt. Gefragt werden u. a. nach

   a) der Qualität der Informationen,

   b) der Anzahl der Personen im Haushalt, die das Magazin lesen und

   c) dem Erscheinungsbild des Magazins.

   Geben Sie für jedes der erhobenen Merkmale die Skalierungsart an. Bestimmen Sie, ob es sich um ein qualitatives oder quantitatives Merkmal handelt.

[1] Siehe *Do you believe in Fairies, Unicorns, or BMI?* von Keith Devlin in `http://www.maa.org/devlin/devlin_05_09.html` (Stand: 18.07.2011)

Lösungen zu den Aufgaben

Kapitel 1

1. Merkmale, ihre Ausprägungen und Eigenschaften

Merkmal	Ausprägungen	qualitativ/ quantitativ	stetig/ diskret	Skala
a) Geschlecht	männl., weibl.	qualitativ	-	nominal
b) Geburtsjahr	1911,...,1997	qualitativ	-	ordinal
c) Religion	ev, rk, keine,...	qualitativ	-	nominal
d) Einkauf	Selbst, Partner, andere Person,...	qualitativ	-	nominal
e) Raucher	ja, nein	qualitativ	-	nominal
f) Gesundheitszustand	gut, mittelmäßig, schlecht	qualitativ	-	ordinal
g) Familienstand	ledig, verh.,...	qualitativ	-	nominal
h) Haushaltsgröße	$\mathbb{N}$	quantitativ	diskret	kardinal
i) Nettoeinkommen des Haushalts	$\mathbb{R}_0^+$	quantitativ	stetig	kardinal
j) Ausgaben für Lebensmittel und Getränke	$\mathbb{R}_0^+$	quantitativ	stetig	kardinal

2. Die ISBN ist ein qualitatives und nominalskaliertes Merkmal.
3. Preis und Größe sind kardinalskaliert. Lage, Bauqualität, nachbarschaftliche Umgebung und Infrastruktur sind jeweils ordinalskaliert.
4. Der BMI ist ein ordinalskaliertes Merkmal.
5. a) Das Merkmal *Qualität der Informationen* ist qualitativ und ordinalskaliert.

   b) Das Merkmal *Anzahl der Personen im Haushalt, die das Magazin lesen* ist quantitativ und kardinalskaliert.

   c) Das Merkmal *Erscheinungsbild des Magazins* ist qualitativ und ordinalskaliert .

2 Eindimensionale Daten

1. Bestimmen Sie für jede der folgenden Datenreihen den (die) Modalwert(e), den Median und das arithmetische Mittel.

   Reihe x : 6 17 10 9 6 6 3 28 10 27 10

   Reihe y : 20 20 11 3 11 11 19 20 11 15 16 11

2. Eine Befragung von 20 Schülern nach ihrem Sternzeichen ergibt folgende Ergebnisse: Krebs, Fische, Krebs, Krebs, Löwe, Waage, Skorpion, Fische, Löwe, Löwe, Löwe, Skorpion, Skorpion, Fische, Skorpion, Löwe, Waage, Skorpion, Skorpion, Fische.

   a) Bilden Sie eine Häufigkeitstabelle für das Merkmal *Sternzeichen.*

   b) Welche(r) Lageparameter sind (ist) zur Beschreibung dieser Daten geeignet? Geben Sie diese(n) an.

3. Laut der Umfrage[1] *Achten Sie beim Kauf von Lebensmitteln auf den Kaloriengehalt?* des Bundesministeriums für Ernährung und Verbraucherschutz achten 39 Prozent der Befragten gar nicht, 30 Prozent gelegentlich, 20 Prozent oft und 7 Prozent immer auf den Kaloriengehalt. Die restlichen 4 Prozent der Befragten geben an, keine Lebensmittel zu kaufen. Bestimmen Sie für ein korrektes Kreisdiagramm die Winkel der Kreissektoren.

4. In vielen Hochschulen ist es verbreitet, dass Studierende am Ende eines Semesters zum Inhalt einer Vorlesung ihre Meinung äußern. Den Inhalt der Statistik-Vorlesung halten 9 Studierende für sehr anspruchsvoll, 80 Studierende halten sie für einfach und 91 Studierende finden das Vorlesungsniveau genau richtig. Geben Sie mindestens einen Lageparameter für das Merkmal *Schwierigkeitsgrad des Vorlesungsinhaltes* an. Begründen Sie Ihre Entscheidung und interpretieren Sie Ihre Ergebnisse.

[1] `http://de.statista.com/statistik/daten/studie/12414/umfrage/beachten-25---34-jaehrige-kalorien-beim-einkauf/` (Stand: 09.03.2011)

5. Eine Umfrage *Wie viele Stunden verbrachten Sie im vergangenen Monat im Internet?* unter zehn zufällig ausgewählten Studierenden einer Hochschule ergibt die folgenden Werte:

   1 0 10 27 13 7 5 12 8 19

   Geben Sie für den gegebenen Zeitraum die mittlere Verweildauer im Internet pro Person an.

6. In einem Fertigungssystem muss ein eingegangener Auftrag in der Regel auf eine freie Bedienungsstation warten. Um sich ein Bild von den derzeit üblichen Wartezeiten zu machen, werden die Wartezeiten (in Tagen) von 20 zufällig ausgewählten Aufträgen bis zur Bearbeitung untersucht:

Wartezeit von... bis unter...	Anzahl
4 – 5	4
5 – 6	3
6 – 7	5
7 – 8	4
8 –11	4
	20

   a) Wählen Sie eine geeignete Grafik für die Darstellung der Daten. Geben Sie die dafür benötigten Größen an.

   b) Wie hoch ist der Anteil der Aufträge, die eine Wartezeit von höchsten 7,5 Tagen haben?

   c) Geben Sie einen Näherungswert für die durchschnittliche Wartezeit pro Auftrag an.

   d) Bestimmen Sie die Modal- und die Medianklasse.

7. Die folgende Tabelle gibt die Wochenumsätze (in 1000 €) von 20 Filialen einer Drogeriekette an:

Klasse	Umsatz von... bis unter...	Häufigkeit	Klassenmittelwert
1	10 – 13	6	12,50
2	13 – 15	5	13,65
3	15 – 17	4	15,80
4	17 – 22	5	20,20

   Berechnen Sie den wöchentlichen Umsatz pro Filiale.

8. Berechen Sie für die Reihe

$$2 \quad 1,25 \quad 1,5 \quad 2 \quad 0,4$$

das arithmetische und das geometrische Mittel.

9. Die folgende Tabelle gibt den Markenwert[2] von fünf Fußballspielern in den Jahren 2010 und 2011 (in Millionen Euro) wieder:

Spieler	Markenwert 2010	Markenwert 2011
B. Schweinsteiger	20,7	25,2
P. Lahm	12,2	24,0
M. Neuer	6,6	22,1
M. Özil	12,2	21,4
T. Müller	10,9	19,5

Geben Sie für jeden Spieler den Wachstumsfaktor bzw. die Wachstumsrate seines Markenwertes an.

10. Die folgende Tabelle gibt die Veränderung des Bruttoinlandsproduktes eines Landes im Zeitraum von 2000 bis 2003 gegenüber dem Vorjahr in Prozent wieder. Mit welcher Rate wuchs das Bruttoinlandsprodukt im Zeitraum von 1999 bis 2003 im Mittel pro Jahr?

Jahr	’00	’01	’02	’03
Veränderung gegenüber dem Vorjahr in %	3,2	1,2	0,0	-0,2

11. Ein Online-Händler freut sich über die Entwicklung seines Unternehmens. Der Umsatz wuchs gegenüber dem Vorjahr im Jahr 2007 um 20%, im Jahr 2008 um 50% und im Jahr 2009 um 80%. Der Weg nach oben setzt sich fort: Im Jahr 2010 hat sich der Umsatz gegenüber 2009 verdoppelt. Ermitteln Sie die jährliche mittlere Wachstumsrate der Umsätze. Geben Sie die Umsatzzahlen des Unternehmens für die Jahre 2007, 2008, 2009 und 2010 an, wenn es im Jahr 2006 einen Umsatz von 2 Mio. € erzielt hatte.

[2] Der Markenwert der Spieler ergibt sich aus Daten wie deren Gehalt, den Werbeeinahmen sowie dem in einer repräsentativen Umfrage ermittelten Image. Quelle: *Süddeutsche Zeitung*, 30./31. Juli 2011

12. Bestimmen Sie für jede der folgenden Datenreihen die Varianz, die Standardabweichung und den Variationskoeffizienten.

Reihe x :	6	12	10	4	6	6	2	18	10	18	10	12
Reihe y :	18	36	30	12	18	18	6	54	30	54	30	36
Reihe z :	10	16	14	8	10	10	6	22	14	22	14	16

13. In den letzten 12 Monaten verzeichnet ein Gebrauchtwagenhändler die folgenden Umsatzzahlen (in 100.000 €):

Monat	Umsatz	Monat	Umsatz
Jan.	5	Juli	1
Feb.	3	Aug.	3
März	6	Sept.	2
April	2	Okt.	2
Mai	4	Nov.	3
Juni	5	Dez.	6

Berechnen Sie die Varianz bzw. die Standardabweichung des Merkmals X: *Umsatz*. Was besagt in diesem Fall die Standardabweichung? Wie groß ist die Spannweite des Umsatzes?

14. Für das Merkmal X: *Wochenumsatz* (in 1000 €) aus der Aufgabe 7 liegt nun für jede Klasse die Standardabweichung vor.

Klasse	Umsatz von... bis unter...	Häufigkeit	Klassen-mittelw.	Standardabw. der Klasse
1	10 − 13	6	12,50	0,8
2	13 − 15	5	13,65	1,2
3	15 − 17	4	15,80	2,5
4	17 − 22	5	20,20	3,0

Wie hoch ist die gesamte Standardabweichung?

Lösungen zu den Aufgaben

Kapitel 2

1. Die geordnete Reihe x:

   $$3 \quad 6 \quad 6 \quad 6 \quad 9 \quad 10 \quad 10 \quad 10 \quad 17 \quad 27 \quad 28$$

   Diese besitzt zwei Modalwerte $x_{mod,1} = 6$ und $x_{mod,2} = 10$.

   Der Median ist $x_{med} = x_{[\frac{11+1}{2}]} = x_{[6]} = 10$.

   Das arithmetische Mittel berechnet sich gemäß:

   $$\begin{aligned}\bar{x} &= \frac{1}{11}\sum_{i=1}^{11} x_i \\ &= \frac{6+17+10+9+6+6+3+28+10+27+10}{11} \\ &= \frac{132}{11} = 12\end{aligned}$$

   bzw.

   $$\begin{aligned}\bar{x} &= \frac{1}{11}\sum_{j=1}^{7} x_j f_j \\ &= \frac{3\cdot 1+6\cdot 3+9\cdot 1+10\cdot 3+17\cdot 1+27\cdot 1+28\cdot 1}{11} = 12\end{aligned}$$

   Die geordnete Reihe y:

   $$3 \quad 11 \quad 11 \quad 11 \quad 11 \quad 11 \quad 15 \quad 16 \quad 19 \quad 20 \quad 20 \quad 20$$

   Der Modus ist $y_{mod} = 11$.

   Der Median ist $y_{med} = \frac{x_{[6]}+x_{[7]}}{2} = \frac{11+15}{2} = 13$.

   Das arithmetische Mittel berechnet sich gemäß:

   $$\begin{aligned}\bar{y} &= \frac{20+20+11+3+11+11+19+20+11+15+16+11}{12} \\ &= \frac{168}{12} = 14\end{aligned}$$

bzw.

$$\bar{y} = \frac{3 \cdot 1 + 11 \cdot 5 + 15 \cdot 1 + 16 \cdot 1 + 19 \cdot 1 + 20 \cdot 3}{12} = 14$$

2. Häufigkeitstabelle:

Sternzeichen	Krebs	Löwe	Waage	Skorpion	Fische
Anzahl	3	5	2	6	4

Das Merkmal *Sternzeichen* ist nominal skaliert. Deshalb ist der Modalwert der einzig geeignete Lageparameter. Für diesen Datensatz ist Skorpion der Modus.

3. Der Winkel bestimmt sich gemäß $\alpha_j = h_j \cdot 360°,\ j = 1, \ldots, 5.$

Antwortmöglichkeiten x_j	Anteil (in %) h_j	Winkel α_j
Achte gar nicht darauf	39	140,4°
Achte gelegentlich darauf	30	108°
Achte oft darauf	20	72°
Achte immer darauf	7	25,2°
Kaufe keine Lebensmittel	4	14,4°
	100	360°

(Geeignet als Lageparameter ist der Modus. Dieser ist x_1= *Achte gar nicht darauf*.)

4. Das Merkmal *Schwierigkeitsgrad des Faches Statistik* ist ordinalskaliert. Die Ausprägungen lassen sich etwa wie folgt ordnen:

$$\text{einfach} = 0 \quad \text{genau richtig} = 1 \quad \text{anspruchsvoll} = 2$$

Deshalb sind Modus und Median als Lageparameter geeignet. Die Mehrheit der Studierenden findet den Vorlesungsinhalt genau richtig. Die Ausprägung *genau richtig* ist auch der Median, denn: Mindestens 50% der 180 Studierenden halten den Inhalt für einfach bis genau richtig und mindestens 50% halten ihn für genau richtig bis anspruchsvoll.

5. Mit X: *Dauer der Internetnutzung* ist

$$\bar{x} = \frac{1 + 0 + 10 + 27 + 13 + 7 + 5 + 12 + 8 + 19}{10} = 10,2.$$

Im Mittel verbrachte im letzten Monat jeder Befragte 10,2 Stunden im Internet.

6. a) Da die Daten klassiert sind, stellt man sie grafisch als Histogramm dar. Um das Histogramm zu zeichnen, benötigt man für jede Klasse ihre Breite $b_j - b_{j-1}$ und ihre Dichte f_j^* (siehe Tabelle). Die Klassendichte f_j^* legt die Säulenhöhe der Klasse j fest.

   b) Der Anteil der Aufträge mit einer Wartezeit von höchstens 7,5 Tagen beträgt 80 Prozent. Dieser ergibt sich gemäß $H(7,5) = H_4 = 0,80$.

   c) Im Mittel beträgt die Wartezeit ca. 6,75 Tage pro Auftrag. Diesen Wert erhält man gemäß

$$\begin{aligned}\bar{m} &= \sum_{j=1}^{5} m_j \cdot h_j \\ &= 4,5 \cdot 0,20 + 5,5 \cdot 0,15 + 6,5 \cdot 0,25 + 7,5 \cdot 0,20 + 9,5 \cdot 0,20 \\ &= 6,75.\end{aligned}$$

   d) Die Modalklasse ist die Klasse [6, 7[(höchste Dichte); die Medianklasse ist ebenso die Klasse [6, 7[.

Wartezeit von… bis unter… $[b_{j-1}, b_j[$	Klassenbreite $b_j - b_{j-1}$	Abs. H. f_j	Klassendichte f_j^*	Rel. H. h_j	Kumul. rel. H. H_j	Klassenmitte m_j
4 − 5	1	4	4	0,20	0,20	4,5
5 − 6	1	3	3	0,15	0,35	5,5
6 − 7	1	5	5	0,25	0,60	6,5
7 − 8	1	4	4	0,20	0,80	7,5
8 −11	3	4	$\frac{4}{3}$	0,20	1,00	9,5
		20		1,00		

7. Mit den Bezeichnungen $\bar{x}_j$: Umsatz pro Filiale in der Klasse j und f_j: Anzahl der Filialen in der Klasse j errechnet sich $\bar{x}_{ges}$, der Umsatz pro Filiale insgesamt, gemäß

$$\begin{aligned}\bar{x}_{ges} &= \frac{1}{20} \sum_{j=1}^{4} f_j \cdot \bar{x}_j \\ &= \frac{6 \cdot 12,50 + 5 \cdot 13,65 + 4 \cdot 15,80 + 5 \cdot 20,20}{20} \\ &= \frac{307,45}{20} = 15,3725.\end{aligned}$$

Der Wochenumsatz beträgt 15.372,50 € pro Filiale.

8. Das arithmetische Mittel ist:

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{2 + 1,25 + 1,5 + 2 + 0,4}{5} = 1,43$$

Das geometrische Mittel ist:

$$x_{geom} = \sqrt[5]{\prod_{i=1}^{5} x_i} = \sqrt[5]{2 \cdot 1,25 \cdot 1,5 \cdot 2 \cdot 0,4} = 1,246$$

Allgemein gilt: $\bar{x} \geq x_{geom}$

9. Sei $x_t =$ *Markenwert zum Zeitpunkt* t, $t = 0$ (2010) und $t = 1$ (2011).

Spieler	Markenwert		W.-Faktor	W.-Rate (in %)
	x_0	x_1	$q_t = \frac{x_1}{x_0}$	$w_t = (q_t - 1)100$
B. Schweinsteiger	20,7	25,2	1,217	21,7
P. Lahm	12,2	24,0	1,967	96,7
M. Neuer	6,6	22,1	3,348	234,8
M. Özil	12,2	21,4	1,754	75,4
T. Müller	10,9	19,5	1,789	78,9

Jeder der fünf Spieler verzeichnet eine positive Wachstumsrate. Philipp Lahm konnte seinen Markenwert nahezu verdoppeln. Den größten Sprung machte Manuel Neuer, der ein Plus von fast 235% erfahren hat.

10. Die jeweiligen Veränderungen gegenüber dem Vorjahr stellen die Wachstumsraten w_t $(t = 1, 2, 3, 4)$ des BIP dar.

Jahr	’00	’01	’02	’03
t	1	2	3	4
Wachstumsrate in % w_t	3,2	1,2	0,0	-0,2
Wachstumsfaktor q_t	1,032	1,012	1,000	0,998

Den durchschnittlichen Wachstumsfaktor q errechnet man gemäß

$$q = \sqrt[T]{\prod_{t=1}^{T} q_t} = \sqrt[4]{1,032 \cdot 1,012 \cdot 1,000 \cdot 0,998} = \sqrt[4]{1,042} = 1,0104.$$

Die durchschnittliche Wachstumsrate (in Prozent) beträgt somit

$$w = (q-1)100 = (1,0104-1)100 = 1,04.$$

Im Mittel wächst das BIP im Zeitraum vom Jahr 1999 bis zum Jahr 2003 um 1,04% jährlich.

11. Für die Ermittlung der durchschnittlichen Wachstumsrate berechnen wir zunächst die Wachstumsfaktoren q_t (4. Zeile der folgenden Tabelle):

Jahr	'07	'08	'09	'10
Zeitpunkt t	1	2	3	4
Wachstumsrate w_t (in %)	20	50	80	100
Wachstumsfaktor q_t	1,2	1,5	1,8	2

Der durchschnittliche Wachstumsfaktor ergibt sich gemäß

$$q = \sqrt[T]{\prod_{t=1}^{T} q_t} = \sqrt[4]{1,2 \cdot 1,5 \cdot 1,8 \cdot 2} = \sqrt[4]{6,48} = 1,5955.$$

Die durchschnittliche Wachstumsrate (in Prozent) beträgt somit

$$w = (q-1)100 = (1,5955-1)100 = 59,55.$$

Im Zeitraum von 2006 bis 2010 wuchs der Umsatz im Mittel um 59,55% jährlich.

Die Umsatzzahlen für die Jahre 2007 bis 2010, wenn im Jahr 2006 der Umsatz 2 Mio. € ($u_0 = 2$) betrug, berechnen sich gemäß

$$u_t = q_t \cdot u_{t-1}.$$

Der Umsatz (in Mio. €) beträgt für das Jahr

$$\begin{aligned}
2007: \quad u_1 &= q_1 \cdot u_0 = 1,2 \cdot 2,00 = 2,40 \\
2008: \quad u_2 &= q_2 \cdot u_1 = 1,5 \cdot 2,40 = 3,60 \\
2009: \quad u_3 &= q_3 \cdot u_2 = 1,8 \cdot 3,60 = 6,48 \\
2010: \quad u_4 &= q_4 \cdot u_3 = 2,0 \cdot 6,48 = 12,96
\end{aligned}$$

12. Arbeitstabelle zur Berechnung der Varianz, der Standardabweichung und des Variationskoeffizienten der Reihe x:

j	x_j	f_j	$x_j f_j$	$(x_j - \bar{x})$	$(x_j - \bar{x})^2 f_j$
1	2	1	2	-7,5	56,25
2	4	1	4	-5,5	30,25
3	6	3	18	-3,5	36,75
4	10	3	30	0,5	0,75
5	12	2	24	2,5	12,5
6	18	2	36	8,5	144,5
		12	114		281

Das arithmetisches Mittel beträgt

$$\bar{x} = \frac{\sum_{j=1}^{6} x_j f_j}{n} = \frac{114}{12} = 9,5.$$

Die Varianz beträgt

$$s_x^2 = \frac{1}{n} \sum_{j=1}^{6} (x_j - \bar{x})^2 f_j = \frac{281}{12} = 23,4167.$$

(Berechnung von s_x^2 mit Hilfe des Verschiebungssatzes:

$$\begin{aligned} s_x^2 &= \frac{1}{12} \sum_{j=1}^{6} x_j^2 f_j - \bar{x}^2 \\ &= \frac{2^2 \cdot 1 + 4^2 \cdot 1 + 6^2 \cdot 3 + 10^2 \cdot 3 + 12^2 \cdot 2 + 18^2 \cdot 2}{12} - 9,5^2 \\ &= \frac{1364}{12} - 90,25 = 23,4167) \end{aligned}$$

Der Variationskoeffizient berechnet sich gemäß

$$VK_x = \frac{s_x}{\bar{x}} = \frac{\sqrt{23,4167}}{9,5} = 0,5094.$$

Für die Werte der Reihe y gilt:

$$y_i = 3x_i,\ i = 1, \ldots, 12$$

Deshalb können wir für die Bestimmung von $\bar{y}$ und s_y^2 deren Eigenschaften nutzen:

$$\bar{y} = 3\bar{x} = 3 \cdot 9,5 = 28,5 \text{ und } s_y^2 = 3^2 s_x^2 = 9 \cdot 23,4167 = 210,75$$

Der Variationskoeffizient ist:

$$VK_y = \frac{s_y}{\bar{y}} = \frac{\sqrt{210,75}}{28,5} = 0,5094$$

Für die Werte der Reihe z gilt:

$$z_i = 4 + x_i,\ i = 1,\dots,12$$

Hier können wir für die Bestimmung von $\bar{z}$ und s_z^2 ebenso deren Eigenschaften nutzen:

$$\bar{z} = 4 + \bar{x} = 4 + 9,5 = 13,5 \text{ und } s_z^2 = s_x^2 = 23,4167$$

Der Variationskoeffizient beträgt:

$$VK_z = \frac{s_z}{\bar{z}} = \frac{\sqrt{23,4167}}{13,5} = 0,3585$$

13. Für die Bestimmung der Varianz stellen wir die folgende Arbeitstabelle auf:

j	x_j	f_j	$x_j f_j$	$x_j - \bar{x}$	$f_j(x_j - \bar{x})^2$
1	1	1	1	-2,5	6,25
2	2	3	6	-1,5	6,75
3	3	3	9	-0,5	0,75
4	4	1	4	0,5	0,25
5	5	2	10	1,5	4,50
6	6	2	12	2,5	12,50
		12	42		31

Der Durchschnittsumsatz pro Monat berechnet sich gemäß

$$\bar{x} = \frac{\sum_{j=1}^{6} x_j f_j}{12} = \frac{42}{12} = 3,5.$$

(Der Umsatz beträgt im Mittel 350.000 € pro Monat.) Die Varianz bzw. die Standardabweichung:

$$s^2 = \frac{31}{12} = 2,5833 \quad \text{bzw.} \quad s = \sqrt{2,5833} = 1,6073$$

Das bedeutet: Im Mittel weicht der monatliche Umsatz um fast 161.000 € von 350.000 € ab.

Die Spannweite (in 1000 €) beträgt

$$R = x_{max} - x_{min} = 600 - 100 = 500.$$

14. Aus der Lösung der Aufgabe 7 entnehmen wir das gesamte mittlere Einkommen von 15.372,50 €.

    Um die gesamte Standardabweichung zu berechnen, muss man zunächst die gesamte Varianz ermitteln. Diese ist die Summe aus der inneren und der äußeren Varianz:

$$s_{ges}^2 = \frac{1}{n}\sum_{j=1}^{k} f_j s_j^2 + \frac{1}{n}\sum_{j=1}^{k} f_j(\bar{x}_j - \bar{x}_{ges})^2$$

    Die innere Varianz errechnet sich gemäß:

$$\frac{1}{n}\sum_{j=1}^{k} f_j s_j^2 = \frac{1}{20}(6\cdot 0,8^2+5\cdot 1,2^2+4\cdot 2,5^2+5\cdot 3^2) = \frac{81,04}{20} = 4,052$$

    Arbeitstabelle zur Berechnung der äußeren Varianz:

j	$[b_{j-1};b_j]$	f_j	$\bar{x}_j$	$(\bar{x}_j - \bar{x}_{ges})^2$	$f_j(\bar{x}_j - \bar{x}_{ges})^2$
1	10 − 13	6	12,50	8,2513	49,5075
2	13 − 15	5	13,65	2,9670	14,8350
3	15 − 17	4	15,80	0,1828	0,7310
4	17 − 22	5	20,20	23,3048	116,5238
		20			181,5974

    Die äußere Varianz beträgt

$$\frac{1}{n}\sum_{j=1}^{k} f_j(\bar{x}_j - \bar{x}_{ges})^2 = \frac{181,5974}{20} = 9,0799.$$

    Somit ergibt sich eine Gesamtvarianz von

$$s_{ges}^2 = 4,052 + 9,0799 = 13,1319.$$

    Daraus ergibt sich eine Standardabweichung von

$$s = \sqrt{13,1319} = 3,6238.$$

    Das heißt: Die Wochenumsätze variieren durchschnittlich um ca. 3624 € vom Mittelwert (15.372,50 €).

3 Zweidimensionale Daten

1. Seien X und Y kardinalskalierte Merkmale mit den Ausprägungen $x_1 = 3,\ x_2 = 5$ bzw. $y_1 = 2,\ y_2 = 4,\ y_3 = 6$. Für X und Y sei die folgende Kontingenztabelle bekannt:

	y		
x	2	4	6
3	15	25	15
5	10	40	115

   Bestimmen Sie für $j = 1, 2$ und $k = 1, 2, 3$:

   a) die relativen Randhäufigkeiten von X bzw. von Y

   b) die bedingten Häufigkeiten von X, gegeben $Y = y_k$

   c) die bedingten Häufigkeiten von Y, gegeben $X = x_j$

2. Berechnen Sie jeweils das arithmetische Mittel und die Varianz der Merkmale X bzw. Y aus der Aufgabe 1. Geben Sie die Kovarianz zwischen X und Y an.

3. Vervollständigen Sie die folgende Kontingenztabelle, wenn die Merkmale X und Y unabhängig sind.

	y			
x	y_1	x_2	y_3	
x_1				60
x_2				90
	50	70	30	150

4. Die folgende Kontingenztabelle zeigt das Ergebnis einer Umfrage *Wo kaufen Sie Ihre Bücher?*. Zusätzlich zum Kaufverhalten (V) wird das Geschlecht (G) der Kunden notiert.

Geschlecht	Wo kaufen Sie Ihre Bücher?				
	Nur im Laden	Eher im Laden	Nur im Internet	Eher im Internet	
Männlich	90	118	50	42	300
Weiblich	90	110	40	60	300
	180	228	90	102	600

a) Wie hoch ist der männliche Anteil unter den Befragten?

b) Wie hoch ist der Anteil der Befragten, die nur im Laden oder eher im Laden ihre Bücher kaufen?

c) Interpretieren Sie die Werte $\frac{f_{1k}}{f_{1\cdot}}$ der bedingten Verteilung des Merkmals V, gegeben $G=$ Männlich ($V|G=$ Männlich).

Nur im L.	Eher im L.	Nur im Int.	Eher im Int.	
0,30	0,39	0,17	0,14	1,00

d) Geben Sie die bedingte Verteilung des Merkmals V, gegeben $G=$ Weiblich ($V|G=$ Weiblich) an.

e) Geben Sie die bedingten Verteilungen des Merkmals G an.

5. Wie sähe die Kontingenztabelle in der Aufgabe 4 aus, wenn das Kaufverhalten vom Geschlecht unabhängig wäre? Wie stark ist die Abhängigkeit zwischen den Merkmalen *Geschlecht* und *Kaufverhalten*?

