Prof. Dr. Tatjana Lange
Prof. Dr. Karl Mosler

Lösungen zu den Übungsaufgaben

Statistik kompakt - Basiswissen für Ökonomen und Ingenieure

1 Summen und Produkte

Aufgabe 1.1

Schreiben Sie ausführlich:

$$\text{a)} \sum_{i=1}^{n} a_i x_i \,, \qquad \text{b)} \sum_{i=1}^{n} a \,, \qquad \text{c)} \sum_{i=1}^{2} \sum_{j=1}^{4} x_i y_j \,, \qquad \text{d)} \sum_{i=1}^{4} x_i y_i \,.$$

Lösung:

a) $a_1 x_1 + ... + a_n x_n$

b) $a + ... + a \quad (n \text{ mal})$

c) $x_1 y_1 + x_1 y_2 + x_1 y_3 + x_1 y_4 + x_2 y_1 + x_2 y_2 + x_2 y_3 + x_2 y_4$

d) $x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4$

Aufgabe 1.2

Berechnen Sie den Wert der folgenden Summen:

$$\text{a)}\ \sum_{i=1}^{11} i\,, \qquad \text{b)}\ \sum_{i=0}^{8}(4i+5)\,, \qquad \text{c)}\ \sum_{i=-2}^{2}\frac{1}{2}i^2\,, \qquad \text{d)}\ \sum_{i=1}^{5}\frac{(i-3)^2}{2}\,.$$

Lösung:

a) $$\sum_{i=1}^{11} i = \frac{11\cdot 12}{2} = 66$$

b) $$\sum_{i=0}^{8}(4i+5) = 4\sum_{i=0}^{8} i + \sum_{i=0}^{8} 5 = 4\frac{8\cdot 9}{2} + 9\cdot 5 = 144+45 = 189$$

c) $$\sum_{i=-2}^{2}\frac{1}{2}i^2 = \frac{1}{2}\sum_{i=-2}^{2} i^2 = \frac{1}{2}\,(4+1+0+1+4) = 5$$

d) $$\sum_{i=1}^{5}\frac{(i-3)^2}{2} = \frac{1}{2}\sum_{i=1}^{5}\left(i^2-6i+9\right) = \frac{1}{2}\left[\sum_{i=1}^{5} i^2 - 6\sum_{i=1}^{5} i + \sum_{i=1}^{5} 9\right]$$
$$= \frac{1}{2}\left[\frac{5\cdot 6\cdot 11}{6} - 6\frac{5\cdot 6}{2} + 5\cdot 9\right] = 5$$

Aufgabe 1.3

Stellen Sie die folgenden Ausdrücke mit Hilfe des Summenzeichens dar:

a) "Die Summe der ersten fünfzig natürlichen Zahlen."

b) "Die Summe der ersten zehn Quadratzahlen."

c) "Die Summe der ersten zehn geraden Zahlen."

d) "Die Summe der Quadrate aller geraden Zahlen von 100 bis 200."

Lösung:

a) $$\sum_{i=1}^{50} i \quad \left(= \frac{50 \cdot 51}{2} = 1275\right)$$

b) $$\sum_{i=1}^{10} i^2 \quad \left(= \frac{10 \cdot 11 \cdot 21}{6} = 385\right)$$

c) $$\sum_{i=1}^{10} 2i \quad \left(= 2\frac{10 \cdot 11}{2} = 110\right)$$

d) $$\sum_{i=50}^{100} (2i)^2$$
$$\left(= 4 \sum_{i=50}^{100} i^2 = 4 \left[\sum_{i=1}^{100} i^2 - \sum_{i=1}^{49} i^2\right] = 4 \left[\tfrac{100 \cdot 101 \cdot 201}{6} - \tfrac{49 \cdot 50 \cdot 99}{6}\right] = 1191700\right)$$

Aufgabe 1.4

Es seien die folgenden $n = 3$ Messreihen der Länge $m = 4$ gegeben:

	$j=1$	$j=2$	$j=3$
x_{1j}	1	3	5
x_{2j}	2	4	6
x_{3j}	1	2	1
x_{4j}	5	4	0

Berechnen Sie die folgenden Summen mit $c = 2$:

a) $\sum_{i=1}^{m}\sum_{j=1}^{n} c\,,$ b) $\sum_{i=1}^{m}\sum_{j=1}^{n} cx_{ij}\,,$ c) $\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}+c)\,,$

d) $\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{1j}+x_{i3})\,,$ e) $\sum_{i=1}^{m}\sum_{j=1}^{n} x_{1j}x_{i2}\,.$

Lösung:

Arbeitstabelle:

	$j=1$	$j=2$	$j=3$	$\sum_j x_{ij}$
x_{1j}	1	3	5	9
x_{2j}	2	4	6	12
x_{3j}	1	2	1	4
x_{4j}	5	4	0	9
$\sum_i x_{ij}$	9	13	12	34

Es ist $c = 2, n = 3, m = 4$.

a) $$\sum_{i=1}^{m}\sum_{j=1}^{n} c = \sum_{i=1}^{4} 3c = 4\cdot 3\cdot c = 12\cdot 2 = 24$$

b) $$\sum_{i=1}^{m}\sum_{j=1}^{n} cx_{ij} = 2\sum_{i=1}^{4}\sum_{j=1}^{3} x_{ij} = 2\cdot 34 = 68$$

c) $$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}+c) = \sum_{i=1}^{4}\sum_{j=1}^{3} x_{ij} + \sum_{i=1}^{4}\sum_{j=1}^{3} 2 = 34 + 24 = 58$$

d) $\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{1j}+x_{i3}) = \sum_{i=1}^{4}\sum_{j=1}^{3} x_{1j} + \sum_{i=1}^{4}\sum_{j=1}^{3} x_{i3} = \sum_{i=1}^{4} 9 + \sum_{j=1}^{3} 12 = 72$

e) $\sum_{i=1}^{m}\sum_{j=1}^{n} x_{1j}x_{i2} = \left(\sum_{i=1}^{4} x_{i2}\right)\cdot\left(\sum_{j=1}^{3} x_{1j}\right) = 13\cdot 9 = 117$

Aufgabe 1.5

Fassen Sie mit Hilfe des Summenzeichens zusammen:

a) $a_3x^3 + a_2x^2 + a_1x + a_0$,

b) $1 + 4 + 9 + 16 + 25$,

c) $3x_1 + 6x_2 + 9x_3 + 2x_1y + 4x_2y^2 + 6x_3y^3$,

d) $x_1y_1 + x_1y_2 + x_1y_3 + x_2y_1 + x_2y_2 + x_2y_3 + x_3y_1 + x_3y_2 + x_3y_3$.

Lösung:

a) $\sum\limits_{i=0}^{3} a_i x^i$

b) $\sum\limits_{i=1}^{5} i^2$

c) $\sum\limits_{i=1}^{3} 3ix_i + \sum\limits_{i=1}^{3} 2ix_iy^i$

d) $\sum\limits_{i=1}^{3}\sum\limits_{j=1}^{3} x_iy_j$

Aufgabe 1.6

Zeigen Sie, dass für beliebige reelle Zahlen $x_1, \ldots, x_5$ gilt

a) $\displaystyle\sum_{i=1}^{5}\left[x_i - \frac{1}{5}\sum_{j=1}^{5} x_j\right] = 0\,,$

b) $\displaystyle\sum_{i=1}^{5}\left[x_i - \frac{1}{5}\sum_{j=1}^{5} x_j\right]^2 = \sum_{i=1}^{5} x_i^2 - \frac{1}{5}\left(\sum_{i=1}^{5} x_i\right)^2.$

Gelten diese Aussagen auch noch, wenn 5 durch eine beliebige natürliche Zahl n ersetzt wird?

Lösung:

a)

$$\sum_{i=1}^{5}\left[x_i - \frac{1}{5}\sum_{j=1}^{5} x_j\right] = \sum_{i=1}^{5} x_i - \sum_{i=1}^{5}\frac{1}{5}\sum_{j=1}^{5} x_j = \sum_{i=1}^{5} x_i - 5\cdot\frac{1}{5}\sum_{j=1}^{5} x_j$$
$$= \sum_{i=1}^{5} x_i - \sum_{j=1}^{5} x_j = 0$$

b) Ja.

$$\begin{aligned}
\sum_{i=1}^{5}\left[x_i - \frac{1}{5}\sum_{j=1}^{5} x_j\right]^2 &= \sum_{i=1}^{5}\left[x_i^2 - 2x_i\left(\frac{1}{5}\sum_{j=1}^{5} x_j\right) + \left(\frac{1}{5}\sum_{j=1}^{5} x_j\right)^2\right] \\
&= \sum_{i=1}^{5} x_i^2 - \sum_{i=1}^{5}\left[2x_i\left(\frac{1}{5}\sum_{j=1}^{5} x_j\right)\right] \\
&\qquad + \sum_{i=1}^{5}\left[\left(\frac{1}{5}\sum_{j=1}^{5} x_j\right)\left(\frac{1}{5}\sum_{j=1}^{5} x_j\right)\right] \\
&= \sum_{i=1}^{5} x_i^2 - 2\cdot\frac{1}{5}\sum_{i=1}^{5} x_i\sum_{j=1}^{5} x_j + \frac{1}{5}\sum_{j=1}^{5} x_j\sum_{j=1}^{5} x_j \\
&= \sum_{i=1}^{5} x_i^2 - \frac{1}{5}\left(\sum_{i=1}^{5} x_i\right)^2
\end{aligned}$$

Aufgabe 1.7

Berechen Sie den Wert der folgenden Produkte:

$$\text{a)} \prod_{i=1}^{5} i\,, \qquad \text{b)} \prod_{i=1}^{5} 2i^2\,, \qquad \text{c)} \prod_{i=1}^{20} \left(1 + \frac{1}{i}\right)\,, \qquad \text{d)} \prod_{i=-2}^{3} (2i-1)\,.$$

Lösung:

a) $$\prod_{i=1}^{5} i = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$$

b) $$\prod_{i=1}^{5} 2i^2 = 2^5 \left(\prod_{i=1}^{5} i^2\right) = 32 \cdot 120^2 = 460800$$

c) $$\prod_{i=1}^{20} \left(1 + \frac{1}{i}\right) = \prod_{i=1}^{20} \left(\frac{i+1}{i}\right) = \frac{2}{1} \cdot \frac{3}{2} \cdot \frac{4}{3} \cdot \ldots \cdot \frac{20}{19} \cdot \frac{21}{20} = 21$$

d) $$\prod_{i=-2}^{3} (2i-1) = (-5) \cdot (-3) \cdot (-1) \cdot 1 \cdot 3 \cdot 5 = (-25) \cdot (-9) \cdot (-1) = -225$$

Aufgabe 1.8

Fassen Sie mit Hilfe des Produktzeichens zusammen:

a) $3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8\,,$

b) $1 \cdot 4 \cdot 9 \cdot 16 \cdot 25\,,$

c) $\dfrac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}\,,$

d) $3 \cdot 6 \cdot 9 \cdot 12 \cdot 15\,.$

Lösung:

a) $\displaystyle\prod_{i=3}^{8} i$

b) $\displaystyle\prod_{i=1}^{5} i^2$

c) $\displaystyle\prod_{i=1}^{6} \frac{50-i}{i} = \prod_{i=1}^{6} \left(\frac{50}{i} - 1\right)$

d) $\displaystyle\prod_{i=1}^{5} 3i$

Aufgabe 1.9

Ein Kunde zahlt zu Jahresbeginn jeweils 1 000 Euro auf sein Sparbuch ein. Pro Jahr erhält er 5% Zinsen.

a) Welcher Betrag steht dem Kunden zu Beginn des 4. Jahres zur Verfügung?

b) Versuchen Sie den mathematischen Zusammenhang für einen Zeitraum von 10 Jahren in allgemeiner Schreibweise zusammenzufassen.

Lösung:

a)

Jahr	Betrag [Euro]	Summe
1.	1000	1000.000
2.	$1000 \cdot 1.05 + 1000$	2050.000
3.	$1000 \cdot 1.05^2 + 1000 \cdot 1.05 + 1000$	3152.500
4.	$1000 \cdot 1.05^3 + 1000 \cdot 1.05^2 + 1000 \cdot 1.05 + 1000$	4310.125

Zu Beginn des vierten Jahres hat er 4 310.125 [Euro].

b) $\sum_{i=1}^{10} 1000 \cdot (1.05)^{i-1}$

Diese Form der Darstellung entspricht der einer endlichen geometrischen Reihe. Demnach kann man den Betrag nach 10 Jahren vereinfacht berechnen mit:

$$\sum_{i=1}^{n} a_i = a_1 \frac{q^n - 1}{q - 1} = 1000 \frac{1.05^{10} - 1}{1.05 - 1} = 12577.89 [\text{Euro}]\,.$$

2 Kombinatorik

Aufgabe 2.1

a) Vor einem Bankschalter stehen sieben Personen und warten in einer Schlange. Wie viele verschiedene Anordnungen innerhalb der Schlange sind möglich?

b) Wenig später öffnet der Nachbarschalter. Daraufhin wechseln vier Personen zum zweiten Schalter. Wie viele Möglichkeiten gibt es nun, vier von den sieben Personen in einer neuen Schlange (vor dem zweiten Schalter) anzuordnen?

Lösung:

a) Ziehen ohne Zurücklegen, Reihenfolge relevant (geordnet), Anordnung aller Objekte, $n = 7$.
Es gibt
$$7! = 5040$$
verschiedene Anordnungen (Permutationen).

b) Ziehen ohne Zurücklegen, Reihenfolge relevant (geordnet), Auswahl von vier Objekten, $n = 7, k = 4$.
Es gibt
$$\frac{7!}{(7-4)!} = \frac{7!}{3!} = 840$$
verschiedene Anordnungen (Variationen ohne Zurücklegen).

Aufgabe 2.2

An einem Judo-Turnier nehmen in der Gewichtsklasse von 70 bis 77 Kilogramm acht Kämpfer teil. Wie viele verschiedene Einzelpaarungen sind möglich?

Lösung:

Ziehen ohne Zurücklegen, Reihenfolge nicht relevant (ungeordnet), Auswahl von zwei Objekten, $n = 8, k = 2$.

Es sind

$$\binom{8}{2} = \frac{8!}{2! \cdot 6!} = 28$$

verschiedene Einzelpaarungen möglich.

Aufgabe 2.3

Vier Würfel werden gleichzeitig geworfen. Wie viele verschiedene Zusammenstellungen von Augenzahlen sind dabei möglich?

Lösung:

Ziehen mit Zurücklegen, Reihenfolge nicht relevant (ungeordnet), $n = 6, k = 4$.

Es sind

$$\binom{6+4-1}{4} = \binom{9}{4} = \frac{9!}{4! \cdot 5!} = 126$$

verschiedene Zusammenstellungen möglich (Kombinationen mit Zurücklegen).

Aufgabe 2.4

Ein Bit kann zwei Zustände (0 oder 1) annehmen. Ein Byte besteht aus 8 Bits (z.B. 01101011). Wie viele verschiedene Bytes gibt es?

Lösung:

Ziehen mit Zurücklegen, Reihenfolge relevant (geordnet), $n = 2, k = 8$.
Es gibt

$$2^8 = 256$$

verschiedene Bytes.

Aufgabe 2.5

In der ersten Fußball-Liga eines Landes spielen in der Saison 1999/2000 15 Mannschaften um die Meisterschaft – darunter die Mannschaften Pechstadt und Glückstein.

a) Wie viele Spiele finden in der Saison, also in Hin- und Rückrunde insgesamt, statt?

b) Wie viele verschiedene Platzierungs-Tabellen der Liga sind nach dem letzten Spieltag der Saison theoretisch möglich?

c) Wie ändert sich die Anzahl aus Teil b), wenn nach dem letzten Spieltag die Mannschaft aus Pechstadt auf Platz 15 und die Mannschaft aus Glückstein auf Platz 1 liegt?

Lösung:

a) Ziehen ohne Zurücklegen, Reihenfolge nicht relevant (ungeordnet), Auswahl von zwei Objekten, $n = 15, k = 2$.
Jede der 15 Mannschaften spielt gegen jede der anderen Mannschaften (aber natürlich nicht gegen sich selbst) zweimal, nämlich in der Hin- und in der Rückrunde, also gibt es

$$2 \cdot \binom{15}{2} = \frac{2 \cdot 15 \cdot 14}{2} = 210$$

Spiele pro Saison.
Alternative Lösung:
Ziehen ohne Zurücklegen, Reihenfolge relevant (geordnet), Auswahl von zwei Objekten.
Dann gibt es insgesamt

$$\frac{15!}{13!} = 15 \cdot 14 = 210$$

Spiele pro Saison.

b) Ziehen ohne Zurücklegen, Reihenfolge relevant (geordnet), Anordnung aller Objekte, $n = 15$.
Es gibt

$$15!$$

verschiedene Platzierungstabellen.

c) Ziehen ohne Zurücklegen, Reihenfolge relevant (geordnet), Anordnung aller Objekte, $n = 13$.
In diesem Fall sind zwei der 15 Tabellenplätze festgelegt, so dass es nur noch

$$13!$$

mögliche Platzierungstabellen gibt.

Aufgabe 2.6

Ein Zahlenschloss besitzt fünf Ringe, die jeweils die Ziffern $0, ..., 9$ tragen.

a) Wie viele verschiedene fünfstellige Zahlencodes sind möglich?

b) Wie ändert sich die Anzahl aus Teil a), wenn in dem Zahlencode jede Ziffer nur einmal vorkommen darf, d.h. der Zahlencode aus fünf verschiedenen Ziffern bestehen soll?

c) Wie ändert sich die Anzahl aus Teil a), wenn der Zahlencode nur aus gleichen Ziffern bestehen soll?

Lösung:

a) Ziehen mit Zurücklegen, Reihenfolge relevant (geordnet), $n = 10, k = 5$.
Es gibt
$$10^5 = 100\,000$$
verschiedene Zahlencodes.

b) Ziehen ohne Zurücklegen, Reihenfolge relevant (geordnet), $n = 10, k = 5$.
Es gibt
$$\frac{10!}{5!} = 30\,240$$
Zahlenkombinationen.

c) Es gibt 10 solcher Zahlencodes (festgelegt durch Auswahl der ersten Stelle).

Aufgabe 2.7

Bei einer Wahl hat jeder Wahlberechtigte zwei Stimmen. Auf dem Wahlschein sind drei Kandidaten A, B und C aufgeführt.

a) Wie viele Möglichkeiten gibt es, den Wahlschein auszufüllen, wenn bei jedem Kandidaten höchstens ein Kreuz stehen darf?

b) Wie viele Möglichkeiten gibt es, den Wahlschein auszufüllen, wenn auch ,,gehäufelt" werden darf, d.h. wenn ein Kandidat auch mehr als nur ein Kreuz erhalten darf?

Lösung:

Man hat zwei Stimmen für drei Kandidaten.

a) Bei jedem Kandidaten darf höchstens nur ein Kreuz stehen, d.h. Ziehen ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge, $n = 3, k = 2$. Somit ist die Anzahl Möglichkeiten, den Wahlschein auszufüllen

$$\binom{3}{2} = 3$$

b) Ein Kandidat darf auch mehr als nur ein Kreuz erhalten, d.h. Ziehen mit Zurücklegen und ohne Berücksichtigung der Reihenfolge, $n = 3, k = 2$. Somit ist die Anzahl Möglichkeiten, den Wahlschein auszufüllen

$$\binom{3+2-1}{2} = \binom{4}{2} = 6$$

Aufgabe 2.8

In einer Partei sind drei verschiedene Positionen zu vergeben. Hierfür stehen sechs Kandidaten zur Verfügung. Wie viele Möglichkeiten gibt es, die drei Positionen zu besetzen, wenn

a) jeder Kandidat höchstens eine Position besetzen darf?

b) jeder Kandidat höchstens zwei Positionen besetzen darf?

c) jeder Kandidat beliebig viele Positionen (d.h. auch alle drei) besetzen darf?

Lösung:

Man hat

6	Kandidaten	(unterscheidbar)	1,2,3,4,5,6
3	Positionen	(unterscheidbar)	A,B,C

a) Jeder Kandidat darf nur eine Position besetzen (Ziehen ohne Zurücklegen, Reihenfolge relevant, $n = 6, k = 3$).

$$\frac{6!}{3!} = 6 \cdot 5 \cdot 4 = 120$$

b) Jeder Kandidat darf höchstens zwei Positionen besetzen.
Man unterscheidet zwei Fälle:

$$\text{jeder Kandidat nur eine Position} : 6 \cdot 5 \cdot 4 = 120 \quad \text{(vgl. a))}$$
$$\text{genau zwei Positionen} : 6 \cdot 5 \cdot 1 \cdot \binom{3}{2} = 90 \text{ (*)}$$
$$\text{Anzahl Möglichkeiten} = 210$$

(*): Für die erste Position hat man 6 Möglichkeiten, für die zweite noch 5, und die dritte Position ist dann festgelegt, da ein Kandidat ja zwei Positionen besetzt. Auf 2 von den 3 Positionen sitzt also derselbe Kandidat, aber auf welchen? Die Anzahl Möglichkeiten, 2 von 3 Positionen auszuwählen ist dann $\binom{3}{2}$.
Die Lösung von b) erhält man alternativ, indem man von der Lösung in Teil c) die Anzahl Möglichkeiten subtrahiert, dass ein Kandidat alle 3 Positionen besetzt, also $216 - 6 = 210$.

c) Jeder Kandidat darf beliebig viele Positionen (also auch alle drei) besetzen (Ziehen mit Zurücklegen, Reihenfolge relevant, $n = 6, k = 3$).

$$6^3 = 216$$

Aufgabe 2.9

Einer Gruppe von 15 Studenten werden drei Eintrittskarten für eine Veranstaltung angeboten. Auf wie viele Arten können die Karten verteilt werden, wenn sie

a) drei nummerierte Sitzplätze sind,

b) drei unnummerierte Stehplätze sind?

Unterscheiden Sie bei den beiden Teilaufgaben auch, ob ein Student

(i) genau eine Karte oder

(ii) mehrere Karten nehmen kann.

Lösung:

Bei allen vier zu untersuchenden Fällen handelt es sich um ein Auswahlproblem. Es werden aus den $n = 15$ Studenten $k = 3$ ausgewählt.

Sind die Sitzplätze nummeriert, so muss die Reihenfolge der Auswahl berücksichtigt werden, bei unnummerierten Stehplätzen nicht!

Kann ein Student mehrere Karten nehmen, so handelt es sich um eine Auswahl mit Wiederholung, ansonsten um eine Auswahl ohne Wiederholung!

a) Drei nummerierte Plätze:

   (i) Mit Reihenfolge, ohne Zurücklegen:

$$\frac{n!}{(n-k)!} = \frac{15!}{12!} = 15 \cdot 14 \cdot 13 = 2730\,.$$

   (ii) Mit Reihenfolge, mit Zurücklegen:

$$n^k = 15^3 = 3375\,.$$

b) Drei unnummerierte Plätze:

   (i) Ohne Reihenfolge, ohne Zurücklegen:

$$\binom{n}{k} = \binom{15}{3} = 2\frac{15!}{2!13!} = \frac{15 \cdot 14 \cdot 13}{1 \cdot 2 \cdot 3} = 455\,.$$

   (ii) Ohne Reihenfolge, mit Zurücklegen:

$$\binom{n+k-1}{k} = \binom{17}{3} = \frac{17 \cdot 16 \cdot 15}{1 \cdot 2 \cdot 3} = 680\,.$$

Aufgabe 2.10

a) Wie viele Möglichkeiten gibt es, die 12 Halbtöne (einer Oktave) ohne Wiederholung zu spielen? (Natürlich ohne den Rhythmus zu beachten.)

b) Jeder Dominostein wird durch zwei Zahlen charakterisiert. Die Steine sind symmetrisch, so dass das Nummernpaar nicht geordnet ist. Wieviele Dominosteine kann man erzeugen, wenn die Nummern $1, 2, ..., n$ benutzt werden?

Lösung:

a) $n = 12$ Elemente (= Töne der Oktave), alle Elemente werden angeordnet

$$12!$$

b) n Elemente (= Zahlen von $1, ..., n$), je $k = 2$ werden ausgewählt, mit Zurücklegen, Reihenfolge egal

$$\binom{n+2-1}{2} = \frac{n(n+1)}{2}$$

Aufgabe 2.11

In einer Fabrikhalle haben acht Werkstätten Platz.

a) Wie viele Möglichkeiten gibt es, in der Halle acht verschiedene Werkstätten einzurichten?

b) Wie viele Möglichkeiten gibt es, in dieser Halle zwei Zuschneidestationen, zwei Drehbänke und drei Lackierstationen einzurichten? (Eine Stelle bleibt also frei.)

Lösung:

$n = 8$ Elemente (= Werkstätten), alle Elemente werden angeordnet

a) alle Elemente sind verschieden

$$8! = 40320$$

b) $n_1 = 2, n_2 = 2, n_3 = 3$ Elemente sind gleich

$$\frac{8!}{2!2!3!} = 1680$$

Aufgabe 2.12

In der Fußball-Bundesliga spielen 18 Mannschaften. Sowohl in der Hin- all auch in der Rückrunde einer Saison spielt jede Mannschaft gegen jede andere Mannschaft.

a) Wie viele Spiele finden an einem Spieltag statt?

b) Wie viele Spieltage gibt es pro Saison?

c) Wie viele Spiele finden ingesamt in einer Saison statt?

Lösung:

a) Pro Spieltag finden $\frac{18}{2} = 9$ Spiele statt.

b) Es gibt $17 \cdot 2 = 34$ Spieltage.

c) Lösung aus (a) multipliziert mit Lösung aus (b) liefert $9 \cdot 34 = 306$ Spiele.
Oder:
$$2 \cdot \binom{18}{2} = 2 \cdot \frac{18!}{16!2!} = \frac{18!}{16!} = 18 \cdot 17 = 306$$

Aufgabe 2.13

a) 20 Personen verabschieden sich voneinander mit Händedruck. Jeder geht alleine nach Hause. Wie oft werden dabei die Hände gedrückt?

b) 15 Ehepaare verabschieden sich voneinander mit Händedruck und gehen paarweise nach Hause. Wie oft werden dabei die Hände gedrückt?

c) Die 15 Ehepaare verabschieden sich folgendermaßen: Die Herren von den Herren mit Händedruck, die Damen von den Damen mit Küsschen auf beide Wangen, die Damen von den Herren mit Händedruck und Küsschen auf die rechte Wange. Die Ehepaare gehen wieder paarweise nach Hause. Wie viele Küsschen werden gegeben? Wie oft werden die Hände gedrückt?

Lösung:

a) Ziehen ohne Zurücklegen, Reihenfolge nicht relevant (ungeordnet), $n = 20, k = 2$.
Je zwei der 20 Personen geben sich die Hand, also werden insgesamt

$$\binom{20}{2} = 190$$

Mal die Hände gedrückt.

b) Wenn sich zwei Ehepaare voneinander verabschieden, werden dabei viermal die Hände gedrückt. Es gibt 15 Ehepaare und

$$\binom{15}{2} = 105$$

Möglichkeiten, zwei Ehepaare (die sich voneinander verabschieden) auszuwählen. Folglich werden

$$4 \cdot \binom{15}{2} = 4 \cdot 105 = 420$$

Mal die Hände gedrückt.

c) Wenn sich zwei Ehepaare voneinander verabschieden, werden dabei dreimal die Hände gedrückt und viermal Wangenküsschen verteilt. Es gibt 15 Ehepaare und

$$\binom{15}{2} = 105$$

Möglichkeiten, zwei Ehepaare (die sich voneinander verabschieden) auszuwählen. Folglich werden

$$3 \cdot \binom{15}{2} = 3 \cdot 105 = 315$$

Mal die Hände gedrückt und

$$4 \cdot \binom{15}{2} = 4 \cdot 105 = 420$$

Mal Küsschen gegeben.

3 Elementare Datenanalyse

Aufgabe 3.1

Bezeichnen Sie für die folgenden Merkmale jeweils das Skalenniveau, auf dem das Merkmal gemessen wird, und nennen Sie mögliche Merkmalsausprägungen.

Geschlecht	Sparguthaben
Rendite von Wertpapieren	Kontonummer
Handelsklasse	Gewicht
Religionszugehörigkeit	Freizeitbeschäftigung
Geburtsdatum	Schulbildung

Lösung:

Merkmal	**Skala**	**Ausprägungen**	**Werteber.**
Geschlecht	Nominal	männlich, weiblich	diskret
Rendite	Absolut	Gewinn von x %, Verlust von x %	stetig
Handelsklasse	Ordinal	Güteklasse bei Eiern: A, B, C	diskret
Religionszugehörigkeit	Nominal	Römisch-Kathol., Islam, Evangel.	diskret
Geburtsdatum	Intervall	1.1.2000	diskret
Sparguthaben	Verhältnis	1000 DM, 70 Euro	(quasi)stetig
Kontonummer	Nominal	135498	diskret
Gewicht	Verhältnis	70 kg	stetig
Freizeitbeschäftigung	Nominal	Lesen, Sport, Kochen	diskret
Schulbildung	Ordinal	Abitur, Realschulabschluss	diskret

Aufgabe 3.2

Geben Sie bei den folgenden Größen an, ob es sich um eine Bestands- oder eine Bewegungsmasse handelt. Nennen Sie für jede Bestandsmasse zugehörige Bewegungsmassen.

Bevölkerung	Vermögen
Umsatz	Zinsen aus Vermögen
Anzahl der Arbeitslosen	Geburten und Sterbefälle
Autoproduktion	Zugelassene Kraftfahrzeuge
Investitionen	Einkommen

Lösung:

Bevölkerung	Bestandsmasse	Zuzüge - Fortzüge Geburten - Sterbefälle
Umsatz	Bewegungsmasse	
Anzahl Arbeitslose	Bestandsmasse	Entlassungen - Einstellungen
Autoproduktion	Bewegungsmasse	
Investitionen	Bewegungsmasse	
Anlagevermögen	Bestandsmasse	Zugänge - Abgänge Investitionen - Abschreibungen
Zinsen aus Vermögen	Bewegungsmasse	
Geburten und Sterbefälle	Bewegungsmasse	
Zugelassene Kraftfahrzeuge	Bestandsmasse	Zulassungen - Abmeldungen
Einkommen	Bewegungsmasse	

Aufgabe 3.3

Der Verband der Tierschützer befragt 40 Haushalte einer Reihenhaussiedlung nach der Anzahl der gehaltenen Haustiere. Dabei ergibt sich folgendes Ergebnis:

$$7, 1, 1, 2, 0, 2, 3, 2, 1, 0, 7, 0, 4, 1, 3, 1, 1, 1, 2, 0,$$
$$5, 1, 1, 0, 2, 0, 0, 1, 2, 1, 6, 1, 1, 0, 0, 1, 1, 0, 3, 8.$$

a) Geben Sie für die Anzahl der Haustiere die diskrete Klassierung mit absoluten und mit relativen Häufigkeiten an und stellen Sie sie graphisch dar.

b) Bestimmen Sie den Modus.

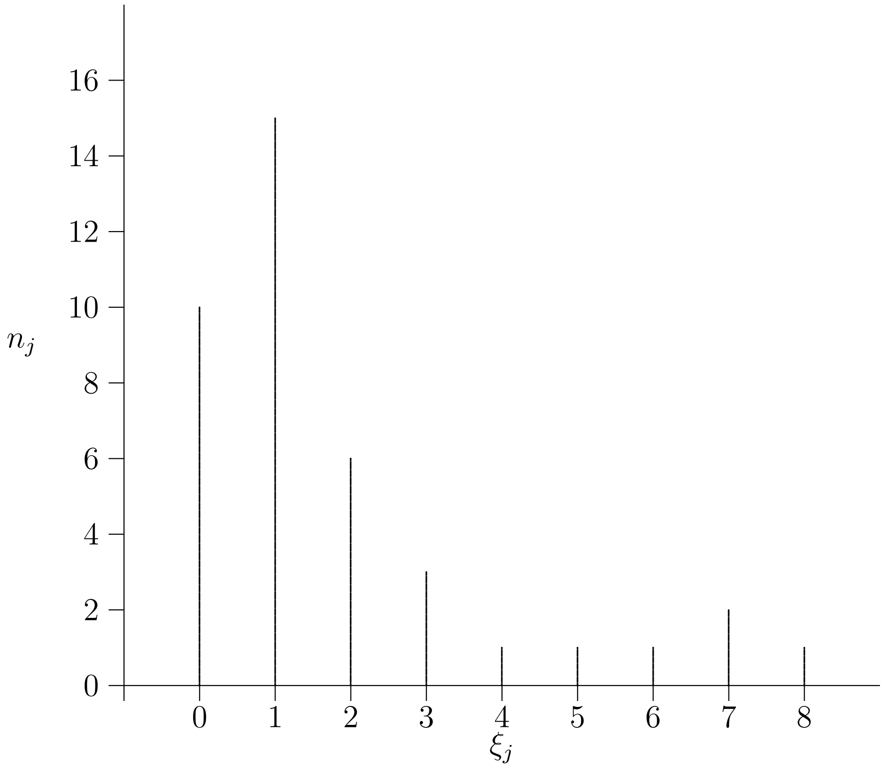
Lösung:

ξ_j : Anzahl der gehaltenen Haustiere
n_j : Anzahl der Haushalte
f_j : Anteil der Haushalte

a) Diskrete Klassierung mit absoluten und relativen Häufigkeiten:

j	ξ_j	n_j	f_j
1	0	10	0.2500
2	1	15	0.3750
3	2	6	0.1500
4	3	3	0.0750
5	4	1	0.0250
6	5	1	0.0250
7	6	1	0.0250
8	7	2	0.0500
9	8	1	0.0250
$\sum$		40	1

Liniendiagramm:



Alternativ ist auch ein Säulendiagramm zur Darstellung geeignet.

b) Gesucht: Modus. $\xi_2 = 1$ tritt am häufigsten auf.

Aufgabe 3.4

In der Betriebskantine werden verschiedene Mittagessen angeboten. Zur Wahl stehen: ein vegetarisches Essen (VE), "Fast Food" (FF), ein Eintopf (E) und ein Komponentenessen (KE). Im letzen Monat wurden von den Mitarbeitern insgesamt 500 Essen wie folgt bestellt:

Mittagessen	VE	FF	E	KE
Anzahl der Mitarbeiter	85	135	92	188

a) Bezeichnen Sie die statistischen Begriffe, die den folgenden Angaben in der Aufgabenstellung entsprechen:

(1) die Zahl 500,

(2) das Essen "Eintopf",

(3) die Zahl 85,

(4) die Gesamtheit der Paare (VE , 85), (FF , 135), (E , 92), (KE , 188)?

b) Berechnen Sie die relativen Häufigkeiten und stellen Sie sie graphisch dar.

Lösung:

a) (1) Gesamtanzahl der Beobachtungen n

(2) eine Merkmalsausprägung ξ_j

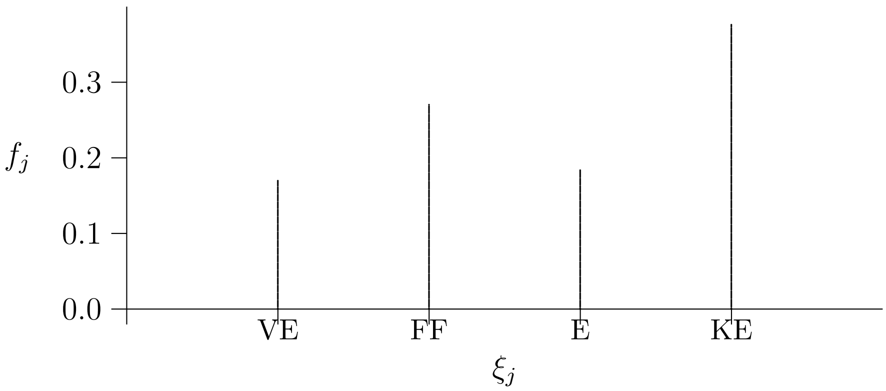
(3) absolute Häufigkeit der Merkmalsausprägung " VE " n_j

(4) diskrete Klassierung / Gruppierung (ξ_j, n_j)

b)

j	ξ_j	n_j	f_j
1	VE	85	0.1700
2	FF	135	0.2700
3	E	92	0.1840
4	KE	188	0.3760
$\sum$		500	1

Liniendiagramm:



Aufgabe 3.5

Zehn Untersuchungseinheiten wurden aus einer bestimmten Grundgesamtheit gezogen und jeweils das Merkmal X gemessen

$$x_1 = 3,\ x_2 = 1,\ x_3 = 4,\ x_4 = 5,\ x_5 = 2,\ x_6 = 6,\ x_7 = 3,\ x_8 = 4,\ x_9 = 1,\ x_{10} = 5.$$

a) Zeichnen Sie die empirische Verteilungsfunktion.

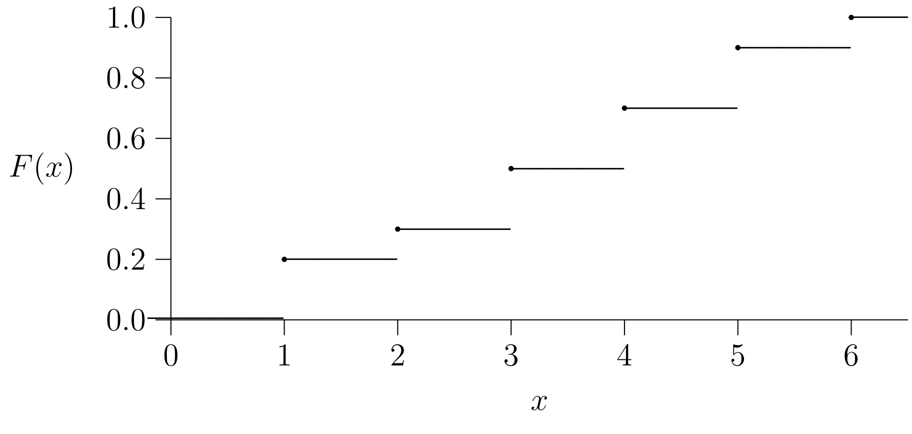
b) Lassen sich aus einer empirischen Verteilungsfunktion die absoluten Häufigkeiten der Merkmalswerte ermitteln?

Lösung:

a)

j	ξ_j	n_j	f_j	$F(\xi_j)$
1	1	2	0.2	0.2
2	2	1	0.1	0.3
3	3	2	0.2	0.5
4	4	2	0.2	0.7
5	5	2	0.2	0.9
6	6	1	0.1	1.0
$\sum$	—	10	1	—

Empirische Verteilungsfunktion:



b) Die auftretenden Merkmalswerte liegen genau an den Sprungstellen der empirischen Verteilungsfunktion $F(x)$. Die Höhe einer Sprungstelle ist gleich der relativen Häufigkeit f_j des entsprechenden Merkmalswertes. Nur falls, wie hier, die Gesamtanzahl n der gemessenen Werte bekannt ist, lassen sich daraus die absoluten Häufigkeiten ermitteln, nämlich mittels $n_j = n \cdot f_j$.

Aufgabe 3.6

Bei einer Abschlussprüfung sind 18 Aufgaben zu bearbeiten, wobei pro Aufgabe ein Punkt erzielt werden kann. Als „nicht bestanden“ gilt die Prüfung, wenn ein Kandidat weniger als sieben Punkte erreicht. Bei sieben bis vierzehn Punkten wird die Note „bestanden“ vergeben, bei mehr als vierzehn Punkten die Note „mit besonderem Erfolg bestanden“.

Die Korrektur der Klausur ergab folgende Häufigkeitsverteilung der erreichten Punktzahlen ξ_j:

ξ_j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
n_j	4	1	1	0	7	5	6	4	7	7	17	22	13	16	1	5	2	2	0

a) Errechnen Sie die diskrete Klassierung gemäß den vergebenen Noten und stellen Sie sie graphisch dar.

b) Inwiefern ist die diskrete Klassierung nach Noten mehr oder weniger informativ als die Klassierung nach Punktzahlen?

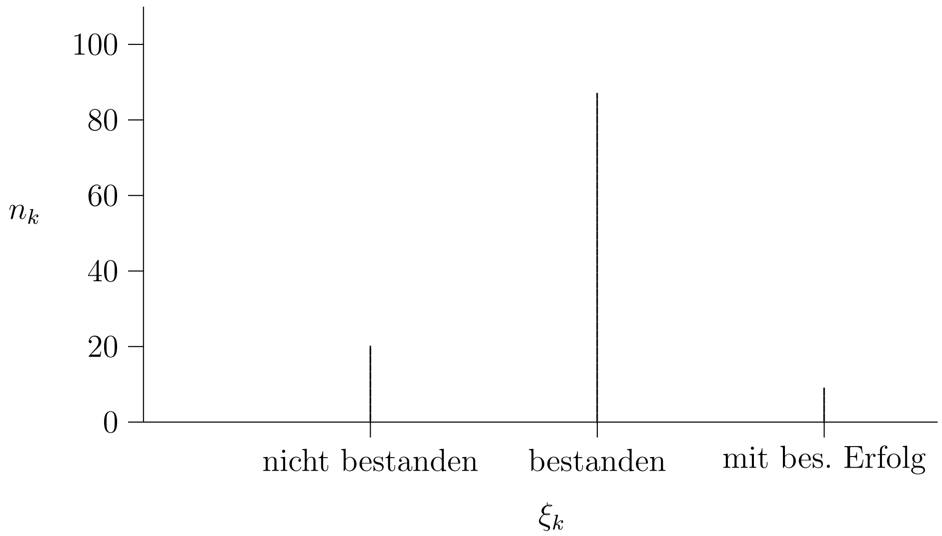
Lösung:

a) Merkmal ξ_k mit Ausprägungen: „nicht bestanden“, „bestanden“ und „mit besonderem Erfolg bestanden“.