6. Eine Modekette, die insbesondere Jugendliche und junge Erwachsene als Zielgruppe hat, bietet Gutscheine im Wert von 25 € und 50 € an. Um die Akzeptanz der Gutscheine zu untersuchen, wird eine Umfrage unter 200 zufällig ausgewählten jungen Kunden im Alter von 12 bis 20 Jahren durchgeführt. Dabei wird das Alter in drei Klassen aufgeteilt. Die folgende Tabelle gibt die Häufigkeiten der Gutscheinkategorie in jeder Altersklasse wieder.

Alter von... bis unter...	Gutschein in Wert von 25 €	50 €
12 – 15	50	20
15 – 18	40	30
18 – 20	30	30

a) Ermitteln Sie für die untersuchten Merkmale den korrigierten Kontingenzkoeffizienten nach Pearson.

b) Gibt es eine weitere Maßzahl, die den Zweck ebenso erfüllt? (Hinweis: Denken Sie an die Skalenart der Merkmale.)

7. Zwischen den Merkmalen *Körpergröße* (in cm) und *Körpergewicht* (in kg) besteht ein linearer Zusammenhang. Welche Maßzahl ist zur Messung des linearen Zusamenhangs geeignet? Berechnen Sie den Wert dieses Maßes für die folgenden Beobachtungswerte:

	Körpergewicht		
Körpergröße	70	75	80
175	5	2	0
180	0	0	3

8. Von sechs zufällig ausgewählten Haushalten einer bestimmten Berufsgruppe liegen die folgenden Jahresbeträge (in Tausend €) der Merkmale X: *Verfügbares Einkommen* und Y: *Ausgaben für Lebensmittel* vor:

Verfügbares Einkommen x_i	160	130	140	140	140	150
Ausgaben y_i	14,4	16	14,8	15	15,2	14,8

Berechnen Sie den Bravais-Pearson-Korrelationskoeffizienten, wenn zwischen den Variablen eine lineare Abhängigkeit unterstellt wird.

9. Das Gesamtrisiko eines Portfolios ist umso geringer, je weniger die einzelnen Anlagen miteinander korrelieren. Ein potentieller Anleger, der besonders vorsichtig ist, möchte ein Portfolio aus zwei Aktien (A und B oder A und C) zusammenstellen. Die Kursentwicklung der Aktien (in €) über sechs Tage wird in der folgenden Tabelle wiedergegeben:

Aktie A	53	54	55	58	60	58
Aktie B	61	58	57	58	61	60
Aktie C	51	45	43	50	53	55

Für welche Aktien wird sich der potentielle Anleger voraussichtlich entscheiden?

10. Bestimmen Sie eine geeignete Maßzahl zur Messung der Abhängigkeit zwischen den Merkmalen *Sportlicher Rang* und *Umsatz der Vereine* der folgenden Fußballmanschaften. Berechnen Sie ihren Wert für die vorliegenden Daten (*Süddeutsche Zeitung 9. Juni 2011*) ohne Bayer Leverkusen.

Fußballmannschaft	Sportlicher Rang 2010/2011	Umsatz
FC Bayern München	3	350,2
Hamburger SV	8	146,2
VFB Stuttgart	12	145,8
FC Schalke 04	14	139,8
SV Werder Bremen	13	126,4
VfL Wolfsburg	15	115,9
Borrusia Dortmund	1	112,2
1. FC Köln	10	72,6
Eintracht Frankfurt	17	68,4
Borrusia Mönchengladbach	16	67,2
Hannover 96	4	54,0
1. FC Nürnberg	6	47,7
1899 Hoffenheim	11	41,8
1. FSV Mainz 05	5	32,2
SC Freiburg	9	31,0
1. FC Kaiserslauten	7	27,7
FC St. Pauli	18	20,5
Bayer Leverkusen	2	keine Angabe

Lösungen zu den Aufgaben

Kapitel 3

1. a) Die relativen Randhäufigkeiten von X:

$$h_{1\bullet} = \frac{f_{1\bullet}}{n} = \frac{1}{n}\sum_{k=1}^{3} f_{1k} = \frac{15+25+15}{220} = \frac{55}{220}$$
$$h_{2\bullet} = \frac{f_{2\bullet}}{n} = \frac{1}{n}\sum_{k=1}^{3} f_{2k} = \frac{10+40+115}{220} = \frac{165}{220}$$

Die relativen Randhäufigkeiten von Y:

$$h_{\bullet 1} = \frac{f_{\bullet 1}}{n} = \frac{1}{n}\sum_{j=1}^{2} f_{j1} = \frac{15+10}{220} = \frac{25}{220}$$
$$h_{\bullet 2} = \frac{f_{\bullet 2}}{n} = \frac{1}{n}\sum_{j=1}^{2} f_{j2} = \frac{25+40}{220} = \frac{65}{220}$$
$$h_{\bullet 3} = \frac{f_{\bullet 3}}{n} = \frac{1}{n}\sum_{j=1}^{2} f_{j3} = \frac{15+115}{220} = \frac{130}{220}$$

b) Die bedingten Häufigkeiten von X, gegeben $Y = y_k$:

$X\|Y=2$	$X\|Y=4$	$X\|Y=6$
$\frac{f_{11}}{f_{\bullet 1}} = \frac{15}{25}$	$\frac{f_{12}}{f_{\bullet 2}} = \frac{25}{65}$	$\frac{f_{13}}{f_{\bullet 3}} = \frac{15}{130}$
$\frac{f_{21}}{f_{\bullet 1}} = \frac{10}{25}$	$\frac{f_{22}}{f_{\bullet 2}} = \frac{40}{65}$	$\frac{f_{23}}{f_{\bullet 3}} = \frac{115}{130}$

c) Die bedingten Häufigkeiten von Y, gegeben $X = x_j$:

$Y\|X=3$	$Y\|X=5$
$\frac{f_{11}}{f_{1\bullet}} = \frac{15}{55}$	$\frac{f_{21}}{f_{2\bullet}} = \frac{10}{165}$
$\frac{f_{12}}{f_{1\bullet}} = \frac{25}{55}$	$\frac{f_{22}}{f_{2\bullet}} = \frac{40}{165}$
$\frac{f_{13}}{f_{1\bullet}} = \frac{15}{55}$	$\frac{f_{23}}{f_{2\bullet}} = \frac{115}{165}$

2. Das arithmetische Mittel von X:

$$\bar{x} = \frac{x_1 f_{1\bullet} + x_2 f_{2\bullet}}{n} = \frac{3 \cdot 55 + 5 \cdot 165}{220} = \frac{990}{220} = 4,5$$

Das arithmetische Mittel von Y:

$$\begin{aligned} \bar{y} &= \frac{y_1 f_{\bullet 1} + y_2 f_{\bullet 2} + y_3 f_{\bullet 3}}{n} = \frac{2 \cdot 25 + 4 \cdot 65 + 6 \cdot 130}{220} \\ &= \frac{1090}{220} = 4,95 \end{aligned}$$

Die Varianz von X:

$$\begin{aligned} s_X^2 &= \frac{1}{n}\sum_{j=1}^{2} x_j^2 f_{j\bullet} - \bar{x}^2 = \frac{1}{220}(3^2 \cdot 55 + 5^2 \cdot 165) - 4,5^2 \\ &= \frac{4620}{220} - 20,25 = 0,75 \end{aligned}$$

Die Varianz von Y:

$$\begin{aligned} s_Y^2 &= \frac{1}{n}\sum_{k=1}^{3} y_k^2 f_{\bullet k} - \bar{y}^2 = \frac{1}{220}(2^2 \cdot 25 + 4^2 \cdot 65 + 6^2 \cdot 130) - 4,95^2 \\ &= \frac{5820}{220} - 24,5025 = 1,95 \end{aligned}$$

Die Kovarianz zwischen X und Y:

$$\begin{aligned} s_{XY} &= \frac{1}{n}\sum_{j=1}^{2}\sum_{k=1}^{3} x_j y_k f_{jk} - \bar{x}\bar{y} \\ &= \frac{1}{220}(3 \cdot 2 \cdot 15 + 3 \cdot 4 \cdot 25 + 3 \cdot 6 \cdot 15 + 5 \cdot 2 \cdot 10 \\ &\quad +5 \cdot 4 \cdot 40 + 5 \cdot 6 \cdot 115) - 4,5 \cdot 4,95 \\ &= \frac{5010}{220} - 22,275 = 0,50 \end{aligned}$$

3. Da X und Y unabhängig sind, gilt $f_{jk} = e_{jk} = \frac{f_{j\bullet} \cdot f_{\bullet k}}{n}$; die Werte sind:

$$\begin{aligned} e_{11} &= \tfrac{f_{1\bullet} \cdot f_{\bullet 1}}{n} = \tfrac{60 \cdot 50}{150} = 20 & e_{21} &= \tfrac{f_{2\bullet} \cdot f_{\bullet 1}}{n} = \tfrac{90 \cdot 50}{150} = 30 \\ e_{12} &= \tfrac{f_{1\bullet} \cdot f_{\bullet 2}}{n} = \tfrac{60 \cdot 70}{150} = 28 & e_{22} &= \tfrac{f_{2\bullet} \cdot f_{\bullet 2}}{n} = \tfrac{90 \cdot 70}{150} = 42 \\ e_{13} &= \tfrac{f_{1\bullet} \cdot f_{\bullet 3}}{n} = \tfrac{60 \cdot 30}{150} = 12 & e_{23} &= \tfrac{f_{2\bullet} \cdot f_{\bullet 3}}{n} = \tfrac{90 \cdot 30}{150} = 18 \end{aligned}$$

Die Kontingenztabelle lautet somit:

	y_1	x_2	y_3	
x_1	20	28	12	60
x_2	30	42	18	90
	50	70	30	150

4. a) Die Hälfte der Befragten ist männlich.

   b) Der Anteil der Befragten, die ihre Bücher ausschließlich oder eher im Laden kaufen, beträgt 68%. Diesen Wert erhalten wir folgendermaßen:

$$h_{\bullet 1} + h_{\bullet 2} = \frac{f_{\bullet 1}}{n} + \frac{f_{\bullet 2}}{n} = \frac{180}{600} + \frac{228}{600} = 0,68$$

   c) Unter den (300) männlichen Befragten kaufen 30% ihre Bücher nur, 39% hauptsächlich im Laden, 17% nur und 14% hauptsächlich im Internet.

   d) Die Werte von $V|G=$ Weiblich sind für $k = 1, 2, 3, 4$ gegeben durch $\frac{f_{2k}}{f_{2\bullet}}$:

Nur im L.	Eher im L.	Nur im Int.	Eher im Int.	
0,30	0,37	0,13	0,20	1,00

   e) Da das Merkmal V vier Ausprägungen besitzt, gibt es vier bedingte Verteilungen für das Merkmal G. Die Ausprägungen von V bezeichnen wir mit $v_1 =$ Nur im Laden, $v_2 =$ Eher im Laden, $v_3 =$ Nur im Internet und $v_4 =$ Eher im Internet.

	$G\|V = v_1$ $\frac{f_{j1}}{f_{\cdot 1}}$	$G\|V = v_2$ $\frac{f_{j2}}{f_{\cdot 2}}$	$G\|V = v_3$ $\frac{f_{j3}}{f_{\cdot 3}}$	$G\|V = v_4$ $\frac{f_{j4}}{f_{\cdot 4}}$
Männlich	0,50	0,52	0,56	0,41
Weiblich	0,50	0,48	0,44	0,59
	1,00	1,00	1,00	1,00

5. Die Kontingenztabelle in der Aufgabe 4, wenn die Merkmale G: *Geschlecht* und V: *Kaufverhalten* unabhängig wären:

Geschlecht	Wo kaufen Sie Ihre Bücher?				
	Nur im Laden	Eher im Laden	Nur im Internet	Eher im Internet	
Männlich	90	114	45	51	300
Weiblich	90	114	45	51	300
	180	228	90	102	600

Die Stärke der Abhängigkeit zwischen den Merkmalen G: *Geschlecht* und V: *Kaufverhalten* wird durch den korrigierten Kontingenzkoeffizienten nach Pearson gemessen, weil die Merkmale nominal skaliert sind. Arbeitstabelle zur Berechnung von P_{korr}:

j	k	f_{jk}	e_{jk}	$f_{jk} - e_{jk}$	$\frac{(f_{jk}-e_{jk})^2}{e_{jk}}$
1	1	90	90	0	0
1	2	118	114	4	0,140
1	3	50	45	5	0,556
1	4	42	51	-9	1,588
2	1	90	90	0	0
2	2	110	114	-4	0,140
2	3	40	45	-5	0,556
2	4	60	51	9	1,588
					4,568 = χ^2

Der Kontingenzkoeffizient nach Pearson errechnet sich gemäß

$$P = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{4,568}{4,568 + 600}} = 0,087.$$

Damit erhalten wir

$$P_{korr} = \frac{P}{P_{max}} = \frac{0,087}{\sqrt{\frac{1}{2}}} = 0,123.$$

Dabei ist $P_{max} = \sqrt{\frac{M-1}{M}}$ und $M = \min\{2, 4\} = 2$.

6. a) Wenn die beiden Merkmale X: *Altersklasse* und Y: *Wert des Gutscheins* voneinander unabhängig wären, dann hätte man die folgende Kontingenztabelle:

Alter von... bis unter...	Gutschein in Wert von 25 €	50 €	
12 − 15	42	28	70
15 − 18	42	28	70
18 − 20	36	24	60
	120	80	200

Arbeitstabelle zur Berechnung von P_{korr}:

j	k	f_{jk}	e_{jk}	$f_{jk}-e_{jk}$	$\frac{(f_{jk}-e_{jk})^2}{e_{jk}}$
1	1	50	42	8	1,524
1	2	20	28	-8	2,286
2	1	40	42	-2	0,095
2	2	30	28	2	0,143
3	1	30	36	-6	1,000
3	2	30	24	6	1,500
					6,548 $=\chi^2$

Der Kontingenzkoeffizient nach Pearson errechnet sich gemäß

$$P=\sqrt{\frac{\chi^2}{\chi^2+n}}=\sqrt{\frac{6,548}{6,548+200}}=0,178.$$

Damit ist $P_{korr}=\frac{P}{P_{max}}=\frac{0{,}178}{\sqrt{\frac{1}{2}}}=0,252.$

($P_{max}=\sqrt{\frac{M-1}{M}}$ und $M=\min\{3,2\}=2.$)

b) Da die Merkmale X: *Altersklasse* und Y: *Gutschein* ordinal skaliert sind, kann man den Rangkorrelationskoeffizienten nach Spearman berechnen.

7. Die Merkmale X: *Körpergröße* (in cm) und Y: *Körpergewicht* (in kg) sind kardinalskaliert; eine geeignete Maßzahl zur Messung des linearen Zusammenhangs zwischen X und Y ist der Bravais-Pearson-Korrelationskoeffizient

$$r=\frac{s_{XY}}{s_X\cdot s_Y}.$$

Für die Berechnung von r benötigen wir die Kovarianz s_{XY}, sowie die Standardabweichungen s_X und s_Y. Wir geben zunächst die Kontingenztabelle mit den Randwerten an:

	y_k			
x_j	70	75	80	
175	5	2	0	7
180	0	0	3	3
	5	2	3	10

Dann bestimmen wir

das arithmetische Mittel von X:

$$\bar{x}=\frac{x_1f_{1\bullet}+x_2f_{2\bullet}}{n}=\frac{175\cdot 7+180\cdot 3}{10}=\frac{1765}{10}=176,5$$

das arithmetische Mittel von Y:

$$\bar{y} = \frac{y_1 f_{\bullet 1} + y_2 f_{\bullet 2} + y_3 f_{\bullet 3}}{n} = \frac{70 \cdot 5 + 75 \cdot 2 + 80 \cdot 3}{10} = \frac{740}{10} = 74$$

die Varianz von X:

$$\begin{aligned} s_X^2 &= \frac{1}{n}\sum_{j=1}^{2} x_j^2 f_{j\bullet} - \bar{x}^2 = \frac{175^2 \cdot 7 + 180^2 \cdot 3}{10} - 176,5^2 \\ &= \frac{311.575}{10} - 31.152,25 = 5,25 \end{aligned}$$

die Varianz von Y:

$$\begin{aligned} s_Y^2 &= \frac{1}{n}\sum_{k=1}^{3} y_k^2 f_{\bullet k} - \bar{y}^2 \\ &= \frac{70^2 \cdot 5 + 75^2 \cdot 2 + 80^2 \cdot 3}{10} - 74^2 \\ &= \frac{54.950}{10} - 5476 = 19 \end{aligned}$$

die Kovarianz zwischen X und Y:

$$\begin{aligned} s_{XY} &= \frac{1}{n}\sum_{j=1}^{2}\sum_{k=1}^{3} x_j y_k f_{jk} - \bar{x}\bar{y} \\ &= \frac{1}{10}(175 \cdot 70 \cdot 5 + 175 \cdot 75 \cdot 2 + 175 \cdot 80 \cdot 0 + 180 \cdot 70 \cdot 0 \\ &\quad +180 \cdot 75 \cdot 0 + 180 \cdot 80 \cdot 3) - 176,5 \cdot 74 \\ &= \frac{130.700}{10} - 13.061 = 9 \end{aligned}$$

Setzen wir die Werte ein, erhalten wir

$$r = \frac{s_{XY}}{s_X \cdot s_Y} = \frac{9}{\sqrt{5,25} \cdot \sqrt{19}} = 0,9011.$$

8. Für die Berechnung des Bravais-Pearson-Korrelationskoeffizienten

$$r = \frac{s_{XY}}{s_X \cdot s_Y}$$

zwischen den Variablen X: *Verfügbares Einkommen* und Y: *Ausgaben für Lebensmittel* bestimmen wir das arithmetische Mittel von X:

$$\bar{x} = \frac{130 + 140 \cdot 3 + 150 + 160}{6} = \frac{860}{6} = 143,3333$$

das arithmetische Mittel von Y:

$$\bar{y} = \frac{14,4 + 14,8 \cdot 2 + 15 + 15,2 + 16}{6} = \frac{90,2}{6} = 15,0333$$

die Varianz von X:

$$\begin{aligned} s_X^2 &= \frac{1}{6}(130^2 + 140^2 \cdot 3 + 150^2 + 160^2) - 143,3333^2 \\ &= \frac{123.800}{6} - 20.544,4444 = 88,8889 \end{aligned}$$

die Varianz von Y:

$$\begin{aligned} s_Y^2 &= \frac{1}{6}(14,4^2 + 14,8^2 \cdot 2 + 15^2 + 15,2^2 + 16^2) - 15,0333^2 \\ &= \frac{1357,48}{6} - 226,0011 = 0,2456 \end{aligned}$$

die Kovarianz zwischen X und Y:

$$\begin{aligned} s_{XY} &= \frac{1}{16}(160 \cdot 14,4 + 130 \cdot 16 + 140 \cdot 14,8 + 140 \cdot 15 \\ &\quad +140 \cdot 15,2 + 150 \cdot 14,8) - 143,3333 \cdot 15,0333 \\ &= \frac{12.904}{6} - 2154,7778 = -4,1111 \end{aligned}$$

Der Korrelationskoeffizient errechnet sich gemäß

$$r = \frac{s_{XY}}{s_X \cdot s_Y} = \frac{-4,1111}{\sqrt{88,8889} \cdot \sqrt{0,2456}} = -0,88.$$

9. Da der Anleger besonders vorsichtig ist, wird er sich für zwei Aktien mit dem kleineren Korrelationskoeffizienten entscheiden. Für die Berechnung der Korrelationskoeffizienten seien X: *Kurs der Aktie* A, Y: *Kurs der Aktie* B und Z: *Kurs der Aktie* C.

   Wir bestimmen zunächst den Korrelationskoeffizienten zwischen X und Y. Für die Berechnung der benötigten Größen stellen wir die folgende Arbeitstabelle auf:

t	x_t	y_t	$(x_t-\bar{x})$	$(x_t-\bar{x})^2$	$(y_t-\bar{y})$	$(y_t-\bar{y})^2$	$(x_t-\bar{x})(y_t-\bar{y})$
1	53	61	-3,3333	11,1111	1,8333	3,3611	-6,1111
2	54	58	-2,3333	5,4444	-1,1667	1,3611	2,7222
3	55	57	-1,3333	1,7778	-2,1667	4,6944	2,8889
4	58	58	1,6667	2,7778	-1,1667	1,3611	-1,9444
5	60	61	3,6667	13,4444	1,8333	3,3611	6,7222
6	58	60	1,6667	2,7778	0,8333	0,6944	1,3889
	338	355	0,0000	37,3333	0,0000	14,8333	5,6667

Das arithmetische Mittel von X bzw. von Y:

$\bar{x} = \frac{\sum_{t=1}^{6} x_t}{6} = \frac{338}{6} = 56,3333$

$\bar{y} = \frac{\sum_{t=1}^{6} y_t}{6} = \frac{355}{6} = 59,1667$

Die Varianzen:

$s_X^2 = \frac{1}{6}\sum_{t=1}^{6}(x_t-\bar{x})^2 = \frac{37{,}3333}{6} = 6,2222$

$s_Y^2 = \frac{1}{6}\sum_{t=1}^{6}(y_t-\bar{y})^2 = \frac{14{,}8333}{6} = 2,4722$

Die Kovarianz:

$s_{XY} = \frac{1}{6}\sum_{t=1}^{6}(x_t-\bar{x})(y_t-\bar{y}) = \frac{5{,}6667}{6} = 0,9444$

Der Korrelationskoeffizient zwischen der Aktie A und der Aktie B bestimmt sich gemäß

$$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y} = \frac{0,9444}{\sqrt{6,2222}\cdot\sqrt{2,4722}} = 0,2408.$$

Nun bestimmen wir den Korrelationskoeffizienten zwischen X und Z. Dazu stellen wir die folgende Arbeitstabelle auf:

t	z_t	$(z_t-\bar{z})$	$(z_t-\bar{z})^2$	$(x_t-\bar{x})(z_t-\bar{z})$
1	51	1,5	2,25	-5,0000
2	45	-4,5	20,25	10,5000
3	43	-6,5	42,25	8,6667
4	50	0,5	0,25	0,8333
5	53	3,5	12,25	12,8333
6	55	5,5	30,25	9,1667
	297	0,0	107,5	37,0000

Wir berechnen:

$$\bar{z} = \frac{297}{6} = 49,5 \quad s_Z^2 = \frac{107,5}{6} = 17,9167 \quad s_{XZ} = \frac{37}{6} = 6,1667$$

Der Korrelationskoeffizient zwischen der Aktie A und der Aktie C bestimmt sich gemäß

$$r_{XZ} = \frac{s_{XZ}}{s_X \cdot s_Z} = \frac{6,1667}{\sqrt{6,2222} \cdot \sqrt{17,9167}} = 0,5840.$$

Der Anleger wird die Aktie A und die Aktie B in sein Portfolio aufnehmen.

10. Das Merkmal *Sportlicher Rang* ist ordinal, das Merkmal *Umsatz* ist kardinal. Deshalb verwenden wir den Rangkorrelationskoeffizienten nach Spearman als Maßzahl für die Abhängigkeit zwischen den beiden Merkmalen. Dazu bestimmen wir (ohne Bayer Leverkusen, siehe nachfolgende Arbeitstabelle) die Rangzahlen X: *Sportlicher Rang* und Y: *Umsatz-Rang*. Für X und Y lautet der Spearmansche Rangkorrelationnskoefizient:

$$r^{SP} = \frac{s_{XY}}{s_X \cdot s_Y}$$

Für die Berechnung der Kovarianz s_{XY} sowie die Standardabweichungen s_X und s_Y bestimmen wir (vgl. Arbeitstabelle)

die arithmetischen Mittel: $\bar{x} = \bar{y} = \frac{153}{17} = 9$

die Varianzen (mit Verschiebungssatz):

$$s_X^2 = s_Y^2 = \frac{1785}{17} - 9^2 = 24$$

die Kovarianz (mit Verschiebungssatz):

$$s_{XY} = \frac{1}{n}\sum_{i=1}^{17} x_i y_i - \bar{x}\bar{y} = \frac{1402}{17} - 9 \cdot 9 = 82,4706 - 81 = 1,4706$$

Der Rangkorrelationskoeffizient nach Spearman ergibt sich gemäß

$$r^{SP} = \frac{s_{XY}}{s_X \cdot s_Y} = \frac{1,4706}{\sqrt{24} \cdot \sqrt{24}} = 0,0613.$$

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	$d_i = x_i - y_i$	d_i^2
FC Bayern München	2	1	4	1	2	1	1
Hamburger SV	7	2	49	4	14	5	25
VFB Stuttgart	11	3	121	9	33	8	64
FC Schalke 04	13	4	169	16	52	9	81
SV Werder Bremen	12	5	144	25	60	7	49
VfL Wolfsburg	14	6	296	36	84	8	64
Borrusia Dortmund	1	7	1	49	7	-6	36
1. FC Köln	9	8	81	64	72	1	1
Eintracht Frankfurt	16	9	256	81	144	7	49
Bor. Mönchengladbach	15	10	225	100	150	5	25
Hannover 96	3	11	9	121	33	-8	64
1. FC Nürnberg	5	12	25	144	60	-7	49
1899 Hoffenheim	10	13	100	169	130	-3	9
1. FSV Mainz 05	4	14	16	196	56	-10	100
SC Freiburg	8	15	64	225	120	-7	49
1. FC Kaiserslauten	6	16	36	256	96	-10	100
FC St. Pauli	17	17	289	289	289	0	0
	153	153	1785	1785	1402		766

Da alle Rangzahlen verschieden sind, können wir r^{SP} wie folgt (s. a. die letzte und die vorletzte Spalte der obigen Tabelle) berechnen:

$$\begin{aligned} r^{SP} &= 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 766}{17(17^2-1)} \\ &= 1 - \frac{4596}{17 \cdot 288} = 1 - 0,9387 = 0,0613 \end{aligned}$$

4 Lineare Regressionsanalyse

1. Berechnen Sie für die unten angegebenen Daten die Regressionsgerade, den Korrelationskoeffizienten sowie das Bestimmtheitsmaß.

i	1	2	3	4	5
x_i	2	5	6	8	4
y_i	4	5	6	7	8

2. Greifen wir auf die Daten in Aufgabe 7, Kapitel 3 zurück:

	Körpergewicht		
Körpergröße	70	75	80
175	5	2	0
180	0	0	3

   Für diesen Datensatz wurde für die Variablen *Körpergröße* (X) und *Körpergewicht* (Y) ein Korrelationskoeffizient von 0,90 berechnet. Mit diesem Ergebnis kann man eine lineare Abhängigkeit zwischen den Variablen unterstellen. Geben Sie nun diese in Form einer Regressionsgeraden an. Nehmen Sie *Körpergröße* als die exogene Variable.

3. In der Makroökonomie wird unter anderem die Abhängigkeit des Sparens (S) von dem Einflussfaktor *Verfügbares Einkommen* (Y) durch eine lineare Funktion $S = a + bY$ beschrieben. Anhand vergangener Werte kann man die lineare Sparfunktion empirisch nach der KQ-Methode schätzen. Die nachfolgende Tabelle[1] gibt *Verfügbares Einkommen* und *Sparen privater Haushalte* von 1999 bis 2008 in Deutschland wieder.

   a) Bestimmen Sie die nach der KQ-Methode geschätzte Sparfunktion.

   b) Interpretieren Sie die Steigung der Regressionsgeraden.

[1] `http://www.boersennotizbuch.de/marktdaten/sparquote-deutschland` (Stand: 25.06.2011)

c) Wie stark ist der lineare Zusammenhang zwischen den Variablen S und Y?

d) Geben Sie das Bestimmtheitsmaß an.

Jahr	Verfügbares Einkommen (Milliarden Euro)	Sparen (Milliarden Euro)
1999	1.298	123
2000	1.337	123
2001	1.390	131
2002	1.403	139
2003	1.432	147
2004	1.459	152
2005	1.483	156
2006	1.516	158
2007	1.541	167
2008	1.554	181

4. In der Aufgabe 3 wurde die Regression von S auf Y untersucht. Die Regressionsfunktion $\hat{s} = \hat{a} + \hat{b}y$ beschreibt, wie das verfügbare Einkommen Y das Sparen S beeinflusst. Auf der anderen Seite verändert das Sparen das verfügbare Einkommen. Bestimmen Sie unter der Annahme, dass $Y = a_1 + b_1 S$ gilt, die KQ-Schätzer für die Regressionskoeffizienten von Y auf S.