Tabelle:

ξ_k	Punktzahlen	n_k	f_k
nicht bestanden	0 - 6	24	$\frac{24}{120} = 0,2$
bestanden	7 - 14	87	$\frac{87}{120} = 0,725$
mit bes. Erfolg	15 - 18	9	$\frac{9}{120} = 0,075$
Σ		120 $(= n)$	1

Liniendiagramm:



b) Durch die Zusammenfassung der Punktzahlen zu Noten gehen einserseits Informationen verloren: Die Abstufung zwischen den einzelnen Kandidaten ist gröber und ein Rückschluss auf die Punktzahlen ist nicht mehr möglich. (Der umgekehrte Weg bei Kenntnis des Notenschemas dagegen schon.) Andererseits werden die Daten durch die Klassierung nach Noten durch eine zusätzliche Information, die Bewertung der Leistung durch den Prüfer, angereichert. Außerdem wird die Menge der Daten reduziert, weniger relevante Informationen fallen weg, so dass Datennutzer in der Lage sind, die relevanten Informationen schneller aufzunehmen.

Aufgabe 3.7

In der Kölner Fußgängerzone wurden 60 zufällig ausgewählte Erwachsene nach ihrem höchsten Schul- oder Studienabschluss befragt. Dabei wurden die folgenden Antworten registriert:

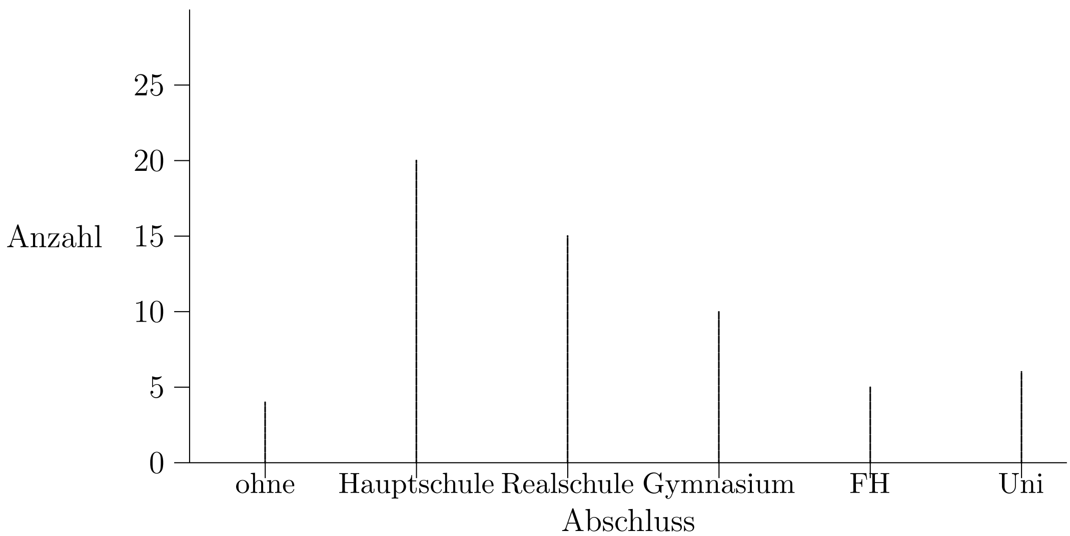
Abschluss	Häufigkeit
ohne	4
Hauptschule	20
Realschule	15
Gymnasium	10
Fachhochschule	5
Universität	6

a) Stellen Sie die diskrete Klassierung graphisch dar.

b) Geben Sie die relativen Häufigkeiten an.

Lösung:

a) Liniendiagramm:



b)

j	ξ_j	n_j	f_j
1	ohne	4	0.0667
2	Hauptschule	20	0.3333
3	Realschule	15	0.2500
4	Gymnasium	10	0.1667
5	Fachhochschule	5	0.0833
6	Universität	6	0.1000
$\sum$	—	60	1.0000

Aufgabe 3.8

Ein Marktforschungsinstitut befragt 40 Kunden eines Teeladens nach der Anzahl der von ihnen durchschnittlich pro Tag getrunkenen Tassen Tee. Dabei ergaben sich die folgenden Zahlen:

Anzahl der Tassen	1	2	3	4	5	6
Anzahl der Kunden	4	8	15	6	5	2

a) Wie hoch ist der Anteil der Kunden, die höchstens 4 Tassen pro Tag trinken?

b) Wie hoch ist der Anteil der Befragten, die mindestens 3 und höchstens 5 Tassen pro Tag trinken?

c) Welche Anzahl an Tassen pro Tag wird von 95% der Kunden nicht überschritten?

d) Bestimmen Sie den Median. Wie verändert sich der Median, wenn einer der Kunden statt 2 Tassen Tee pro Tag 7 Tassen angeben würde?

Lösung:

ξ_j : Anzahl der Tassen
n_j : Anzahl der Kunden

j	ξ_j	n_j	f_j	$F(\xi_j)$
1	1	4	0.100	0.100
2	2	8	0.200	0.300
3	3	15	0.375	0.675
4	4	6	0.150	0.825
5	5	5	0.125	0.950
6	6	2	0.050	1.000
$\sum$	—	40	1	—

a) 4 oder weniger: $F(4) = 0.825 = 82.5\%$

b) 3, 4 oder 5: $f(3) + f(4) + f(5) = F(5) - F(2) = 0.65 = 65\%$

c) 95%-Quantil: $F(\xi_4) < 0.95 \leq F(\xi_5)$, deshalb $\tilde{x}_{0,95} = \xi_5 = 5$

d) 50%-Quantil: $F(\xi_2) < 0.5 \leq F(\xi_3)$, deshalb $\tilde{x}_{0,5} = \xi_3 = 3$.

Wenn ein Kunde statt 2 Tassen 7 Tassen trinken würde, würden sich die Funktionswerte $F(x)$ von $x = 2$ bis $x = 6$ um $\frac{1}{40} = 0.025$ verringern, d.h.

j	ξ_j	n_j	f_j	$F(\xi_j)$
1	1	4	0.100	0.100
2	2	7	0.175	0.275
3	3	15	0.375	0.650
4	4	6	0.150	0.800
5	5	5	0.125	0.925
6	6	2	0.050	0.975
7	7	1	0.025	1.000
$\sum$	—	40	1	—

Der Median wäre weiterhin $\tilde{x}_{0,5} = 3$.

Aufgabe 3.9

Für die 25 Tage eines Monats, an dem ein Einzelhändler sein Geschäft geöffnet hatte, wurden folgende Verkaufszahlen eines bestimmten Artikels registriert:

an 3 Tagen:	0 Stück ,
an 2 Tagen:	1 Stück ,
an 4 Tagen:	2 Stück ,
an 3 Tagen:	3 Stück ,
an 8 Tagen:	4 Stück ,
an 4 Tagen:	6 Stück ,
an 1 Tagen:	8 Stück .

a) Geben Sie die Grundgesamtheit und das Merkmal an.

b) Stellen Sie die Daten in einer Häufigkeitstabelle mit absoluten und relativen Häufigkeiten dar.

c) Geben Sie die empirische Verteilungsfunktion explizit an und stellen Sie diese graphisch dar.

d) Betrachten Sie jeweils den Anteil der Tage, an denen

 (1) höchstens 4,
 (2) weniger als 4,
 (3) höchstens 2.5,
 (4) mehr als 2.5,
 (5) mindestens 2,
 (6) mehr als 2 und höchstens 6.3

 Stück verkauft wurden. Drücken Sie diese Anteile durch die empirische Verteilungsfunktion aus und berechnen Sie diese.

e) Berechnen Sie den Median und interpretieren Sie ihn.

f) Finden Sie das kleinste x, für das $F(x) \geq 0.3$ ist.

Lösung:

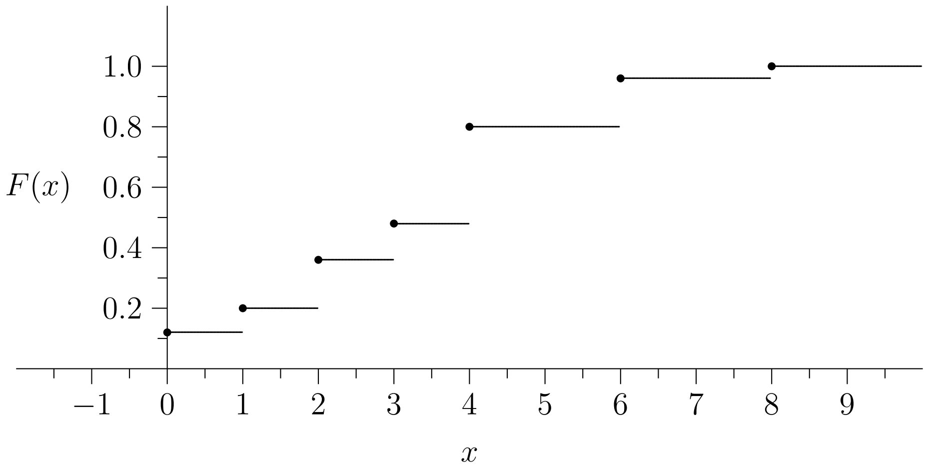
a) Grundgesamtheit: Die 25 Verkaufstage des betrachteten Monats.
Merkmal: Die Verkaufszahl des betrachteten Artikels.

b)

Merkmalswert	abs. Häufigkeit	rel. Häufigkeit
0	3	0.12
1	2	0.08
2	4	0.16
3	3	0.12
4	8	0.32
6	4	0.16
8	1	0.04
$\sum$	25	1.00

c)

$$F(x) = \begin{cases} 0 & \text{für } x < 0, \\ 0.12 & \text{für } 0 \leq x < 1, \\ 0.20 & \text{für } 1 \leq x < 2, \\ 0.36 & \text{für } 2 \leq x < 3, \\ 0.48 & \text{für } 3 \leq x < 4, \\ 0.80 & \text{für } 4 \leq x < 6, \\ 0.96 & \text{für } 6 \leq x < 8, \\ 1.00 & \text{für } x \geq 8. \end{cases}$$



d) (1) $F(4) = 0.80$.

(2) $F(4-) = \lim_{x\to 4, x<4} F(x) = 0.48$ bzw. auch $F(3) = 0.48$.

(3) $F(2.5) = 0.36$.

(4) $1 - F(2.5) = 0.64$.

(5) $1 - F(2-) = 0.8$ bzw. auch $1 - F(1) = 0.8$.

(6) $F(6.3) - F(2) = 0.96 - 0.36 = 0.6$.

e) Gesucht ist das kleinste x, für das die empirische Verteilungsfunktion $F(x) \geq 0.5$ ist. Da $F(3) < 0.5$ und $F(4) > 0.5$, liegt der Median bei $\tilde{x} = 4$. An jenen 50% der Verkaufstage mit den niedrigsten Verkaufszahlen wurden nicht mehr als 4 Stück pro Tag verkauft.

f) Das gesuchte x ist $x = 2$ .Es handelt sich um das 30%-Quantil.

Aufgabe 3.10

Die Umfrage des Tierschützerverbandes in der Reihenhaussiedlung aus Aufgabe 3.3 ergab folgendes Ergebnis:

a) Berechnen Sie die empirische Verteilungsfunktion und stellen Sie sie graphisch dar.

b) Bestimmen Sie für die Anzahl der Haustiere rechnerisch und graphisch den Median und das 0.7–Quantil und interpretieren Sie diese Werte.

c) Welche Haustieranzahl wird nur von 10% der Befragten überschritten?

Lösung:

ξ_j : Anzahl der gehaltenen Haustiere
n_j : Anzahl der Haushalte
f_j : Anteil der Haushalte

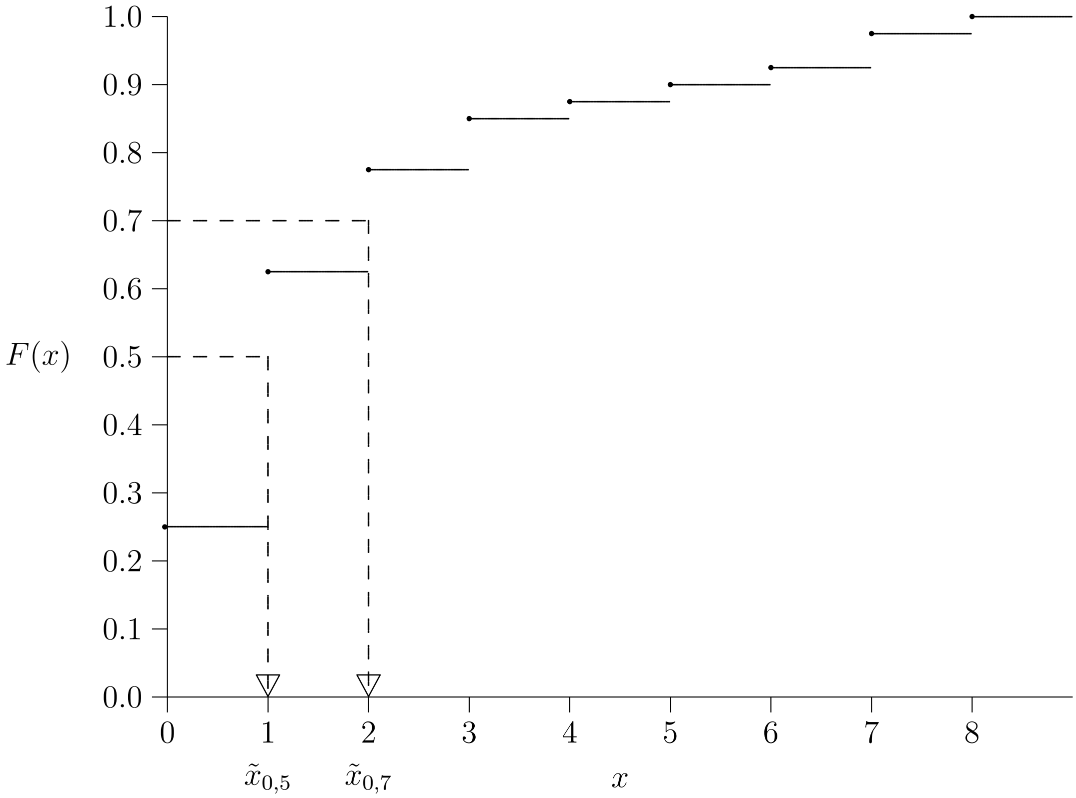
a) Empirische Verteilungsfunktion $F(x) = \sum_{\{r|\xi_j \leq x\}} f_r$

j	ξ_j	n_j	f_j	$F(\xi_j)$
1	0	10	0.250	0.250
2	1	15	0.375	0.625
3	2	6	0.150	0.775
4	3	3	0.075	0.850
5	4	1	0.025	0.875
6	5	1	0.025	0.900
7	6	1	0.025	0.925
8	7	2	0.050	0.975
9	8	1	0.025	1.000
$\sum$		40	1.000	

Die f_j sind aus Aufgabe 2.1 bekannt.

Die Verteilungsfunktion ist auf der ganzen Zahlengeraden definiert:

$$F(x) = \begin{cases} 0 & \text{für } x < 0, \\ 0.25 & \text{für } 0 \leq x < 1, \\ 0.625 & \text{für } 1 \leq x < 2, \\ 0.775 & \text{für } 2 \leq x < 3, \\ 0.85 & \text{für } 3 \leq x < 4, \\ 0.875 & \text{für } 4 \leq x < 5, \\ 0.9 & \text{für } 5 \leq x < 6, \\ 0.925 & \text{für } 6 \leq x < 7, \\ 0.975 & \text{für } 7 \leq x < 8, \\ 1 & \text{für } x \geq 8. \end{cases}$$



b) (1) Median $\tilde{x}_{0.5}$:

$$\begin{aligned} \tilde{x}_{0.5} &= \xi_2 = 1, \\ \text{denn } F(\xi_1) &= 0.25 < 0.5 \\ \text{und } F(\xi_2) &= 0.625 \geq 0.5 \end{aligned}$$

(2) $\tilde{x}_{0.7}$:

$$\begin{aligned} \tilde{x}_{0.7} &= \xi_3 = 2, \\ \text{denn } F(\xi_2) &= 0.625 < 0.7 \\ \text{und } F(\xi_3) &= 0.775 \geq 0.7 \end{aligned}$$

Interpretation:
Median: Mindestens 50 % der Haushalte haben höchstens 1 Haustier.
$\tilde{x}_{0.7}$: Mindestens 70 % der Haushalte haben höchstens 2 Haustiere.

c) Gesucht ist $\tilde{x}_{0,9}$ derart, dass $F(\tilde{x}_{0,9}) \geq 0.9$.
Tabelle (oder Zeichnung) liefert $\tilde{x}_{0,9} = \xi_6 = 5$.

Aufgabe 3.11

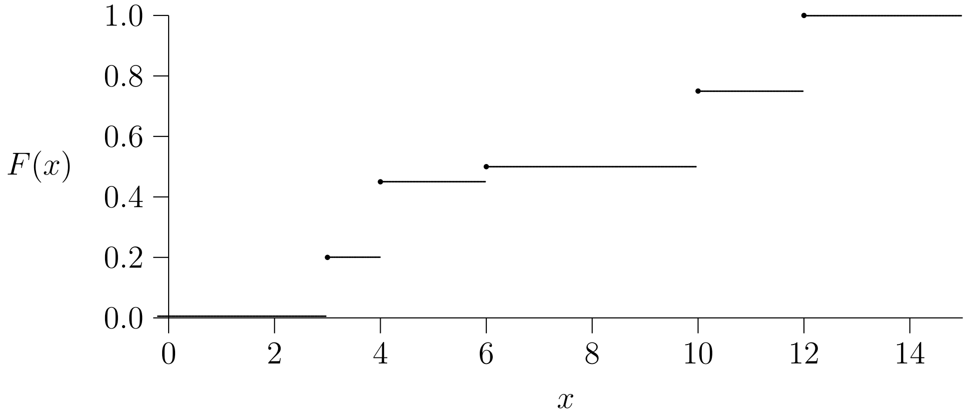
Bei einer statistischen Untersuchung wurden für ein metrisches Merkmal X die Daten $x_1, x_2, \ldots, x_{100}$ gemessen und daraus die empirische Verteilungsfunktion berechnet:

$$F(x) = \begin{cases} 0 & \text{für } x < 3\,, \\ 0.20 & \text{für } 3 \leq x < 4\,, \\ 0.45 & \text{für } 4 \leq x < 6\,, \\ 0.50 & \text{für } 6 \leq x < 10\,, \\ 0.75 & \text{für } 10 \leq x < 12\,, \\ 1 & \text{für } x \geq 12\,. \end{cases}$$

Stellen Sie die empirische Verteilungsfunktion graphisch dar und bestimmen Sie anhand der Graphik das untere und obere Quartil.

Lösung:

Empirische Verteilungsfunktion:



An der Stelle $x = 4$ überschreitet die Verteilungsfunktion erstmals den Wert 0.25 und an der Stelle $x = 10$ erstmals den Wert 0.75. Daher erhält man für das untere Quartil $\tilde{x}_{0.25} = 4$ und für das obere Quartil $\tilde{x}_{0.75} = 10$.

Aufgabe 3.12

Eine Autozeitschrift will zwei neue Kleinwagen vergleichen. Von Interesse ist unter anderem der Benzinverbrauch auf 100 km. Mit Wagen A werden 50 Fahrten gemacht. Das Ergebnis der Messungen zeigt die folgende Tabelle:

Benzinverbrauch (in Litern je 100 km)	Anteil der Fahrten
5.0	0.02
5.5	0.06
6.0	0.16
6.4	0.30
6.9	0.34
7.5	0.08
7.8	0.04

Da der Redaktionsschluss naht, können mit Wagen B nur noch 20 Fahrten unternommen werden. Die Testergebnisse sind die folgenden:

$$4.2, 4.8, 4.8, 4.8, 5.4, 5.4, 5.4, 5.4, 5.4, 5.4,$$
$$5.9, 5.9, 5.9, 5.9, 5.9, 5.9, 6.0, 6.0, 6.0, 6.5.$$

a) Berechnen Sie für Wagen B die relativen Häufigkeiten und stellen Sie sie graphisch dar.

b) Bestimmen Sie für den Benzinverbrauch bei Wagen A und B jeweils den Median.

c) Auf wie vielen Fahrten haben Wagen A bzw. Wagen B weniger als 6.5 Liter verbraucht?

Lösung:

a) Benzinverbrauch [in Litern je 100 km]

Wagen A			Wagen B		
ξ_j	f_j	$F(\xi_j)$	ξ_j	f_j	$F(\xi_j)$
5.0	0.02	0.02	4.2	0.05	0.05
5.5	0.06	0.08	4.8	0.15	0.20
6.0	0.16	0.24	5.4	0.30	0.50
6.4	0.30	0.54	5.9	0.30	0.80
6.9	0.34	0.88	6.0	0.15	0.95
7.5	0.08	0.96	6.5	0.05	1.00
7.8	0.04	1.00			
$\sum$	1	—	$\sum$	1	—

Stabdiagramm!

b) Wagen A: $\tilde{x}_{0,5} = 6.4$ [l/100km]
Wagen B: $\tilde{x}_{0,5} = 5.4$ [l/100km]

c) Weniger als 6.5 [l/100km] haben verbraucht

Wagen A: auf $F_A(6.4) \cdot n_A = 50 \cdot 0.54 = 27$ Fahrten,
Wagen B: auf $F_B(6.0) \cdot n_B = 20 \cdot 0.95 = 19$ Fahrten.

Aufgabe 3.13

Der Verband der Tierschützer, der auch die Umfrage in der Reihenhaussiedlung gemacht hat (vgl. Aufgabe 3.3 und 3.10), befragt zusätzlich die 20 Parteien eines Hochhauses nach der Anzahl der Haustiere. Dabei ergibt sich folgendes Ergebnis:

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

a) Berechnen Sie für die Anzahl der Haustiere den Median, das 1. und 3. Quartil und das 0.33–Quantil.

b) Zeichnen Sie den zugehörigen Boxplot.

Lösung:

Das Merkmal X bezeichnet die Anzahl der Haustiere, Es liegen $n = 20$ Einzeldaten vor.

a) Gesucht: $\tilde{x}_{0.5}$, $\tilde{x}_{0.25}$, $\tilde{x}_{0.75}$, $\tilde{x}_{0.33}$ mit

$$\tilde{x}_p = \begin{cases} x_{np}\,, & \text{falls } np \text{ ganzzahlig,} \\ x_{\lfloor np \rfloor + 1} & \text{sonst.} \end{cases}$$

Aufsteigend geordnete Daten:

i	x_i(geordnet)		i	x_i(geordnet)	
1	0		11	1	
2	0		12	1	
3	0		13	1	
4	0		14	1	
5	0	$\tilde{x}_{0.25}$	15	2	$\tilde{x}_{0.75}$
6	0		16	2	
7	0	$\tilde{x}_{0.33}$	17	3	
8	0		18	4	
9	1		19	5	
10	1	Median	20	6	

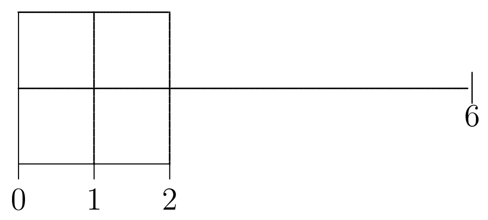
$\tilde{x}_{0.5}$: $n \cdot p = 20 \cdot 0.5 = 10$, d.h. $\tilde{x}_{0.5} = x_{10} = 1$
$\tilde{x}_{0.25}$: $n \cdot p = 20 \cdot 0.25 = 5$, d.h. $\tilde{x}_{0.25} = x_5 = 0$
$\tilde{x}_{0.75}$: $n \cdot p = 20 \cdot 0.75 = 15$, d.h. $\tilde{x}_{0.75} = x_{15} = 2$
$\tilde{x}_{0.33}$: $n \cdot p = 20 \cdot 0.33 = 6.6$, $\lfloor 6.6 \rfloor = 6$, d.h. $\tilde{x}_{0.33} = x_{\lfloor 6.6 \rfloor + 1} = x_7 = 0$

Alternative Lösung: Werte an den Werten der Verteilungsfunktion ablesen: $\tilde{x}_p$ ist der Wert, so dass erstmalig $F(\tilde{x}_p) \geq p$

ξ_j	n_j	f_j	$F(\xi_j)$
0	8	0.40	0.40
1	6	0.30	0.70
2	2	0.10	0.80
3	1	0.05	0.85
4	1	0.05	0.90
5	1	0.05	0.95
6	1	0.05	1
$\sum$	20	1	—

Man liest ab: $\tilde{x}_{0.25} = 0$; $\tilde{x}_{0.33} = 0$; $\tilde{x}_{0.50} = 1$; $\tilde{x}_{0.75} = 2$

b) Boxplot:
$\min_{i=1,\ldots,20} x_i = 0$, $\max_{i=1,\ldots,20} x_i = 6$
$\tilde{x}_{0.25} = 0$, $\tilde{x}_{0.5} = 1$, $\tilde{x}_{0.75} = 2$



Aufgabe 3.14

In einer Automobilfabrik wurden die Höchstgeschwindigkeiten von 400 Kraftfahrzeugen eines bestimmten Typs gemessen. Dabei ergaben sich folgende Messergebnisse:

Höchstgeschwindigkeit (in km/h)	absolute Häufigkeit
[135,140]	18
]140,142]	38
]142,144]	82
]144,146]	105
]146,148]	89
]148,150]	46
]150,155]	22

a) Bestimmen Sie die relativen Häufgkeiten der stetigen Klassierung.

b) Stellen Sie die Daten anhand eines Histogramms graphisch dar.

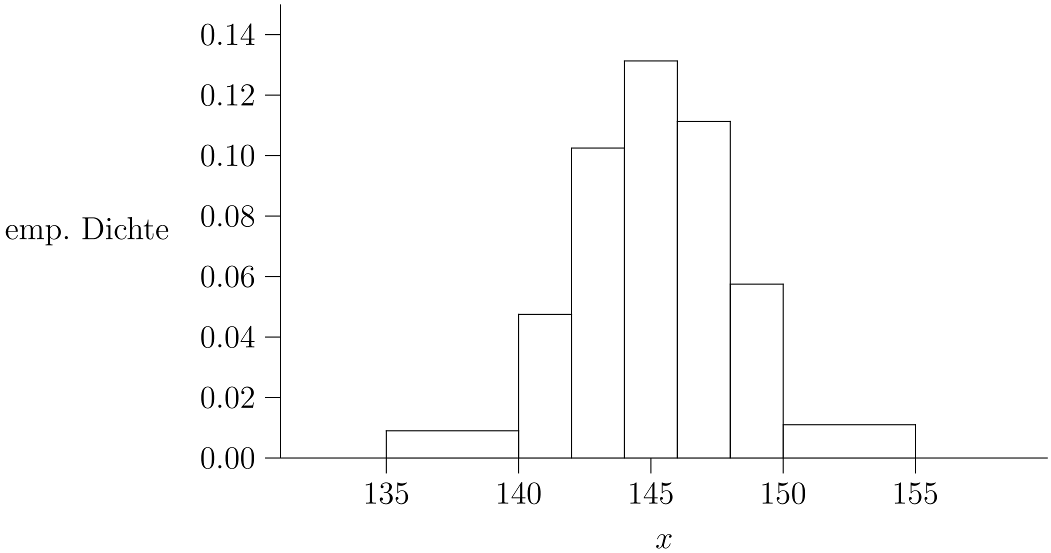
c) Zeichnen Sie die empirische Verteilungsfunktion.

Lösung:

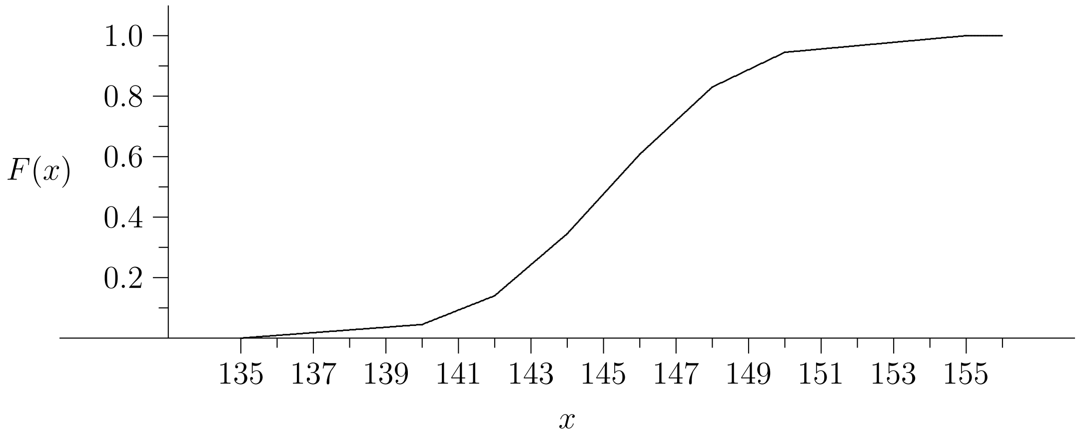
a) Relative Häufigkeiten f_j und empirische Dichten $\frac{f_j}{x_j^o - x_j^u}$:

j	K_j	n_j	f_j	$\frac{f_j}{x_j^o - x_j^u}$	$F(x_j^o)$
1	[135,140]	18	0.0450	0.0090	0.0450
2	]140,142]	38	0.0950	0.0475	0.1400
3	]142,144]	82	0.2050	0.1025	0.3450
4	]144,146]	105	0.2625	0.1313	0.6075
5	]146,148]	89	0.2225	0.1113	0.8300
6	]148,150]	46	0.1150	0.0575	0.9450
7	]150,155]	22	0.0550	0.0110	1
$\sum$	—	400	1	—	—

b) Histogramm:



b) Empirische Verteilungsfunktion:



Aufgabe 3.15

Gegeben ist die Bevölkerung Deutschlands nach Altersklassen im Jahr 1997:

Altersklasse (in Jahren)	Bevölkerung in Millionen östliche Bundesländer	westliche Bundesländer
0 - 15	2.28	10.82
15 - 21	1.28	4.17
21 - 40	4.21	19.45
40 - 60	4.26	17.65
60 - 65	0.99	3.97
65 - ...	2.35	10.62

a) In welchen Bundesländern ist der Anteil der Menschen bis 18 Jahre höher? In welchen Bundesländern ist der Anteil der Menschen, die über 62 Jahre alt sind, höher?

b) Zeichnen Sie die empirische Verteilungsfunktion $F(x)$ für die östlichen Bundesländer.

c) Zeichnen Sie für die Altersverteilung der östlichen Bundesländer ein Histogramm.

Hinweis: Nehmen Sie an, dass die Obergrenze der letzten Klasse $x_6^o = 100$ ist.

Lösung:

X : Alter in Jahren

a) Stetige Klassierung

$$F(x) \approx F(x_j^u) + \frac{f_j}{x_j^o - x_j^u}(x - x_j^u)$$

Östliche Bundesländer:

$]x_j^u, x_j^o]$	n_j	f_j	$F(x_j^o)$
$0\,,15$	2.28	0.1483	0.1483
$15\,,21$	1.28	0.0833	0.2316
$21\,,40$	4.21	0.2739	0.5055
$40\,,60$	4.26	0.2772	0.7827
$60\,,65$	0.99	0.0644	0.8471
$65\,\ldots$	2.35	0.1529	1.0000
$\sum$	15.37	1.0000	—

Westliche Bundesländer:

$]x_j^u, x_j^o]$	n_j	f_j	$F(x_j^o)$
0 , 15	10.82	0.1623	0.1623
15 , 21	4.17	0.0625	0.2248
21 , 40	19.45	0.2917	0.5165
40 , 60	17.65	0.2647	0.7812
60 , 65	3.97	0.0595	0.8407
65 ...	10.62	0.1593	1.0000
$\sum$	66.68	1.0000	—

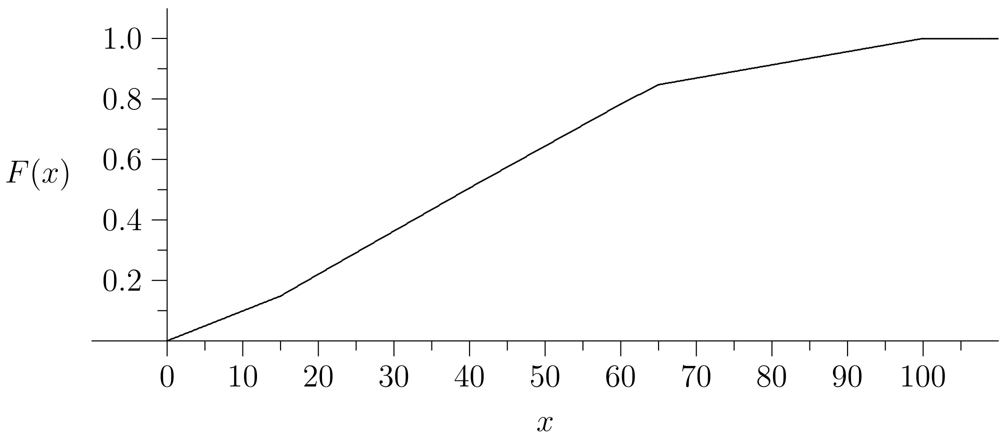
$$F_{ost}(18) \approx 0.1483 + \frac{0.0833}{21-15} \cdot (18-15) = 0.1900\,,$$
$$F_{west}(18) \approx 0.1623 + \frac{0.0625}{21-15} \cdot (18-15) = 0.1936\,.$$

Der Anteil der Menschen, die nicht älter als 18 Jahre sind, ist also in den westlichen Bundesländern etwas höher.

$$1 - F_{ost}(62) \approx 1 - [0.7827 + \frac{0.0644}{65-60} \cdot (62-60)] = 1 - 0.8085 = 0.1915\,,$$
$$1 - F_{west}(62) \approx 1 - [0.7812 + \frac{0.0595}{65-60} \cdot (62-60)] = 1 - 0.8050 = 0.1950\,.$$

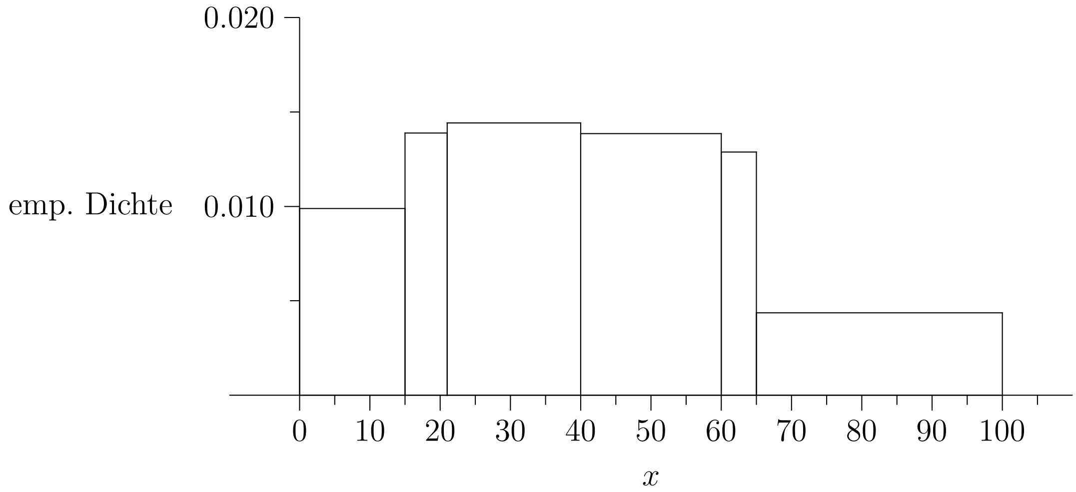
Der Anteil der Menschen, die älter als 62 Jahre sind, ist also ebenfalls in den westlichen Bundesländern etwas höher.

b) Empirische Verteilungsfunktion der Altersverteilung in den östlichen Bundesländern:



c) Histogramm der Altersverteilung der östlichen Bundesländer:
Arbeitstabelle:

j	x_j^u	x_j^o	n_j	f_j	$\frac{f_j}{b_j}$
1	0	15	2,28	0,1483	0,0099
2	15	21	1,28	0,0833	0,0139
3	21	40	4,21	0,2739	0,0144
4	40	60	4,26	0,2772	0,0139
5	60	65	0,99	0,0644	0,0129
6	65	100	2,35	0,1529	0,0044
$\sum$			15,37	1	—



Aufgabe 3.16

In einer Gemeinde wurde die Haushaltsgröße der privaten Haushalte ermittelt. Dabei ergaben sich die folgenden Daten:

a) Berechnen Sie die durchschnittliche Haushaltsgröße in der Gemeinde.

b) Bestimmen Sie die Quartile der Haushaltsgröße.

Lösung:

ξ_j: Anzahl der Personen
n_j : Anzahl der Haushalte

ξ_j	n_j	$\xi_j n_j$	f_j	$F(\xi_j)$
1	1 319	1 319	0.3540	0.3540
2	1 203	2 406	0.3229	0.6769
3	577	1 731	0.1549	0.8318
4	455	1 820	0.1221	0.9539
5	119	595	0.0319	0.9858
6	32	192	0.0086	0.9944
7	12	84	0.0032	0.9976
8	5	40	0.0013	0.9989
9	3	27	0.0008	0.9997
10	1	10	0.0003	1.0000
$\sum$	3 726	8 224	1.000	—

a) Gesucht ist das arithmetische Mittel $\overline{x} = \frac{1}{n}\sum_{j=1}^{J} \xi_j n_j$. Die durchschnittliche Haushaltsgröße beträgt $\overline{x} = \frac{8\,224}{3\,726} = 2,2072$ Personen.

b) Der Median $\tilde{x}_{0,5}$ hat den Wert 2, da $F(1) < 0.5$ und $F(2) \geq 0.5$.
Das untere Quartil $\tilde{x}_{0,25}$ hat den Wert 1, da $F(0) < 0.25$ und $F(1) \geq 0.25$.
Das obere Quartil $\tilde{x}_{0,75}$ hat den Wert 3, da $F(2) < 0.75$ und $F(3) \geq 0.75$.

Aufgabe 3.17

Aus einer Urliste von Daten wurde die folgende empirische Verteilungsfunktion berechnet:

$$F(x) = \begin{cases} 0 & \text{für } x < 2 \ , \\ 0.15 & \text{für } 2 \leq x < 3 \ , \\ 0.35 & \text{für } 3 \leq x < 5 \ , \\ 0.60 & \text{für } 5 \leq x < 8 \ , \\ 0.80 & \text{für } 8 \leq x < 10 \ , \\ 1 & \text{für } x \geq 10 \ . \end{cases}$$

a) Bestimmen Sie den Median der Daten.

b) Berechnen Sie das arithmetische Mittel.

c) Was können Sie über die Anzahl der Daten in der Urliste sagen?

Lösung:

Arbeitstabelle:

j	ξ_j	$F(\xi_j)$	f_j	$\xi_j \cdot f_j$
1	2	0.15	0.15	0.30
2	3	0.35	0.20	0.60
3	5	0.60	0.25	1.25
4	8	0.80	0.20	1.60
5	10	1.00	0.20	2.00
$\sum$	—	—	1.00	5.75

a) Gesucht ist das kleinste x, für das $F(x) \geq 0.5$ ist. Somit ist $\tilde{x}_{0.5} = \xi_3 = 5$.

b) $\overline{x} = \dfrac{1}{n} \sum\limits_{j=1}^{J} \xi_j n_j = \sum\limits_{j=1}^{J} \xi_j f_j = 5.75$

c) Jede Sprunghöhe, d.h. jede relative Häufigkeit, ist ein ganzzahliges Vielfaches von 0.05. Daher ist die Anzahl der Daten in der Urliste ein ganzzahliges Vielfaches von 20.

Aufgabe 3.18

An zwei Registrierkassen eines Supermarktes wurden eine Stunde lang jeweils die Bedienungszeiten (in Sekunden) gemessen:

Kasse 1	35	45	15	36	68	75	12	9	35	23	45	25	28	67	46					
Kasse 2	76	21	49	63	47	48	69	62	52	41	68	79	45	32	11	12	16	45	23	7

a) Berechnen Sie die arithmetischen Mittel der Bedienungszeiten für die zwei Kassen getrennt sowie für beide Kassen insgesamt.

b) Bestimmen Sie für jede Kasse den Median sowie das untere und obere Quartil der Bedienungszeiten. Zeichnen Sie für beide Kassen die Boxplots der Bedienungszeiten.