5. Preis-Absatz-Funktionen beschreiben, wie sich die Preise auf die Nachfrage auswirken. Bei einer linearen Preis-Absatz-Funktion unterstellt man, dass zwischen der Variablen Preis p und der Variablen Absatz x die Beziehung $x(p) = a + bp,\ b < 0$ gilt. Die Steigung ist negativ, da eine Preiserhöhung in der Regel zu einem Absatzrückgang führt. Die lineare Preis-Absatz-Funktion kann man empirisch mit Hilfe der KQ-Methode ermitteln. Geben Sie für die folgenden Beobachtungen die geschätzte Preis-Absatz-Funktion an.

i	1	2	3	4	5	6	7	8	9	10
Preis p_i	1	2	3	4	5	6	7	8	9	10
Absatz x_i	12	10	10	7	8	6	5	5	3	2

a) Interpretieren Sie die Steigung der Regressionsgeraden $\hat{b}$.

b) Wie stark ist der lineare Zusammenhang zwischen den Variablen *Preis* und *Absatz*?

c) Berechnen und interpretieren Sie das Bestimmtheitsmaß.

6. In der 1980er Jahren wurde zwischen dem Benzin- und dem Dieselpreis ein Korrelationskoeffizient von 0,98 ermittelt. Die Preisschwankungen der beiden Kraftstoffe wurden durch ihre jeweilige Varianz angegeben. Mit X: *Benzinpreis* und Y: *Dieselpreis* betrugen die Varianzen $s_X^2 = 0,0359$ und $s_Y^2 = 0,0086$. Mit welcher Erhöhung des Dieselpreises musste man rechnen, wenn der Benzinpreis um DM 1,- erhöht wurde?

7. Aus langjährigen Erfahrungen weiß ein Gebrauchtwagenhändler, dass zwischen dem Verkaufspreis und der gefahrenen Kilometerzahl eine lineare Beziehung besteht. Für eine bestimmte Wagenkategorie möchte er die lineare Abhängigkeit der Variablen Y: *Preis* (in Tausend Euro) von der exogenen Variablen X: *Gefahrene Kilometer* (in Tausend) funktional angeben. Dazu wählte er zufällig 10 Wagen gleichen Typs und notierte für jeden Wagen jeweils den Preis und die gefahrenen Kilometer:

i	1	2	3	4	5	6	7	8	9	10
Kilometer x_i	127	92	173	6	12,7	237	59,7	49,6	34	120
Preis y_i	12,2	19,5	11	38,7	24,7	3	10,7	28,9	26,5	5

Welchen Verkaufspreis kann der Händler für einen Wagen mit 10.000 km erwarten?

8. Wenn man die Körpergrößen von Vater-Sohn-Paaren in ein kartesisches Koordinatensystem einträgt, so dass die Körpergrößen der Väter auf der x-Achse und die zugehörigen Körpergrößen der Söhne auf der y-Achse stehen, sieht man, dass die Werte in der Diagonalrichtung gehäuft auftreten. Dies lässt vermuten, dass zwischen der Variablen X: *Vatergröße* und Y: *Sohngröße* eine lineare Beziehung besteht. Ausgehend von dieser Annahme bestimmen Sie für die folgenden Beobachtungen die Regressionsgerade von Y auf X (d. h. Y in Abhängigkeit von X).

i	1	2	3	4	5	6	7	8	9	10
x_i (in cm)	162	163	175	180	180	190	182	178	170	170
y_i (in cm)	165	167	165	178	180	180	188	178	171	188

Lösungen zu den Aufgaben

Kapitel 4

1. Die nach der KQ-Methode geschätzte Regressionsgerade lautet:

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

Dabei sind die KQ-Schätzer der Koeffizienten:

$$\hat{b} = \frac{s_{XY}}{s_X^2} \quad \text{und} \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

Arbeitstabelle:

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	2	4	-3	-2	9	4	6
2	5	5	0	-1	0	1	0
3	6	6	1	0	1	0	0
4	8	7	3	1	9	1	3
5	4	8	-1	2	1	4	-2
	25	30	0	0	20	10	7

Das arithmetische Mittel von X: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{25}{5} = 5$

Das arithmetische Mitel von Y: $\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{30}{5} = 6$

Die Varianz von X: $s_X^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{20}{5} = 4$

Die Varianz von Y: $s_Y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{10}{5} = 2$

Die Kovarianz: $s_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{7}{5} = 1,4$

Die KQ-Schätzung für b : $\hat{b} = \frac{1,4}{4} = 0,35$

Die KQ-Schätzung für a : $\hat{a} = 6 - 0,35 \cdot 5 = 4,25$

Wir erhalten die Regressionsgerade:

$$\hat{y}_i = 4,25 + 0,35 \cdot x_i$$

den Bravais-Pearson-Korrelationskoeffizienten:

$$r = \frac{s_{XY}}{s_X \cdot s_Y} = \frac{1,4}{\sqrt{4} \cdot \sqrt{2}} = 0,495$$

und das Bestimmtheitsmaß: $R^2 = r^2 = 0,495^2 = 0,245$

2. Aus der Lösung zur Aufgabe 7, Kapitel 3 entnehmen wir:

$$\bar{x} = 176,5 \qquad \bar{y} = 74 \qquad s_X^2 = 5,25 \qquad s_{XY} = 9$$

Die Steigung der Regressionsgeraden ist

$$\hat{b} = \frac{s_{XY}}{s_X^2} = \frac{9}{5,25} = 1,7143$$

und der Achsenabschnitt ist

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 74 - \frac{9}{5,25} \cdot 176,5 = -228,5714.$$

Die Regressionsgerade lautet: $\hat{y} = -228,5714 + 1,7143 \cdot x$

3. a) Die nach der KQ-Methode geschätzte lineare Sparfunktion lautet

$$\hat{s} = \hat{a} + \hat{b}y.$$

Dabei sind die KQ-Schätzer für b bzw. a:

$$\hat{b} = \frac{s_{YS}}{s_Y^2} \qquad\qquad \hat{a} = \bar{s} - \hat{b}\bar{y}$$

Arbeitstabelle zur Berechnung der KQ-Schätzer $\hat{b}$ und $\hat{a}$:

	y_i	s_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$s_i - \bar{s}$	$(s_i - \bar{s})^2$	$(y_i - \bar{y})(s_i - \bar{s})$
1999	1298	123	-143,3	20534,89	-24,7	610,09	3539,51
2000	1337	123	-104,3	10878,49	-24,7	610,09	2576,21
2001	1390	131	-51,3	2631,69	-16,7	278,89	856,71
2002	1403	139	-38,3	1466,89	-8,7	75,69	333,21
2003	1432	147	-9,3	86,49	-0,7	0,49	6,51
2004	1459	152	17,7	313,29	4,3	18,49	76,11
2005	1483	156	41,7	1738,89	8,3	68,89	346,11
2006	1516	158	74,7	5580,09	10,3	106,09	769,41
2007	1541	167	99,7	9940,09	19,3	372,49	1924,21
2008	1554	181	112,7	12701,29	33,3	1108,89	3752,91
	14413	1477	0	65872,10	0	3250,10	14180,90

Das arithmetische Mittel von Y:

$$\bar{y} = \frac{\sum_{i=1}^{10} y_i}{10} = \frac{14413}{10} = 1441,30$$

Das arithmetische Mitel von S: $\bar{s} = \frac{\sum_{i=1}^{10} s_i}{10} = \frac{1477}{10} = 147,70$

Die Varianz von Y: $s_Y^2 = \frac{\sum_{i=1}^{10}(y_i - \bar{y})^2}{10} = \frac{65872{,}1}{10} = 6587,21$

Die Varianz von S: $s_S^2 = \frac{\sum_{i=1}^{10}(s_i - \bar{s})^2}{10} = \frac{3250{,}1}{10} = 325,01$

Die Kovarianz: $s_{YS} = \frac{\sum_{i=1}^{10}(y_i - \bar{y})(s_i - \bar{s})}{10} = \frac{14180{,}9}{10} = 1418,09$

Die KQ-Schätzung für b : $\hat{b} = \frac{1418{,}09}{6587{,}21} = 0,2153$

Die KQ-Schätzung für a:

$$\hat{a} = 147,7 - \frac{1418,09}{6587,21} \cdot 1441,3 = -162,5821$$

Die nach der KQ-Methode geschätzte Sparfunktion lautet:

$$\hat{s}_i = -162,5821 + 0,2153 \cdot y_i$$

b) $\hat{b} = 0,2153$ bedeutet, dass jeder zusätzliche Euro eine Erhöhung des Sparens um 21,53 Cents mit sich bringt.

c) Die Stärke des linearen Zusammenhangs wird durch den Korrelationskoeffizienten nach Bravais-Pearson wiedergegeben. Dieser berechnet sich gemäß

$$r = \frac{s_{YS}}{s_Y \cdot s_S} = \frac{1418,09}{\sqrt{6587,21} \cdot \sqrt{325,01}} = 0,9692.$$

d) Bestimmtheitsmaß: $R^2 = r^2 = 0,9692^2 = 0,94$. Das heißt ca. 94 % der Variationen in Sparverhalten werden durch das lineare Modell erfasst.

4. Die Rollen der Variablen werden hier vertauscht (Umkehrregression). Das Merkmal Y: *Einkommen* ist nun die endogene und das Merkmal S: *Sparen* die exogene Variable. Die Regressionsgerade lautet: $\hat{y}_i = \hat{a}_1 + \hat{b}_1 s_i$. Dabei sind

$$\hat{b}_1 = \frac{s_{YS}}{s_S^2} \quad \text{und} \quad \hat{a}_1 = \bar{y} - \hat{b}_1 \bar{s}.$$

Ergebnisse aus Aufgabe 3 sind:

$s_{YS} = 1418,09$ $\quad s_S^2 = 325,01$ $\quad \bar{y} = 1441,30$ $\quad \bar{s} = 147,70$

Daraus errechnen wir:

$\hat{b}_1 = \frac{1418{,}09}{325{,}01} = 4,3632$

$\hat{a}_1 = 1441,3 - \frac{1418{,}09}{325{,}01} \cdot 147,7 = 796,8525$

Somit lautet die Umkehrregressionsgerade

$$\hat{y}_i = 796,8525 + 4,3632 \cdot s_i.$$

5. Die nach der KQ-Methode geschätzte Preis-Absatz-Funktion ist

$$\hat{x} = \hat{a} + \hat{b}p.$$

Arbeitstabelle zur Schätzung der Koeffizienten a und b:

i	p_i	x_i	$p_i - \bar{p}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(p_i - \bar{p})^2$	$(p_i - \bar{p})(x_i - \bar{x})$
1	1	12	-4,5	5,2	27,04	20,25	-23,4
2	2	10	-3,5	3,2	10,24	12,25	-11,2
3	3	10	-2,5	3,2	10,24	6,25	-8,0
4	4	7	-1,5	0,2	0,04	2,25	-0,3
5	5	8	-0,5	1,2	1,44	0,25	-0,6
6	6	6	0,5	-0,8	0,64	0,25	-0,4
7	7	5	1,5	-1,8	3,24	2,25	-2,7
8	8	5	2,5	-1,8	3,24	6,25	-4,5
9	9	3	3,5	-3,8	14,44	12,25	-13,3
10	10	2	4,5	-4,8	23,04	20,25	-21,6
	55	68	0	0	93,60	82,5	-86

Die arithmetischen Mittel:

$$\bar{p} = \frac{55}{10} = 5,5 \quad \text{bzw.} \quad \bar{x} = \frac{68}{10} = 6,8$$

Die Varianz der Variablen *Preis*:

$$s_p^2 = \frac{1}{10}\sum_{i=1}^{10}(p_i - \bar{p})^2 = \frac{82,5}{10} = 8,25$$

Die Kovarianz zwischen den Variablen *Preis* und *Absatz*:

$$s_{px} = \frac{1}{10}\sum_{i=1}^{10}(p_i - \bar{p})(x_i - \bar{x}) = \frac{-86}{10} = -8,6$$

Die Regressionskoeffizienten (auf ein Hundertstel gerundet):

$$\hat{b} = \frac{s_{px}}{s_p^2} = \frac{-8,6}{8,25} = -1,04$$

und

$$\hat{a} = \bar{x} - \hat{b}\bar{p} = 6,8 - \left(\frac{-8,6}{8,25}\right) \cdot 5,5 = 12,53$$

Die geschätzte Preis-Absatz-Funktion lautet

$$\hat{x} = 12,53 - 1,04 \cdot p.$$

a) Die Steigung von $\hat{b} = -1,04$ bedeutet: Mit jeder Preissteigerung um eine Währungseinheit geht der Absatz um ca. 1,04 Einheiten zurück.

b) Der Korrelationskoeffizient zwischen den beiden Variablen Preis und Absatz beträgt

$$r = \frac{s_{px}}{s_p \cdot s_x} = \frac{-8,6}{\sqrt{8,25} \cdot \sqrt{9,36}} = -0,979.$$

$(s_x^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x})^2 = \frac{93,6}{10} = 9,36)$

c) Das Bestimmtheitsmaß beträgt

$$R^2 = r^2 = (-0,979)^2 = 0,958.$$

Das heißt: Fast 96% der Variationen in Absatz werden durch das Modell erklärt.

6. Die Erhöhung des Dieselpreises bei einer Erhöhung des Benzinpreises um DM 1,- wird durch die Steigung $\hat{b}$ der Regressionsgeraden $\hat{y} = \hat{a} + \hat{b}x$ gegeben. Dabei bezeichnen X: *Benzinpreis* und Y: *Dieselpreis*. Die Steigung berechnet sich gemäß

$$\hat{b} = r \cdot \frac{s_Y}{s_X} = 0,98 \cdot \frac{\sqrt{0,0086}}{\sqrt{0,0359}} = 0,48.$$

Das heißt: Eine Erhöhung des Benzinpreises um DM 1,- bewirkt eine Steigerung des Dieselpreises um 48 Pfennige.

7. Die Regressionsgerade von Y: *Preis* auf X: *Gefahrene Kilometer* lautet $\hat{y} = \hat{a} + \hat{b}x$. Arbeitstabelle zur Bestimmung von $\hat{a}$ bzw. $\hat{b}$:

i	x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	127	12,2	35,9	1288,81	-5,83	33,9889	-209,297
2	92	19,5	0,9	0,81	1,47	2,1609	1,323
3	173	11	81,9	6707,61	-6,93	48,0249	-567,567
4	6	38,7	-85,1	7242,01	20,67	427,2489	-1759,017
5	12,7	24,7	-78,4	6146,56	6,67	44,4889	-522,928
6	237	3	145,9	21286,81	-15,03	225,9009	-2192,877
7	59,7	10,7	-31,4	985,96	-7,33	53,7289	230,162
8	49,6	28,9	-41,5	1722,25	10,87	118,1569	-451,105
9	34	26,5	-57,1	3260,41	8,47	71,7409	-483,637
10	120	5	28,9	835,21	-13,03	169,7809	-376,567
	911	180,3	0,00	49476,44	0,00	1195,22	-6331,51

Die arithmetischen Mittel:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{911}{10} = 91,1 \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{180,3}{10} = 18,03$$

Die Varianzen:

$$s_X^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{49.476,44}{10} = 4947,644$$

und

$$s_Y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} = \frac{1195,22}{10} = 119,522$$

Die Kovarianz

$$s_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{-6331,51}{10} = -633,151$$

Die Steigung der Regressionsgeraden ist

$$\hat{b} = \frac{s_{XY}}{s_X^2} = \frac{-633,151}{4947,644} = -0,128$$

und der Achsenabschnitt beträgt

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 18,03 - \left(\frac{-633,151}{4947,644}\right) \cdot 91,1 = 29,688.$$

Somit lautet die Regressionsgerade:

$$\hat{y} = \hat{a} + \hat{b}x = 29,688 - 0,128 \cdot x$$

Für einen Wagen mit 10.000 km kann man einen Verkaufspreis von etwa 28.400 € ($\approx 29,688 - 0,128 \cdot 10$) erwarten.

8. Die Regressionsgerade von Y (Sohngröße) auf X (Vatergröße) (in cm) lautet

$$\hat{y} = \hat{a} + \hat{b}x.$$

Dabei sind

$$\hat{b} = \frac{s_{XY}}{s_X^2} \quad \text{und} \quad \hat{a} = \bar{y} - \hat{b}\bar{x}.$$

Für die Berechnung der Regressionskoeffizienten bestimmen wir

$\bar{x} = \frac{1750}{10} = 175, \quad \bar{y} = \frac{1760}{10} = 176, \quad s_X^2 = \frac{696}{10} = 69,6$ und

$s_{XY} = \frac{396}{10} = 39,6$ (vgl. Arbeitstabelle).

Wir erhalten die Steigung der Regressionsgeraden

$$\hat{b} = \frac{39,6}{69,6} = 0,57$$

und das absolute Glied

$$\hat{a} = 176 - \frac{39,6}{69,6} \cdot 175 = 76,43$$

Somit lässt sich die lineare Beziehung zwischen Vater-Sohn-Größen durch die folgende Funktion beschreiben:

$$\hat{y} = 76,43 + 0,57 \cdot x$$

Arbeitstabelle:

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	162	165	-13	-11	169	121	143
2	163	167	-12	-9	144	81	108
3	175	165	0	-11	0	121	0
4	180	178	5	2	25	4	10
5	180	180	5	4	25	16	20
6	190	180	15	4	225	16	60
7	182	188	7	12	49	144	84
8	178	178	3	2	9	4	6
9	170	171	-5	-5	25	25	25
10	170	188	-5	12	25	144	-60
	1750	1760	0	0	696	676	396

5 Verhältniszahlen

1. Auf dem Oktoberfest in München wurden die folgenden durchschnittlichen Bierpreise in Euro pro Maß in den Jahren 2000 bis 2009 regristriert:

Jahr	’00	’01	’02	’03	’04	’05	’06	’07	’08	’09
Preis	6,14	6,33	6,59	6,59	6,90	7,04	7,32	7,73	8,13	8,44

   a) Stellen Sie die Bierpreise als Preismesszahlen dar. Nehmen Sie das Jahr 2000 als Basisjahr. Interpretieren Sie die Preismesszahl für das Jahr 2007.

   b) Berechnen Sie für die Jahre 2001 bis 2009 jeweils die Preiserhöhung zum Vorjahr (in Prozent).

   c) Geben Sie die jährliche durchschnittliche Preiserhöhung in den Jahren 2000 bis 2009 an.

2. Berechnen Sie für die folgenden Mengen- und Preisvektoren

$$q_0 = \begin{pmatrix} 2 & 4 & 6 \end{pmatrix} \quad q_1 = \begin{pmatrix} 3 & 5 & 7 \end{pmatrix} \quad p_0 = \begin{pmatrix} 7 & 5 & 3 \end{pmatrix} \quad p_1 = \begin{pmatrix} 6 & 4 & 2 \end{pmatrix}$$

   die Ausgaben zum Zeitpunkt $t = 0$ bzw. $t = 1$, die Preisindizes sowie die Mengenindizes nach Laspeyres und nach Paasche. Wie hoch ist die nominale Veränderung der Ausgaben? Wie hoch die reale?

3. Eine Gruppe von Studenten in München möchte einen Preisindex für ihre Lebenshaltungskosten berechnen. Nach einer Umfrage unter den Kommilitonen haben sie herausgefunden, dass folgende Güter als repräsentativ angesehen werden können: Kommunikation (Internet, Handy usw.), Mensa und Kaffee. Für die Monate Mai 2009 und Mai 2010 haben sie die folgenden Werte ermittelt:

	Mai 2009		Mai 2010	
Gut	Menge	Preis	Menge	Preis
Kommunikation	100	0,50	110	0,25
Mensa-Essen	20	1,50	20	2,50
Kaffee	1	5,00	2	6,00

Für Kommunikation werden die Preise in €/Std., fürs Essen in der Mensa in €/Essen und für Kaffee in €/kg gemessen. (Die Mengen werden entsprechend in Stunden, Essen und Kilogramm angegeben.) Berechnen Sie den Preisindex nach Laspeyres bzw. den nach Paasche für Mai 2010 zur Basis Mai 2009. Interpretieren Sie die Ergebnisse.

4. Die Ergebnisse der Aufgabe 3 zeigen, dass die Preise sowohl nach Laspeyres als auch nach Paasche im Mai 2010 gegenüber Mai 2009 zurückgingen. Dennoch haben die Studenten im Mai 2010 mehr Ausgaben (89,50 €) als im gleichen Vorjahresmonat (85,00 €). Dies kann nur durch Mehrkonsum erklärt werden. Geben Sie die Veränderung der konsumierten Mengen in Prozent an.

5. Ein junges Unternehmen produziert vier verschiedene Güter. Die Umsatzzahlen jedes Produktes (in Mio €) und die Stückpreise (in 100 €) in den ersten drei Jahren sind in der folgenden Tabelle gegeben:

	Umsatz im Jahr			Preise im Jahr		
Produkt	0	1	2	0	1	2
1	2,5	2	3	2,00	2,20	4,00
2	0,5	1	1	1,00	1,10	2,00
3	3,0	4	5	1,25	1,30	2,50
4	2,0	3	3	1,50	1,50	3,00

Berechnen und interpretieren Sie jeweils für $t = 1$ und $t = 2$ zur Basis 0:

a) den Umsatzindex

b) den Laspeyres-Preisindex

c) den Mengenindex nach Paasche

6. Der vom Statistischen Bundesamt veröffentlichte Verbraucherpreisindex VPI beschreibt die Preisentwicklung des für Deutschland als typisch geltenden Warenkorbes. Um Veränderungen in Verbraucherverhalten zu berücksichtigen, wird der Warenkorb in der Regel alle fünf Jahre aktualisiert. Der Zeitpunkt, an dem der Warenkorb angepasst wird, dient gleichzeitig als neues Basisjahr. So löst das Jahr 2005 das bis dahin geltende Basisjahr 2000 ab. Durch diesen Vorgang entstehen die folgenden

Verbraucherpreisindex-Reihen[1] VPI_{0t} zur Basis 2000 und VPI_{5t} zur Basis 2005:

Jahr	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09	'10
VPI_{0t}	100	102	103,4	104,5	106,2	108,3					
VPI_{5t}						100	101,6	103,9	106,6	107	108

Geben Sie eine durchgehende VPI-Reihe vom Jahr 2000 bis 2010 an, indem Sie die Reihe zur Basis 2000 fortführen bzw. die Reihe zur Basis 2005 rückrechnen. Interpretieren Sie den Indexwert im Jahr 2010 zur Basis 2000.

7. Die folgende Tabelle zeigt für jedes Jahr von 2004 bis 2008 das nominale verfügbare Einkommen (in Mio. €) der privaten Haushalte und den Verbraucherpreisindex zur Basis 2005.

Jahr	t	Y_t	$VPI_{5,t}$
2004	4	1.435.650	98,6
2005	5	1.463.670	100,0
2006	6	1.493.320	101,8
2007	7	1.517.090	103,9
2008	8	1.558.110	107,3

Will man nun feststellen, ob mit den Einkommenszuwächsen die Kaufkraft zunimmt, so muss man das reale verfügbare Einkommen untersuchen.

a) Ermitteln Sie die realen Einkommen zu Preisen von 2005 für die Jahre 2004 bis 2008.

b) Geben Sie jeweils die nominale und die reale Einkommensveränderung in den Jahren 2005 bis 2008 gegenüber dem Vorjahr an. Vergleichen Sie Ihre Ergebnisse.

c) Geben Sie die jährliche durchschnittliche Veränderungsrate des nominalen bzw. des realen Einkommens im Zeitraum von 2004 bis 2008 an.

d) Mit welcher Rate hat sich das Einkommen nominal bzw. real im Jahr 2008 gegenüber dem Jahr 2005 verändert? Mit welcher Rate haben sich die Preise im selben Zeitraum verändert?

[1] Quelle: Statistisches Bundesamt, Fachserie 17, Reihe 7 (Verbraucherpreisindizes für Deutschland) in `http://www.gug-aktuell.de/Aktuelle_Daten/Verbraucherpreisindex_Basis_20/verbraucherpreisindex_basis_20.html` (Basis 2000) (Stand: 24.11.2011) bzw. `http://www.gug-aktuell.de/Aktuelle_Daten/Verbraucherpreisindex_Basis_20/verbraucherpreisindex_basis_201.html` (Basis 2005) (Stand: 24.11.2011)

Lösungen zu den Aufgaben

Kapitel 5

1. a) Die Preismesszahlen (2000=100) bestimmen sich gemäß

$$\frac{y_t}{y_0} \cdot 100 = \frac{y_t}{6,14} \cdot 100.$$

Dabei bedeuten y_t der Bierpreis im Jahr t, $t = 0, 1, \ldots, 9$.

Jahr	Preismesszahl	Jahr	Preismesszahl
'00	100	'05	114,66
'01	103,09	'06	119,22
'02	107,33	'07	125,90
'03	107,33	'08	132,41
'04	112,38	'09	137,46

Im Jahr 2007 beträgt die Preismesszahl 125,90. Das bedeutet: Im Vergleich zum Basisjahr 2000 ist der Bierpreis auf dem Oktoberfest um 25,9% gestiegen.

b) Die Preiserhöhung zum Vorjahr (in Prozent) berechnet sich gemäß

$$\left(\frac{y_t}{y_{t-1}} - 1\right) \cdot 100, \; t = 1, \ldots, 9.$$

Jahr	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09
Preiserhöhung zum Vorjahr	-	3,09	4,11	0,00	4,71	2,03	3,98	5,60	5,17	3,81

c) Der Bierpreis auf dem Oktoberfest wächst in den Jahren 2000 bis 2009 im Mittel um 3,6% pro Jahr. Das bestimmt sich gemäß

$$\left(\sqrt[9]{\frac{y_9}{y_0}}-1\right)\cdot 100 = \left(\sqrt[9]{\frac{8,44}{6,14}}-1\right)\cdot 100 = (1,036-1)\cdot 100 = 3,6.$$

2. Die Ausgaben zum Zeitpunkt $t = 0$ ergeben sich gemäß:

$$p_0 q_0 = \sum_{i=1}^{3} p_{i0} q_{i0} = 7 \cdot 2 + 5 \cdot 4 + 3 \cdot 6 = 52$$

Die Ausgaben zum Zeitpunkt $t = 1$ berechnen sich gemäß:

$$p_1 q_1 = \sum_{i=1}^{3} p_{i1} q_{i1} = 6 \cdot 3 + 4 \cdot 5 + 2 \cdot 7 = 52$$

Der Preisindex nach Laspeyres ergibt sich wie folgt:

$$P_{01}^{L} = \frac{p_1 q_0}{p_0 q_0} \cdot 100 = \frac{6 \cdot 2 + 4 \cdot 4 + 2 \cdot 6}{52} \cdot 100 = 76,92$$

Der Preisindex nach Paasche ergibt sich wie folgt:

$$P_{01}^{P} = \frac{p_1 q_1}{p_0 q_1} \cdot 100 = \frac{52}{7 \cdot 3 + 5 \cdot 5 + 3 \cdot 7} \cdot 100 = 77,61$$

Nach Laspeyres (gemessen an dem Warenkorb der Basisperiode) sind die Preise um ca. 23% gesunken; nach Paasche (bezüglich des aktuellen Warenkorbes) sind die Preise zum Zeitpunkt $t = 1$ um ca. 22% niedriger als zum Zeitpunkt $t = 0$.

Der Mengenindex nach Laspeyres ergibt sich wie folgt:

$$Q_{01}^{L} = \frac{p_0 q_1}{p_0 q_0} \cdot 100 = \frac{67}{52} \cdot 100 = 128,85$$

Der Mengenindex nach Paasche ergibt sich wie folgt:

$$Q_{01}^{P} = \frac{p_1 q_1}{p_1 q_0} \cdot 100 = \frac{52}{40} \cdot 100 = 130$$

Nach Laspeyres (bewertet mit den Preisen der Basisperiode) wurden fast 29% mehr Waren produziert (konsumiert); nach Paasche (bewertet mit den aktuellen Preisen) notiert man eine Mengenzunahme um 30%.

Um die nominale Veränderung zu messen, berechnet man den Wertindex

$$I_{01}^{W} = \frac{p_1 q_1}{p_0 q_0} \cdot 100 = \frac{52}{52} \cdot 100 = 100.$$

Das heißt: Nominal haben sich die Ausgaben nicht geändert.

Nimmt man den Laspeyres-Preisindex als Deflator, ergibt sich eine reale Veränderung um 30%. Deflationiert man mit dem Paasche-Preisindex, so ergibt sich eine reale Veränderung um 28,85%.

3. Arbeitstabelle für die Berechnung der Preisindizes (2009=0 und 2010=1):

Gut i	q_{i0}	p_{i0}	q_{i1}	p_{i1}	$p_{i0}q_{i0}$	$p_{i1}q_{i1}$	$p_{i1}q_{i0}$	$p_{i0}q_{i1}$
Kommunikation	100	0,50	110	0,25	50,00	27,50	25,00	55,00
Mensa-Essen	20	1,50	20	2,50	30,00	50,00	50,00	30,00
Kaffee	1	5,00	2	6,00	5,00	12,00	6,00	10,00
					85,00	89,50	81,00	95,00

Preisindex nach Laspeyres für Mai 2010 zur Basis Mai 2009 beträgt

$$P_{01}^{L} = \frac{p_1 q_0}{p_0 q_0} \cdot 100 = \frac{81}{85} \cdot 100 = 95,29.$$

Die Preise sind von Mai 2009 bis Mai 2010 um ca. 4,7% gesunken.

Preisindex nach Paasche für Mai 2010 zur Basis Mai 2009 beträgt

$$P_{01}^{P} = \frac{p_1 q_1}{p_0 q_1} \cdot 100 = \frac{89,50}{95,00} \cdot 100 = 94,21.$$

Nach Paasche fielen die Preise im Zeintraum von Mai 2009 bis Mai 2010 um ca. 5,8%.

4. Mengenveränderungen werden durch Mengenindizes gemessen. Nach Laspeyres berechnet sich der Mengenindex gemäß

$$Q_{01}^{L} = \frac{p_0 q_1}{p_0 q_0} \cdot 100 = \frac{95}{85} \cdot 100 = 111,76$$

und nach Paasche gemäß

$$Q_{01}^{P} = \frac{p_1 q_1}{p_1 q_0} \cdot 100 = \frac{89,5}{81} \cdot 100 = 110,49.$$

Nach Laspeyres wurden im Mai 2010 fast 12% mehr Güter als in demselben Vorjahresmonat konsumiert; nach Paasche beläuft sich die Zunahme auf ca. 10,5%.

5. a) Für jedes $t = 0, 1, 2$ ergibt sich der Gesamtumsatz wie folgt:

$$u_t = \sum_{i=1}^{4} p_{it} q_{it}$$

Der Gesamtumsatz (in Mio. €) im Jahr $t = 0, 1, 2$:

$u_0 = \sum_{i=1}^{4} p_{i0} q_{i0} = 2,5 + 0,5 + 3 + 2 = 8$

$u_1 = \sum_{i=1}^{4} p_{i1}q_{i1} = 2+1+4+3 = 10$

$u_2 = \sum_{i=1}^{4} p_{i2}q_{i2} = 3+1+5+3 = 12$

Für $t = 1$ ist der Umsatzindex

$$I_{01}^W = \frac{u_1}{u_0} \cdot 100 = \frac{10}{8} \cdot 100 = 125.$$

Das heißt: Der Umsatz ist im Jahr 1 gegenüber dem Jahr 0 wertmäßig (nominal) um 25% gestiegen.

Für $t = 2$ beträgt der Umsatzindex

$$I_{02}^W = \frac{u_2}{u_0} \cdot 100 = \frac{12}{8} \cdot 100 = 150.$$

Dies entspricht einer nominalen Umsatzsteigerung von 50% im Jahr 2 gegenüber dem Jahr 0.

b) Den Laspeyres-Preisindex zur Basis 0 berechnen wir als gewogenes Mittel der Preismesszahlen $\frac{p_{it}}{p_{i0}}$:

$$P_{0t}^L = \sum_{i=1}^{4} \frac{p_{it}}{p_{i0}} \cdot g_i, \quad g_i = \frac{p_{i0}q_{i0}}{u_0}$$

Arbeitstabelle zur Berechnung von P_{0t}^L:

i	$p_{io}q_{io}$	$\frac{p_{i1}}{p_{i0}}$	$\frac{p_{i2}}{p_{i0}}$	g_i	$\frac{p_{i1}}{p_{i0}}g_i$
1	2,5	1,10	2	0,3125	0,34375
2	0,5	1,10	2	0,0625	0,06875
3	3,0	1,04	2	0,3750	0,39000
4	2,0	1,00	2	0,2500	0,25000
	8,0			1,0000	1,0525

Der Laspeyres-Preisindex (in Prozent) für $t = 1$:

$$P_{01}^L = 1,0525 \cdot 100 = 105,25$$

Das heißt: Die Preise sind im Jahr 1 gegenüber dem Jahr 0 um 5,25% gestiegen.

Da die Preismesszahlen $\frac{p_{i2}}{p_{i0}} = 2$ für $i = 1, \ldots, 4$, gilt

$$P_{02}^L = \sum_{i=1}^{4} \frac{p_{i2}}{p_{i0}} \cdot g_i \cdot 100 = \sum_{i=1}^{4} 2 \cdot g_i \cdot 100 = 200 \sum_{i=1}^{4} g_i = 200.$$

Das bedeutet: Die Preise haben sich im Jahr 2 gegenüber dem Jahr 0 verdoppelt.

c) Für $t = 1$ bzw. $t = 2$ sind:

$Q_{01}^{P} = \frac{I_{01}^{W}}{P_{01}^{L}} \cdot 100 = \frac{125}{105{,}25} \cdot 100 = 118,76$

$Q_{02}^{P} = \frac{I_{02}^{W}}{P_{02}^{L}} \cdot 100 = \frac{150}{200} \cdot 100 = 75$

Im ersten Jahr ($t = 1$) stieg der Umsatz real um 18,76%. Im 2. Jahr ($t = 2$) schrumpfte der Umsatz real um 25%.

6. Für die Verkettung der beiden Reihen muss für $t = 0, 1, \ldots, 10$ gelten:

$$\frac{\text{VPI}_{0t}}{108,3} = \frac{\text{VPI}_{5t}}{100}$$

Die Fortführung von VPI_{0t} für $t = 6, \ldots, 10$ erfolgt gemäß:

$$\text{VPI}_{0t} = \frac{108,3}{100} \cdot \text{VPI}_{5t}$$

Die Rückrechnung von VPI_{5t} für $t = 0, \ldots, 4$ erfolgt gemäß:

$$\text{VPI}_{5t} = \frac{100}{108,3} \cdot \text{VPI}_{0t}$$

Die Ergebnisse stehen in der nachfolgenden Tabelle:

	Verbraucherpreisindex	
Jahr	(2000 = 100)	(2005 = 100)
2000	100	$\frac{100}{108{,}3} \cdot 100 = 92,34$
2001	102,0	$\frac{100}{108{,}3} \cdot 102,0 = 94,18$
2002	103,4	$\frac{100}{108{,}3} \cdot 103,4 = 95,48$
2003	104,5	$\frac{100}{108{,}3} \cdot 104,5 = 96,49$
2004	106,2	$\frac{100}{108{,}3} \cdot 106,2 = 98,06$
2005	108,3	100
2006	$\frac{108{,}3}{100} \cdot 101,6 = 110,03$	101,6
2007	$\frac{108{,}3}{100} \cdot 103,9 = 112,52$	103,9
2008	$\frac{108{,}3}{100} \cdot 106,6 = 115,45$	106,6
2009	$\frac{108{,}3}{100} \cdot 107,0 = 115,88$	107,0
2010	$\frac{108{,}3}{100} \cdot 108,0 = 116,96$	108,0

Der Verbraucherpreisindex im Jahr 2010 zur Basis 2000 von 116,96 bedeutet, dass die Verbraucherpreise sich im Jahr 2010 im Vergleich zum Jahr 2000 um 16,96% erhöht haben.

7. a) Die realen Einkommen für $t = 4, 5, 6, 7, 8$ berechnet zu Preisen von 2005:

$$Y_4^{real} = \frac{Y_4}{\mathrm{VPI}_{5,4}} \cdot 100 = \frac{1.435.650}{98{,}6} \cdot 100 = 1.456.034,48$$

$$Y_5^{real} = 1.463.670$$

$$Y_6^{real} = \frac{Y_6}{\mathrm{VPI}_{5,6}} \cdot 100 = \frac{1.493.320}{101{,}8} \cdot 100 = 1.466.915,52$$

$$Y_7^{real} = \frac{Y_7}{\mathrm{VPI}_{5,7}} \cdot 100 = \frac{1.517.090}{103{,}9} \cdot 100 = 1.460.144,37$$

$$Y_8^{real} = \frac{Y_8}{\mathrm{VPI}_{5,8}} \cdot 100 = \frac{1.558.110}{107{,}3} \cdot 100 = 1.452.106,24$$

b) Die nominalen Einkommensveränderungen (in %) für $t = 5, 6, 7, 8$ gegenüber dem Vorjahr sind:

$$\left(\frac{Y_5}{Y_4} - 1\right) \cdot 100 = \left(\frac{1.463.670}{1.435.650} - 1\right) \cdot 100 = 1,95$$

$$\left(\frac{Y_6}{Y_5} - 1\right) \cdot 100 = \left(\frac{1.493.320}{1.463.670} - 1\right) \cdot 100 = 2,03$$

$$\left(\frac{Y_7}{Y_6} - 1\right) \cdot 100 = \left(\frac{1.517.090}{1.493.320} - 1\right) \cdot 100 = 1,59$$

$$\left(\frac{Y_8}{Y_7} - 1\right) \cdot 100 = \left(\frac{1.558.110}{1.517.090} - 1\right) \cdot 100 = 2,70$$

Die realen Einkommensveränderungen (in %) für $t = 5, 6, 7, 8$ gegenüber dem Vorjahr sind:

$$\left(\frac{Y_5^{real}}{Y_4^{real}} - 1\right) \cdot 100 = \left(\frac{1.463.670}{1.456.034{,}48} - 1\right) \cdot 100 = 0,52$$

$$\left(\frac{Y_6^{real}}{Y_5^{real}} - 1\right) \cdot 100 = \left(\frac{1.466.915{,}52}{1.463.670} - 1\right) \cdot 100 = 0,22$$

$$\left(\frac{Y_7^{real}}{Y_6^{real}} - 1\right) \cdot 100 = \left(\frac{1.460.144{,}37}{1.466.915{,}52} - 1\right) \cdot 100 = -0,46$$

$$\left(\frac{Y_8^{real}}{Y_7^{real}} - 1\right) \cdot 100 = \left(\frac{1.452.106{,}24}{1.460.144{,}37} - 1\right) \cdot 100 = -0,55$$

Gegenüberstellung der nominalen und der realen Einkommensveränderung (in %) gegenüber dem Vorjahr:

Jahr	nominal	real
2005	1,95	0,52
2006	2,03	0,22
2007	1,59	-0,46
2008	2,70	-0,55

c) Nominal wächst das Einkommen im Zeitraum von 2004 bis 2008 durchschnittlich um 2,07% pro Jahr. Die durchschnittliche Zuwachsrate berechnet sich folgendermaßen:

$$\left(\sqrt[4]{\frac{1.558.110}{1.435.650}} - 1\right) \cdot 100 = (1,0207 - 1) \cdot 100 = 2,07$$

Das reale Einkommen schrumpft im Zeitraum von 2004 bis 2008 durchschnittlich um 0,07% pro Jahr. Das ergibt sich gemäß

$$\left(\sqrt[4]{\frac{1.452.106,24}{1.456.034,48}} - 1\right) \cdot 100 = (0,9993 - 1) \cdot 100 = -0,07.$$

d) Im Jahr 2008 stieg das Einkommen gegenüber 2005 nominal um ca. 6,45%. Das ergibt sich gemäß:

$$\left(\frac{Y_8}{Y_5} - 1\right) \cdot 100 = \left(\frac{1.558.110}{1.463.670} - 1\right) \cdot 100 = 6,45$$

Real schrumpfte es jedoch um 0,79%. Das berechnet sich gemäß:

$$\left(\frac{Y_8^{real}}{Y_5^{real}} - 1\right) \cdot 100 = \left(\frac{1.452.106,24}{1.463.670} - 1\right) \cdot 100 = -0,79$$

(In demselben Zeitraum erhöhten sich die Preise um 7,3%.)

6 Einführung

1. Entscheiden Sie, ob es sich im Folgenden um Zufallsexperimente handelt
   a) Eine Bäckerei produziert 500 Brezen täglich.
   b) Die Anzahl der verkauften Brezen pro Tag
   c) Die Rentabilität einer Investition
   d) Die Größe einer neuerworbenen Waldparzelle beträgt 1500 m^2.
2. *Gottfried Wilhelm Leibniz* (1646-1716) fragte sich: „Bei zwei Würfen kommt die Augensumme neun durch zwei Zahlenpaare: 3 und 6 bzw. 4 und 5 zustande, ebenso die Augensumme zehn, nämlich durch die Zahlenpaare 4 und 6 sowie 5 und 5. Warum soll dabei die Augensumme neun wahrscheinlicher sein als die Augensumme zehn? “ Versuchen Sie Leibniz eine Antwort zu geben.
3. Zeigen Sie, dass die Gleichung $A = (A \backslash B) \cup (A \cap B)$ gilt.
4. In einem Gefäß befinden sich 10 rote (r), 5 grüne (g) und 15 blaue (b) Kugeln. Eine Kugel wird blind gezogen. Geben Sie an:
   a) die Ergebnismenge Ω
   b) die Ereignismenge $\mathfrak{P}(\Omega)$

   Drücken Sie die Ereignisse in $\mathfrak{P}(\Omega)$ in Ihren eigenen Worten aus.
5. Geben Sie die Ergebnismenge des Zufallsexperiments *Ein Würfel wird zweimal geworfen* an und schreiben Sie die folgenden Ereignisse als Mengen:

   W_1: *Der 1. Wurf ist gerade*

   W_2: *Der 2. Wurf ist kleiner oder gleich zwei*

   W_3: *Der 1. Wurf ist eine Primzahl*

   Drücken Sie jedes der Ereignisse $\overline{W}_1$ und $W_1 \cap W_3$ in Ihren eigenen Worten aus.

Lösungen zu den Aufgaben

Kapitel 6

1. Zufallsexperimente sind *Die Anzahl der verkauften Brezen pro Tag* und *Die Rentabilität einer Investition*. Bei den anderen beiden Aussagen handelt es sich nicht um Zufallsexperimente, da die Anzahl der Brezen bzw. die Größe der Waldparzelle feste Größen sind.

2. Seien

   N: *Die Augensumme Neun* und Z: *Die Augensumme Zehn*

   Schreibt man die Ereignisse als Mengen

   $$N = \{(3,6),(6,3),(4,5),(5,4)\} \quad \text{und} \quad Z = \{(4,6),(6,4),(5,5)\},$$

   dann sieht man, dass $|N| = 4 > 3 = |Z|$.

3. Da $(A \backslash B) = A \cap \bar{B}$, können wir die zu zeigende Beziehung wie folgt schreiben:

   $$A = (A \cap \bar{B}) \cup (A \cap B)$$

   Weiter gilt

   $$(A \cap \bar{B}) \cup (A \cap B) = [(A \cap \bar{B}) \cup A] \cap [(A \cap \bar{B}) \cup B]$$

   Wegen $(A \cap \bar{B}) \subset A$ ist $(A \cap \bar{B}) \cup A = A$ und

   $$(A \cap \bar{B}) \cup B = (A \cup B) \cap \underbrace{(B \cup \bar{B})}_{=\Omega} = A \cup B.$$

   Insgesamt gilt die Gleichung

   $$(A \backslash B) \cup (A \cap B) = A \cap (A \cup B) = A,$$

   weil $A \subset (A \cup B)$.

4. a) Die Ergebnismenge lautet $\Omega = \{\text{r, g, b}\}$

   b) Die Ereignismenge ist die Potenzmenge von Ω:

   $$\mathfrak{P}(\Omega) = \{\emptyset, \{\text{r}\}, \{\text{g}\}, \{\text{b}\},\{\text{r,g}\},\{\text{r,b}\},\{\text{b,g}\},\{\text{r,g,b}\}\}.$$

Ereignis	In Worten
$\emptyset$	Weder rote noch grüne noch blaue Kugel
$R = \{\mathrm{r}\} = \bar{G} \cap \bar{B}$	Eine rote Kugel wird gezogen. $\Leftrightarrow$ Keine grüne *und* keine blaue Kugel wird gezogen.
$G = \{\mathrm{g}\} = \bar{R} \cap \bar{B}$	Eine grüne Kugel wird gezogen. $\Leftrightarrow$ Keine rote *und* keine blaue Kugel wird gezogen.
$B = \{\mathrm{b}\} = \bar{R} \cap \bar{G}$	Eine blaue Kugel wird gezogen. $\Leftrightarrow$ Keine rote *und* keine grüne Kugel wird gezogen.
$R \cup G = \{\mathrm{r,g}\} = \bar{B}$	Eine rote *oder* eine grüne Kugel wird gezogen. $\Leftrightarrow$ Keine blaue Kugel wird gezogen.
$R \cup B = \{\mathrm{r,b}\} = \bar{G}$	Eine rote *oder* eine blaue Kugel wird gezogen. $\Leftrightarrow$ Keine grüne Kugel gezogen.
$B \cup G = \{\mathrm{b,g}\} = \bar{R}$	Eine blaue *oder* eine grüne Kugel wird gezogen. $\Leftrightarrow$ Keine rote Kugel gezogen.
$R \cup G \cup B = \Omega$	Eine rote *oder* eine grüne *oder* eine blaue Kugel wird gezogen.

5. Die Ergebnismenge ist $\Omega = \{(i,j) | 1 \leq i,j \leq 6\}$ bzw.

		2. Wurf j					
		1	2	3	4	5	6
1. Wurf i	1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Ereignis	In Mengenschreibweise
W_1: *Der 1. Wurf ist gerade*	$\{(2,j),(4,j),(6,j) \| 1 \leq j \leq 6\}$
W_2: *Der 2. Wurf ist kleiner oder gleich zwei*	$\{(i,1),(i,2) \| 1 \leq i \leq 6\}$
W_3: *Der 1. Wurf ist eine Primzahl*	$\{(2,j),(3,j),(5,j) \| 1 \leq j \leq 6\}$
$\overline{W}_1$: *Der 1. Wurf ist ungerade*	$\{(1,j),(3,j),(5,j) \| 1 \leq j \leq 6\}$
$W_1 \cap W_3$: *Der 1. Wurf ist die Zwei*	$\{(2,j) \| 1 \leq j \leq 6\}$

7 Der Begriff der Wahrscheinlichkeit

1. Für die Ereignisse A und B seien:

   $$P(A \cup B) = 0,8 \quad P(\bar{B}) = 0,3 \quad P(A \cap B) = 0,4$$

   Berechnen Sie die Wahrscheinlichkeit, dass

   a) das Ereignis A eintritt.

   b) genau eines der Ereignisse eintritt.

   c) keines der Ereignisse eintritt.

2. Ein Laplace-Würfel wird zweimal geworfen. Berechnen Sie die Wahrscheinlichkeit für das Ereignis:

   W_1: *Der 1. Wurf ist eine gerade Augenzahl*

   W_2: *Der 2. Wurf ist kleiner oder gleich zwei*

3. Während der Sommerzeit an der Isar ist bekannt, dass man mit einer Wahrscheinlichkeit von 65% einen Sonnenbrand bekommt, mit einer Wahrscheinlichkeit von 30% von Insekten gestochen wird, und mit einer Wahrscheinlichkeit von 15% einen Sonnenbrand und Insektenstich haben kann. Wie groß ist die Wahrscheinlichkeit, dass man nur Sonnenbrand bekommt? Wie groß ist die Wahrscheinlichkeit, dass man von mindestens einer dieser Plagen heimgesucht wird?

4. Im Rahmen einer Sonderaktion plant ein Einzelhändler drei Produkte A, B und C anzubieten. Er weiß, dass 80% der Kunden mindestens eines der drei Produkte, 40% nur das Produkt B und ebenfalls 40% nur das Produkt C kaufen werden. Nach einer näheren Analyse der vergangenen Verkaufszahlen erfährt er zudem, dass 5% seiner Kunden alle drei Produkte, 25% die Produkte A und B, 6% die Produkte A und C, 4% die Produkte B und C gekauft haben. Wie viele Einheiten von Produkt A soll er täglich in seinem Laden bereit stellen, wenn er 100 Kunden pro Tag erwartet?

5. Bei einer Abschlussprüfung sind 15% der Prüflinge im Fach Englisch, 10% im Fach Deutsch und 8% in beiden Fächern durchgefallen. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Prüfling
   a) nur im Fach Englisch (Deutsch) durchgefallen ist?
   b) in genau einem Fach durchgefallen ist?
   c) in mindestens einem Fach durchgefallen ist?
   d) in beiden Fächern bestanden hat?
6. Wie oft muss man einen fairen Würfel mindestens werfen, um mit einer Wahrscheinlichkeit von mehr als 50% mindestens einmal eine 6 zu erhalten?
7. Peter glaubt, dass er mit einer Wahrscheinlichkeit von 25% die Statistik-Klausur mit Note 2 oder besser bestehen kann. Er überlegt sich, ob er durch eine Teilnahme an den zusätzlich angebotenen Übungen seine Chancen verbessern kann. Nach seiner Recherche fand er heraus, dass 96% der Kommilitonen, die die Note 2 oder besser erreicht haben, an den Übungen teilgenommen haben. Er stellte aber auch fest, dass 10% der Studierenden mit Noten schlechter als 2 Übungsteilnehmer waren. Peter wird nur die Zusatzübungen besuchen, wenn er dadurch seine Chancen mindestens verdoppeln kann. Wie wird er sich nach seinen Ermittlungen entscheiden?

Lösungen zu den Aufgaben

Kapitel 7

1. a) Aus $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ erhält man

$$\begin{aligned} P(A) &= P(A \cup B) - P(B) + P(A \cap B) \\ &= 0,8 - (1 - 0,3) + 0,4 = 0,5 \end{aligned}$$

b) Das Ereignis *Genau eines der Ereignisse tritt ein*, als Menge geschrieben, lautet $A \backslash B \cup B \backslash A$. Es gilt die Zerlegung

$$A \cup B = (A \backslash B \cup B \backslash A) \cup (A \cap B).$$

Nach dem Additionssatz für disjunkte Ereignisse erhalten wir die Eintrittswahrscheinlichkeit:

$$P(A \backslash B \cup B \backslash A) = P(A \cup B) - P(A \cap B) = 0,8 - 0,4 = 0,4$$

c) Das Ereignis *Keines der Ereignisse tritt ein* in Mengenschreibweise lautet $\overline{A \cup B}$. Die Wahrscheinlichkeit dafür berechnet sich gemäß

$$P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0,8 = 0,2.$$

2. Für die Ergebnismenge $\Omega = \{(i,j) | 1 \leq i,j \leq 6\}$ gilt $|\Omega| = 36$.

Für das Ereignis W_1: *Der 1. Wurf ist eine gerade Augenzahl* gibt es 18 günstige Fälle:

		2. Wurf					
		1	2	3	4	5	6
1. Wurf	2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
	4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
	6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Die Wahrscheinlichkeit für das Ereignis W_1 beträgt nach Laplace

$$P(W_1) = \frac{|W_1|}{|\Omega|} = \frac{18}{36}$$

Für das Ereignis W_2: *Der 2. Wurf ist kleiner oder gleich zwei* sind 12 Ergebnisse günstig:

		2. Wurf	
		1	2
1. Wurf	1	(1,1)	(1,2
	2	(2,1)	(2,2)
	3	(3,1)	(3,2)
	4	(4,1)	(4,2)
	5	(5,1)	(5,2)
	6	(6,1)	(6,2)

Somit ergibt die Wahrscheinlichkei für W_2 nach Laplace:

$$P(W_2) = \frac{|W_2|}{|\Omega|} = \frac{12}{36}$$

3. Seien B: *Sonnenbrand* und S: *Insektenstich*. Bekannt sind:

$$P(B) = 0,65 \qquad P(S) = 0,30 \qquad P(B \cap S) = 0,15$$

Da die Ereignisse $B \backslash S$ und $B \cap S$ eine Zerlegung[1] von B bilden, berechnet sich die Wahrscheinlichkeit, dass man nur Sonnenbrand bekommt, gemäß

$$P(B \backslash S) = P(B) - P(B \cap S) = 0,65 - 0,15 = 0,5.$$

Die Wahrscheinlichkeit, dass man von mindestens einer dieser Plagen heimgesucht wird, beträgt

$$P(B \cup S) = P(B) + P(S) - P(B \cap S) = 0,65 + 0,3 - 0,15 = 0,8.$$

4. Es bezeichnen die Ereignisse:

   A: *Kunde kauft Produkt* A

   B: *Kunde kauft Produkt* B, $P(B) = 0,4$

   C: *Kunde kauft Produkt* C, $P(C) = 0,4$

   $A \cap B$: *Kunde kauft Produkt* A *und* B, $P(A \cap B) = 0,25$

   $A \cap C$: *Kunde kauft Produkt* A *und* C, $P(A \cap C) = 0,06$

[1] Zur Erinnerung: Eine Zerlegung bedeutet $B = (B \backslash S) \cup (B \cap S)$ und $(B \backslash S) \cap (B \cap S) = \emptyset$

$B \cap C$: *Kunde kauft Produkt* B *und* C, $P(B \cap C) = 0,04$

$A \cap B \cap C$: *Kunde kauft* A, B *und* C, $P(A \cap B \cap C) = 0,05$

$A \cup B \cup C$: *Kunde kauft mindestens eines der Produkte* A, B *und* C, $P(A \cup B \cup C) = 0,8$

Es gilt:

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Setzen wir die angegebenen Werte ein, erhalten wir

$$0,8 = P(A) + 0,4 + 0,4 - 0,25 - 0,06 - 0,04 + 0,05.$$

Daraus ergibt sich

$$P(A) = 0,8 - 0,5 = 0,3.$$

Der Anteil der Kunden, die das Produkt A kaufen, beträgt 30%. Wenn 100 Kunden zu erwarten sind, dann wird der Händler 30 Einheiten des Produkts A bereit stellen.

5. Es bezeichnen die Ereignisse:

   E: *Prüfling ist im Fach Englisch durchgefallen*

   D: *Prüfling ist im Fach Deutsch durchgefallen*

   Bekannt sind:

   $$P(E) = 0,15 \qquad P(D) = 0,10 \qquad P(E \cap D) = 0,08$$

   a) Die Wahrscheinlichkeit für das Ereignis $E \cap \bar{D}$: *Nur im Fach Englisch durchgefallen* berechnet sich gemäß

   $$P(E \cap \bar{D}) = P(E) - P(E \cap D) = 0,15 - 0,08 = 0,07.$$

   Die Wahrscheinlichkeit für das Ereignis $\bar{E} \cap D$: *Nur im Fach Deutsch durchgefallen* ergibt sich gemäß

   $$P(\bar{E} \cap D) = P(D) - P(E \cap D) = 0,10 - 0,08 = 0,02.$$

   b) Die Wahrscheinlichkeit für das Ereignis $(E \cap \bar{D}) \cup (\bar{E} \cap D)$: *In genau einem Fach durchgefallen* beträgt

   $$\begin{aligned} P((E \cap \bar{D}) \cup (\bar{E} \cap D)) &= P(E \cap \bar{D}) + P(\bar{E} \cap D) \\ &= 0,07 + 0,02 = 0,09. \end{aligned}$$

c) Die Wahrscheinlichkeit für das Ereignis $E \cup D$: *In mindestens einem Fach durchgefallen* ergibt sich gemäß:

$$\begin{aligned} P(E \cup D) &= P(E) + P(D) - P(E \cap D) \\ &= 0,15 + 0,10 - 0,08 = 0,17 \end{aligned}$$

d) Die Wahrscheinlichkeit für das Ereignis $\bar{E} \cap \bar{D}$: *In beiden Fächern bestanden* berechnet sich gemäß:

$$P(\bar{E} \cap \bar{D}) = P(\overline{E \cup D}) = 1 - P(E \cup D) = 1 - 0,17 = 0,83$$

6. Sei A_i: *Der i-te Wurf ist eine* 6. Weiter bezeichnen wir die Anzahl der benötigten Versuche mit n. Da der Würfel fair ist, gilt für jedes $i = 1, \ldots, n$:

$$P(A_i) = \frac{1}{6} \quad \text{und} \quad P(\bar{A}_i) = \frac{5}{6}$$

Das Ereignis *Bei n Würfen fällt mindestens einmal eine* 6 kann man wie folgt darstellen: $\bigcup\limits_{i=1}^{n} A_i$

Es gilt: $P\Big(\bigcup\limits_{i=1}^{n} A_i\Big) = 1 - P\Big(\bigcap\limits_{i=1}^{n} \bar{A}_i\Big)$

Da die Würfe voneinander unabhängig sind, gilt:

$$P\Big(\bigcap_{i=1}^{n} \bar{A}_i\Big) = \prod_{i=1}^{n} P(\bar{A}_i) = \Big(\frac{5}{6}\Big)^n$$

Damit gilt die Gleichung:

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 - \Big(\frac{5}{6}\Big)^n$$

Diese Wahrscheinlichkeit soll mindestens 0,5 betragen, d. h.