Lösung:

X = Bedienungszeit in Sekunden

a) Kasse 1: $n = 15\,,\ \overline{x} = 37.6$ [Sekunden]
Kasse 2: $n = 20\,,\ \overline{x} = 43.3$ [Sekunden]
Kasse 1 und 2 zusammen: $n = 35$, $\overline{x} = 40.9$ [Sekunden]

b) Kasse 1: $\tilde{x}_{0.25} = 23$, $\tilde{x}_{0.5} = 35$, $\tilde{x}_{0.75} = 46$
Kasse 2: $\tilde{x}_{0.25} = 21$, $\tilde{x}_{0.5} = 45$, $\tilde{x}_{0.75} = 62$
Boxplots!

Aufgabe 3.19

Ein Haus wird von zehn Personen bewohnt. Fünf dieser Personen haben ein Monatseinkommen von je 2 500 Euro, die übrigen Personen haben Monatseinkommen von 2 600, 2 700, 2 800, 2 900 und 3 000 Euro. In das Haus zieht eine weitere Person ein, deren Monatseinkommen 100 000 Euro beträgt. Welche Auswirkungen ergeben sich dadurch auf den Modus, den Median und das arithmetische Mittel der Monatseinkommen aller Einwohner des Hauses?

Lösung:

Es liegen folgende Messwerte vor:

$$\begin{array}{lllll} x_1 = 2\,500\,, & x_2 = 2\,500\,, & x_3 = 2\,500\,, & x_4 = 2\,500, & x_5 = 2\,500\,, \\ x_6 = 2\,600\,, & x_7 = 2\,700\,, & x_8 = 2\,800\,, & x_9 = 2\,900\,, & x_{10} = 3\,000\,. \end{array}$$

Modus: 2 500 [Euro]
Median: $\tilde{x}_{0,5} = x_{0.5 \cdot 10} = x_5 = 2\,500$ [Euro]
Arithmetisches Mittel: $\bar{x} = \frac{1}{10} \sum\limits_{i=1}^{10} x_i = \frac{26500}{10} = 2\,650$ [Euro]

Nun komme ein Wert $x_{11} = 100\,000$ hinzu. Dann folgt:

Modus: 2 500 [Euro]
Median: $\tilde{x}_{0,5} = x_{[0.5 \cdot 11]+1} = x_6 = 2\,600$ [Euro]
Arithmetisches Mittel: $\bar{x} = \frac{1}{11} \sum\limits_{i=1}^{11} x_i = \frac{126500}{11} = 11\,500$ [Euro]

Aufgabe 3.20

Im Rahmen einer Flurbereinigung werden die Nutzflächen (in ha) von 40 landwirtschaftlichen Betrieben ermittelt:

3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 8, 9, 9, 9, 9, 10, 10, 14,
16, 16, 18, 21, 25, 28, 33, 36, 40, 40, 48, 60, 69, 77, 77, 80, 99.

a) Berechnen Sie aus diesen Einzelwerten die durchschnittliche Nutzfläche und die Varianz der Nutzfläche.

b) Bilden Sie eine stetige Klassierung der Daten (Klassen: 2 bis 4, über 4 bis 9, über 9 bis 20, über 20 bis 50, über 50 bis 100 ha), und berechnen Sie daraus die durchschnittliche Nutzfläche. Begründen Sie den Unterschied zu a).

c) Berechnen Sie die Varianz der Nutzfläche aus der stetigen Klassierung in b). Begründen Sie den Unterschied zu a).

d) Berechnen Sie zunächst für jede Klasse aus b) den Durchschnitt und aus den Klassendurchschnitten den Gesamtdurchschnitt der Nutzfläche.

Lösung:

X = Nutzfläche in ha

a) $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{40}(3 + 3 + \cdots + 80 + 99) = \frac{1}{40}923 = 23.075$ [ha]

$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{40}(9 + \cdots + 9\,801) - (23.075)^2 = 668.8194$ [ha^2]

b) Arbeitstabelle

j	$K_j =]x_j^u, x_j^o]$	n_j	ξ_j	$\xi_j n_j$	$\xi_j^2 n_j$
1	2 , 4	10	3.0	30	90.0
2	4 , 9	10	6.5	65	422.5
3	9 , 20	6	14.5	87	1 261.5
4	20 , 50	8	35.0	280	9 800.0
5	50 , 100	6	75.0	450	33 750.0
$\sum$		40		912	45 324.0

Die Klassenmittel werden durch die Klassenmittelpunkte $\xi_j = \dfrac{x_j^u + x_j^o}{2}$ approximiert, da die einzelnen $\bar{x}_j$ unbekannt sind.

$$\overline{x} \approx \frac{1}{n} \sum_{j=1}^{J} \xi_j n_j = \frac{1}{40} 912 = 22.8 \text{ [ha]}$$

Der Unterschied zu a) liegt im Informationsverlust durch die Bildung stetiger Klassen. Die Einzeldaten werden dabei nicht berücksichtigt, da die repräsentativen Klassenmittelpunkte die eigentliche Verteilung in der Regel nicht wiedergeben.
(Man könnte die Werte $\overline{x}_j$ jedoch berechnen, da die Einzeldaten vorliegen – dies wird in Teil d) gemacht.)

c)

$$s^2 \approx \frac{1}{n} \sum_{j=1}^{J} \xi_j^2 n_j - \overline{x}^2 = \frac{1}{40} 45\,324 - (22.8)^2 = 613.26 \text{ [ha}^2\text{]}$$

Hier haben wir erneut durch die stetige Klassierung einen Informationsverlust.

d) Arbeitstabelle

j	$K_j =]x_j^u, x_j^o]$	$\bar{x}_j$	$\bar{x}_j n_j$
1	2 , 4	3.500	35
2	4 , 9	7.100	71
3	9 , 20	14.000	84
4	20 , 50	33.875	271
5	50 , 100	77.000	462
$\sum$			923

Mit dem Additionssatz für arithmetische Mittel erhält man:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{J} \bar{x}_j n_j = \frac{1}{40} 923 = 23.075 \text{ [ha]}$$

(also wie in Teil a)). Die $\overline{x}_j$ berechnen sich aus den gegebenen Einzelwerten, die in die jeweilige Klasse K_j fallen.

Aufgabe 3.21

Gegeben sind die drei nachstehenden stetigen Klassierungen I, II und III:

j	K_j	I n_j	II n_j	III n_j
1	[0,5]	5	3	2
2	]5,10]	6	4	3
3	]10,15]	4	6	4
4	]15,20]	3	4	6
5	]20,25]	2	3	5

Bestimmen Sie für jede Verteilung die Lagemaße Median und arithmetisches Mittel und vergleichen Sie die Verteilungen.

Lösung:

			I			II			III		
j	K_j	ξ_j	n_j	f_j	$F(x_j^o)$	n_j	f_j	$F(x_j^o)$	n_j	f_j	$F(x_j^o)$
1	[0,5]	2.5	5	0.25	0.25	3	0.15	0.15	2	0.10	0.10
2	]5,10]	7.5	6	0.30	0.55	4	0.20	0.35	3	0.15	0.25
3	]10,15]	12.5	4	0.20	0.75	6	0.30	0.65	4	0.20	0.45
4	]15,20]	17.5	3	0.15	0.90	4	0.20	0.85	6	0.30	0.75
5	]20,25]	22.5	2	0.10	1.00	3	0.15	1.00	5	0.25	1.00
$\sum$	—	—	20	1	—	20	1	—	20	1	—

- Klassierung I (rechtsschief / linkssteil) :
  - $\bar{x} \approx \frac{1}{20} \sum\limits_{i=1}^{5} \xi_j n_j = 10.25.$
  - Da $F(x_2^u) < 0.5 \leq F(x_2^o)$, gilt:

$$\tilde{x}_{0,5} \approx x_2^u + \frac{p - F(x_2^u)}{f_2}(x_2^o - x_2^u) = 5 + \frac{0.5 - 0.25}{0.3}(10 - 5) \approx 9.17$$

  - Modus = 7.5
- Klassierung II (symmetrisch) :
  - $\bar{x} \approx 12.5.$

- Da $F(x_3^u) < 0.5 \leq F(x_3^o)$, gilt:

$$\tilde{x}_{0,5} \approx x_3^u + \frac{p - F(x_3^u)}{f_3}(x_3^o - x_3^u) = 10 + \frac{0.5 - 0.35}{0.3}(15 - 10) = 12.5$$

- Modus = 12.5

• Klassierung III (linksschief / rechtssteil) :

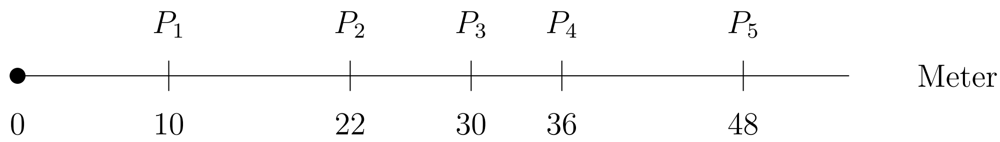
- $\bar{x} \approx 14.75$.
- Da $F(x_4^u) < 0.5 \leq F(x_4^o)$, gilt:

$$\tilde{x}_{0,5} \approx x_4^u + \frac{p - F(x_4^u)}{f_4}(x_4^o - x_4^u) = 15 + \frac{0.5 - 0.45}{0.3}(20 - 15) \approx 15.83$$

- Modus = 17.5

Aufgabe 3.22

In einer Fertigungshalle eines Betriebs seien fünf Produktionsstätten $P_1, \ldots, P_5$ auf einer Fertigungsstraße entlang einer Geraden hintereinander in folgenden Abständen angeordnet (Start der Straße bei Meter Null):



Produktionsstätte	P_1	P_2	P_3	P_4	P_5
Anzahl der Arbeiter	8	5	2	5	18

Um die Produktionsstätten mit Material zu beliefern, soll ein Zwischenlager auf dieser Straße gebaut werden. Die Summe der von den Arbeitern zurückgelegten Wegstrecken zum Lager soll minimal werden. (Nehmen Sie dafür an, dass jeder Arbeiter einer Produktionsstätte einmal am Tag zum Lager muss.) Wo würden Sie, wenn jeder Platz an der Straße ansonsten in gleicher Weise geeignet ist, das Lager bauen?

Lösung:

ξ_j = Abstand der Produktionsstätte vom Anfang der Straße (in Metern).
n_j = Anzahl der Arbeiter

Auf Grund der Minimumeigenschaft des Median gilt für jedes a:

$$\sum |\xi_j - \tilde{x}_{0,5}| n_j \leq \sum |\xi_j - a| n_j$$

Da wir die absoluten Abstände minimieren wollen, suchen wir den Median $\tilde{x}_{0,5}$.

j	ξ_j	n_j	f_j	$F(\xi_j)$
1	10	8	0.2105	0.2105
2	22	5	0.1316	0.3421
3	30	2	0.0526	0.3947
4	36	5	0.1316	0.5263
5	48	18	0.4737	1.0000
$\sum$		38	1.0000	

Median $\tilde{x}_{0.5}$:

$$\begin{aligned}
\tilde{x}_{0.5} &= \xi_4 = 36, \\
\text{denn } F(\xi_3) &= 0.3947 < 0.5 \\
\text{und } F(\xi_4) &= 0.5263 \geq 0.5
\end{aligned}$$

Das Lager sollte bei Produktionsstätte P4 im Abstand von 36 Metern vom Anfang der Straße gebaut werden.

Aufgabe 3.23

Es sei $x_1, \ldots, x_n$ eine Messreihe $\overline{x}$ das zugehörige arithmetische Mittel.

a) Man zeige: Werden die Werte der Messreihe gemäß $y_i = a + b \cdot x_i$, $i = 1, \ldots n$, linear transformiert, so gilt für das arithmetische Mittel $\overline{y}$ der transformierten Werte

$$\overline{y} = a + b \cdot \overline{x},$$

d.h. das arithmetische Mittel der transformierten Werte ist gleich dem transformierten arithmetischen Mittel der ursprünglichen Werte.

b) Auf einer Touristeninsel in der Karibik wurden in den letzten beiden Juliwochen jeweils morgens zur gleichen Zeit die folgenden Lufttemperaturen in Grad Fahrenheit (°F) gemessen:

$$76,\ 84,\ 82,\ 76,\ 77,\ 85,\ 81,$$
$$79,\ 80,\ 81,\ 82,\ 81,\ 79,\ 77\,.$$

Man berechne die Durchschnittstemperatur, d. h. das arithmetische Mittel der gemessenen Temperaturen, in Grad Fahrenheit (°F) und in Grad Celsius (°C).

Hinweis: $x[°F]$ entsprechen $y = \frac{5}{9} \cdot (x - 32)[°C]$.

Lösung:

a)

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \;= \frac{1}{n}\sum_{i=1}^{n}(a + bx_i) \;= \frac{1}{n}\left(na + b\sum_{i=1}^{n} x_i\right) \;= a + b\frac{1}{n}\sum_{i=1}^{n} x_i \;= a + b\bar{x}$$

b) Seien die angegebenen, in °Fahrenheit gemessenen Temperaturen mit $x_i, i = 1, \ldots, 14$ bezeichnet. Dann gilt für den Temperaturmittelwert in °Fahrenheit:

$$\bar{x} = \frac{1}{14}\sum_{i=1}^{14} x_i = \frac{1}{14}1120\ [\ °\mathrm{F}] = 80\ [\ °\mathrm{F}]$$

$y_i, i = 1, \ldots, 14$ sollen die gleichen Temperaturen in °C darstellen. Man erhält sie wie folgt aus den x_i $(i = 1 \ldots, 14)$:

$$y_i = \frac{5}{9} \cdot (x - 32) = -\frac{160}{9} + \frac{5}{9}x_i.$$

Mit der Formel aus a) ergibt sich für den Temperaturmittelwert in °C:

$$\bar{y} = -\frac{160}{9} + \frac{5}{9}\bar{x} = 26.7\ [\ °\mathrm{C}].$$

Aufgabe 3.24

Die folgende Tabelle enthält die Anzahl (in 100 000) und das durchschnittliche Nettoeinkommen je Monat (in 1 000 Euro) privater Haushalte im früheren Bundesgebiet für das Jahr 2006:

Nettoeinkommen über ... bis einschließlich ...	Anzahl der Haushalte	Durchschnittliches Nettoeinkommen
0 – 1.5	16	1.12
1.5 – 2.5	50	1.98
2.5 – 5	206	3.13
5 – 10	18	6.75
Summe	290	

a) Zeichnen Sie das Histogramm und die empirische Verteilungsfunktion.

b) Berechnen Sie für alle privaten Haushalte

(1) das durchschnittliche Nettoeinkommen,

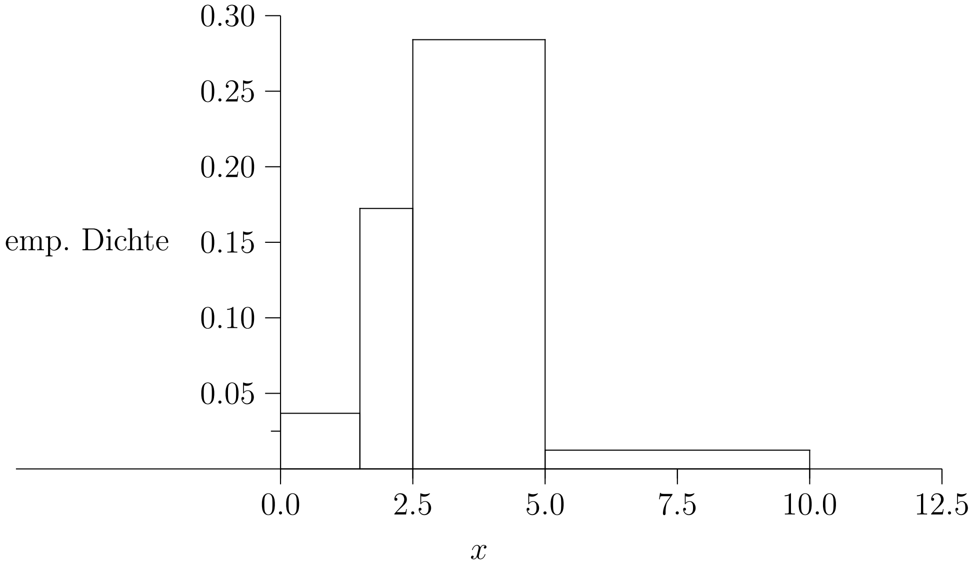
(2) die Standardabweichung des Nettoeinkommens.

Lösung:

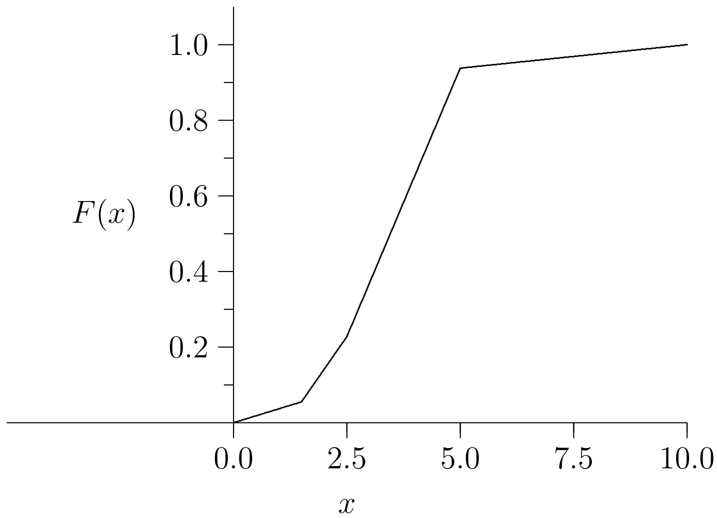
Arbeitstabelle:

$K_j =]x_j^u, x_j^o]$	n_j	$\bar{x}_j$	f_j	$F(x_j^o)$	$\bar{x}_j f_j$	$(\bar{x}_j - \bar{x})^2 f_j$	$\frac{f_j}{x_j^o - x_j^u}$
0 – 1.5	16	1.12	0.0552	0.0552	0.0618	0.2047	0.0368
1.5 – 2.5	50	1.98	0.1724	0.2276	0.3414	0.1958	0.1724
2.5 – 5	206	3.13	0.7103	0.9379	2.2232	0.0051	0.2841
5 – 10	18	6.75	0.0621	1	0.4192	0.8522	0.0124
$\sum$	290	—	1	—	3.0456	1.2578	—

a) Histogramm:



Empirische Verteilungsfunktion:



b) (1) $\bar{x} = \sum_{j=1}^{J} \bar{x}_j f_j = 3.0456$
Durchschnittliches Nettoeinkommen 3045,60 Euro.

(2) $s_x = \sqrt{\sum_{j=1}^{J} (\bar{x}_j - \bar{x})^2 f_j} = 1.1215$
Standardabweichung des Nettoeinkommens 1121,50 Euro.

Aufgabe 3.25

Bruttoinlandsprodukt (BIP) 1998 in den fünf neuen Bundesländern in Mrd. DM:

Bundesland	BIP
Brandenburg	77.8
Mecklenburg–Vorpommern	48.4
Sachsen	125.3
Sachsen–Anhalt	71.4
Thüringen	66.5

a) Berechnen Sie für das Bruttoinlandsprodukt das arithmetische Mittel und den Median.

b) Berechnen Sie die Spannweite, den Quartilabstand, die mittlere absolute Abweichung vom Median und die Varianz.

c) Wie verändern sich die in a) und b) berechneten Größen, falls das Bruttoinlandsprodukt in Euro gemessen wird? (1.95583 DM = 1 Euro)

Lösung:

Arbeitstabelle:
(Daten aufsteigend sortieren zur Berechnung der Quantile)

i	x_i	$\lvert x_i - \tilde{x}_{0,5}\rvert$	x_i^2
1	48.4	23.00	2 342.56
2	66.5	4.90	4 422.25
3	71.4	0.00	5 097.96
4	77.8	6.40	6 052.84
5	125.3	53.90	15 700.09
$\sum$	389.4	88.20	33 615.70

a) Arithmetische Mittel: $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{1}{5}389,4 = 77,88$ [Mrd.DM].