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 - \Big(\frac{5}{6}\Big)^n \geq 0,5 \quad \Leftrightarrow \quad 1 - 0,5 \geq \Big(\frac{5}{6}\Big)^n$$

Logarithmieren der Ungleichung führt zu

$$ln(0,5) \geq n \cdot ln\Big(\frac{5}{6}\Big) \quad \Leftrightarrow \quad \frac{ln(0,5)}{ln\big(\frac{5}{6}\big)} \leq n$$

Damit ist $n \geq 3,8$. Da n eine ganze Zahl darstellt, muss man mindestens viermal würfeln, um mit einer Wahrscheinlichkeit von mindestens 0,5 mindestens einmal eine 6 zu erhalten.

7. Seien die Ereignisse:

   Z: *Klausurnote* 2 *oder besser*

   U: *Besuch der Zusatzübungen*

   Bekannt sind:

$P(Z) = 0,25$	Peters A-priori-Wahrscheinlichkeit
$P(U\|Z) = 0,96$	Anteil der Übungsbesucher unter Studierenden mit Klausurnoten 2 oder besser
$P(U\|\bar{Z}) = 0,10$	Anteil der Übungsbesucher unter Studierenden mit Klausurnoten schlechter als 2
$P(Z\|U) =?$	

   Die Wahrscheinlichkeit für eine Note 2 oder besser, wenn man an den Zusatzübungen teilnimmt, berechnet sich nach Bayes:

   $$\begin{aligned} P(Z|U) &= \frac{P(U|Z) \cdot P(Z)}{P(U|Z) \cdot P(Z) + P(U|\bar{Z}) \cdot P(\bar{Z})} \\ &= \frac{0,96 \cdot 0,25}{0,96 \cdot 0,25 + 0,10 \cdot 0,75} \\ &= \frac{0,24}{0,24 + 0,075} = 0,76 \end{aligned}$$

   Peter wird die Zusatzübungen besuchen, denn er kann dadurch seine Chancen sogar verdreifachen.

8 Kombinatorik

1. Ein Lottospieler hat die Auswahl zwischen zwei Spielkategorien „6 aus 49“ oder „7 aus 50“. Die Spielregeln sind bei den beiden Kategorien identisch: Man kreuzt sechs bzw. sieben Zahlen aus einer Zahlenreihe 1 bis 49 bzw. 1 bis 50 auf einem dafür vorgesehenen Tippschein an. Welche der Spielkategorie bietet mehr Chancen für *Alle Richtige*?
2. Wie viele Kombinationen für vier Richtige sind beim Lottospiel „6 aus 49“ möglich?
3. In einem Marionettentheater nimmt eine Familie mit drei Kindern in einer Reihe mit fünf Stühlen Platz. Die Kinder möchten, dass mindestens ein Elternteil neben ihnen sitzt. Wie viele Sitzordnungen gibt es dafür? Wie viele Möglichkeiten gibt es, wenn die Familienmitglieder keine bestimmte Sitzordnung bevorzugen?
4. Für eine Musical-Aufführung in einer Schule werden ein Hauptdarsteller, eine Hauptdarstellerin und zwei Musiker gesucht. Zudem sind drei Nebenrollen zu besetzen. Für den Hauptdarsteller stehen drei Schüler und für die Hauptdarstellerin zwei Schülerinnen zur Verfügung. Genau zwei Musiker haben sich bereit erklärt, die Aufführung zu begleiten. Für die Nebenrollen gibt es sechs Kandidaten. Wie viele Musical-Gruppen hat der Musiklehrer insgesamt zur Auswahl?
5. Um sich auf die Prüfung vorzubereiten, können die Kandidaten zusätzliche Kurse besuchen. Im Fach Statistik gibt es für 30 Personen zwei Tutoriengruppen (Gruppe 1 und Gruppe 2). Jede Person trägt sich unabhängig von der Wahl der Kommilitonen in eine der beiden ausgehängten Listen ein. Wie viele Gruppen-Einteilungen sind nach dieser Vorschrift möglich?
6. In einer Filiale eines Schokoladenherstellers können Kunden eine Schokoladensorte: *Vollmilch* oder *Halbbitter* und weitere Zutaten *Haselnüsse, Mandeln, Erdnüsse, Cornflakes, Chili, Rosinen,*

Marshmallows nach ihren Vorstellungen kombinieren. Die gewünschte Tafel wird anschließend von einem Chocolatier angefertigt. Ein Kunde möchte eine Tafel mit drei weiteren Zutaten herstellen lassen. Wie viele Schokoladenkombinationen stehen ihm zur Verfügung? Wie viele Schokoladenkombinationen gibt es, wenn er auf Mandeln verzichtet?

7. Ein Skatblatt besteht aus 32 Karten in vier Farben (Herz, Pik, Kreuz und Karo) zu je acht Karten (Sieben, Acht, Neun, Zehn, Bube, Dame, König und As). Die Karten werden gemischt und verdeckt auf dem Tisch gelegt. Nacheinander werden zwei Karten (ohne Zurücklegen) gezogen. Wie viele Ergebnisse sind insgesamt möglich? Wie viele Ergebnisse führen zum folgenden Ereignis?

   A: *Die gezogenen Karten haben den gleichen Wert* (z.B. zwei Damen)

   B: *Die gezogenen Karten sind von gleicher Farbe* (z.B. Karo)

Lösungen zu den Aufgaben

Kapitel 8

1. Im Spiel „6 aus 49“ gibt es
$$\binom{49}{6} = 13.983.816$$
mögliche Tippreihen. Im Spiel „7 aus 50“ kann man auf
$$\binom{50}{7} = 99.884.400$$
Arten, seine sieben Kreuze zu verteilen. Für *Alle Richtige* gibt es bei jeder Kategorie nur eine Kombination. Da die Anzahl der Tippkombinationen beim Spiel „7 aus 50“ größer ist, ist die Chance für *Alle Richtige* hier geringer.
2. Für vier Richtige aus sechs Zahlen gibt es $\binom{6}{4}$ Möglichkeiten und aus den restlichen 43 Zahlen $\binom{43}{2}$ Möglichkeiten, zwei auszuwählen. Insgesamt gibt es
$$\binom{6}{4} \cdot \binom{43}{2} = \frac{6!}{4!2!} \cdot \frac{43!}{2!41!} = 13.545$$
Kombinationen für vier Richtige.
3. Drei Kinder, Vater und Mutter nehmen in einer Reihe mit fünf Stühlen Platz derart, dass neben jedem Kind mindestens ein Elternteil sitzt. Der erste Platz muss also von einem Kind (Kind 1, Kind 2 oder Kind 3) besetzt werden, auf dem zweiten muss ein Elternteil (Vater oder Muter) Platz nehmen; auf dem dritten kann ein der verbleibenden zwei Kinder sitzen; auf dem vierten nimmt der andere Elternteil Platz und schließlich sitzt das verbleibende Kind auf dem 5. Stuhl. Also gibt es für diese Sitzordnung insgesamt
$$3 \cdot 2 \cdot 2 \cdot 1 \cdot 1 = 12$$
Möglichkeiten. Gibt es keine Vorliebe für eine bestimmte Sitzordnung, können sich fünf Personen auf
$$5! = 120$$
Arten auf fünf Stühle verteilen.

4. Für den Hauptdarsteller gibt es $\binom{3}{1}$, für die Hauptdarstellerin $\binom{2}{1}$ und für die Nebenrollen $\binom{6}{3}$ Besetzungsmöglichkeiten. Zusammen mit der einzigen Möglichkeit für die Musiker ergeben sich
$$\binom{3}{1} \cdot \binom{2}{1} \cdot \binom{6}{3} \cdot 1 = 120$$
verschiedene Musical-Gruppen zur Auswahl.

5. Der Vorgang entspricht einem 30-maligen Werfen einer fairen Münze. Deshalb gibt es $2^{30} = 1.073.741.824$ Möglichkeiten, 30 Teilnehmer auf zwei Gruppen zu verteilen.

6. Es gibt $\binom{7}{3} = 35$ Möglichkeiten, aus einer siebenelementigen Menge von Zutaten drei verschiedene auszuwählen. Diese kann man mit zwei verschiedenen Schokoladensorten kombinieren. Insgesamt gibt es $35 \cdot 2 = 70$ verschiedene Schokoladenkombinationen. Ohne Mandeln hat der Kunde $\binom{6}{3} = 20$ verschiedene Zusammenstellungen von Zutaten, die beliebig mit zwei möglichen Schokoladensorten kombinierbar sind. Insgesamt hat er $20 \cdot 2 = 40$ Kombinationen zur Auswahl.

7. Es gibt $\binom{32}{2} = 496$ Möglichkeiten, aus 32 Karten 2 zu ziehen.

   Für jeden Wert gibt es $\binom{4}{2} = 6$ Möglichkeiten, zwei verschiedene Karten zu ziehen. Da es 8 verschiedene Werte gibt, gibt es insgesamt $8 \cdot 6 = 48$ verschiedene Ergebnisse für das Ereignis A.

   Für jede Farbe gibt es $\binom{8}{2} = 28$ Möglichkeiten, aus 8 Karten 2 verschiedene zu ziehen. Da es 4 Farben gibt, gibt es $4 \cdot 28 = 112$ verschiedene Ergebnisse für das Ereignis B.

9 Zufallsvariablen

1. Bei einem Zufallsexperiment wird eine Laplace-Münze dreimal unabhängig hintereinander geworfen. Die Zufallsvariable X_i wird für $i = 1, 2, 3$ folgendermaßen definiert:

   $$X_i = \begin{cases} 1 & \text{der } i\text{-te Wurf zeigt Kopf} \\ 0 & \text{sonst} \end{cases}$$

   Geben Sie die Ergebnismenge des Zufallsexperiments an. Bestimmen Sie für die Zufallsvariablen $S = X_1 + X_2 + X_3$ und $S_{23} = X_2 + X_3$ jeweils den Wertebereich und die Wahrscheinlichkeitsverteilung.

2. Betrachten Sie das Zufallsexperiment *Dreimaliges Werfen einer Laplace-Münze* erneut. Bestimmen Sie für die Zufallsvariable X: *Anzahl von „Zahl“* die Wahrscheinlichkeitsverteilung, den Erwartungwert und die Varianz.

3. Nun verwendet man das Zufallsexperiment *Dreimaliges Werfen einer Laplace-Münze* als Grundlage für das folgende Glücksspiel: Wird einmal oder dreimal „Zahl“ geworfen, erhält der Spieler 2 € als Auszahlung. Ansonsten verliert er den Einsatz von 1 €. Stellen Sie fest, ob es sich um ein faires Spiel handelt.

4. Betrachten Sie das Zufallsexperiment *Ein Laplace-Würfel wird zweimal geworfen.* Überprüfen Sie, ob die folgenden Ereignisse voneinander unabhängig sind:

   A: *Der 1. Wurf ist eine 4*

   B: *Die Augensumme ist gerade*

   C: *Die Augensumme ist 7 oder 8 oder 12*

5. In einer Urne befinden sich vier rote und sechs schwarze Kugeln. Man zieht ohne Zurücklegen solange Kugeln, bis die erste rote Kugel gezogen wird. Mit der Zufallsvariablen X bezeichnet

man *Anzahl der dafür benötigten Züge*. Ermitteln Sie die Wahrscheinlichkeitsverteilung von X und berechnen Sie die erwartete Anzahl der Züge, bis die erste rote Kugel gezogen wird.

6. Größere Projekte bestehen in der Regel aus Projektsequenzen, die von unterschiedlicher Dauer sein können. Die Zeit bis zum Abschluss einer Projektsequenz kann man als eine Zufallsvariable auffassen. Die gesamte Projektdauer ergibt sich aus einer Addition der einzelnen Projektsequenzen. Betrachten wir ein Großprojekt, das aus vier voneinander unabhängigen Projektsequenzen besteht. Die mittlere Dauer und die Varianz der einzelnen Sequenzen seien in der folgenden Tabelle gegeben:

Projektsequenz i	Mittlere Dauer (Wochen) μ_i	Varianz σ^2
1	15	4
2	10	5
3	16	4
4	12	3

Wie lange dauert das gesamte Projekt im Mittel? Geben Sie die Standardabweichung der gesamten Pojektdauer an.

7. Das amerikanische Roulette besteht aus 38 Feldern, die mit den Ziffern 1 bis 36 (18 rote und 18 schwarze Felder), 0 und 00 gekennzeichnet sind. Mit einem Einsatz von 10$ kann ein Spieler 10$, 20$ oder 350$ gewinnen. Einen Gewinn von 10$ kann man erzielen, wenn man auf *Rot* oder *Schwarz*, *Gerade* oder *Ungerade*, 1-18 oder 19-36 setzt; ein Einsatz auf 1-12 (*1. Douze*) oder 13-24 (*2. Douze*) oder $25-36$ (*3. Douze*) kann einen Gewinn von 20$ erbringen. Der größte Gewinn von 350$ ist möglich, wenn man auf eine einzelne Zahl zwischen 1-36 setzt. Welche Kategorie bietet eine größere Gewinnchance? (Hinweis: Berechnen Sie für jede Spielkategorie die Gewinnerwartung.)

8. **St. Petersburger-Spiel**[1]. Eine faire Münze wird solange geworfen, bis zum ersten Mal *Zahl* erscheint. Erscheint *Zahl* beim ersten Wurf, dann gibt es einen Gewinn von 2 Euro. Der Gewinn verdoppelt sich bei jedem weiteren Wurf, d. h., fällt *Zahl* beim zweiten Mal, dann beträgt der Gewinn 4 Euro, beim dritten Mal

[1]Diese Aufgabe war am Anfang des achtzehnten Jahrhunderts – natürlich ohne Euro – ein Diskussionsthema von Nikolaus und Daniel Bernoulli an der Petersburger Akademie der Wissenschaften. Daher stammt der Name.

8 Euro usw.. Allgemein: Fällt *Zahl* beim n-ten Wurf, dann gibt es 2^n Euro zu gewinnen. Geben Sie, falls möglich, den erwarteten Gewinn an.

9. Zwei Zufallsvariablen X und Y besitzen die folgenden Wahrscheinlichkeitsverteilungen:

x	1	5
$P(X = x)$	0,7	0,3

y	2	4	6
$P(Y = y)$	0,3	0,2	0,5

Geben Sie ihre gemeinsame Wahrscheinlichkeitsverteilung an, wenn X und Y unabhängig sind.

10. Die Trefferwahrscheinlichkeit bei einem Bernoulli-Experiment sei p. Geben Sie für die Bernoulli-Variable

$$X = \begin{cases} 1, & \text{wenn Treffer} \\ 0 & \text{sonst} \end{cases}$$

den Erwartungswert und die Varianz an.

11. Zeigen Sie, dass np und $np(1 - p)$ der Erwartungswert und die Varianz einer $B(n; p)$-verteilten Zufallsvariablen X darstellen. Verwenden Sie dabei das Ergebnis der Aufgabe 10.

12. Gegeben sei eine $B(5; 0, 3)$-verteilte Zufallsvariable X. Berechnen Sie die folgenden Wahrscheinlichkeiten mit der Formel der Binomial-Verteilung:

a)$P(X = 0)$ b)$P(X = 2)$ c)$P(X = 5)$

13. Laut einer Studie laden 5% der Deutschen Musik, Filme und digitale Bücher illegal aus dem Internet herunter. Wie groß ist die Wahrscheinlichkeit, dass sich unter fünf zufällig ausgewählten Downloads
    a) keine illegale Kopie befindet?
    b) höchstens eine illegale Kopie befindet?
    c) genau vier illegale Kopien befinden?

14. In einer Urne befinden sich vier rote und sechs weiße Kugeln. Drei Kugeln werden zufällig mit einem Griff gezogen. Bestimmen Sie die Wahrscheinlichkeit, dass sich darunter

a) keine rote Kugel befindet.

b) genau eine rote Kugel befindet.

c) zwei oder drei rote Kugeln befinden.

15. In einem Gefäß mit 4 roten und 6 schwarzen Kugeln werden 6 Kugeln zufällig hintereinander ohne Zurücklegen gezogen. Die Zufallsvariable X bezeichnet die *Anzahl der schwarzen unter den gezogenen Kugeln*. Geben Sie den Wertebereich und die Wahrscheinlichkeitsverteilung der Zufallsvariablen X an.

16. Ein Skatblatt besteht aus 32 Karten in vier Farben (Karo, Pik, Herz und Kreuz). Zu jeder Farbe gehören acht Karten (Sieben, Acht, Neun, Zehn, Bube, Dame, König und As). Die Karten werden gemischt und verdeckt auf den Tisch gelegt. Nacheinander zieht man (ohne Zurücklegen) sechs Karten. Berechnen Sie die Wahrscheinlichkeit, dass

    a) sich unter den 6 gezogenen Karten 3 Könige befinden.

    b) die 6 gezogenen Karten von gleicher Farbe (z.B. Karo) sind.

17. Ein Automobilzulieferer liefert einem Autohersteller Bauteile in Kartons zu je 60 Stück. Laut Liefervertrag darf ein Karton höchstens 5% Ausschuss enthalten. Aus jedem Karton werden 10 Bauteile zufällig und ohne Zurücklegen entnommen und überprüft. Jeder Karton wird nur dann angenommen, wenn alle 10 Geräte einwandfrei sind. Wie groß ist die Wahrscheinlichkeit, dass ein Karton zurückgewiesen wird? Gehen Sie davon aus, dass der Ausschussanteil genau 5% beträgt.

18. Bei einem Optiker wird festgestellt, dass ein Messgerät im fünfjährigen Mittel einen falschen Messwert abliefert. Wie groß ist die Wahrscheinlichkeit, dass es in einem Jahr

    a) keinen falschen Messwert gibt?

    b) höchstens zwei Messfehler gibt?

19. Die Verkehrsbetriebe einer Großstadt gehen davon aus, dass 3% ihrer Kunden schwarz fahren. Wie groß ist die Wahrscheinlichkeit, dass in einer U-Bahn, die mit 100 Personen besetzt ist, sich

    a) genau zwei Schwarzfahrer,

    b) zwei oder mehr Schwarzfahrer befinden?

20. Die Anzahl der Tippfehler innerhalb eines bestimmten Zeitraumes kann man als eine Poisson-verteilte Zufallsvariable auffassen. Erfahrungsgemäß beträgt die Wahrscheinlichkeit, dass einem Aktienhändler innerhalb von fünf Jahren kein Tippfehler unterläuft, 85%. Bestimmen Sie die Wahrscheinlichkeit, dass

    a) innerhalb von fünf Jahren höchstens ein Fehler vorkommt.

    b) innerhalb eines Jahres der Händler keinen Fehler macht.

    c) innerhalb eines Jahres der Händler genau einen Fehler macht.

21. In einer Risikostudie für Kernkraftwerke wird die Wahrscheinlichkeit für einen Kernschmelzunfall als 10^{-4} pro Reaktorjahr berechnet. Nehmen wir an, dass 450 Kernkraftwerke weltweit im Betrieb sind. Wie hoch ist die Wahrscheinlichkeit, dass in einem Reaktorjahr ein Kernschmelzunfall stattfindet? Verwenden Sie für die Berechnung zunächst die Binomial- und anschließend die Poisson-Verteilung. Vergleichen Sie Ihre Ergebnisse.

22. Sei X eine normalverteilte Zufallsvariable mit dem Erwartungswert 0,5 und Standardabweichung 0,5. Bestimmen Sie:

    a) $P(X \leq 2)$ b) $P(1 \leq X \leq 2)$ c)$P(X \geq 4)$

23. Bestimmen Sie das α-Quantil z_α der $N(0;1)$-Verteilung für:

    a) $\alpha = 0,95$ b) $\alpha = 0,05$ c) $\alpha = 0,90$ d) $\alpha = 0,10$

    Was bedeutet z_α, beispielsweise, wenn $\alpha = 0,95$?

24. Nach einer Studie schlafen Personen in Deutschland im Durchschnitt 7,4 Stunden (444 Minuten) pro Nacht. Die Standardabweichung beträgt 50 Minuten. Ausgehend von der Normalverteilungsannahme der Schlafdauer X bestimmen Sie die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person

    a) länger als 8 Stunden (480 Minuten) schläft.

    b) weniger als 3 Stunden (180 Minuten) schläft.

    Geben Sie für die Zufallvariable X die Ein-, Zwei- und Drei-Sigma-Bereiche an.

25. Die Lebensdauer von Energiesparlampen kann man als eine normalverteilte Zufallsvariable mit dem Erwartungswert 10.000 Stunden und Standardabweichung 4000 Stunden ansehen. Geben Sie die Wahrscheinlichkeit an, dass eine zufällig ausgewählte Lampe

    a) weniger als 6000 Stunden leuchtet.

    b) um mehr als 3000 Stunden über dem Mittelwert leuchtet.

    c) eine Lebensdauer zwischen 12.000 und 16.000 Stunden besitzt.

Lösungen zu den Aufgaben

Kapitel 9

1. Die Ergebnismenge ist:

$$\Omega = \{(0,0,0),(0,0,1),(0,1,0),(1,0,0),(0,1,1),(1,0,1),(1,1,0),(1,1,1)\}$$

Der Wertebereich von S ist $\{0,1,2,3\}$; die W.-verteilung lautet:

$$P(S=0)=\frac{1}{8} \quad P(S=1)=\frac{3}{8} \quad P(S=2)=\frac{3}{8} \quad P(S=3)=\frac{1}{8}$$

Der Wertebereich von S_{23} ist $\{0,1,2\}$; die W.-verteilung lautet:

$$P(S_{23}=0)=\frac{2}{8} \quad P(S_{23}=1)=\frac{4}{8} \quad P(S_{23}=2)=\frac{2}{8}$$

2. Die Zufallsvariable X hat den Wertebereich $\{0,1,2,3\}$ und die W.-verteilung:

x	0	1	2	3
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Der Erwartungswert von X lautet:

$$E(X)=\sum_{x=0}^{3} x\cdot P(X=x) = 0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+2\cdot\frac{3}{8}+3\cdot\frac{1}{8}=1,5$$

Die Varianz von X lautet:

$$\begin{aligned}
V(X) &= \sum_{x=0}^{3}(x-E(X))^2\cdot P(X=x)\\
&= (0-1,5)^2\cdot\frac{1}{8}+(1-1,5)^2\cdot\frac{3}{8}+(2-1,5)^2\cdot\frac{3}{8}+(3-1,5)^2\cdot\frac{1}{8}\\
&= 0,75
\end{aligned}$$

3. Sei X: *Anzahl von „Zahl"*. Als Gewinn bezeichnet man Auszahlung abzüglich Einsatz. Bei diesem Glückspiel kann man den Gewinn als Funktion von X wie folgt schreiben:

$$g(X) = \begin{cases} 2-1, & x = 1 \text{ oder } 3 \\ -1, & x = 0 \text{ oder } 2 \end{cases}$$

Das Spiel ist fair, weil die Gewinnerwartung gleich Null ist:

$$E[g(X)] = \sum_{x=0}^{3} g(x) \cdot P(X = x) = 1 \cdot \left(\frac{3}{8} + \frac{1}{8}\right) + (-1) \cdot \left(\frac{1}{8} + \frac{3}{8}\right) = 0$$

4. Da $A \cap B \cap C = \{(4,4)\}$, ergibt sich:

$$P(A \cap B \cap C) = \frac{|A \cap B \cap C|}{|\Omega|} = \frac{1}{36}$$

Auf der anderen Seite gilt:

$$P(A) \cdot P(B) \cdot P(C) = \frac{|A|}{|\Omega|} \cdot \frac{|B|}{|\Omega|} \cdot \frac{|C|}{|\Omega|} = \frac{6}{36} \cdot \frac{18}{36} \cdot \frac{12}{36} = \frac{1}{36}$$

Wir stellen fest:

$$P(A) \cdot P(B) \cdot P(C) = \frac{1}{36} = P(A \cap B \cap C)$$

Das bedeutet: Die Ereignisse A, B, C sind unabhängig.

5. Die Wahrscheinlichkeitsverteilung ist gegeben durch:

$P(X = 1) = \frac{4}{10} = 0,4$

$P(X = 2) = \frac{6}{10} \cdot \frac{4}{9} = 0,2667$

$P(X = 3) = \frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} = 0,1667$

$P(X = 4) = \frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} \cdot \frac{4}{7} = 0,0952$

$P(X = 5) = \frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{4}{6} = 0,0476$

$P(X = 6) = \frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{4}{5} = 0,0190$

$P(X = 7) = \frac{6}{10} \cdot \frac{5}{9} \cdot \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} \cdot \frac{4}{4} = 0,0048$

Der Erwartungswert bestimmt sich gemäß:

$$\begin{aligned}
E(X) &= \sum_{x=1}^{7} x \cdot P(X=x) \\
&= 1 \cdot 0,4 + 2 \cdot 0,2667 + 3 \cdot 0,1667 + 4 \cdot 0,0952 + 5 \cdot 0,0476 \\
&\quad +6 \cdot 0,0190 + 7 \cdot 0,0048 \\
&= 0,4 + 0,5333 + 0,5001 + 0,3808 + 0,2381 + 0,1143 + 0,0333 \\
&= 2,2
\end{aligned}$$

Im Mittel benötigt man 3 Züge, bis zum ersten Mal eine rote Kugel gezogen wird.

6. Mit X_i: *Dauer der Sequenz* i, $i = 1, 2, 3, 4$ lässt sich die gesamte Projektdauer als die Summe der einzelnen Projektsequenzen $X = \sum_{i=1}^{4} X_i$ darstellen. Somit berechnet sich die mittlere Projektdauer insgesamt gemäß

$$E(X) = \sum_{i=1}^{4} \mu_i = 15 + 10 + 16 + 12 = 53$$

und, da die Projektsequenzen voneinander unabhängig sind, ergibt sich die Varianz wie folgt:

$$V(X) = \sigma^2 = \sum_{i=1}^{4} \sigma_i^2 = 4 + 5 + 4 + 3 = 16$$

Daraus berechnen wir die Standardabweichung $\sigma = \sqrt{16} = 4$ (Wochen).

7. Wir bezeichnen mit der Zufallsvariablen X das Feld, auf das die Roulette-Kugel fällt. Der Einsatz auf *Rot* oder *Schwarz*, *Gerade* oder *Ungerade*, 1-18 oder 19-36 bringt den gleichen Gewinn von 10$; die Gewinnwahrscheinlichkeit ist für jede Auswahl gleich, nämlich $\frac{18}{38}$. Folglich beträgt die Wahrscheinlichkeit für einen Verlust $\frac{20}{38}$. Ohne Einschränkung setzen wir auf *Rot*. Die Gewinnfunktion lässt sich wie folgt formulieren:

$$g(X) = \begin{cases} 10, & \text{wenn Rot} \\ -10 & \text{sonst} \end{cases}$$

Die Gewinnerwartung ist

$$E[g(X)] = 10 \cdot \frac{18}{38} + (-10) \cdot \frac{20}{38} = -0,53.$$

Die Gewinnwahrscheinlichkeit für den Eisatz auf 1-12 (*1. Douze*) oder 13-24 (*2. Douze*) oder 25-36 (*3. Douze*) beträgt $\frac{12}{38}$; die Wahrscheinlichkeit für einen Verlust beträgt somit $\frac{26}{38}$. Ohne Einschränkung setzt man auf die 1. Douze. Die Gewinnfunktion lautet:

$$g(X) = \begin{cases} 20, & \text{wenn 1. Douze} \\ -10 & \text{sonst} \end{cases}$$

Die Gewinnerwartung berechnet sich gemäß

$$E[g(X)] = 20 \cdot \frac{12}{38} + (-10) \cdot \frac{26}{38} = -0,53.$$

Für die Berechnung der Gewinnerwartung bei der 3. Spielkategorie setzt man auf eine Zahl $j \in \{1, \dots, 36\}$. Die Wahrscheinlichkeit für einen Gewinn beträgt $\frac{1}{38}$. Dadurch ergibt sich eine Verlustwahrscheinlichkeit von $\frac{37}{38}$. Die Gewinnfunktion lautet:

$$g(X) = \begin{cases} 350, & \text{wenn } j \\ -10 & \text{sonst} \end{cases}$$

Mit dieser Gewinnfunktion erhält man die folgende Gewinnerwartung:

$$E[g(X)] = 350 \cdot \frac{1}{38} + (-10) \cdot \frac{37}{38} = -0,53$$

Egal für welche Version des Spiels man sich entscheidet, verliert man langfristig 0,53\$.

8. Sei $\{X = n\}$ das Ereignis *Zahl fällt zum 1. Mal beim n-ten Wurf.* Es gilt:

$$P(X = n) = \left(\frac{1}{2}\right)^n$$

Somit berechnet sich die Gewinnerwartung gemäß

$$E[g(X)] = \sum_{n=1}^{\infty} 2^n \cdot \left(\frac{1}{2}\right)^n = \sum_{n=1}^{\infty} 1 = \infty.$$

9. Wegen der Unabhängigkeit berechnet sich die gemeinsame Verteilung gemäß:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

Die Werte sind:

$$P(X=1,Y=2)=0,7\cdot 0,3=0,21$$

$$P(X=1,Y=4)=0,7\cdot 0,2=0,14$$

$$P(X=1,Y=6)=0,7\cdot 0,5=0,35$$

$$P(X=5,Y=2)=0,3\cdot 0,3=0,09$$

$$P(X=5,Y=4)=0,3\cdot 0,2=0,06$$

$$P(X=5,Y=6)=0,3\cdot 0,5=0,15$$

10. Der Erwartungswert und die Varianz von X sind:

$$\begin{aligned}
E(X) &= 1\cdot P(X=1)+0\cdot P(X=0)=1\cdot p+0\cdot(1-p)=p\\
V(X) &= (1-E(X))^2\cdot P(X=1)+(0-E(X))^2\cdot P(X=0)\\
&= (1-p)^2\cdot p+(0-p)^2\cdot(1-p)\\
&= (1-p)^2\cdot p+p^2\cdot(1-p)\\
&= (1-p)\big((1-p)p+p^2\big)\\
&= p(1-p)
\end{aligned}$$

11. Die $B(n;p)$-Verteilung kann man als das Ergebnis einer n-maligen Wiederholung eines Bernoulli-Experiments auffassen. Die Wiederholung des Experiments erfolgt dabei unabhängig voneinander. Mit n Bernoulli-Variablen $X_i,\ i=1,\dots,n$ gilt:

$$X=\sum_{i=1}^{n}X_i\sim B(n;p)$$

Nach der Additionseigenschaft des Erwartungswertes ergibt sich:

$$E(X)=E\Big(\sum_{i=1}^{n}X_i\Big)=\sum_{i=1}^{n}E(X_i)=\sum_{i=1}^{n}p=np$$

Da die Wiederholungen voneinander unabhängig sind, folgt nach der Additionseigenschaft für die Varianz:

$$V(X)=V\Big(\sum_{i=1}^{n}X_i\Big)=\sum_{i=1}^{n}V(X_i)=\sum_{i=1}^{n}p(1-p)=np(1-p)$$

12. Für $X\sim B(5;0,3)$ sind:

a) $P(X=0)=\binom{5}{0}0,3^0\cdot 0,7^5=0,16807$

b) $P(X=2)=\binom{5}{2}0,3^2\cdot 0,7^3=0,3087$

c) $P(X=5)=\binom{5}{5}0,3^5\cdot 0,7^0=0,00243$

13. Sei X: *Anzahl der illegalen Kopien in der Auswahl*; $X\sim B(5;0,05)$. Aus der Tabelle entnehmen wir:

    a) $P(X=0)=0,7738$

    b) $P(X\leq 1)=0,9774$

    c) $P(X=4)=P(X\leq 4)-P(X\leq 3)=1-1=0$

14. Da die Ziehung mit einem Griff erfolgt, gilt $X\sim H(3;10;4)$. Dabei bezeichnet die Zufallsvariable X: *Anzahl der roten unter den drei gezogenen Kugeln*. Der Wertebereich von X lautet $\{0,1,2,3\}$.

    a) $P(X=0)=\frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}=\frac{20}{120}$

    b) $P(X=1)=\frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}}=\frac{60}{120}$

    c) $P(X=2)+P(X=3)=\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}+\frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}}=\frac{40}{120}$

15. Die Zufallsvariable X ist hypergeometrisch verteilt mit den Parametern $n=6$, $N=10$ und $M=6$. Der Wertebereich von X lautet $\{2,\ldots,6\}$.

    $P(X=2)=\frac{\binom{6}{2}\binom{4}{4}}{\binom{10}{6}}=\frac{15}{210}$ $\qquad P(X=3)=\frac{\binom{6}{3}\binom{4}{3}}{\binom{10}{6}}=\frac{80}{210}$

    $P(X=4)=\frac{\binom{6}{4}\binom{4}{2}}{\binom{10}{6}}=\frac{90}{210}$ $\qquad P(X=5)=\frac{\binom{6}{5}\binom{4}{1}}{\binom{10}{6}}=\frac{24}{210}$

    $P(X=6)=\frac{\binom{6}{6}\binom{4}{0}}{\binom{10}{6}}=\frac{1}{210}$

16. Insgesamt gibt es $\binom{32}{6}$ Möglichkeiten, aus 32 Karten 6 verschiedene Karten zu ziehen.

    a) Bezeichnen wir *die Anzahl der Könige unter den sechs gezogenen Karten* mit X, gilt $X\sim H(6;32;4)$.

    $$P(X=3)=\frac{\binom{4}{3}\binom{28}{3}}{\binom{32}{6}}=\frac{13.104}{906.192}=0,0145$$

b) Es gibt $\binom{8}{6}$ Möglichkeiten, aus einer bestimmten Farbe (z.B. Herz) 6 verschiedene Karten zu ziehen. Wir bezeichnen das Ereignis *Die 6 gezogenen Karten sind von gleicher Farbe* mit B. Da es 4 Farben gibt, erhalten wir:

$$P(B) = 4 \cdot \frac{\binom{8}{6}\binom{24}{0}}{\binom{32}{6}} = 4 \cdot \frac{28}{906.192} = 0,0001$$

17. Die Zufallsvariable X: *Anzahl der Ausschussstücke in einem Karton* ist hypergeometrisch verteilt mit den Parametern

$$N = 60 \quad M = 0,05 \cdot 60 = 3 \quad n = 10$$

und dem Wertebereich $W_X = \{0, 1, 2, 3\}$. Ein Karton wird angenommen, wenn $\{X = 0\}$ ist. Die Wahrscheinlichkeit dafür ist

$$P(X = 0) = \frac{\binom{3}{0}\binom{57}{10}}{\binom{60}{10}} = 0,5728.$$

Die Lieferung wird verweigert, wenn $\{X > 0\}$ ist, und die Wahrscheinlichkeit dafür beträgt

$$P(X > 0) = 1 - P(X = 0) = 1 - 0,5728 = 0,4272.$$

18. Sei X: *Anzahl der Messfehler in einem Jahr*. Die Zufallsvariable X kann man als Poisson-verteilt mit der Intensitätsrate $\lambda = 0,2$ (innerhalb von 5 Jahren tritt im Mittel 1 Fehler auf) auffassen. Die Wahrscheinlichkeitsverteilung von X bestimmt sich gemäß:

$$P(X = x) = \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda} = \frac{0,2^x}{x!}\mathrm{e}^{-0,2}, \; x = 0, 1, \ldots$$

a) Die Wahrscheinlichkeit, dass innerhalb eines Jahres alle Messwerte richtig sind, berechnet sich folgendermaßen

$$P(X = 0) = \mathrm{e}^{-0,2} = 0,8187.$$

b) Die Wahrscheinlichkeit, dass innerhalb eines Jahres höchstens zwei Messfehler auftreten, berechnet sich gemäß

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \mathrm{e}^{-0,2} + 0,2\mathrm{e}^{-0,2} + \frac{0,2^2}{2}\mathrm{e}^{-0,2} \\ &= \mathrm{e}^{-0,2}(1 + 0,2 + 0,02) = 0,9989 \approx 1. \end{aligned}$$

19. Die Zufallsvariable X: *Anzahl der Schwarzfahrer unter 100 Fahrgästen* ist $B(100; 0,03)$-verteilt. Da $p = 0,03 < 0,1$ und $n = 100$, kann man die $B(100; 0,03)$- durch die Poisson-Verteilung mit der Intensitätsrate $\lambda = 100 \cdot 0,03 = 3$ annähern. Die Wahrscheinlichkeitsverteilung von X bestimmt sich gemäß:

$$P(X = x) = \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda} = \frac{3^x}{x!}\mathrm{e}^{-3}, \ x = 0, 1, \ldots$$

    a) $P(X = 2) = \frac{3^2}{2!}\mathrm{e}^{-3} = 0,2240$

    b) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - (\frac{3^0}{0!}\mathrm{e}^{-3} + \frac{3^1}{1!}\mathrm{e}^{-3}) = 0,8000$

20. Die Zufallsvariable X: *Anzahl der Tippfehler innerhalb von fünf Jahren* ist Poisson-verteilt mit der Intensitätsrate λ. Aus

$$P(X = 0) = e^{-\lambda} = 0,85$$

ergibt sich die Intensitätsrate $\lambda = -\ln 0,85 = 0,1625$.

    a) Die Wahrscheinlichkeit, dass innerhalb von 5 Jahren der Händler höchstens einen Tippfehler macht, berechnet sich gemäß

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= 0,85 + \frac{0,1625^1}{1}\mathrm{e}^{-0,1625} \\ &= 0,9881. \end{aligned}$$

    b) Die Zufallsvariable Y: *Anzahl der Tippfehler innerhalb eines Jahres* ist wieder Poisson-verteilt mit der Intensitätsrate $\lambda_1 = \frac{0,1625}{5} = 0,0325$. Die Wahrscheinlichkeit, dass innerhalb eines Jahres der Händler keinen Fehler macht, errechnet man gemäß

$$P(Y = 0) = \mathrm{e}^{-\lambda_1} = 0,9680.$$

    c) Die Wahrscheinlichkeit, dass der Händler innerhalb eines Jahres genau einen Tippfehler macht, ergibt sich gemäß

$$P(Y = 1) = \lambda_1\mathrm{e}^{-\lambda_1} = 0,0315.$$

21. Sei X: *Anzahl der Kernschmelzunfälle in einem Reaktorjahr.*

a) $X \sim B(450; 10^{-4})$. Die Wahrscheinlichkeit, dass in einem Reaktorjahr ein Kernschmelzunfall passiert, ergibt sich gemäß

$$\begin{aligned} P(X=1) &= \binom{450}{1} 10^{-4}(1-10^{-4})^{449} \\ &= 450 \cdot 10^{-4} \cdot 0,956 = 0,04302409. \end{aligned}$$

b) Die $B(450; 10^{-4})$- lässt sich durch die Poisson-Verteilung mit der Intensitätsrate $\lambda = 450 \cdot 10^{-4} = 0,045$ annähern. Die Wahrscheinlichkeit, dass in einem Reaktorjahr ein Kernschmelzunfall passiert, ergibt sich gemäß

$$P(X=1) = \lambda \mathrm{e}^{-\lambda} = 0,045\mathrm{e}^{-0,045} = 0,04301989.$$

22. Damit man die Standardnormalverteilungstabelle anwenden kann, wird $X \sim N(0,5; 0,5)$ standardisiert gemäß:

$$Z = \frac{X-\mu}{\sigma} = \frac{X-0,5}{0,5}$$

$$\begin{aligned} P(X \leq 2) &= P\Big(Z \leq \frac{2-0,5}{0,5}\Big) \\ &= \Phi(3) = 0,9987 \end{aligned}$$

$$\begin{aligned} P(1 \leq X \leq 2) &= P(X \leq 2) - P(X \leq 1) \\ &= \Phi(3) - \Phi(1) \\ &= 0,9987 - 0,8413 \\ &= 0,1574 \end{aligned}$$

$$\begin{aligned} P(X \geq 4) &= 1 - P(X \leq 4) \\ &= 1 - P\Big(Z \leq \frac{4-0,5}{0,5}\Big) \\ &= 1 - \Phi(7) = 1 - 1 = 0 \end{aligned}$$

23. Das 0,95-Quantil $z_{0,95}$ ist derjenige Wert, für den $\Phi(z_{0,95}) = 0,95$ gilt. Aus der Tabelle entnehmen wir

$$z_{0,95} = 1,65, \text{ d. h. } \quad \Phi(1,65) = P(Z \leq 1,65) = 0,95.$$

Das 0,05-Quantil $z_{0,05}$ ist derjenige Wert, für den $\Phi(z_{0,05}) = 0,05$ gilt. Die $N(0;1)$-Tabelle gibt es für $\alpha \geq 0,5$. Wegen der Symmetrie gilt $\Phi(z_{0,05}) = \Phi(-z_{0,95})$, deshalb erhalten wir

$$z_{0,05} = -z_{0,95} = -1,65.$$

Für $\alpha = 0,90$ und $\alpha = 0,10$ bestimmt man entsprechend

$$z_{0,90} = 1,28 \text{ und } z_{0,10} = -z_{0,90} = -1,28.$$

24. Für die Zufallsvariable X gilt $X \sim N(444; 50)$. Die Standardisierung erfolgt gemäß:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 444}{50}$$

a) Die Wahrscheinlichkeit, dass die Schlafdauer länger als 8 Stunden (480 Minuten) beträgt, berechnet sich folgendermaßen:

$$\begin{aligned} P(X > 480) &= 1 - P(X \leq 480) \\ &= 1 - P\Big(Z \leq \frac{480 - 444}{50}\Big) \\ &= 1 - \Phi(0,72) = 1 - 0,58 = 0,42 \end{aligned}$$

b) Die Wahrscheinlichkeit, dass die Schlafdauer weniger als 3 Stunden (180 Minuten) beträgt, ergibt sich gemäß:

$$P(X \leq 180) = P\Big(Z \leq \frac{180 - 444}{50}\Big) = \Phi(-5,28) \approx 0$$

Nach den Sigma-Regeln schlafen

a) ca. 68% der Bevölkerung zwischen 6,57 und 8,23 Stunden.

b) ca. 95% der Bevölkerung zwischen 5,73 und 9,07 Stunden.

c) fast alle Bundesbürger zwischen 4,9 und 9,9 Stunden.

25. Sei X: *Lebensdauer von Energiesparlampen* (in 1000 Stunden), $X \sim N(10; 4)$. Die Standardisierung erfolgt gemäß:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 10}{4}$$

a) Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Lampe weniger als 6000 Stunden leuchtet, berechnet sich gemäß

$$\begin{aligned} P(X \leq 6) &= P\Big(Z \leq \frac{6-10}{4}\Big) \\ &= \Phi(-1) = 1 - \Phi(1) \\ &= 1 - 0,8413 \\ &= 0,1587. \end{aligned}$$

b) Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Lampe um mehr als 3000 Stunden über dem Mittelwert leuchtet, ergibt sich gemäß

$$\begin{aligned} P(X \geq 13) &= 1 - P(X \leq 13) \\ &= 1 - P\Big(Z \leq \frac{13-10}{4}\Big) \\ &= 1 - \Phi(0,75) \\ &= 1 - 0,7734 \\ &= 0,2266. \end{aligned}$$

c) Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Lampe eine Lebensdauer zwischen 12.000 und 16.000 Stunden besitzt, ergibt sich gemäß

$$\begin{aligned} P(12 \leq X \leq 16) &= P(X \leq 16) - P(X \leq 12) \\ &= P\Big(Z \leq \frac{16-10}{4}\Big) - P\Big(Z \leq \frac{12-10}{4}\Big) \\ &= \Phi(1,5) - \Phi(0,5) \\ &= 0,9332 - 0,6915 \\ &= 0,2417. \end{aligned}$$

10 Die wichtigsten Grenzwertsätze

1. Formulieren Sie die Ungleichung von Tschebyscheff für die Zufallsvariable $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, wenn X_i Bernoulli-Variablen sind.
2. Wie oft muss man eine faire Münze werfen, damit man mit einer Wahrscheinlichkeit von mindestens 95 Prozent sicher sein kann, dass die relative Häufigkeit von *Kopf* um höchstens 2 Prozent vom wahren Wert (0,5) abweicht? (Hinweis: Tschebyscheff-Ungleichung)
3. Eine Molkerei wird von einer sehr großen Anzahl von Milchbauern versorgt, die unabhängig voneinander liefern. Die gelieferte Milchmenge beträgt im Mittel 180.000 Liter pro Tag mit einer Standardabweichung von 10.000 Litern. Für eine optimale Produktion darf die Milchmenge die Marke 160.000 Liter nicht unterschreiten. Berechnen Sie die Wahrscheinlichkeit, dass an einem Tag die Produktion nicht reibungslos läuft.
4. Die Stadtwerke einer Großstadt betreiben insgesamt 800 Rolltreppen. Pro Tag werden im Mittel 20 Störungen registriert. Wie groß ist die Wahrscheinlichkeit, dass an einem beliebigen Tag bei weniger oder gleich 30 Rolltreppen Probleme auftreten?
5. In einer Waldparzelle wurden 100 Baumstämme geerntet. Ein erfahrener Förster schätzt, dass ein Zehntel der Ernte eher als Holz der mittleren Kategorie B einzustufen sind. Für seine Kalkulation möchte er die Wahrscheinlichkeit, dass die Anzahl der Stämme mit B-Qualität zwischen 6 und 14 liegt, ermitteln.
   a) Schätzen Sie die oben genannte Wahrscheinlichkeit mit Hilfe der Tschebyscheff-Ungleichung nach unten ab.
   b) Berechnen Sie einen Näherungswert derselben Wahrscheinlichkeit unter Anwendung des zentralen Grenzwertsatzes.

Lösungen zu den Aufgaben

Kapitel 10

1. Für die Bernoulli-Variablen $X_1, \ldots, X_n$ mit
$$X_i = \begin{cases} 1 & \text{Treffer} \\ 0 & \text{sonst} \end{cases}$$
sei $P(X_i = 1) = p$. Für die relative Häufigkeit $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ gilt:
$$E(\bar{X}) = p \quad \text{und} \quad V(\bar{X}) = \frac{p(1-p)}{n}$$
Die Ungleichung von Tschebyscheff für die relative Häufigkeit $\bar{X}$ lautet somit:
$$P(|\bar{X} - p| \geq c) \leq \frac{p(1-p)}{nc^2}$$

2. Sei
$$X_i = \begin{cases} 1 & \text{Kopf} \\ 0 & \text{sonst} \end{cases}$$
mit $P(X_i = 1) = 0,5$. Die Zufallsvariable
$$\bar{X}_{(n)} = \frac{1}{n}\sum_{i=1}^n X_i$$
stellt die relative Häufigkeit der Köpfe bei n Würfen dar. Der Erwartungswert und die Varianz der Zuvallsvariablen $\bar{X}_{(n)}$ sind:
$$E(\bar{X}_{(n)}) = 0,5 \quad \text{und} \quad V(\bar{X}_{(n)}) = \frac{1}{4n}$$
Die Wahrscheinlichkeit, dass $\bar{X}_{(n)}$ um höchstens 0,02 vom 0,5 abweicht, soll mindestens 0,95 betragen, d.h.
$$P(|\bar{X}_{(n)} - 0,5| \leq 0,02) \geq 0,95.$$
Es gilt:
$$\begin{aligned} P(|\bar{X}_{(n)} - 0,5| \leq 0,02) &= 1 - P(|\bar{X}_{(n)} - 0,5| > 0,02) \\ &\geq 1 - P(|\bar{X}_{(n)} - 0,5| \geq 0,02) \quad (\star) \\ &\geq 1 - \frac{1}{0,0016n} \quad (\star\star) \\ &\geq 0,95 \quad (\text{Bedingung}) \end{aligned}$$

Aus der letzten Ungleichung $1 - \frac{1}{0{,}0016n} \geq 0{,}95$ folgt

$$0{,}05 \geq \frac{1}{0{,}0016n} \Leftrightarrow n \geq \frac{1}{0{,}0016 \cdot 0{,}05} = 12.500.$$

Bei mindestens 12.500 Würfen können wir zu 95% sicher sein, dass der Anteil von *Kopf* um nicht mehr als 2% von 0,5 abweicht.

($\star$) $P(|\bar{X}_{(n)} - 0{,}5| > 0{,}02) \leq P(|\bar{X}_{(n)} - 0{,}5| \geq 0{,}02)$

($\star\star$) Nach Tschebyscheff gilt:

$$P(|\bar{X}_{(n)} - 0{,}5| \geq 0{,}02) \leq \frac{V(\bar{X}_{(n)})}{0{,}02^2} = \frac{1}{4n \cdot 0{,}02^2} = \frac{1}{0{,}0016n}$$

3. Die Zufallsvariable X: *Tägliche Milchmenge* lässt sich als eine Summe von sehr vielen, unabhängigen Zufallsvariablen X_i: *Milchmenge des Bauern i* auffassen. Nach dem zentralen Grenzwertsatz gilt (in 10.000 Liter):

$$X \underset{a}{\sim} N(18\,;\,1) \quad \text{bzw.} \quad Z = X - 18 \underset{a}{\sim} N(0\,;\,1)$$

Die Produktion läuft nicht optimal, wenn $X \leq 16$. Die Wahrscheinlichkeit dafür berechnen wir gemäß:

$$\begin{aligned} P(X \leq 16) &= P(Z \leq 16 - 18) = \Phi(-2) \\ &= 1 - \Phi(2) \\ &= 1 - 0{,}9772 \\ &= 0{,}0228 \end{aligned}$$

4. Sei X: *Die Anzahl der Störungen an einem Tag.* Die Zufallsvariable X ist binomialverteilt mit dem Erwartungswert $E(X) = np = 20$. Aus $np = 20$ und $n = 800$ erhalten wir $p = 0{,}025$. Folglich beträgt die Varianz $V(X) = np(1-p) = 20 \cdot 0{,}975 = 19{,}5$.

Nach dem Grenzwertsatz von de Moivre gilt:

$$\frac{X - 20}{\sqrt{19{,}5}} \underset{a}{\sim} N(0\,;\,1)$$

Damit ist

$$P(X \leq 30) = P\Big(\frac{X - 20}{\sqrt{19{,}5}} \leq \frac{30 - 20}{\sqrt{19{,}5}}\Big) = \Phi(2{,}26) = 0{,}9881.$$

Um zu zeigen, dass dies ein guter Näherungswert für die gesuchte Wahrscheinlichkeit darstellt, ermitteln wir den genauen Wert von $P(X \leq 30)$ folgendermaßen:

$$P(X \leq 30) = \sum_{x=0}^{30} \binom{800}{x} 0,025^x \cdot 0,975^{800-x}$$

Die Funktion

BINOMVERT(Zahl_Erfolge;Versuche;Erfolgswahrsch;Kumuliert)

in Excel berechnet für Zahl_Erfolge = 30, Versuche = 800, Erfolgswahrsch = 0,025 und Kumuliert = WAHR die Wahrscheinlichkeit $P(X \leq 30)$. Das Ergebnis ist

$$P(X \leq 30) = \text{BINOMVERT(30;800;0,025;WAHR)} = 0{,}9876.$$

Die Abweichung von $0,9881 = P(X \leq 30) = \Phi(2,26)$ ist vernachlässigbar klein.

5. Sei X: *Die Anzahl der Stämme mit B-Qualität.* Die exakte Verteilung der Zufallsvariablen X ist die $B(100; 0,1)$-Verteilung mit dem Erwartungswert $E(X) = 100 \cdot 0,1 = 10$ und der Varianz $V(X) = 100 \cdot 0,1 \cdot 0,9 = 9$.

   a) Gesucht ist die Wahrscheinlichkeit $P(|X - 10| \leq 4)$.

   $$\begin{aligned} P(|X-10| \leq 4) &= 1 - P(|X-10| > 4) \\ &\geq 1 - P(|X-10| \geq 4) \quad (\star) \\ &\geq 1 - \frac{9}{4^2} = 0,4375 \end{aligned}$$

   Mit einer Wahrscheinlichkeit von mindestens 43,75% liegt die Anzahl der Baumstämme mit B-Qualität zwischen 6 und 14.

   $(\star)$ Nach der Ungleichung von Tschebyscheff gilt:

   $$P(|X-10| \geq 4) \leq \frac{V(X)}{4^2} = \frac{9}{16}$$

   b) Nach dem zentralen Grenzwertsatz von de Moivre gilt:

   $$X \underset{a}{\sim} N(10\,;\,3) \quad \text{bzw.} \quad Z = \frac{X-10}{3} \underset{a}{\sim} N(0\,;\,1)$$

Wir berechnen

$$
\begin{aligned}
P(6 \leq X \leq 14) &= P(X \leq 14) - P(X \leq 6) \\
&= P\Big(\frac{14-10}{3}\Big) - P\Big(\frac{6-10}{3}\Big) \\
&= \Phi\Big(\frac{4}{3}\Big) - \Phi\Big(\frac{-4}{3}\Big) \\
&= 2 \cdot \Phi\Big(\frac{4}{3}\Big) - 1 \\
&= 2 \cdot 0,9082 - 1 = 0,8164
\end{aligned}
$$

Mit einer Wahrscheinlichkeit von ca. 81,64% beträgt die Anzahl der B-Qualität-Baumstämme zwischen 6 und 14. (Den exakten Wert 0,8103 kann man mit Excel ausrechnen.)

11 Statistische Schätzverfahren

1. Aus einem Gefäß mit 3000 Kugeln wurde eine Stichprobe vom Umfang 50 gezogen. Darunter findet man 20 rote Kugeln. Wie viele rote Kugeln befinden sich schätzungsweise in dem Gefäß?

2. In der Absatzlogistik kann man davon ausgehen, dass die Lieferzeit (die Zeit zwischen dem Auftragseingang bis zur Ankunft der bestellten Ware) normalverteilt ist mit einem unbekannten Erwartungswert und einer bekannten Standardabweichung von 5 Tagen. Für die mittlere Lieferzeit sollen zweiseitige Intervallschätzer zu unterschiedlichen Niveaus von 90%, 95% und 99% gebildet werden. Dazu werden 20 Lieferzeiten notiert:

20	25	32	26	26	30	30	28	42	40
25	25	35	38	26	34	33	29	24	32

   Wie verhält sich die Genauigkeit der Schätzungen in Abhängigkeit von den Konfidenzniveaus?

3. Die Ludwig-Maximilian-Universität in München liegt nah am Englischen Garten. Dort befindet sich ein Biergarten, in dem sich viele Studenten im Sommer gern aufhalten. Während eines Sommers haben einige von ihnen die Biermenge in ihren Bierkrügen nachgemessen. Die Ergebnisse (in ml) seien:

   1000 1008 990 1010 985 980 995 997 975 985

   Bilden Sie ein Konfidenzintervall für den mittleren Bierinhalt jeweils zum Niveau 80%, 90% und 95%. Gehen Sie dabei von der Annahme aus, dass die Biermenge in den Maßkrügen normalverteilt mit der Varianz 100 ml ist. Vergleichen Sie die Genauigkeit der Schätzungen.

4. Wiederholen Sie die Aufgabe 3, wenn die Varianz unbekannt ist.

5. Die Dauer des Fernsehens von Jugendlichen kann man als normalverteilt ansehen. Eine Befragung von 16 Jugendlichen nach ihrem Fernsehkonsum ergibt die folgenden Werte (in Minuten):

120	100	120	80	180	150	140	120
100	0	120	120	0	50	100	0

Ermitteln Sie die durchschnittliche Fernsehdauer von Jugendlichen zum Konfidenzniveau 99%.

6. Bestimmen Sie ein 95% Konfidenzintervall für die Varianz der Zufallsvariablen X: *Biermenge* aus der Aufgabe 4.

7. Im Rahmen einer Studie zu den Ernährungsgewohnheiten der Bevölkerung in Deutschland wurde u.a. der Body Mass Index (BMI) der Teilnehmer gemessen. Laut dem Ergebnisbericht[1] haben 2005 der 13.207 Teilnehmer im Alter von 18 bis 80 Jahren einen BMI von 30,0 bis 34,9. Sie werden somit als leicht fettleibig eingestuft (Adipositas Grad I). Bestimmen Sie das 95% Konfidenzintervall für den als leicht adipös eingestuften Bevölkerungsanteil.

8. Aus dem Ergebnis der in der Aufgabe 7 genannten Studie wurde das mittlere Körpergewicht 17-jähriger Mädchen 62,0 kg mit einer Standardabweichung von 0,97 kg berechnet. Bestimmen Sie das 95% Konfidenzintervall für den Erwartungswert der Variablen *Körpergewicht*, wenn insgesamt 138 Mädchen in dem genannten Alter an der Studie teilgenommen haben.

   Wiederholen Sie die Aufgabe für den Erwartungswert des Merkmals *Körpergröße*, wenn man bei den 138 Mädchen eine mittlere Körpergröße von 166,8 cm und eine Standardabweichung von 0,56 gemessen hat.

9. Um Verbraucher zu informieren, wird die mittlere Ladezeit von Smartphone-Akkus zum Vertrauensniveau 99% geschätzt. Dafür geht man von einer normalverteilten Ladezeit mit einer Standardabweichung von 5 Minuten aus. Wie viele Akkus sollen mindestens in die Stichprobe aufgenommen werden, damit der Schätzfehler die Marke 2 Minuten nicht überschreitet?

10. Der Raucheranteil von Gaststättenbesuchern soll zum Vertrauensniveau 95% geschätzt werden. Wie viele Gaststättenbesucher müssen befragt werden, um eine Fehlerobergrenze von 1% nicht zu überschreiten?