Median: $\tilde{x}_{0,5} = x_{[5\cdot 0,5]+1} = x_{[2,5]+1} = x_3 = 71,4$ [Mrd.DM] $\widehat{=}$ Sachsen-Anhalt.

b) Spannweite: $R = 125,3 - 48,4 = 76.9$ [Mrd.DM]
Unteres/oberes Quartil: $\tilde{x}_{0,25} = x_2 = 66,5$; $\tilde{x}_{0,75} = x_4 = 77,8$
Quartilabstand: $Q = 77,8 - 66,5 = 11,3$ [Mrd.DM].
Mittlere absolute Abweichung vom Median: $d = \frac{1}{n} \sum |x_j - \tilde{x}_{0,5}| = \frac{1}{5} 88,20 = 17,64$ [Mrd.DM].
Varianz: $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{5} 33615,70 - 77,88^2 = 657,85$ [Mrd.DM^2].

c) Arithmetisches Mittel, Median, Spannweite, Quartilabstand, mittlere absolute Abweichung vom Median und Standardabweichung müssen durch 1.95583 dividiert werden. Die Varianz wird durch 1.95583^2 dividiert.
Es folgt:
$\bar{x} = 39.82$ [Mrd. Euro]
$\tilde{x}_{0,5} = 36.51$ [Mrd. Euro]
$R = 39.32$ [Mrd. Euro]
$Q = 5.778$ [Mrd. Euro]
$d = 9.01$ [Mrd. Euro]
$s^2 = 171.97$ [Mrd. Euro^2]
$s = 13.1139$ [Mrd. Euro]

Aufgabe 3.26

Ein junges Kraftfahrt-Versicherungsunternehmen will die Vollkasko-Schäden des vergangenen Geschäftsjahres auswerten. Folgende Daten liegen vor [Angaben in Tsd. Euro]:

$5,5$	$5,9$	$6,1$	$6,4$	$6,4$	$6,7$	$6,7$
$7,1$	$7,1$	$7,2$	$7,5$	$7,6$	$7,8$	$8,2$
$8,7$	$9,2$	$9,5$	$10,1$	$24,0$	$27,5$	$71,0$

a) Berechnen Sie das arithmetische Mittel $\overline{x}$ sowie das α-getrimmte Mittel $\overline{x}_\alpha$ für $\alpha = 0,1$ und $\alpha = 0,2$. Was fällt dabei auf?

b) Berechnen Sie die Standardabweichung sowie die mittlere absolute Abweichung vom Median.

c) Berechnen Sie für die erste Zeile der Daten Ginis mittlere Differenz.

Lösung:

a) Arithm. Mittel

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{21} \cdot 256,2 = 12,2 \mathrel{\hat{=}} 12200 \text{ Euro}$$

Zur Berechnung von

$$\overline{x}_\alpha = \frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} x_i$$

müssen die Daten aufsteigend geordnet sein (das ist hier schon geschehen). Also hat man

$$\begin{aligned}
\alpha &= 0,1 \Longrightarrow n\alpha = 2,1 \Longrightarrow [n\alpha] = 2 \text{ (d.h. auf jeder Seite 2 Werte weg)} \\
\Longrightarrow & \quad \overline{x}_{0,1} = \frac{1}{17}\sum_{i=3}^{19} x_i = \frac{1}{17} \cdot 146,3 = 8,606 \mathrel{\hat{=}} 8606 \text{ Euro} \\
\alpha &= 0,2 \Longrightarrow n\alpha = 4,2 \Longrightarrow [n\alpha] = 4 \text{ (d.h. auf jeder Seite 4 Werte weg)} \\
\Longrightarrow & \quad \overline{x}_{0,2} = \frac{1}{13}\sum_{i=5}^{17} x_i = \frac{1}{13} \cdot 99,7 = 7,669 \mathrel{\hat{=}} 7669 \text{ Euro}
\end{aligned}$$

Man sieht, dass sich - durch Nicht-Berücksichtigung der "Ausreißer" (z.B. "Totalschaden" 71,0) - die mittlere Schadenhöhe stark reduziert (was für die Kunden u.a. Auswirkungen auf die zu zahlende Prämie hat).

b) Standardabweichung

$$\begin{aligned} s &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2-\overline{x}^2} = \sqrt{\frac{1}{21}\cdot 7394,56-12,2^2} \\ &= \sqrt{203,282} = 14,258 \mathrel{\hat{=}} 14258 \text{ Euro} \end{aligned}$$

Median

$$\tilde{x}_{0,5} = x_{[21\cdot 0,5]+1} = x_{11} = 7,5$$

Mittlere absolute Abweichung vom Median

$$d = \frac{1}{n}\sum_{i=1}^{n}|x_i-\tilde{x}_{0,5}| = \frac{118,5}{21} = 5,643 \mathrel{\hat{=}} 5643 \text{ Euro}$$

(Zum Vergleich noch: $\Delta = 6,839 \mathrel{\hat{=}} 6839$ Euro.)

c) Ginis mittlere Differenz

$$\Delta = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-x_j| = \frac{1}{7^2}\cdot 22 = 0,44898 \mathrel{\hat{=}} 449 \text{ Euro}$$

Arbeitstabelle:

$\lvert x_i-x_j\rvert$	5,5	5,9	6,1	6,4	6,4	6,7	6,7
5,5	0	0,4	0,6	0,9	0,9	1,2	1,2
5,9	0,4	0	0,2	0,5	0,5	0,8	0,8
6,1	0,6	0,2	0	0,3	0,3	0,6	0,6
6,4	0,9	0,5	0,3	0	0	0,3	0,3
6,4	0,9	0,5	0,3	0	0	0,3	0,3
6,7	1,2	0,8	0,6	0,3	0,3	0	0
6,7	1,2	0,8	0,6	0,3	0,3	0	0

Aufgabe 3.27

70% der Belegschaft einer Firma sind Arbeiter, 25% Angestellte und 5% leitende Angestellte. Die folgende Tabelle enthält die arithmetischen Mittel und Standardabweichungen der Monatslöhne bzw. -gehälter (in Euro) dieser Gruppen:

	Arithmetisches Mittel	Standardabweichung
Arbeiter	3 000	1 000
Angestellte	5 000	2 000
Leitende Angestellte	11 000	4 000

a) Berechnen Sie das arithmetische Mittel und die Standardabweichung der Monatslöhne bzw. -gehälter aller Beschäftigten.

b) Wie groß ist der Anteil der durch die Aufteilung der Belegschaft in Arbeiter, Angestellte und leitende Angestellte erklärten Varianz an der Gesamtvarianz der Löhne bzw. Gehälter?

c) Tariferhöhungen ergeben:

- für Arbeiter 5% lineare Erhöhung und 200 Euro Sockelbetrag,
- für Angestellte 3% lineare Erhöhung und keinen Sockelbetrag,
- für leitende Angestellte keine Erhöhungen.

Berechnen Sie die Werte für a) und b) nach der Tariferhöhung.

Lösung:

X = Monatslöhne vor der Tariferhöhung in EURO

	j	$\bar{x}_j$	s_j	f_j	$\bar{x}_j f_j$	$s_j^2 f_j$	$(\bar{x}_j - \bar{x})^2 f_j$
Arb.	1	3 000	1 000	0.70	2 100	700 000	567 000
Ang.	2	5 000	2 000	0.25	1 250	1 000 000	302 500
Leit. Ang.	3	11 000	4 000	0.05	550	800 000	2 520 500
	$\sum$	19 000	—	1.00	3 900	$s_{int}^2 = 2\,500\,000$	$s_{ext}^2 = 3\,390\,000$

a) Der Mittelwert wird durch den Additionssatz für arithmetische Mittel berechnet: $\bar{x} = \frac{1}{n}\sum \bar{x}_j n_j = \sum \bar{x}_j f_j = 3900$ [EURO].
Die Standardabweichung wird über den Varianzzerlegungssatz berechnet:
$s = \sqrt{\sum s_j^2 f_j + \sum(\bar{x}_j - \bar{x})^2 f_j} = \sqrt{2\,500\,000 + 3\,390\,000} = 2426.9322$ [EURO].

b) Der Anteil der erklärten Varianz beträgt:
$(\frac{s^2_{ext}}{s^2_{ges}})100 = (\frac{3\,390\,000}{5\,890\,000})100 = 57.56\%$.

Y= Monatslöhne nach der Tariferhöhung in EURO.

j	$\bar{y}_j$	$(bs_j)^2 f_j$	$(\bar{y}_j - \bar{y})^2 f_j$
1	3 350	771 750	485 139.3750
2	5 150	1 060 900	234 014.0625
3	11 000	800 000	2 323 915.3125
$\sum$		$s^2_{int} = 2\,632\,650$	$s^2_{ext} = 3\,043\,068.7500$

c) Die lineare Lohnerhöhung lässt sich durch eine multiplikative Transformation der x_j mit b und der Sockelbetrag durch Addition von a darstellen. $\bar{y}_j = a + b\bar{x}_j$. Für $j = 1$ beträgt $b = 1.05$, $a = 200$. Für $j = 2$ beträgt $b = 1.03$ und $j = 3$ erfährt keine Transformation. $\bar{y} = \sum \bar{y}_j f_j = 4182.50$ [EURO].
Für die Varianz gilt: $s^2 = \sum(bs_j)^2 f_j + \sum(\bar{y}_j - \bar{y})^2 f_j$, da die Lage a keinen Einfluss auf die Varianz hat. Somit ist die veränderte Standardabweichung $s = \sqrt{2\,632\,650 + 3\,043\,068,75} = 2382.3767$ [EURO] und der Anteil der erklärten Varianz beträgt 53,62%.

Aufgabe 3.28

Bruttojahreseinkommen von Erwerbstätigen in einem Kölner Stadtteil:

Bruttojahreseinkommen (in Tsd. Euro über ... bis einschließlich ...	Anzahl der Erwerbstätigen
0 – 20	400
20 – 50	1 983
50 – 100	423

a) Stellen Sie die empirische Verteilungsfunktion graphisch dar.

b) Wie hoch ist der Anteil der Erwerbstätigen, die im Jahr höchstens 12 000 Euro brutto verdienen?

c) Wie hoch ist der Anteil der Erwerbstätigen, deren Bruttojahreseinkommen zwischen 65 000 und 85 000 Euro liegt?

d) Welches Bruttojahreseinkommen überschreiten die oberen 50% der aufgeführten Personen?

e) Wie hoch ist im Durchschnitt das Bruttojahreseinkommen der Erwerbstätigen in dem Stadtteil?

f) Berechnen Sie Quartilabstand, Varianz und Standardabweichung des Einkommens.

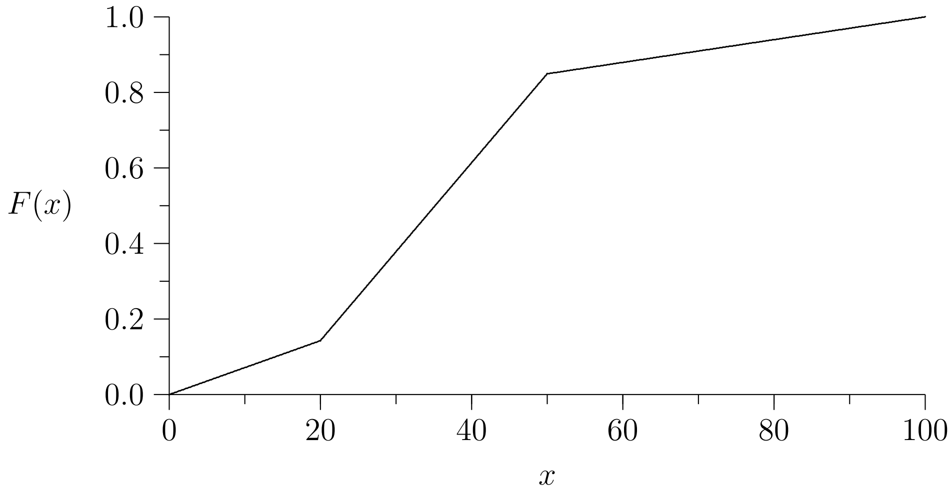
Lösung:

X = Bruttojahreseinkommen (in 1 000 Euro)
Das Merkmal ist stetig klassiert.

a) Die folgende Tabelle enthält die absoluten Häufigkeiten, die relativen Häufigkeiten, sowie die Werte der Verteilungsfunktion an den Klassenobergrenzen:

x_j^u	x_j^o	n_j	f_j	$F(x_j^o)$
0	20	400	0.1426	0.1426
20	50	1 983	0.7067	0.8493
50	100	423	0.1507	1.0000
$\sum$	—	2806	1	—

Empirische Verteilungsfunktion:



b)

$$F(x) \approx F(x_j^u) + \frac{f_j}{x_j^o - x_j^u}(x - x_j^u)$$
$$F(12) = 0 + \frac{0.1426}{20-0} \cdot (12 - 0) = 0.0856.$$

Der gesuchte Anteil ist also 8.56%.

c) Gesucht ist $F(85) - F(65)$:

$$F(85) = 0.8493 + \frac{0.1507}{100-50} \cdot (85 - 50) = 0.9548$$
$$F(65) = 0.8493 + \frac{0.1507}{100-50} \cdot (65 - 50) = 0.8945$$

Der gesuchte Anteil ist also $F(85) - F(65) = 0.0603$, d.h. 6.03%.

d) Gesucht ist der Median bzw. das 50%-Quantil:

$$\tilde{x}_{0.5} \approx x_j^u + \frac{p - F(x_j^u)}{f_j}(x_j^o - x_j^u)$$
$$\tilde{x}_{0.5} \approx 20 + \frac{0.5 - 0.1426}{0.7067} \cdot (50 - 20)$$
$$\tilde{x}_{0.5} = 35\,171,90[\text{Euro}]$$

e),f) Arbeitstabelle zur Berechnung von $\bar{x}$ und s^2:

x_j^u	x_j^o	n_j	ξ_j	$\xi_j n_j$	$(\xi_j - \bar{x})^2 n_j$
0	20	400	10	4 000	301 754.6597
20	50	1 983	35	69 405	12 059.9104
50	100	423	75	31 725	595 919.7136
$\sum$	—	2 806	—	105 130	909 734.2837

Durchschnittliches Bruttojahreseinkommen:

$$\bar{x} \approx \frac{1}{n} \sum \xi_j n_j$$

mit den Klassenmittelpunkten $\xi_j = \frac{x_j^o + x_j^u}{2}$.

$$\bar{x} = \frac{105\,130}{2\,806} = 37.4661.$$
$$\bar{x} = 37\,466,10[\text{Euro}].$$

Quartilabstand:

$$\tilde{x}_{0.75} \approx 20 + \frac{0.75 - 0.1426}{0.7067}(50 - 20) = 45,7846[1000\text{Euro}].$$
$$\tilde{x}_{0.25} \approx 20 + \frac{0.25 - 0.1426}{0.7067}(50 - 20) = 24,5592[1000\text{Euro}].$$

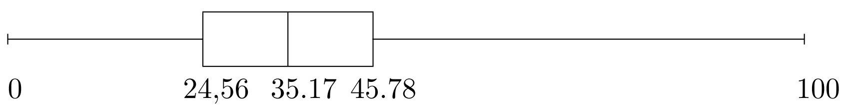
Der Quartilabstand beträgt also $Q = 21\,225.40[\text{Euro}]$.

Varianz und Standardabweichung:

$$s^2 \approx \frac{1}{n} \sum_{j=1}^{J} (\xi_j - \bar{x})^2 n_j = 324.2104$$
$$s \approx 18.0058$$

Die Standardabweichung beträgt 18 005,80[Euro]

Boxplot zur Veranschaulichung:



Aufgabe 3.29

Die folgende Tabelle enthält für die Länder der EU im Jahr 1998 jeweils die Einwohnerzahl, das Bruttoinlandsprodukt (BIP) sowie die Anzahl der Sitze im Europäischen Parlament:

Land	Einwohner (in Mio.)	BIP (in Mrd. DM)	Sitze
A	8.1	391	21
B	10.2	494	25
D	82.1	3 758	99
DK	5.3	281	16
E	39.3	1 326	64
F	58.6	2 614	87
FIN	5.1	225	16
GB	59.0	2 517	87
GR	10.5	305	25
I	57.5	2 522	87
IRL	3.7	168	15
L	0.4	30	6
NL	15.6	729	31
P	9.8	305	25
S	8.9	380	22

a) Bestimmen Sie die Gini–Koeffizienten zu den drei Verteilungen.

b) Ändert sich die Disparität des Bruttoinlandsprodukts wesentlich, wenn man Luxemburg von der Betrachtung ausschließt?

Lösung:

a) $D_G = \sum_{i=1}^{n} \frac{2i - n - 1}{n} \cdot h_i$

i	x_i	h_i	$(2i-16)h_i$
1	0.4	0.001	-0.015
2	3.7	0.010	-0.119
3	5.1	0.014	-0.136
4	5.3	0.014	-0.113
5	8.1	0.022	-0.130
6	8.9	0.024	-0.095
7	9.8	0.026	-0.052
8	10.2	0.027	0.000
9	10.5	0.028	0.056
10	15.6	0.042	0.167
11	39.3	0.105	0.630
12	57.5	0.154	1.230
13	58.6	0.157	1.566
14	59.0	0.158	1.893
15	82.1	0.219	3.072
$\sum$	374.1	1.000	7.953

i	x_i	h_i	$(2i-16)h_i$
1	30	0.002	-0.026
2	168	0.010	-0.126
3	225	0.014	-0.140
4	281	0.018	-0.140
5	305	0.019	-0.114
6	305	0.019	-0.076
7	380	0.024	-0.047
8	391	0.024	0.000
9	494	0.031	0.062
10	729	0.045	0.182
11	1326	0.083	0.496
12	2517	0.157	1.255
13	2522	0.157	1.572
14	2614	0.163	1.955
15	3758	0.234	3.279
$\sum$	16045	1.000	8.130

i	x_i	h_i	$(2i-16)h_i$
1	6	0.010	-0.134
2	15	0.024	-0.288
3	16	0.026	-0.256
4	16	0.026	-0.204
5	21	0.034	-0.201
6	22	0.035	-0.141
7	25	0.040	-0.080
8	25	0.040	0.000
9	25	0.040	0.080
10	31	0.050	0.198
11	64	0.102	0.613
12	87	0.139	1.112
13	87	0.139	1.390
14	87	0.139	1.668
15	99	0.158	2.214
$\sum$	626	1.000	5.970

$$D_G(Einw.) = 0.53, \quad D_G(BIP) = 0.54, \quad D_G(EP) = 0.40.$$

b) Daten ohne Luxemburg

i	x_i	h_i	$(2i-15)h_i$
1	168	0.010	-0.136
2	225	0.014	-0.155
3	281	0.018	-0.158
4	305	0.019	-0.133
5	305	0.019	-0.095
6	380	0.024	-0.071
7	391	0.024	-0.024
8	494	0.031	0.031
9	729	0.046	0.137
10	1 326	0.083	0.414
11	2 517	0.157	1.100
12	2 522	0.157	0.417
13	2 614	0.163	1.795
14	3 758	0.235	3.051
$\sum$	16 015	1.000	7.168

$$D_G(BIP) = 0.51$$

Aufgabe 3.30

Fünf Unternehmen teilen sich einen Markt. Die Gewinne vor Steuern (in 1 000 Euro) sind:

$$10\,, \quad 120\,, \quad 10\,, \quad 40\,, \quad 20\,.$$

Bezeichne $t(x)$ die Steuer auf den Gewinn x. Betrachten Sie die drei folgenden Steuertarife:

a) proportionaler Steuertarif, $t(x) = 0.3x$,

b) progressiver Steuertarif,

$$t(x) = \begin{cases} 0 & \text{für} \quad 0 \leq x \leq 20\,, \\ \dfrac{x-10}{200}\,x & \text{für} \quad 20 < x \leq 110\,, \\ 0.5x & \text{für} \quad x > 110\,, \end{cases}$$

c) Kopfsteuer, $t(x) = 1$.

Berechnen Sie für jeden der drei Steuertarife die Lorenzkurve und den Gini–Koeffizienten für die Gewinne vor Steuern sowie nach Steuern. Interpretieren Sie die Ergebnisse.