[1] Tab. A 6, S. 130 im Ergebnisbericht, Teil 1. Nationale Verzehrsstudie II. Herausgeber: Max Rubner-Institut. Bundesforschungsinstitut für Ernährung und Lebensmittel, 2008.

Lösungen zu den Aufgaben

Kapitel 11

1. Der Anteil der roten Kugeln in der Stichprobe beträgt $\frac{20}{50}$. Mit diesem Ergebnis schätzt man die Anzahl der roten Kugeln im Gefäß auf $\frac{2}{5} \cdot 3000 = 1200$.

2. Sei X: *Lieferzeit*; $X \sim N(\mu; 5)$. Ein erwartungstreuer und konsistenter Schätzer für μ ist das Stichprobenmittel

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i;\ \bar{X} \sim N\Big(\mu; \frac{\sigma}{\sqrt{n}}\Big).$$

Das $(1-\alpha)$-Konfidenzintervall für μ lautet

$$\Big[\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\ ;\ \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\Big].$$

Für die vorliegenden Daten berechnen wir

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{20} = \frac{600}{20} = 30 \text{ und } \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{20}} = \sqrt{1,25}.$$

Bestimmung der Konfidenzintervalle für $\alpha = 0,10;\ 0,05;\ 0,01$:

$1-\alpha$	$z_{1-\alpha/2}$	$z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$	KI-Untergrenze $\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$	KI-Obergrenze $\bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$
0,90	1,65	$1,65 \cdot \sqrt{1,25} = 1,8448$	28,1552	31,8448
0,95	1,96	$1,96 \cdot \sqrt{1,25} = 2,1913$	27,8087	32,1913
0,99	2,58	$2,58 \cdot \sqrt{1,25} = 2,8845$	27,1155	32,8845

Je kleiner α ist, umso ungenauer wird die Schätzung.

3. Den Inhalt der Maßkrüge kann man als eine Zufallsvariable X auffassen; $X \sim N(\mu; 10)$. Ein erwartungstreuer und konsistenter Schätzer für μ ist das Stichprobenmittel

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i;\ \bar{X} \sim N\Big(\mu; \frac{\sigma}{\sqrt{n}}\Big).$$

Mit $n = 10$ erhalten wir $\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{10}} = \sqrt{10}$ und

$$\bar{x} = \frac{1000 + 1008 + 990 + 1010 + 985 + 980 + 995 + 997 + 975 + 985}{10} = 992,5.$$

Bestimmung der Konfidenzintervalle für $1 - \alpha = 0,8;\ 0,9;\ 0,95$:

$z_{1-\alpha/2}$	$z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$	KI-Untergrenze $\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$	KI-Obergrenze $\bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$	Schätzgenauigkeit $2 \cdot z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$
$z_{0,90} = 1,28$	4,0477	988,4523	996,5477	8,0954
$z_{0,95} = 1,65$	5,2178	987,2822	997,7178	10,4356
$z_{0,975} = 1,96$	6,1981	986,3019	998,6981	12,3962

Die Schätzgenauigkeit sinkt mit steigendem Konfidenzniveau.

4. Ist die Varianz unbekannt, dann ermittelt man das $(1 - \alpha)$-Konfidenzintervall gemäß

$$\left[\bar{X} - t_{1-\alpha/2\,;n-1}\frac{S}{\sqrt{n}}\ ;\ \bar{X} + t_{1-\alpha/2\,;n-1}\frac{S}{\sqrt{n}}\right].$$

Dabei dient $S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{(n-1)}}$ als Schätzer für σ. Für die vorliegenden Daten beträgt die Stichprobenstandardabweichung:

$$s = \sqrt{\frac{1210,5}{10-1}} = \sqrt{134,5} \quad \text{(siehe Tabelle)}$$

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	1000	7,5	56,25	6	980	-12,5	156,25
2	1008	15,5	240,25	7	995	2,5	6,25
3	990	-2,5	6,25	8	997	4,5	20,25
4	1010	17,5	306,25	9	975	-17,5	306,25
5	985	-7,5	56,25	10	985	-7,5	56,25
					9925	0	1210,5

Mit $\bar{x} = 992,5$ und $\frac{s}{\sqrt{n}} = \frac{\sqrt{134,5}}{\sqrt{10}} = \sqrt{13,45}$ bestimmen wir für $\alpha = 0,2;\ 0,1;\ 0,05$:

$t_{1-\alpha/2;9}$	$t_{1-\alpha/2;9}\frac{s}{\sqrt{n}}$	KI-Untergrenze $\bar{x} - t_{1-\alpha/2;9}\frac{s}{\sqrt{n}}$	KI-Obergrenze $\bar{x} + t_{1-\alpha/2;9}\frac{s}{\sqrt{n}}$	Schätzgenauigkeit $2 \cdot t_{1-\alpha/2;9}\frac{s}{\sqrt{n}}$
1,3830	5,0720	987,4280	997,5720	10,1440
1,8331	6,7228	985,7772	999,2228	13,4456
2,2622	8,2964	984,2036	1000,7964	16,5928

Die Unkenntnis der Varianz führt zum breiteren Konfidenzintervall $\bar{X} \pm t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}}$ gegenüber $\bar{X} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$.

5. Sei X: *Fernsehdauer von Jugendlichen*, $X \sim N(\mu;\sigma)$. Da σ unbekannt und der Stichprobenumfang $n = 16$ klein ist, bestimmt man das 99% Konfidenzintervall folgendermaßen

$$\left[\bar{X} - t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}} \,;\, \bar{X} + t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}}\right].$$

Mit dem Konfidenzniveau $1-\alpha = 0,99$ ist $t_{1-\alpha/2;n-1} = 2,9467$. Arbeitstabelle:

j	x_j	f_j	x_jf_j	$(x_j-\bar{x})$	$(x_j-\bar{x})^2$	$f_j(x_j-\bar{x})^2$
1	0	3	0	-93,75	8789,0625	26367,1875
2	50	1	50	-43,75	1914,0625	1914,0625
3	80	1	80	-13,75	189,0625	189,0625
4	100	3	300	6,25	39,0625	117,1875
5	120	5	600	26,25	689,0625	3445,3125
6	140	1	140	46,25	2139,0625	2139,0625
7	150	1	150	56,25	3164,0625	3164,0625
8	180	1	180	86,25	7439,0625	7439,0625
		16	1500			44.775

Das Stichprobenmittel beträgt $\bar{x} = \frac{\sum_{j=1}^{8} x_jf_j}{16} = \frac{1500}{16} = 93,75$. Die Stichprobenstandardabweichung ergibt sich gemäß

$$s = \sqrt{\frac{\sum_{j=1}^{8} f_j(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{44.775}{15}} = \sqrt{2985}.$$

Mit $\frac{s}{\sqrt{n}} = \frac{\sqrt{2985}}{\sqrt{16}} = \sqrt{186,5625}$ ergeben sich

die Untergrenze: $93,75 - 2,9467 \cdot \sqrt{186,5625} = 53,5016$

die Obergrenze: $93,75 + 2,9467 \cdot \sqrt{186,5625} = 133,9984$

Zum Konfidenzniveau 99% liegt die mittlere Fernsehdauer der Jugendlichen näherungsweise zwischen 54 und 134 Minuten pro Tag.

6. Das $(1-\alpha)$-Konfidenzintervall für die Varianz σ^2 lautet:

$$\left[\frac{(n-1)S^2}{c_{1-\alpha/2;n-1}}\ ;\ \frac{(n-1)S^2}{c_{\alpha/2;n-1}}\right]$$

In der Lösung zur Aufgabe 4 wurde $s^2 = 134,5$ berechnet. Für das 95% Konfidenzintervall bestimmen wir

das 0,975-Quantil der $\chi^2(9)$-Verteilung $c_{0,975;9} = 19,0228$,

die Untergrenze $\frac{(n-1)S^2}{c_{1-\alpha/2;n-1}} = \frac{9\cdot 134,5}{19,0228} = 63,6342$,

das 0,025-Quantil der $\chi^2(9)$-Verteilung $c_{0,025;9} = 2,7004$ und

die Obergrenze $\frac{(n-1)S^2}{c_{\alpha/2;n-1}} = \frac{9\cdot 134,5}{2,7004} = 448,2669$.

Wir erhalten das 95% Konfidenzintervall für die Varianz:

$$[63,6342\ ;\ 448,2669]$$

7. Das $(1-\alpha)$-Konfidenzintervall für p bestimmt man gemäß

$$\left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ;\ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right].$$

Unter den 13.207 Teilnehmern wurden 2005 als leicht adipös eingestuft, dies entspricht einem Anteilswert von

$$\hat{p} = \frac{2005}{13.207} = 0,1518.$$

Für das 95% Konfidenzintervall bestimmen wir die Grenzen:

$$\hat{p} - z_{1-\alpha/2}\sqrt{\tfrac{\hat{p}(1-\hat{p})}{n}} = 0,1518 - 1,96 \cdot \sqrt{\tfrac{0,1518\cdot 0,8472}{13.207}} = 0,1457$$

$$\hat{p} + z_{1-\alpha/2}\sqrt{\tfrac{\hat{p}(1+\hat{p})}{n}} = 0,1518 + 1,96 \cdot \sqrt{\tfrac{0,1518\cdot 0,8472}{13.207}} = 0,1579$$

Das 95% Konfidenzintervall für p lautet somit

$$[0,1457\ ;\ 0,1579].$$

Das heißt: Zum Konfidenzniveau 95% sind zwischen 14,6% und 15,8% der Bevölkerung leicht fettleibig.

8. Mit der Zufallsvariablen X bezeichnen wir das Körpergewicht von 17-jährigen Mädchen. Der Erwartungswert dieser Zufallsvariablen sei μ_X. Der Stichprobenumfang $n = 138$ ist groß genug für ein approximatives Konfidenzintervall

$$\left[\bar{X} - z_{1-\alpha/2}\frac{S}{\sqrt{n}}\,;\,\bar{X} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right].$$

   Für $\alpha = 0,05$ ist das $1 - \alpha/2$-Quantil der $N(0;1)$-Verteilung $z_{1-\alpha/2} = z_{0,975} = 1,96$. Setzen wir diesen Wert, $n = 138$, $\bar{x} = 62$ und $s_X = 0,97$ ein, erhalten wir das 95% Konfidenzintervall für μ_X:

$$62 \pm 1,96 \cdot \frac{0,97}{\sqrt{138}} = 62 \pm 0,1618$$

   Bezeichnen wir die Körpergröße von 17-jährigen Mädchen mit Y und den Erwartungswert mit μ_Y, lautet das 95% Konfidenzintervall für μ_Y entsprechend ($\bar{y} = 166,8$ und $s_Y = 0,56$):

$$166,8 \pm 1,96 \cdot \frac{0,56}{\sqrt{138}} = 166,8 \pm 0,0934$$

9. Mit dem Konfidenzniveau $1-\alpha = 0,99$ ist $z_{1-\alpha/2} = z_{0,995} = 2,58$. Setzen wir diesen Wert, $e = 2$ und $\sigma = 5$ ein, ergibt sich

$$n \geq \left(\frac{z_{1-\alpha/2} \cdot \sigma}{e}\right)^2 = \left(\frac{2,58 \cdot 5}{2}\right)^2 = 6,45^2 = 41,6.$$

   Um mit einem Vertrauensniveau von 99% und der Fehlerobergrenze von 2 Minuten die mittlere Ladezeit zu schätzen, muss man mindestens 42 Akkus in die Stichprobe aufnehmen.

10. Das Konfidenzniveau beträgt 95%. Damit ist $z_{1-\alpha/2} = z_{0,975} = 1,96$. Für die Bestimmung des Stichprobenumfangs geht man von einem gleichen Anteil von Rauchern p und Nichtrauchern $1-p$ aus, also ist $p = 0,5$. Setzen wir diesen Wert und $e = 0,01$ ein, ergibt sich

$$n \geq \left(\frac{z_{1-\alpha/2}}{e}\right)^2 p(1-p) = \left(\frac{1,96}{0,01}\right)^2 \cdot 0,5 \cdot 0,5 = 9604.$$

    Damit die geforderten Voraussetzungen erfüllt sind, müssen mindestens 9604 Gaststättenbesucher befragt werden.

12 Statistische Testverfahren

1. Laut einer Umfrage kommen 30 Prozent der Studenten mit dem neuen Bachelor-Studiengang nicht zurecht. Diese Vermutung wollen die Studentenvertreter mit Hilfe eines statistischen Tests widerlegen. Sie glauben, dass der Anteil höher als 30 Prozent liegt. Dazu werden 25 Studenten zufällig ausgewählt und befragt. Bei dieser Umfrage gaben 13 von ihnen an, mit dem neuen System nicht zurechtzukommen. Kann man damit das Umfrageergebnis (30 Prozent) zum Niveau 5% widerlegen?

2. Jedes Jahr zur Oktoberfestzeit in München wiederholt sich die Klage: „Die Maßkrüge werden zu wenig gefüllt“. Es wird sogar behauptet, dass mindestens 15 Prozent der Maßkrüge schlecht eingeschenkt seien. Die Stadt München möchte diese Hypothese widerlegen und behauptet, dass dieser Anteil weniger als 15 Prozent beträgt. Bei 40 zufällig ausgewählten Maßkrügen findet man zwei Krüge mit zu wenig Inhalt. Kann die Stadt München mit diesem Ergebnis und mit einem Signifikanzniveau 5% die Hypothese widerlegen? Wie lautet die Entscheidung, wenn das Signifikanzniveau auf 1% (10%) gesenkt (erhöht) wird?

3. Ein Bauunternehmer reklamiert, dass mindestens sechs Prozent der ihm gelieferten Keramikfliesen beschädigt sind. Der Lieferant will diese Behauptung nicht einfach hinnehmen. Aus seiner langjährigen Erfahrung weiß er, dass die Ausschussquote seiner Waren weniger als sechs Prozent beträgt. Um die Behauptung des Bauunternehmers zu widerlegen, wird ein statistischer Test durchgeführt. Das Signifikanzniveau 5% wird festgelegt und eine Stichprobe vom Umfang 150 wird gezogen. Darunter befinden sich 10 fehlerhafte Stücke. Kann der Lieferant damit die Behauptung des Bauunternehmers widerlegen?

4. Ein Einzelhändler wird ein neues Produkt in sein Sortiment nur dann aufnehmen, wenn Kunden auch bereit sind, es zu kaufen. Um die Akzeptanz eines neuen Produktes zu untersuchen, wird

nach einer Testphase eine Umfrage durchgeführt. Der Händler beschließt, das neue Produkt auf Dauer anzubieten, wenn mehr als die Hälfte seiner Kunden das Produkt akzeptieren. Von 80 zufällig ausgewählten Käufern gaben 48 an, das neue Produkt weiter zu beziehen. Wie wird sich der Händler nach diesem Ergebnis entscheiden? Wird er das neue Produkt auf Dauer anbieten? Als Entscheidungsinstrument wählt er einen statistischen Test zum Niveau 10%.

5. In Deutschland gibt es noch viele Einfamilienhäuser aus dem Jahr 1970, die nicht ausreichend saniert sind. Der Energieverbrauch solcher Gebäude mit alter Heizung wird im Mittel auf 3000 Liter Heizöl jährlich geschätzt. Ein Unternehmer, der sich auf besonders sparsame Heizungen spezialisiert, behauptet gar, der durchschnittliche Ölverbrauch betrage mehr als 3000 Liter pro Jahr. Seine Hypothese wird durch das Ergebnis eines statistischen Tests untermauert. Der Test wurde zum Signifikanzniveau 10% durchgeführt. Aus einer Stichprobe von 900 Einfamilienhäusern gleicher Art wurde ein mittlerer Verbrauch von 3009 Litern Heizöl jährlich ermittelt. Den jährlichen Ölverbrauch kann man als normalverteilt mit einer Standardabweichung von 100 (Litern) ansehen. Führen Sie den Test durch und bestimmen Sie den p-Wert dieses Tests.

6. Eine wichtige Einheit eines Fernsehers ist die Leiterplatte. Diese verwandelt Empfangssignale in Bild und Ton. Um zu überprüfen, ob der Sollwert der Leiterplattenstärke von 1,5 mm eingehalten wird, wird ein statistischer Test zum Niveau 5% durchgeführt. Dazu werden der Produktion 16 Platten entnommen. Eine Nachmessung der Platten ergibt einen Mittelwert von 1,41 mm. Kann man mit diesem Ergebnis die Vermutung *Die Leiterplattenstärke weicht vom Sollwert ab* widerlegen? Gehen Sie davon aus, dass die Leiterplattenstärke normalverteilt ist mit $\sigma = 0,04$ mm.

7. Die für die Produktion von Zucker angebauten Zuckerrüben werden in 10 Kilogramm-Säcken verkauft. Die Säcke werden von Maschinen zu diesem Sollgewicht abgefüllt. Von Zeit zu Zeit werden die Maschinen auf ihre korrekte Funktion mit Hilfe eines statistischen Tests überprüft. Zu diesem Zweck werden jeder Abfüllmaschine zufällig 9 Zuckerrüben-Säcke entnommen. Nachwiegen der Säcke aus einer der Maschinen ergibt die folgenden Werte:

10,69 10,03 10,95 10,28 10,31 10,27 10,12 10,66 10,95

Kann man zum Signifikanzniveau 0,05 dieses Ergebnis als zufällig ansehen? Das Abfüllgewicht kann man als normalverteilt mit einer Standardabweichung von 0,1 kg ansehen.

8. Bevor Holz in einer Waldparzelle geerntet wird, überprüft man, ob die Baumstämme einen Durchmesser von mindestens 60 cm haben. Die Überprüfung erfolgt mittels eines statistischen Tests. Bei einer Stichprobe von 10 Baumstämmen wurden die folgenden Durchmesser (in Zentimetern) gemessen:

   59 60 62 55 61 58 59 60 54 59

   Führen Sie den Test zum Niveau 0,10 durch. Den Durchmesser von Baumstämmen kann man als normalverteilt ansehen.

9. Tendenziell beobachtet man, dass, je länger ein Kunde in einem Geschäft bleibt, er umso eher etwas kauft. Um Kunden zum Verweilen aufzufordern, bringt ein Ladeninhaber seinen Laden auf den neuesten Stand. Nach der Wiedereröffnung hat man bei 20 zufällig ausgewählten Kunden die folgenden Werte (in Minuten) beobachtet:

35	34	45	48	5	50	15	45	40	30
30	28	45	58	48	45	25	66	22	10

   Überprüfen Sie mittels eines statistischen Tests zum Niveau 5%, ob die bisherige mittlere Verweildauer der Kunden von 25 Minuten signifikant übertroffen wird. Gehen Sie von einer normalverteilten Zufallsvariablen *Kundenverweildauer* aus.

10. Eine Maschine füllt Kaffeebohnen zu je 1000 Gramm ab. Die Standardabweichung des Abfüllgewichts von 4,5 Gramm soll dabei eingehalten werden. In regelmäßigen Abständen wird überprüft, ob diese Vorschrift eingehalten wird. Im Rahmen einer solchen Überprüfung werden 10 Packungen der Abfüllmaschine zufällig entnommen und nachgewogen. Die Ergebnisse sind:

    1000 998 990 997 1000 995 995 1008 1010 1005

    Stellen Sie fest, ob man bei einem Signifikanzniveau von 10% die Hypothese *Die Standardabweichung des Abfüllgewichts beträgt 4,5 Gramm* verwerfen kann. Gehen Sie dabei von einer Normalverteilung des Abfüllgewichts aus. Wie lautet die Entscheidung, wenn das Signifikanzniveua auf 5% reduziert wird?

Lösungen zu den Aufgaben

Kapitel 12

1. Die zu testende Hypothese und die Alternative sind

$$H_0 : p = 0,3 \quad \text{und} \quad H_1 : p > 0,3.$$

Dabei bezeichnet p den Anteil der Studenten mit Schwierigkeiten. Ihre Anzahl X in der Stichprobe ist unter der Nullhypothese $B(25; 0,3)$-verteilt. Der Wertebereich von X ist die Menge $W_X = \{0, 1, \ldots, 25\}$. Große X-Werte sprechen für die Alternative. Deshalb wird man H_0 ablehnen, wenn sich in der Stichprobe viele Personen mit Schwierigkeiten befinden. Zum Signifikanzniveau $\alpha = 0,05$ wird der Ablehnungsbereich

$$B = \{k, \ldots, 25\}$$

so bestimmt, dass

$$P(X \in B | p = 0,3) \leq 0,05.$$

Wegen

$$\begin{aligned} P(X \in B | p = 0,3) &= P(X \geq k | p = 0,3) \\ &= 1 - P(X \leq k - 1 | p = 0,3) \end{aligned}$$

können wir k so bestimmen, dass

$$P(X \leq k - 1 | p = 0,3) \geq 0,95.$$

Aus der $B(25; 0,3)$-Verteilungstabelle entnehmen wir:

$$P(X \leq 10 | p = 0,3) = 0,9022 < 0,95$$

$$P(X \leq 11 | p = 0,3) = 0,9558 > 0,95$$

Das heißt $k - 1 = 11 \Leftrightarrow k = 12$. Folglich lautet der Ablehnungsbereich

$$B = \{12, 13, \ldots, 25\}.$$

In der vorliegenden Stichprobe gaben $x = 13$ Personen an, mit dem neuen Bachelor-Studiengang nicht zurecht zu kommen. Da $x = 13 \in B$, wird H_0 abgelehnt. Der Anteil der Studenten mit Schwierigkeiten ist signifikant (zum Niveau 5%) größer als 30 Prozent.

2. Wir bezeichnen den Anteil der schlecht eingeschenkten Maßkrüge mit p. Die Behauptung $p \geq 0,15$ soll widerlegt werden, deswegen wird der Test für

$$H_0 : p \geq 0,15 \quad \text{gegen} \quad H_1 : p < 0,15$$

durchgeführt. Als Testgröße dient die Zufallsvariable X: *Anzahl der schlecht eingeschenkten Bierkrüge in der Stichprobe*. Unter H_0, an der Stelle $p = 0,15$, ist $X \sim B(40; 0,15)$.

Die Nullhypothese wird abgelehnt, wenn in der Stichprobe weniger als k $(0 \leq k \leq 40)$ Maßkrüge nicht ordnungsgemäß gefüllt sind. Die Grenze k wird für $\alpha = 0,05$ so bestimmt, dass

$$P(X \leq k|p = 0,15) = \sum_{x=0}^{k} \binom{40}{x} 0,15^x \cdot 0,85^{40-x} \leq 0,05.$$

Wir berechnen die einzelnen Wahrscheinlichkeiten

$$P(X = x|p = 0,15) = \binom{40}{x} 0,15^x \cdot 0,85^{40-x}$$

und kumulieren diese:

x	$P(X = x\|p = 0,15)$	$P(X \leq x\|p = 0,15)$
0	$\binom{40}{0} 0,15^0 \cdot 0,85^{40} = 0,0015$	0,0015
1	$\binom{40}{1} 0,15^1 \cdot 0,85^{39} = 0,0106$	0,0121
2	$\binom{40}{2} 0,15^2 \cdot 0,85^{38} = 0,0365$	$0,0486 < 0,05$
3	$\binom{40}{3} 0,15^3 \cdot 0,85^{37} = 0,0816$	$0,1302 > 0,05$

Wegen $P(X \leq 2|p = 0,15) = 0,0486 < 0,05$ und $P(X \leq 3|p = 0,15) = 0,1302 > 0,05$ erhalten wir $k = 2$ bzw. den Ablehnungsbereich

$$B = \{0, 1, 2\}.$$

In der Stichprobe befinden sich genau zwei schlecht eingeschenkte Bierkrüge. Da $x = 2 \in B$, wird die Nullhypothese abgelehnt.

Wird das Signifikanzniveau auf $\alpha = 0,01$ reduziert, lautet die Bedingung

$$P(X \leq k|p = 0,15) \leq 0,01.$$

Da $P(X \leq 0|p = 0,15) = 0,0015 < 0,01$ und $P(X \leq 1|p = 0,15) = 0,0121 > 0,01$ ist $B = \{0\}$. Wegen $x = 2 \notin B$ wird die Nullhypothese nicht abgelehnt.

Wird das Signifikanzniveau auf $\alpha = 0,10$ erhöht, erhält man den gleichen Ablehnungsbereich wie bei $\alpha = 0,05$. Deshalb führt $x = 2$ ebenfalls zur Ablehnung der Nullhypothese.

3. Sei p der Anteil der fehlerhaften Fliesen in der Lieferung. Die Hypothesen lauten:

$$H_0 : p \geq 0,06 \qquad\qquad H_1 : p < 0,06$$

Die Testgröße X: *Anzahl der fehlerhaften Fliesen in der Stichprobe* ist unter H_0 annähernd normalverteilt mit dem Erwartungswert

$$E(X) = np_0 = 150 \cdot 0,06 = 9$$

und der Varianz

$$V(X) = np_0(1 - p_0) = 150 \cdot 0,06 \cdot 0,94 = 8,46.$$

Folglich gilt unter H_0:

$$Z = \frac{X - 9}{\sqrt{8,46}} \underset{a}{\sim} N(0;1)$$

Die Nullhypothese wird abgelehnt, wenn die Realisation von Z in

$$B = \{z : z < -z_{1-\alpha}\}$$

liegt. Für $\alpha = 0,05$ beträgt das 0,95-Fraktil der $N(0;1)$-Verteilung $z_{1-\alpha} = z_{0,95} = 1,65$. Also ist

$$B = \{z : z < -1,65\}.$$

Aus der Stichprobe entnehmen wir $x = 10$. Folglich beträgt

$$z = \frac{10 - 9}{\sqrt{8,46}} = 0,34.$$

Da $0,34 \notin B$, wird die Nullhypothese nicht abgelehnt.

Selbstverständlich kann man einen exakten Binomialtest für das oben genannte Hypothesenpaar durchführen. Unter H_0, an der

Stelle $p = 0,06$, ist $X \sim B(150; 0,06)$. Der Ablehnungsbereich $B = \{0, 1, \ldots, k\}$ wird so bestimmt, dass

$$P(X \leq k|p = 0,06) = \sum_{x=0}^{k} \binom{150}{x} 0,06^x \cdot 0,94^{150-x} \leq 0,05.$$

Mit der Excel-Funktion BINOMVERT errechnen wir:

BINOMVERT(4;150;0,06;WAHR) = 0,0499 $< 0,05$

BINOMVERT(5;150;0,06;WAHR) = 0,1083 $> 0,05$

Damit ist $k = 4$. Da $x = 10 \notin B = \{0, 1, \ldots, 4\}$, wird H_0 nicht abgelehnt.

4. Die Aussage *Mehr als die Hälfte der Kunden akzeptieren das neue Produkt* soll statistisch zum Signifikanzniveau 10% gesichert werden. Deshalb wird der Test für

$$H_0 : p \leq 0,5 \quad \text{gegen} \quad H_1 : p > 0,5$$

durchgeführt. Dabei bezeichnet p den unbekannten Anteil der Kunden, die das neue Produkt akzeptieren. Unter H_0 ist die Zufallsvariable X: *Anzahl der Kunden, die das neue Produkt akzeptieren* annähernd normalverteilt mit dem Erwartungswert und der Varianz ($n = 80,\ p_0 = 0,5$)

$$E(X) = np_0 = 40 \quad \text{und} \quad V(X) = np_0(1 - p_0) = 20.$$

Mit anderen Worten: Die Testgröße

$$Z = \frac{X - 40}{\sqrt{20}}$$

ist unter H_0 asymptotisch standardnormalverteilt. Die Nullhypothese lehnt man ab, wenn die Realisation von Z in den Ablehnungsbereich

$$B =]z_{1-\alpha}; \infty[$$

fällt. Für $\alpha = 0,10$ ist das 0,90-Quantil der Standardnormalverteilung $z_{0,90} = 1,28$. Damit lautet der Ablehnungsbereich

$$B =]1,28; \infty[.$$

In der Stichprobe befinden sich $x = 48$ Kunden, die das neue Produkt akzeptieren. Das ergibt den Wert

$$z = \frac{48 - 40}{\sqrt{20}} = 1,79.$$

Da $1,79 \in B$, wird H_0 abgelehnt. Der Anteil der Kundschaft, die bereit ist, das neue Produkt zu kaufen, ist signifikant (zum Niveau 0,10) höher als 50 Prozent. Der Händler wird das Produkt in sein Sortiment aufnehmen.

5. Mit X bezeichnen wir den Heizölverbrauch des genannten Gebäudetyps. Die Zufallsvariable X ist normalverteilt mit dem Erwartungswert μ und der Standardabweichung $\sigma = 100$. Da σ bekannt ist, kann man einen Gauß-Test für

$$H_0 : \mu = 3000 \quad \text{gegen} \quad H_1 : \mu > 3000$$

durchführen. Wenn H_0 vorliegt, gehorcht das Stichprobenmittel $\bar{X}$ (Stichprobenumfang $n = 900$) einer Normalverteilung mit dem Erwartungswert und der Standardabweichung

$$\mu_0 = 3000 \quad \text{und} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{900}}.$$

Mit anderen Worten: Unter H_0 gilt

$$Z = \frac{\bar{X} - 3000}{100}\sqrt{900} \sim N(0;1).$$

Die Nullhypothese wird abgelehnt, wenn die Realisation von Z im Ablehnungsbereich

$$B =]z_{1-\alpha}; \infty[$$

liegt. Für $\alpha = 0,10$ ist das $(1-\alpha)$-Fraktil der $N(0;1)$-Verteilung $z_{0,90} = 1,28$. Folglich lautet der Ablehnungsbereich

$$B =]1,28; \infty[.$$

Aus der vorliegenden Stichprobe wurde ein Mittelwert $\bar{x} = 3009$ ermittelt. Das heißt

$$z = \frac{3009 - 3000}{100}\sqrt{900} = 2,7.$$

Da $z = 2,7 \in B$, wird die Nullhypothese abgelehnt. Der Heizölverbrauch ist (zum Niveau 10%) signifikant höher als 3000 Liter pro Jahr. Der p-Wert dieses Tests errechnet sich folgendermaßen:

$$P(Z > 2,7) = 1-P(Z \leq 2,7) = 1-\Phi(2,7) = 1-0,9965 = 0,0035$$

Der p-Wert ist kleiner als $\alpha = 0,10$; die Nullhypothese wird verworfen.

6. Sei X: *Leiterplattenstärke*; X ist eine normalverteilte Zufallsvariable mit dem Erwartungswert μ und der Standardabweichung $\sigma = 0,04$. Zu testen ist die Hypothese

$$H_0 : \mu = 1,5 \quad \text{gegen} \quad H_1 : \mu \neq 1,5.$$

Da σ bekannt ist, wird zur Überprüfung dieses Hypothesenpaares ein Gauß-Test durchgeführt. Die Testgröße lautet

$$Z = \frac{\bar{X} - 1,5}{\sigma}\sqrt{n}.$$

Unter H_0 gilt $Z \sim N(0;1)$. Die Nullhypothese wird abgelehnt, wenn die Realisation von Z im Ablehnungsbereich

$$B =]-\infty; -z_{1-\alpha/2}[\cup]z_{1-\alpha/2}; \infty[$$

liegt. Für $\alpha = 0,05$ ist das $(1-\alpha/2)$-Fraktil der $N(0;1)$-Verteilung $z_{0,975} = 1,96$. Folglich lautet der Ablehnungsbereich

$$B =]-\infty; -1,96[\cup]1,96; \infty[.$$

Die Stichprobe vom Umfang $n = 16$ liefert einen Mittelwert von $\bar{x} = 1,41$. Setzen wir diese beiden Werte und $\sigma = 0,04$ ein, ergibt sich

$$z = \frac{1,41 - 1,5}{0,04}\sqrt{16} = -9 \in B.$$

Die Daten sprechen gegen die Nullhypothese. Die Stärke der Leiterplatten weicht signifikant (zum Niveau 5%) von 1,5 mm ab.

7. Sei X: *Abfüllgewicht der Zuckerrüben-Säcke*; X ist eine normalverteilte Zufallsvariable mit dem Erwartungswert μ und der

Standardabweichung $\sigma = 0,1$. Da σ bekannt ist, kann man zur Überprüfung einen Gauß-Test für

$$H_0 : \mu = 10 \quad \text{gegen} \quad H_1 : \mu \neq 10$$

durchführen. Unter H_0 ist $\bar{X}$ normalverteilt mit dem Erwartungswert $\mu_0 = 10$ und der Standardabweichung $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0,1}{\sqrt{9}}$. Die Nullhypothese wird abgelehnt, wenn das Stichprobenmittel $\bar{X}$ im Ablehnungsbereich

$$B = \left] -\infty; \mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \right[\cup \left] \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \infty \right[$$

liegt. Für $\alpha = 0,05$ ist das $(1-\alpha/2)$-Fraktil der $N(0;1)$-Verteilung $z_{0,975} = 1,96$. Setzen wir diesen Wert, $\mu_0 = 10$, $\sigma = 0,1$ und $n = 9$ ein, lautet der Ablehnungsbereich

$$B =]-\infty; 9,935[\cup]10,065; \infty[.$$

Die vorliegende Stichprobe vom Umfang $n = 9$ liefert einen Mittelwert von

$$\begin{aligned} \bar{x} &= \frac{10,69 + 10,03 + 10,95 + 10,28 + 10,31 + 10,27 + 10,12 + 10,66 + 10,95}{9} \\ &= 10,47 \in B. \end{aligned}$$

Die Nullhypothese wird abgelehnt. Das Abfüllgewicht der Zuckerrüben-Säcke weicht signifikant (zum Niveau 5%) von 10 kg ab.

8. Sei X: *Durchmesser von Baumstämmen*; X ist eine normalverteilte Zufallsvariable mit dem Erwartungswert μ und der Standardabweichung σ. Da σ unbekannt ist, führen wir einen t-Test durch. Die Hypothesen sind:

$$H_0 : \mu \geq 60 \quad \text{gegen} \quad H_1 : \mu < 60$$

Die Testgröße lautet

$$T = \frac{\bar{X} - 60}{S}\sqrt{n}.$$

Dabei ist $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$. Unter H_0 gilt $T \sim t(n-1)$.

Die Nullhypothese wird abgelehnt, wenn die Realisation von T im Ablehnungsbereich

$$B =]-\infty; -t_{1-\alpha;n-1}[$$

liegt. Für $\alpha = 0,10$ und $n = 10$ ist das 0,90-Fraktil der $t(9)$-Verteilung $t_{0,90;9} = 1,3830$. Folglich lautet der Ablehnungsbereich

$$B =]-\infty; -1,3830[.$$

Für die Berechnung von T ermitteln wir zunächst die Stichprobenstandardabweichung S. Dazu stellen wir die folgende Arbeitstabelle auf:

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	59	0,3	0,09	6	58	-0,7	0,49
2	60	1,3	1,69	7	59	0,3	0,09
3	62	3,3	10,89	8	60	1,3	1,69
4	55	-3,7	13,69	9	54	-4,7	22,09
5	61	2,3	5,29	10	59	0,3	0,09
					587	0,00	56,10

Wir erhalten

$$s = \sqrt{\frac{56,10}{10-1}} = 2,5 \quad \text{und} \quad \bar{x} = \frac{587}{10} = 58,7.$$

Setzen wir diese Werte und $n = 10$ ein, ergibt sich

$$t = \frac{58,7 - 60}{2,5}\sqrt{10} = -1,65 \in B.$$

Die Nullhypothese wird abgelehnt. Die Durchmesser der Baumstämme sind signifikant (zum Niveau 10%) kleiner als 60 cm.

9. Die Zufallsvariable X: *Kundenverweildauer* ist normalverteilt mit dem Erwartungswert μ und der Standardabweichung σ. Da σ unbekannt ist, wird ein t-Test durchgeführt. Das zu testende Hypothesenpaar ist

$$H_0 : \mu = 25 \quad \text{und} \quad H_1 : \mu > 25.$$

Mit $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ ist die Testgröße

$$T = \frac{\bar{X} - 25}{S}\sqrt{n}$$

bei Gültigkeit der Nullhypothese $t(n-1)$-verteilt. Die Nullhypothese wird abgelehnt, wenn die Realisation von T im Ablehnungsbereich

$$B =]t_{1-\alpha;n-1}; \infty[$$

liegt. Für $\alpha = 0,05$ und $n = 20$ ist das 0,95-Fraktil der $t(19)$-Verteilung $t_{0,95;19} = 1,7291$. Folglich lautet der Ablehnungsbereich

$$B =]1,7291; \infty[.$$

Um den Wert der Testgröße zu berechnen, ermitteln wir zunächst die Stichprobenstandardabweichung S. Dazu stellen wir die folgende Arbeitstabelle auf:

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	35	-1,2	1,44	11	30	-6,2	38,44
2	34	-2,2	4,84	12	28	-8,2	67,24
3	45	8,8	77,44	13	45	8,8	77,44
4	48	11,8	139,24	14	58	21,8	475,24
5	5	-31,2	973,44	15	48	11,8	139,24
6	50	13,8	190,44	16	45	8,8	77,44
7	15	-21,2	449,44	17	25	-11,2	125,44
8	45	8,8	77,44	18	66	29,8	888,04
9	40	3,8	14,44	19	22	-14,2	201,64
10	30	-6,2	38,44	20	10	-26,2	686,44
					724	0,0	4743,20

Mit

$$\bar{x} = \frac{724}{20} = 36,2 \quad \text{und} \quad s = \sqrt{\frac{4743,20}{20-1}} = 15,80$$

erhalten wir

$$t = \frac{36,2 - 25}{15,80}\sqrt{20} = 3,17 \in B.$$

Die Nullhypothese wird abgelehnt. Man kann davon ausgehen, dass die Mordernisierungsmaßnahmen sich gelohnt haben.

10. Wir bezeichnen das Abfüllgewicht mit X; $X \sim N(1000; \sigma)$. Die zu überprüfende Hypothese und die Alternative sind:

$$H_0 : \sigma^2 = 4,5^2 = 20,25 \quad \text{und} \quad H_1 : \sigma^2 \neq 20,25$$

Wenn die Nullhypothese vorliegt, ist die Teststatistik

$$\chi^2 = \frac{n-1}{20,25} S^2$$

$\chi^2(n-1)$-verteilt. Dabei ist $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Die Nullhypothese wird abgelehnt, wenn die Realisation der Testgröße in

$$B =]0; c_{\alpha/2;n-1}[\cup]c_{1-\alpha/2;n-1}; \infty[$$

fällt. Für $n = 10$ und $\alpha = 0,10$ sind das 0,05- und das 0,95-Quantil der $\chi^2(9)$-Verteilung

$$c_{0,05;9} = 3,3251 \quad \text{und} \quad c_{0,95;9} = 16,9190.$$

Der Ablehnungsbereich lautet somit

$$B =]0\ ;\ 3,3252[\ \cup\ ]16,9190\ ;\ \infty[.$$

Um die Realisation von χ^2 zu berechnen, bestimmen wir für die erhobenen $n = 10$ Werte zunächst (vgl. nachfolgende Tabelle):

$$s^2 = \frac{\sum_{i=1}^{10}(x_i - \bar{x})^2}{9} = \frac{351,60}{9} = 39,0667$$

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	1000	0,2	0,04	6	995	-4,8	23,04
2	998	-1,8	3,24	7	995	-4,8	23,04
3	990	-9,8	96,04	8	1008	8,2	67,24
4	997	-2,8	7,84	9	1010	10,2	104,04
5	1000	0,2	0,04	10	1005	5,2	27,04
					9998	0,00	351,60

Setzen wir $s^2 = 39,07$ und $n = 10$ ein, erhalten wir

$$\chi^2 = \frac{9}{20,25} \cdot 39,0667 = 17,3630.$$

Da $\chi^2 = 17,3630 \in B$, wird H_0 abgelehnt. Die Daten sprechen für die Alternative; das Ergebnis ist zum Niveau 5% signifikant.

Für $\alpha = 0,05$ lauten die entsprechenden Quantile

$$c_{0,025;9} = 2,7004 \quad \text{und} \quad c_{0,975;9} = 19,0228.$$

Somit ergibt sich der Ablehnungsbereich

$$B =]0\ ;\ 2,7004[\ \cup\ ]19,0228\ ;\ \infty[.$$

Da $\chi^2 = 17,3630 \notin B$, wird H_0 nicht abgelehnt. Gegen die Nullhypothese ist nichts einzuwenden; das Ergebnis ist nicht signifikant.

13 Chi-Quadrat-Tests

1. Ein Tetraeder trägt auf jeder seiner Begrenzungsflächen eine der Zahlen 1, 2, 3 und 4. Es soll überprüft werden, ob der Tetraeder symmetrisch ist. Dazu wurde der Tetraeder 500-mal geworfen. (Eine Zahl gilt als geworfen, wenn der Tetraeder auf die Fläche mit dieser Zahl fällt.) Die Wurfergebnisse sind:

Zahl j	1	2	3	4
Häufigkeit n_j	118	122	125	135

   Kann man mit diesen Häufigkeiten die Vermutung *Der Tetraeder ist symmetrisch* verwerfen? Für eine Entscheidung führen Sie einen statistischen Test zum Signifikanzniveau $\alpha = 0,10$ durch.

2. Zur Vorbereitung einer Multiple-Choice-Klausur werden frühere Klausuraufgaben samt Lösungen verteilt. Jede Aufgabe hat fünf Lösungsvorschläge, von denen genau eine richtig ist. Ein Prüfungskandidat vermutet, dass der Dozent die richtigen Antworten zufällig verteilt. Seine Hypothese will er anhand eines statistischen Tests überprüfen. Unter 30 zufällig ausgewählten Aufgaben beobachtet er die folgenden Häufigkeiten der richtigen Antworten:

Richtige Antwort j	1	2	3	4	5
Häufigkeit n_j	9	3	7	7	4

   Wird die Nullhypothese nach diesen Beobachtungen abgelehnt? Führen Sie den Test durch.

3. Die bisherige Ernte eines Waldstückes besteht zu 35% aus Holz der Kategorie A, zu 40% der Kategorie B und zu 25% der Kategorie C. Von Zeit zu Zeit werden diese Werte von den zuständigen Förstern überprüft. Die zuletzt geführte Überprüfung hat ergeben, dass es sich bei 54 Baumstämmen um Holz der Güteklasse A, bei 46 der Güteklasse B und bei 20 der Güteklasse C handelt. Die Förster möchten nun feststellen, ob dieses Ergebnis eher durch Zufall zustande gekommen ist oder ob die

Zusammensetzung der Ernte sich verändert hat. Dazu wird ein χ^2-Anpassungstest zum Signifikanzniveau 5% durchgeführt.

4. Überprüfen Sie die Hypothese *Die Merkmale* Schulabschluss *und* Körpergewicht *bei Frauen sind unabhängig* mit Hilfe eines statistischen Tests zum vorgegebenen Signifikanzniveau 0,05. Verwenden Sie dazu die folgenden Daten[1]:

	Untergewicht	Normalgewicht	Übergewicht bis Adipositas
Volks-/Hauptschulabschluss	18	726	1525
Realschulabschluss/ POS	39	1227	1223
Fachhochschul-/ Hochschulreife	57	1336	658

5. Führen Sie den Test wie in der Aufgabe 4 für die Hypothese *Die Merkmale* Schulabschluss *und* Körpergewicht *bei Männern sind unabhängig* durch. Das Signifikanzniveau bleibt bei 0,05. Die beobachteten Häufigkeiten[2] sind in der folgenden Kontingenztabelle gegeben:

	Untergewicht	Normalgewicht	Übergewicht bis Adipositas
Volks-/Hauptschulabschluss	12	470	1664
Realschulabschluss/ POS	12	499	1181
Fachhochschul-/ Hochschulreife	12	838	1161

6. (Ergänzung) Vergleichen Sie die Stärke des Zusammenhangs zwischen den Merkmalen in der Aufgabe 4 bzw. 5, indem Sie für jedes der Geschlechter den korrigierten Kontingenzkoeffizienten P_{korr} (aus der deskriptiven Statistik) berechnen.

[1] Ergebnisbericht, Teil 1. Nationale Verzehrsstudie II. Herausgeber: Max Rubner-Institut. Bundesforschungsinstitut für Ernährung und Lebensmittel, 2008.

[2] ibid

Lösungen zu den Aufgaben

Kapitel 13

1. Die Hypothese *Der Tetraeder ist symmetrisch* wird mit Hilfe eines χ^2-Anpassungstests zum Niveau $\alpha = 0,10$ überprüft. Wir formulieren die Nullhypothese und die Alternative:

$$H_0 : p_j = \frac{1}{4} \quad \text{für } j = 1, \ldots, 4 \qquad H_1 : p_j \neq \frac{1}{4} \quad \text{für mind. ein } j$$

Wenn H_0 vorliegt, erwartet man bei $n = 500$ Würfen, dass jede Augenzahl 125-mal ($np_j = 500 \cdot \frac{1}{4}$) vorkommt. Die Teststatistik

$$\chi^2 = \sum_{j=1}^{4} \frac{(N_j - np_j)^2}{np_j} = \sum_{j=1}^{4} \frac{(N_j - 125)^2}{125}$$

ist unter H_0 annähernd $\chi^2(3)$-verteilt. (Die Zufallsvariable N_j bezeichnet die Häufigkeit, mit der die Augenzahl j erscheint.) Die Nullhypothese wird abgelehnt, wenn die Realisation der Testgröße im Ablehnungsbereich

$$B = \left]c_{1-\alpha;4-1}\ ;\ \infty\right[$$

liegt. Für $\alpha = 0,10$ beträgt das 0,90-Fraktil der $\chi^2(3)$-Verteilung $c_{0,9;3} = 6,2514$. Folglich lautet der Ablehnungsbereich

$$B = \left]6,2514\ ;\ \infty\right[.$$

Arbeitstabelle zur Berechnung der Prüfgröße χ^2:

Augenzahl j	n_j	np_j	$(n_j - np_j)^2$	$\frac{(n_j-np_j)^2}{np_j}$
1	118	125	49	0,392
2	122	125	9	0,072
3	125	125	0	0,000
4	135	125	100	0,800
				1,264

Da $\chi^2 = 1,264 \notin B$, wird H_0 nicht abgelehnt. Die Vermutung *Der Tetraeder ist symmetrisch* wird nicht verworfen. Das bedeutet jedoch nicht, dass sie bestätigt wird.

2. Wir führen einen χ^2-Anpassungstest für

$$H_0 : p_j = \frac{1}{5} \text{ für } j = 1, \ldots, 5 \quad \text{gegen} \quad H_1 : p_j \neq \frac{1}{5} \text{ für mind. ein } j$$

durch. Wenn die Nullhypothese richtig ist, erwartet man, dass jeder Lösungsvorschlag j ($np_j = 30 \cdot 0,2$) 6-mal vorkommt. Die Teststatistik

$$\chi^2 = \sum_{j=1}^{5} \frac{(N_j - np_j)^2}{np_j} = \sum_{j=1}^{5} \frac{(N_j - 6)^2}{6}$$

ist unter der Nullhypothese annähernd $\chi^2(4)$-verteilt. Zum Signifikanzniveau $\alpha = 0,05$ lautet der Ablehnungsbereich

$$B = \left]c_{1-\alpha;k-1} \; ; \; \infty\right[= \left]c_{0,95;4} \; ; \; \infty\right[= \left]9,49 \; ; \; \infty\right[.$$

Arbeitstabelle für die Berechnung des χ^2-Wertes:

Antwort j	n_j	np_j	$(n_j - np_j)$	$\frac{(n_j-np_j)^2}{np_j}$
1	9	6	3	$\frac{9}{6}$
2	3	6	-3	$\frac{9}{6}$
3	7	6	1	$\frac{1}{6}$
4	7	6	1	$\frac{1}{6}$
5	4	6	-2	$\frac{4}{6}$
	30			4

$\chi^2 = 4 \notin B \;\Rightarrow\; H_0$ wird nicht abgelehnt. Gegen die Vermutung ist nichts einzuwenden. Das Ergebnis ist aber nicht signifikant.

3. Sei p_j der Anteil der Güteklasse j, $j = 1$ (A), 2 (B), 3 (C). Mit diesen Bezeichnungen lautet die Hypothese:

$$H_0 : p_1 = 0,35 \quad p_2 = 0,40 \quad p_3 = 0,25$$

Wenn die Nullhypothese vorliegt, erwartet man unter den $n = 120$ $(54 + 46 + 20)$ Baumstämmen 42 $(120 \cdot 0,35)$ mit A-, 48 $(120 \cdot 0,40)$ mit B- und 30 $(120 \cdot 0,25)$ mit C-Qualität. Unter der Nullhypothese ist die Teststatistik

$$\chi^2 = \sum_{j=1}^{3} \frac{(N_j - np_j)^2}{np_j} \underset{a}{\sim} \chi^2(2).$$

Zum Signifikanzniveau $\alpha = 0,05$ wird die Nullhypothese abgelehnt, wenn der χ^2-Wert in

$$B = \left]c_{1-\alpha;2}\ ;\ \infty\right[= \left]5,99\ ;\ \infty\right[$$

liegt. Die Arbeitstabelle für die Berechnung des χ^2-Wertes:

Güteklasse j	n_j	np_j	$(n_j - np_j)^2$	$\frac{(n_j-np_j)^2}{np_j}$
1	54	42	144	3,43
2	46	48	4	0,08
3	20	30	100	3,33
				6,84

Wir erhalten $\chi^2 = 6,84 \in B \Rightarrow H_0$ wird abgelehnt. Zum Signifikanzniveau 5% kann man nicht mehr von der bisherigen Zusammensetzung der Ernte ausgehen.

4. Sei X: *Schulabschluss*; die Ausprägungen von X sind

   x_1 = Volks-/Hauptschulabschluss,

   x_2 = Realschulabschluss/POS und

   x_3 = Fachhochschul-/Hochschulreife.

   Weiter sei Y: *Gewichtskategorie* mit den Ausprägungen

   y_1 = Untergewicht,

   y_2 = Normalgewicht und

   y_3 = Übergewicht bis Adipositas.

   Für die Durchführung des Tests formulieren wir die Nullhypothese X *und* Y *sind unabhängig* wie folgt:

$$H_0 : p_{jk} = p_{j\bullet} \cdot p_{\bullet k} \quad \text{für alle } j, k = 1, 2, 3$$

   Dabei bezeichnen

$$p_{jk} = P(X = x_j, Y = y_k),\ p_{j\bullet} = P(X = x_j) \text{ und } p_{\bullet k} = P(Y = y_k).$$

   Bei Gültigkeit der Nullhypothese ist die Testgröße

$$\chi^2 = \sum_{j=1}^{3} \sum_{k=1}^{3} \frac{(N_{jk} - e_{jk})^2}{e_{jk}}$$

asymptotisch (annähernd) χ^2-verteilt mit $(3-1)(3-1) = 4$ Freiheitsgraden. Die Zufallsvariable N_{jk} gibt die Anzahl der Frauen mit dem Schulabschluss x_j und einer Gewichtskategorie y_k an. Ihre Anzahl, wenn X und Y unabhängig sind, bezeichnet e_{jk}; die Werte ergeben sich gemäß

$$e_{jk} = \frac{n_{j\bullet} \cdot n_{\bullet k}}{n}.$$

Dabei sind:

$n_{j\bullet}$ = Anzahl der Frauen mit dem Schulabschluss x_j

$n_{\bullet k}$ = Anzahl der Frauen in der Gewichtskategorie y_k

Um e_{jk} zu berechnen, ergänzen wir die Kontingenztabelle durch die Randhäufigkeiten:

	y_1	y_2	y_3	$n_{j\bullet}$
x_1	18	726	1525	2269
x_2	39	1227	1223	2489
x_3	57	1336	658	2051
$n_{\bullet k}$	114	3289	3406	6809

Im Falle der Unabhängigkeit hätte man die folgende Kontingenztabelle (gerundet):

	y_1	y_2	y_3	$n_{j\bullet}$
x_1	37,99	1096,01	1135	2269
x_2	41,67	1202,28	1245,05	2489
x_3	34,34	990,71	1025,95	2051
$n_{\bullet k}$	114	3289	3406	6809

Die Nullhypothese wird für $\alpha = 0,05$ abgelehnt, wenn der aus den Daten errechnete χ^2-Wert im Ablehnungsbereich

$$B =]c_{1-\alpha;4}\ ;\ \infty[\ =\ ]c_{0,95;4}\ ;\ \infty[\ =\ ]9,4877\ ;\ \infty[$$

liegt. Der Wert der Testgröße $\chi^2 = 537,77 \in B$ (siehe Arbeitstabelle). Folglich wird die Vermutung *Die Merkmale* Schulabschluss *und* Körpergewicht *bei Frauen sind unabhängig* widerlegt. Das Ergebnis ist zum Niveau 5% signifikant.

Arbeitstabelle zur Berechnung von χ^2:

j	k	n_{jk}	e_{jk}	$n_{jk} - e_{jk}$	$(n_{jk} - e_{jk})^2$	$\frac{(n_{jk}-e_{jk})^2}{e_{jk}}$
1	1	18	37,99	-19,99	399,55	10,52
1	2	726	1096,01	-370,01	136.908,37	124,92
1	3	1525	1135	390,00	152.100,12	134,00
2	1	39	41,67	-2,67	7,14	0,17
2	2	1227	1202,28	24,72	611,10	0,51
2	3	1223	1245,05	-22,05	486,13	0,39
3	1	57	34,34	22,66	513,52	14,95
3	2	1336	990,71	345,29	119.225,73	120,34
3	3	658	1025,95	-367,95	135.388,55	131,96
						537,77

5. Mit den entsprechenden Bezeichnungen können wir analog der Aufgabe 4 vorgehen. Für die Berechnung des χ^2-Wertes geben wir zunächst die Kontingenztabelle mit den Randhäufigkeiten an:

	y_1	y_2	y_3	$n_{j\bullet}$
x_1	12	470	1664	2146
x_2	12	499	1181	1692
x_3	12	838	1161	2011
$n_{\bullet k}$	36	1807	4006	5849

Im Falle der Unabhängigkeit hätten wir die folgende Kontingenztabelle (gerundet):

	y_1	y_2	y_3	$n_{j\bullet}$
x_1	13,21	662,99	1469,80	2146
x_2	10,41	522,73	1158,86	1692
x_3	12,38	621,28	1377,34	2011
$n_{\bullet k}$	36	1807	4006	5849

Arbeitstabelle zur Berechnung von χ^2:

j	k	n_{jk}	e_{jk}	$n_{jk}-e_{jk}$	$(n_{jk}-e_{jk})^2$	$\frac{(n_{jk}-e_{jk})^2}{e_{jk}}$
1	1	12	13,21	-1,21	1,46	0,11
1	2	470	662,99	-192,99	37244,71	56,18
1	3	1664	1469,80	194,20	37712,59	25,66
2	1	12	10,41	1,59	2,52	0,24
2	2	499	522,73	-23,73	563,08	1,08
2	3	1181	1158,86	22,14	490,33	0,42
3	1	12	12,38	-0,38	0,14	0,01
3	2	838	621,28	216,72	46966,80	75,60
3	3	1161	1377,34	-216,34	46803,32	33,98
						193,28

Der Wert der Testgröße $\chi^2 = 193,28 \in B =]9,4877\ ;\ \infty[$. Die Vermutung *Die Merkmale* Schulabschluss *und* Körpergewicht *bei Männern sind unabhängig* wird ebenfalls widerlegt.

6. Der korrigierte Kontingenzkoeffizient lautet $P_{korr} = \frac{P}{P_{max}}$. Dabei sind

$$P = \sqrt{\frac{\chi^2}{\chi^2+n}} \quad \text{und} \quad P_{max} = \sqrt{\frac{M-1}{M}},\ M = \min\{m, \ell\}.$$

Da $m = \ell = 3$, ist $M = 3$. Das ergibt einen maximalen Wert

$$P_{max} = \sqrt{\frac{2}{3}}.$$

Bei „Frauen" beträgt $\chi^2 = 537,77$. Zusammen mit dem Stichprobenumfang von $n = 6809$ ergibt sich der folgende Wert des Kontingenzkoeffizienten nach Pearson:

$$P = \sqrt{\frac{\chi^2}{\chi^2+n}} = \sqrt{\frac{537,77}{537,77+6809}} = \sqrt{\frac{537,77}{7346,77}}.$$

Setzt man diesen Wert und $P_{max} = \sqrt{\frac{2}{3}}$ ein, beträgt der korrigierte Kontingenzkoeffizient

$$P_{korr} = \sqrt{\frac{537,77}{7346,77}} \cdot \sqrt{\frac{3}{2}} = 0,3314.$$

Bei „Männern“ wurde $\chi^2 = 193,28$ berechnet. Mit dem Stichprobenumfang $n = 5849$ ist

$$P = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{193,28}{193,28 + 5849}} = \sqrt{\frac{193,28}{6042,28}}.$$

Zusammen mit $P_{max} = \sqrt{\frac{2}{3}}$ erhält man den korrigierten Kontingenzkoeffizienten

$$P_{korr} = \sqrt{\frac{193,28}{6042,28}} \cdot \sqrt{\frac{3}{2}} = 0,2190.$$

Der korrigierte Kontingenzkoeffizient bei Frauen ist größer als der bei Männern, d.h. der Zusammenhang zwischen den Merkmalen *Schulabschluss* und *Körpergewicht* ist bei Frauen ausgeprägter als bei Männern.