Lösung:

Gewinne vor Steuern (in 1 000 Euro):

i	x_i	h_i	$\sum_{r=1}^{i} h_r$	$(2i-n-1)h_i$
1	10	0.05	0.05	-0.2
2	10	0.05	0.10	-0.1
3	20	0.10	0.20	0.0
4	40	0.20	0.40	0.4
5	120	0.60	1.00	2.4
	200	1.00		2.5

$$n = 5 : D_G \in [0\ ;\ 0,8]$$

$$
\begin{aligned}
D_G &= \frac{1}{n}\sum_{i=1}^{n}(2i-n-1)h_i = \frac{1}{5}[(-)4\cdot 0.05 - 2\cdot 0.05 + 0.1\cdot 0 + 0.2\cdot 0.2 + 0.6\cdot 4] \\
&= \frac{1}{5}2.5 = 0.5
\end{aligned}
$$

Gewinn nach Steuern: $y_i = x_i - t(x_i)$

a) Proportionaler Steuertarif:

i	y_i	h_i	$\sum_{r=1}^{i} h_r$	$(2i-n-1)h_i$
1	7	0.05	0.05	-0.2
2	7	0.05	0.10	-0.1
3	14	0.10	0.20	0.0
4	28	0.20	0.40	0.4
5	84	0.60	1.00	2.4
$\sum$	140	1.00		2.5

Da die h_i sich nicht geändert haben, bleibt $D_G = 0.5$ ebenfalls unverändert.

b) Progressiver Steuertarif:

i	y_i	h_i	$\sum_{r=1}^{i} h_r$	$(2i-n-1)h_i$
1	10	0.0746	0.0746	-0.2985
2	10	0.0746	0.1493	-0.1493
3	20	0.1493	0.2985	0.0000
4	34	0.2537	0.5522	0.5074
5	60	0.4478	1.0000	1.7910
$\sum$	134	1.0000		1.8506

$$D_G = \frac{1}{5}1.8506 = 0.3701$$

c) Kopfsteuer:

i	y_i	h_i	$\sum_{r=1}^{i} h_r$	$(2i-n-1)h_i$
1	9	0.0462	0.0462	-0.1846
2	9	0.0462	0.0923	-0.0923
3	19	0.0974	0.1897	0.0000
4	39	0.2000	0.3897	0.4000
5	119	0.6103	1.0000	2.4410
$\sum$	195	1.0000		2.5641

$$D_G = 0.5128$$

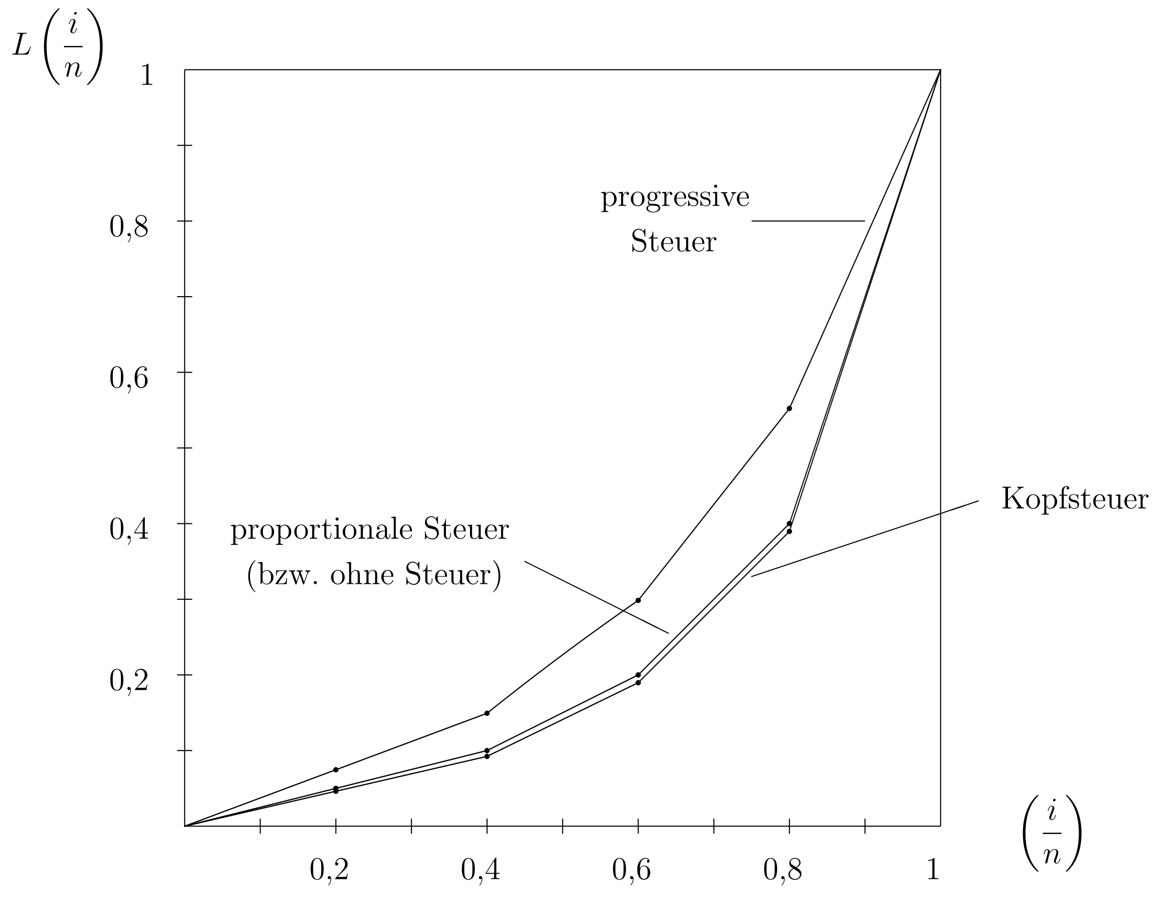
Die Lorenzkurve ergibt sich aus der Verbindung der Punkte $(0,0), (\frac{1}{n}, L(\frac{1}{n})), ((\frac{2}{n}), L(\frac{2}{n})), \cdots, ((\frac{n-1}{n}), L(\frac{n-1}{n})), (1,1)$ mit $L(\frac{i}{n}) = h_i + \cdots + h_i$.

$\frac{i}{n}$	Prop.St.$L(\frac{i}{n})$	Progr.St.$L(\frac{i}{n})$	Kopfst.$L(\frac{i}{n})$
0.00	0.00	0.0000	0.0000
0.20	0.05	0.0746	0.0462
0.40	0.10	0.1493	0.0923
0.60	0.20	0.2985	0.1897
0.80	0.40	0.5522	0.3897
1.00	1.00	1.0000	1.0000

	vor Steuern	Prop.st.	Progr.st.	Kopfst.
Gini-Koeffizient	0.5000	0.5000	0.3701	0.5128

Mit der Progressiven Steuer sinkt die Disparität und die "Reichen werden benachteiligt". Mit der Kopfsteuer steigt die Disparität und die "Reichen werden bevorzugt".

Bildchen:



$L\left(\frac{i}{n}\right)$
1
0,8
0,6
0,4
0,2
progressive Steuer
Kopfsteuer
proportionale Steuer (bzw. ohne Steuer)
0,2
0,4
0,6
0,8
1
$\left(\frac{i}{n}\right)$

Aufgabe 3.31

Gegeben ist die Anzahl (in 100 000) und das durchschnittliche Nettoeinkommen je Monat (in 1 000 Euro) privater Haushalte im früheren Bundesgebiet für das Jahr 2006, geschichtet nach der Höhe des Haushaltsnettoeinkommens (vgl. Aufgabe 3.10):

Nettoeinkommen über ... bis einschließlich ...	Anzahl der Haushalte	Durchschnittliches Nettoeinkommen
0 – 1.5	16	1.12
1.5 – 2.5	50	1.98
2.5 – 5	206	3.13
5 – 10	18	6.75
Summe	290	

a) Berechnen Sie den Gini–Koeffizienten.

b) Zeichnen Sie die Lorenzkurve.

c) Wenn die Angaben über das durchschnittliche Nettoeinkommen fehlen würden, könnten Sie dann trotzdem eine Aussage über die Disparität treffen?

Lösung:

X = Nettoeinkommen in 1 000 Euro je Monat
n_j = Anzahl der Haushalte in 100 000

j	x_j^u	x_j^o	n_j	f_j	$\overline{x}_j$	$\overline{x}_j f_j$	h_j
1	0	1.5	16	0.0552	1.12	0.0618	0.0203
2	1.5	2.5	50	0.1724	1.98	0.3414	0.1121
3	2.5	5	206	0.7103	3.13	2.2232	0.7300
4	5	10	18	0.0621	6.75	0.4192	0.1376
$\sum$	—	—	290	1	—	3.0456 $(= \overline{x})$	1.0000

$\sum_{r=1}^{j} h_r$	$(\sum_{r=1}^{j-1} h_r + \sum_{r=1}^{j} h_r)$	$(\sum_{r=1}^{j-1} h_r + \sum_{r=1}^{j} h_r) \cdot f_j$
0.0203	0.0203	0.0011
0.1324	0.1527	0.0263
0.8624	0.9948	0.7066
1.0000	1.8624	0.1157
—	—	0.8497

a) Gesucht ist der Gini-Koeffizient

$$D_G = 1 - \sum_{j=1}^{4} \left(\sum_{r=1}^{j-1} h_r + \sum_{r=1}^{j} h_r \right) f_j$$

mit $h_r = \dfrac{\bar{x}_r n_r}{n\bar{x}} = \dfrac{\bar{x}_r}{\bar{x}} f_r$ und $\bar{x} = \sum\limits_{j=1}^{J} \bar{x}_j f_j = 3.0456.$

$D_G = 1 - 0.8497 = 0.1503$

b) Zur Darstellung der Lorenzkurve fehlen noch die Werte $\sum\limits_{r=1}^{j} f_r$:

j	$\sum\limits_{r=1}^{j} f_r$	$\sum\limits_{r=1}^{j} h_r$
1	0.0552	0.0203
2	0.2276	0.1324
3	0.9379	0.8624
4	1.0000	1.0000

c) Obwohl die $\bar{x}_j$ nicht mehr gegeben sind, kann eine approximative Aussage mit Hilfe der Klassenmittelpunkte ξ_j gemacht werden.

$\overline{x}_r$ wird durch $\xi_r = \frac{x_r^u + x_r^0}{2}$ approximiert. Dann ergibt sich $h_r = \frac{\xi_r}{\bar{x}} f_r$

j	ξ_j	$\xi_j f_j$	h_j	$\sum\limits_{r=1}^{j} h_r$	$\left(\sum\limits_{r=1}^{j-1} h_r + \sum\limits_{r=1}^{j} h_r \right) f_j$
1	0.75	0.0414	0.0118	0.0118	0.0007
2	2	0.3448	0.0981	0.1099	0.0210
3	3.75	2.6636	0.7577	0.8676	0.6943
4	7.5	0.4658	0.1325	1.0001	0.1160
$\sum$		3.5156 ($= \overline{x}$)	1.0001	—	0.8320

$D_G = 1 - 0.8320 = 0.1680$

Aufgabe 3.32

Bei einer Untersuchung der monatlichen Haushaltseinkommen in Deutschland ergab sich folgende Übersicht:

Monatseinkommen in Euro	Anzahl Haushalte
0 – 750	347
750 – 1250	436
1250 – 1750	216
1750 – 5000	97

Berechnen Sie für diese Einkommensverteilung den Gini-Koeffizient.

Lösung:

Arbeitstabelle:

j	x_j^u	x_j^o	n_j	f_j	ξ_j	$\xi_j n_j$	h_j
1	0	750	347	0,3166	375	130125	0,10688
2	750	1250	436	0,3978	1000	436000	0,35811
3	1250	1750	216	0,1971	1500	324000	0,26612
4	1750	5000	97	0,0885	3375	327375	0,26889
$\sum$	—	—	1096	1	—	1217500	1

j	$\sum\limits_{r=1}^{j} h_r$	$\left(\sum\limits_{r=1}^{j-1} h_r + \sum\limits_{r=1}^{j} h_r\right)$	$\left(\sum\limits_{r=1}^{j-1} h_r + \sum\limits_{r=1}^{j} h_r\right) f_j$
1	0,10688	0,10688	0,0338
2	0,46499	0,57187	0,2275
3	0,73111	1,19610	0,2357
4	1	1,73111	0,1532
$\sum$	—	—	0,6502

Gini-Koeffizient:

$$\begin{aligned} D_G &= 1 - \sum_{j=1}^{4} \left(\sum_{r=1}^{j-1} h_r + \sum_{r=1}^{j} h_r \right) f_j \\ &= 1 - 0,6502 \\ &= 0,3498 \end{aligned}$$

Aufgabe 3.33

Die folgende Tabelle enthält die Schlusskurse zweier Aktien A und B an elf aufeinanderfolgenden Börsentagen:

Tag	1	2	3	4	5	6	7	8	9	10	11
A	85	87	90	88	86	86	87	90	92	93	91
B	103	102	104	105	99	97	96	96	99	102	101

Messen Sie den Zusammenhang der Kurse der Aktien A und B mit einer geeigneten Maßzahl.

Lösung:

Da es sich um zwei metrische Merkmale handelt, berechnet man den Korrelationskoeffizienten von Bravais-Pearson:

$$r_{XY} = \frac{s_{xy}}{\sqrt{s_x^2}\sqrt{s_y^2}} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i\, y_i - \bar{x}\,\bar{y}}{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n} y_i^2 - \bar{y}^2}}$$

$$= \frac{\sum\limits_{i=1}^{n} x_i\, y_i - n\,\bar{x}\,\bar{y}}{\sqrt{\sum\limits_{i=1}^{n} x_i^2 - n\,\bar{x}^2}\sqrt{\sum\limits_{i=1}^{n} y_i^2 - n\,\bar{y}^2}}$$

Arbeitstabelle:

i	x_i Kurs A	y_i Kurs B	x_i^2	y_i^2	$x_i \cdot y_i$
1	85	103	7225	10609	8755
2	87	102	7569	10404	8874
3	90	104	8100	10816	9360
4	88	105	7744	11025	9240
5	86	99	7396	9801	8514
6	86	97	7396	9409	8342
7	87	96	7569	9216	8352
8	90	96	8100	9216	8640
9	92	99	8464	9801	9108
10	93	102	8649	10404	9486
11	91	101	8281	10201	9191
$\sum$	975	1104	86493	110902	97862

$$\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{975}{11} = 88.6364 \\
\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1104}{11} = 100.3636
\end{aligned}$$

Berechnung:

$$\begin{aligned}
r_{XY} &= \frac{97.862 - 11 \cdot 88.6364 \cdot 100.3636}{\sqrt{86493 - 11 \cdot 88.6364^2}\sqrt{110.902 - 11 \cdot 100.3636^2}} \\
&= \frac{7.49}{8.5132 \cdot 10.0312} \\
&= 0,0877
\end{aligned}$$

Bei der Berechnung mit Brüchen würde man folgendes Ergebnis erhalten

$$r_{XY} = \frac{97862 - 11 \cdot \frac{975}{11} \cdot \frac{1104}{11}}{\sqrt{86493 - 11 \cdot \left(\frac{975}{11}\right)^2}\sqrt{110902 - 11 \cdot \left(\frac{1104}{11}\right)^2}} = 0.0873$$

Aufgabe 3.34

Nach einem Trainingslager für die 200 Nachwuchsspieler eines Schachvereins stellen die Trainer folgenden Zusammenhang zwischen der Vereinszugehörigkeit in Jahren X und der Anzahl der Siege bei den Probeturnieren während des Lagers Y fest:

Vereins-zugehörigkeit X über ... bis einschl. ...	Anzahl Siege Y 0 – 1	2 – 4	5 – 6
0 – 1	40	10	0
1 – 2	20	30	5
2 – 4	10	25	15
4 – 8	0	15	30

a) Bestimmen Sie die Randverteilung von X und Y.

b) Geben Sie die Verteilung von X für "fünf bis sechs Siege" an, und errechnen Sie das arithmetische Mittel und die Varianz dieser Verteilung.

c) Berechnen Sie die mittlere Anzahl der Siege bei den Nachwuchspielern, die länger als ein Jahr und höchstens zwei Jahre im Verein sind.

d) Sind die beiden Merkmale unabhängig? Begründen Sie Ihre Aussage, indem Sie eine Maßzahl für die Abhängigkeit berechnen.

Lösung:

a)

Vereins-zugehörigkeit X über ... bis einschl. ...	Anzahl Siege Y 0 – 1	2 – 4	5 – 6	$n_{j\cdot}$	$f_{j\cdot}$
0 – 1	40	10	0	50	0.250
1 – 2	20	30	5	55	0.275
2 – 4	10	25	15	50	0.250
4 – 8	0	15	30	45	0.225
$n_{\cdot k}$	70	80	50	200	
$f_{\cdot k}$	0.35	0.40	0.25		1

b)

j	K_j	ξ_j	n_{j3}	$f_{j\|3}$
1	[0,1]	0.5	0	0.0
2	]1,2]	1.5	5	0.1
3	]2,4]	3.0	15	0.3
4	]4,8]	6.0	30	0.6
$\sum$			50	1.0

$$\bar{x}_3 = \sum_{j=1}^{J} \xi_j f_{j|3} = 4.65$$

$$s^2_{X|3} = \sum_{j=1}^{J} \xi_j^2 f_{j|3} - \bar{x}_3^2 = 24.525 - 21.6225 = 2.9025$$

c)

k	K_k	η_k	n_{2k}
1	[0,2]	0.5	20
2	]2,4]	3	30
3	]4,6]	5.5	5

$$\overline{y}_2 = \frac{1}{n_{2\cdot}} \sum_{k=1}^{K} \eta_k n_{2k} = \frac{1}{55} 127.5 = 2.3182$$

d) Korrelationskoeffizient r_{XY}:

		Y				
	$\xi_j \eta_k n_{jk}$	0,5	3	5,5	$n_{j\cdot}$	ξ_j
	0,5	10	15	0	50	0.5
X	1,5	15	135	41.25	55	1.5
	3	15	225	247.5	50	3.0
	6	0	270	990	45	6.0
	$n_{\cdot k}$	70	80	50	1963.75	
	η_k	0.5	3	5.5		

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{4} \xi_j n_{j\cdot} = 2.6375$$

$$\bar{y} = \frac{1}{n} \sum_{k=1}^{3} \eta_k n_{\cdot k} = 2.75$$

$$
\begin{aligned}
r_{XY} &= \frac{\sum_{j=1}^{4}\sum_{k=1}^{3}\xi_j\eta_k n_{jk} - 200\bar{x}\bar{y}}{\sqrt{\left(\sum_{j=1}^{4}\xi_j^2 n_{j\cdot} - 200\bar{x}^2\right)\left(\sum_{k=1}^{3}\eta_k^2 n_{\cdot k} - 200\bar{y}^2\right)}} \\
&= \frac{1963.25 - 200\cdot 2.6375\cdot 2.75}{\sqrt{(2206.25 - 200\cdot 6.9564)(2250 - 200\cdot 7.5625)}} \\
&= \frac{513.125}{775.268} \\
&= 0.6619
\end{aligned}
$$

Kontingenzkoeffizient C (Pearson):

	$\frac{n_{jk}^2}{n_{j\cdot}n_{\cdot k}}$	Y 1	2	3
X	1	$\frac{1600}{3500}$	$\frac{100}{4000}$	$\frac{0}{2500}$
	2	$\frac{400}{3850}$	$\frac{900}{4400}$	$\frac{25}{2750}$
	3	$\frac{100}{3500}$	$\frac{625}{4000}$	$\frac{225}{2500}$
	4	$\frac{0}{3150}$	$\frac{225}{3600}$	$\frac{900}{2250}$

$$\chi^2 = n\left(\sum_{j=1}^{4}\sum_{k=1}^{3}\frac{n_{jk}^2}{n_{j\cdot}n_{\cdot k}} - 1\right) = 200(1.5370 - 1) = 107.40$$

$$C = \sqrt{\frac{\chi^2}{\chi^2+n}\cdot\frac{\min\{J,K\}}{\min\{J,K\}-1}} = \sqrt{\frac{107.40}{107.40+200}\frac{3}{3-1}} = \sqrt{0.5241} = 0.7239$$

Sowohl r_{XY} als auch C zeigen, dass X und Y nicht unabhängig voneinander sind.

4 Zufallsvorgänge und Wahrscheinlichkeiten

Aufgabe 4.1

Geben Sie für die folgenden Zufallsexperimente die Ergebnismenge in geeigneter Form an:

a) Eine Münze (mit $K =$ „Kopf" und $Z =$ „Zahl") wird zweimal geworfen. Erscheint mindestens einmal „Zahl", so wird die Münze ein drittes Mal geworfen.

b) Ein Würfel wird einmal geworfen. Falls die Augenzahl „Sechs" ist, wird der Würfel ein weiteres Mal geworfen.

c) Eine Münze (mit $K =$ „Kopf" und $Z =$ „Zahl") wird solange geworfen, bis zum ersten Mal „Zahl" zweimal hintereinander erscheint. Die Münze soll jedoch nicht öfter als viermal geworfen werden.

d) Ein Würfel wird solange geworfen, bis die Summe aller geworfenen Augenzahlen mindestens vier ist.

Lösung:

Die Ergebnismengen können wie folgt dargestellt werden:

a) $\Omega = \{KK, KZK, KZZ, ZKK, ZKZ, ZZK, ZZZ\}$, $|\Omega| = 7$.

b) $\Omega = \{1, 2, 3, 4, 5, (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$, $|\Omega| = 11$.

c)

$$\begin{aligned}\Omega = \{&KKKK, KKKZ, KKZK, KKZZ, KZKK, KZKZ,\\ &KZZ, ZKKK, ZKKZ, ZKZK, ZKZZ, ZZ\}\\ |\Omega| = 12&\end{aligned}$$

d)

$$\begin{aligned}\Omega = \{&6, 5, 4, (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),\\ &(2, 1, 1), (2, 1, 2), (2, 1, 3), (2, 1, 4), (2, 1, 5), (2, 1, 6), (1, 3), (1, 4), (1, 5), (1, 6),\\ &(1, 2, 1), (1, 2, 2), (1, 2, 3), (1, 2, 4), (1, 2, 5), (1, 2, 6),\\ &(1, 1, 2), (1, 1, 3), (1, 1, 4), (1, 1, 5), (1, 1, 6),\\ &(1, 1, 1, 1), (1, 1, 1, 2), (1, 1, 1, 3), (1, 1, 1, 4), (1, 1, 1, 5), (1, 1, 1, 6)\}\\ |\Omega| = 41&\end{aligned}$$

Aufgabe 4.2

Aus einem Skat-Kartenspiel mit 32 Blatt wird zufällig eine Karte gezogen. Wie groß ist die Wahrscheinlichkeit

a) einen König oder ein As zu ziehen,

b) eine rote (Herz- oder Karo-) Sieben zu ziehen,

c) eine Kreuz-Karte zu ziehen, die nicht Figur (König, Dame, Bube) ist,

d) keinen Buben zu ziehen?

Lösung:

Es handelt sich um ein Laplace-Experiment, da alle Karten die gleiche Wahrscheinlichkeit haben, gezogen zu werden. Es ist $|\Omega| = 32$.

a) $A = \{\text{König oder As}\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{8}{32} = \frac{1}{4} = 0,25$$

b) $A = \{\text{rote 7}\} = \{\text{Herz 7, Karo 7}\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2}{32} = \frac{1}{16} = 0,0625$$

c) $A = \{\text{Kreuz 7, 8, 9, 10 oder As}\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{5}{32} = 0,15625$$

d) $A = \{\text{keinen Buben}\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{28}{32} = \frac{7}{8} = 0,875$$

Aufgabe 4.3

Betrachten Sie die folgenden Laplace-Experimente.

a) Jemand würfelt sechsmal. Wie groß ist die Wahrscheinlichkeit, dass keine „Eins“ dabei ist?

b) Jemand würfelt dreimal. Wie groß ist die Wahrscheinlichkeit, dass genau die Augenzahlen „Eins“, „Zwei“ und „Drei“ dabei sind?

Lösung:

a) $A = \{\text{keine „Eins“ bei sechsmaligem Würfeln}\}$
Variation: $n = 5$ Elemente (= Augen auf Würfel außer der „Eins“), $k = 6$ werden ausgewählt, Ziehen mit Zurücklegen, Reihenfolge relevant (geordnet)
$=> |A| = 5^6$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{5^6}{6^6} = \left(\frac{5}{6}\right)^6 = 0,3349$$

b) $A = \{\text{„Eins“, „Zwei“, „Drei“ bei Werfen von drei Würfeln}\}$
Permutation: $n = 3$ Elemente, alle werden angeordnet, alle verschieden
$=> |A| = 3!$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3!}{6^3} = 0,0278$$

Aufgabe 4.4

Das Amtsgericht in Astadt sucht vier Schöffen. Der zuständige Beauftragte hat dazu eine Gruppe von sieben Personen zusammengestellt, von denen drei Personen weiblichen und vier Personen männlichen Geschlechts sind. Die Auswahl der Schöffen aus dieser Gruppe erfolgt zufällig.
Wie groß ist die Wahrscheinlichkeit dafür, dass

a) alle Schöffen Männer sind?

b) sich jeweils zwei Frauen und zwei Männer unter den Schöffen befinden?

c) mindestens ein Schöffe weiblich ist?

Lösung:

Die Auswahl erfolgt ohne Beachtung der Reihenfolge und ohne Zurücklegen.

Hinweis: Die gesuchten Wahrscheinlichkeiten können (später) auch mittels der hypergeometrischen Verteilung berechnet werden.

a) $A = \{\text{alle Schöffen sind Männer}\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{3}{0}\binom{4}{4}}{\binom{7}{4}} = \frac{1}{35} = 0,0286$$

b) $B = \{\text{jeweils zwei Frauen und zwei Männer unter den Schöffen}\}$

$$P(B) = \frac{|B|}{|\Omega|} = \frac{\binom{3}{2}\binom{4}{2}}{\binom{7}{4}} = \frac{18}{35} = 0,5143$$

c) $C = \{\text{mindestens ein Schöffe weiblich}\}$

$$P(C) = \frac{|C|}{|\Omega|} = 1 - P(A) = 1 - \frac{\binom{3}{0}\binom{4}{4}}{\binom{7}{4}} = 1 - 0,0286 = 0,9714$$

Oder:

$$P(C) = \frac{\binom{3}{1}\binom{4}{3} + \binom{3}{2}\binom{4}{2} + \binom{3}{3}\binom{4}{1}}{\binom{7}{4}} = 0,9714$$

Aufgabe 4.5

Betrachten Sie das Lotto-Spiel „6 aus 49". Vernachlässigen Sie im Folgenden die Zusatzzahl sowie die Superzahl.

a) Wie groß ist die Wahrscheinlichkeit, dass genau k Richtige ($k = 0, 1, ..., 6$) getippt werden?

b) Bei diesem Lottospiel erhält man einen Gewinn, wenn man mindestens drei Richtige getippt hat. Wie groß ist folglich die Wahrscheinlichkeit, irgendeinen Gewinn zu erzielen?

Lösung:

Hinweis: Die gesuchten Wahrscheinlichkeiten können (später) auch mittels der hypergeometrischen Verteilung berechnet werden.

a) $A = \{k \text{ Richtige getippt}\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}$$

k	$P(\{k \text{ Richtige getippt }\})$
6	0.000 000 07151
5	0.000 018 45
4	0.000 968 6
3	0.017 65
2	0.132 4
1	0.413
0	0.436

b) $B = \{\text{irgendeinen Gewinn}\} = \{3, 4, 5 \text{ oder } 6 \text{ Richtige getippt}\}$

$$P(B) = \sum_{k=3}^{6} \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}} = \frac{20 \cdot 12341 + 15 \cdot 903 + 6 \cdot 43 + 1}{13983816}$$
$$= \frac{260624}{13983816} = 0,01864$$

Aufgabe 4.6

Eine Münze (mit K = „Kopf" und Z = „Zahl") wird so lange geworfen, bis zum zweiten Mal „Kopf" erscheint, jedoch höchstens fünfmal.

a) Geben Sie die Ergebnismenge Ω dieses Zufallsexperimentes explizit an.

b) Geben Sie die folgenden Ereignisse explizit an.

   (i) A : Bei den ersten beiden Würfen erscheint „Kopf" nicht.

   (ii) B : Beim ersten Wurf erscheint „Zahl" nicht.

   (iii) C : Der erste „Kopf" erscheint frühestens beim dritten Wurf.

Lösung:

a)

$$\begin{aligned}\Omega = \{&KK, KZK, KZZK, KZZZK, KZZZZ,\\ &ZKK, ZKZK, ZKZZK, ZKZZZ,\\ &ZZKK, ZZKZK, ZZKZZ,\\ &ZZZKK, ZZZKZ, ZZZZK, ZZZZZ\}\end{aligned}$$

b) Explizite Ereignisse:

   (i)

$$\begin{aligned}A = \{&ZZKK, ZZKZK, ZZKZZ, ZZZKK,\\ &ZZZKZ, ZZZZK, ZZZZZ\}\end{aligned}$$

   (ii)

$$B = \{KK, KZK, KZZK, KZZZK, KZZZZ\}$$

   (iii)

$$\begin{aligned}C = \{&ZZKK, ZZKZK, ZZKZZ, ZZZKK,\\ &ZZZKZ, ZZZZK, ZZZZZ\}\end{aligned}$$

Aufgabe 4.7

In einer Stadt erscheinen die drei Lokalblätter a, b und c. Das Ereignis A sei definiert durch

A : „Ein zufällig ausgewählter erwachsener Einwohner der Stadt liest Blatt a."

Entsprechend seien die Ereignisse B und C definiert. Es ist bekannt, dass

$$\begin{aligned} P(A) &= 0,4 & P(B) &= 0,50 & P(C) &= 0,20 \\ P(A \cap B) &= 0,1 & P(B \cap C) &= 0,05 & P(A \cap C) &= 0,15 \\ & & P(A \cap B \cap C) &= 0,04 & & \end{aligned}$$

Berechnen Sie die Wahrscheinlichkeit dafür, dass ein zufällig ausgewählter erwachsener Einwohner der Stadt

a) mindestens eine der drei Zeitungen liest,

b) ausschließlich b liest,

c) weder a noch b liest,

d) nur b und c liest,

e) höchstens zwei der drei Zeitungen liest.

Lösung:

a)

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \\ &= 0.4 + 0.5 + 0.2 - 0.1 - 0.05 - 0.15 + 0.04 \\ &= 0.84 \end{aligned}$$

b)

$$\begin{aligned} P(\bar{A} \cap B \cap \bar{C}) &= P(B \setminus (A \cup C)) \\ &= P(B) - P\left(B \cap (A \cup C)\right) \\ &= P(B) - P\left((A \cap B) \cup (B \cap C)\right) \\ &= P(B) - [P(A \cap B) + P(B \cap C) - P(A \cap B \cap C)] \\ &= 0.5 - [0.1 + 0.05 - 0.04] \\ &= 0.39 \end{aligned}$$

c)

$$
\begin{aligned}
P(\overline{A \cup B}) &= 1 - P(A \cup B) \\
&= 1 - (P(A) + P(B) - P(A \cap B)) \\
&= 1 - [0.4 + 0.5 - 0.1] \\
&= 0.2
\end{aligned}
$$

d)

$$
\begin{aligned}
P(\bar{A} \cap B \cap C) &= P\left((B \cap C) \setminus A\right) \\
&= P(B \cap C) - P(A \cap B \cap C) \\
&= 0.05 - 0.04 \\
&= 0.01
\end{aligned}
$$

e)

$$
\begin{aligned}
P(\overline{A \cap B \cap C}) &= 1 - P(A \cap B \cap C) \\
&= 1 - 0.04 \\
&= 0.96
\end{aligned}
$$

Aufgabe 4.8

An der Haltestelle einer Regionalverkehrsbahn warten zwölf Personen. Die Bahn besteht aus drei Wagen. Nach Eintreffen der Bahn steigen die Wartenden in die Bahn ein. Jede Person wählt zufällig und unabhängig von den anderen Personen einen Wagen. Wie groß ist (unter der Annahme, dass es sich um ein Laplace-Experiment handelt) die Wahrscheinlichkeit dafür, dass

a) genau fünf Personen in den ersten Wagen steigen?

b) jeweils vier Personen in jeden Wagen steigen?

c) die zwölf Personen sich in Gruppen zu drei, vier und fünf Personen aufteilen?

Lösung:

a) $A = \{\text{genau fünf Personen steigen in den ersten Wagen}\}$

$$P(A) = \frac{\binom{12}{5} \cdot 2^7}{3^{12}} = 0,1908$$

b) $B = \{\text{in jeden Wagen steigen genau vier Personen}\}$

$$P(B) = \frac{\binom{12}{4}\binom{8}{4}\binom{4}{4}}{3^{12}} = \frac{34650}{531441} = 0,0652$$

c) $C = \{\text{Aufteilung in Gruppen zu drei, vier und fünf Personen}\}$

$$P(C) = \frac{\binom{12}{5}\binom{7}{4}\binom{3}{3} \cdot 3!}{3^{12}} = \frac{166320}{531441} = 0,3130$$

Aufgabe 4.9

Einer Urne, die 3 rote und 4 blaue gleichartige Kugeln enthält, werden nacheinander ohne Zurücklegen einzelne Kugeln entnommen. Geben Sie die folgenden Ereignisse explizit an.

a) Spätestens nach dem 4. Zug sind alle roten Kugeln gezogen worden.

b) Es wird fünfmal gezogen und nach dem 5. Zug ist nur noch eine blaue Kugel in der Urne.

Lösung:

Es ist $r =$ rote Kugel und $b =$ blaue Kugel

a) $A = \{rrr, rrbr, rbrr, brrr\}$

b) $B = \{rrbbb, rbrbb, brrbb, rbbrb, rbbbr, bbbrr, bbrbr, bbrrb, brbbr, brbrb\}$

Aufgabe 4.10

Die Studenten Frank, Ralph und Stephan kandidieren als Vorsitzender der WiSo-Fachschaft bzw. als dessen Vertreter. Geben Sie die folgenden Ereignisse und deren Gegenereignisse explizit an.

a) Stephan wird zum Vorsitzenden gewählt.

b) Frank wird Vorsitzender oder Stellvertreter.

c) Ralph wird nicht Vorsitzender.

Lösung:

Wir nehmen an F : Frank, R : Ralph und S = Stephan. Die erste Komponente gibt den Vorsitzenden, die zweite den Stellvertreter an. Es ist $\Omega = \{(F|R), (F|S), (R|F), (R|S), (S|F), (S|R)\}$.

a) $A = \{(S|F), (S|R)\}$

b) $B = \{(F|R), (F|S), (R|F), (S|F)\}$

c) $C = \{(F|R), (F|S), (S|F), (S|R)\}$

Die jeweiligen Gegenereignisse ergeben sich aus $\overline{A} = \Omega - A$ usw.

Aufgabe 4.11

a) Geben Sie die folgenden Verknüpfungen der Ereignisse A und B formal an.

   (i) Es tritt das Ereignis A, aber nicht das Ereignis B ein.

   (ii) Es tritt das Ereignis „A oder B“ ein, aber nicht das Ereignis „A und B“.

b) Von zwei Ereignissen A und B weiß man, dass $A \cap B = \emptyset$ und $A \cup B = \Omega$. Was kann man über die Ereignisse A und B aussagen?

Lösung:

a) Es ist

   (i) $A \cap \overline{B}$

   (ii) $(A \cup B) \cap (\overline{A \cap B}) = (\overline{A} \cap B) \cup (A \cap \overline{B})$

b) B ist das Gegenereignis zu A. Auf einen Beweis sei hier verzichtet.

Aufgabe 4.12

Ein Würfel wird einmal geworfen. A sei das Ereignis „die Augenzahl ist kleiner als 3“, B bedeute das Ereignis „die Augenzahl ist eine ungerade Zahl“. Ermitteln Sie die folgenden Ereignisse! Welche der Ereignisse sind gleich?

a) $A \cap B$

b) $A \cup B$

c) $\overline{A} \cap \overline{B}$

d) $\overline{A} \cup \overline{B}$

e) $\overline{A \cap B}$

f) $\overline{A \cup B}$

g) $A \cap \overline{B}$

h) $\overline{A} \cap B$

i) $A \cup \overline{B}$

j) $\overline{\overline{A} \cup B}$

Lösung:

Es ist $A = \{1, 2\}$, $B = \{1, 3, 5\}$, $\overline{A} = \{3, 4, 5, 6\}$ und $\overline{B} = \{2, 4, 6\}$.

a) $A \cap B = \{1\}$

b) $A \cup B = \{1, 2, 3, 5\}$

c) $\overline{A} \cap \overline{B} = \{4, 6\}$

d) $\overline{A} \cup \overline{B} = \{2, 3, 4, 5, 6\}$

e) $\overline{A \cap B} = \overline{A} \cup \overline{B} = \{2, 3, 4, 5, 6\}$

f) $\overline{A \cup B} = \overline{A} \cap \overline{B} = \{4, 6\}$

g) $A \cap \overline{B} = \{2\}$

h) $\overline{A} \cap B = \{3, 5\}$

i) $A \cup \overline{B} = \{1, 2, 4, 6\}$

j) $\overline{\overline{A} \cup B} = A \cap \overline{B} = \{2\}$

Gleiche Ereignisse: c) und f); d) und e); g) und j).

Aufgabe 4.13

In einer Bank sind zwei unabhängig voneinander arbeitende Geldautomaten aufgestellt. Es ist bekannt, dass während einer Woche die Ausfallwahrscheinlichkeiten für die beiden Automaten $0,3$ bzw. $0,2$ betragen. Wie groß ist die Wahrscheinlichkeit, dass im Laufe einer Woche

a) mindestens ein Geldautomat ausfällt?

b) alle beiden Geldautomaten ausfallen?

c) kein Geldautomat ausfällt?

d) genau ein Geldautomat ausfällt?

Lösung:

Man hat

$$\begin{aligned} P(A := \text{Automat 1 fällt aus}) &= 0,3 = P(A) \\ P(B := \text{Automat 2 fällt aus}) &= 0,2 = P(B) \end{aligned}$$

a)

$$\begin{aligned} &P(\text{mind. 1 Automat fällt aus}) = P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ &\underset{\text{unabhängig}}{=} 0,3 + 0,2 - 0,3 \cdot 0,2 = 0,44 \end{aligned}$$

Oder:

$$\begin{aligned} P(\text{mind. 1 Automat fällt aus}) &= 1 - P(\text{kein Automat fällt aus}) \\ &= 1 - P(\overline{A} \cap \overline{B}) = 1 - 0,7 \cdot 0,8 \\ &= 1 - 0,56 = 0,44 \end{aligned}$$

b)

$$\begin{aligned} P(\text{beide Automaten fallen aus}) &= P(A \cap B) \\ &\underset{\text{unabhängig}}{=} P(A) \cdot P(B) = 0,3 \cdot 0,2 = 0,06 \end{aligned}$$

c) Da Unabhängigkeit auch für Komplementärereignisse gilt, ist

$$\begin{aligned} P(\text{kein Automat fällt aus}) &= P(\overline{A} \cap \overline{B}) = P(\overline{A}) \cdot P(\overline{B}) \\ &= 0,7 \cdot 0,8 = 0,56 \end{aligned}$$

Alternativ kann man auch berechnen:

$$\begin{aligned} P(\text{kein Automat fällt aus}) &= P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) \\ &\underset{\text{Teil a)}}{=} 1 - P(A \cup B) = 1 - 0,44 = 0,56 \end{aligned}$$

d)

$$\begin{aligned} P(\text{genau 1 Automat fällt aus}) &= P\left((A \cap \overline{B}) \cup (\overline{A} \cap B)\right) \\ &\underset{\text{disjunkt}}{=} P(A \cap \overline{B}) + P(\overline{A} \cap B) \\ &\underset{\text{unabhängig}}{=} P(A) \cdot P(\overline{B}) + P(\overline{A}) \cdot P(B) \\ &= 0,3 \cdot 0,8 + 0,7 \cdot 0,2 = 0,38 \end{aligned}$$

Aufgabe 4.14

7% der Produktion eines Artikels besitzen den Fehler F_1; 5% besitzen den Fehler F_2. 90% der Produktion sind fehlerfrei. Wie groß ist die Wahrscheinlichkeit, dass ein aus der Gesamtproduktion zufällig ausgewählter Artikel

a) beide Fehler besitzt?

b) beide Fehler besitzt, unter der Bedingung, dass er den Fehler F_1 besitzt?

c) beide Fehler besitzt, unter der Bedingung, dass er mindestens einen Fehler besitzt?

d) den Fehler F_2 besitzt, unter der Bedingung, dass er den Fehler F_1 nicht besitzt?

e) genau einen Fehler besitzt, unter der Bedingung, dass er den Fehler F_1 besitzt?

Lösung:

Gegeben:

$$\begin{aligned} P(F_1) &= 0,07 \\ P(F_2) &= 0,05 \\ P(\overline{F_1 \cup F_2}) &= 0,9 = 1 - P(F_1 \cup F_2) = P(\overline{F}_1 \cap \overline{F}_2) \end{aligned}$$

Vier-Felder Tafel (Fett gedruckte Ziffern = Werte aus Aufgabenstellung):

$\cap$	F_2	$\overline{F}_2$	
F_1	$0,02$	$0,05$	$\mathbf{0,07}$
$\overline{F}_1$	$0,03$	$\mathbf{0,9}$	$0,93$
	$\mathbf{0,05}$	$0,95$	$\mathbf{1}$

a) Gesucht: $P(F_1 \cap F_2)$

$$\begin{aligned} P(F_1 \cup F_2) &= P(F_1) + P(F_2) - P(F_1 \cap F_2) \\ P(F_1 \cap F_2) &= P(F_1) + P(F_2) - P(F_1 \cup F_2) \\ &= 0,07 + 0,05 - 0,1 \\ &= 0,02 \end{aligned}$$

b) Gesucht: $P(F_1 \cap F_2 | F_1)$

$$\begin{aligned} P(F_1 \cap F_2 | F_1) &= \frac{P(F_1 \cap F_2 \cap F_1)}{P(F_1)} = \frac{P(F_2 \cap F_1)}{P(F_1)} \\ &= \frac{0,02}{0,07} = 0,2857 \end{aligned}$$

c) Gesucht: $P(F_1 \cap F_2 | F_1 \cup F_2)$

$$\begin{aligned} P(F_1 \cap F_2 | F_1 \cup F_2) &= \frac{P(F_1 \cap F_2) \cap (F_1 \cup F_2))}{P(F_1 \cup F_2)} \\ &= \frac{P(F_1 \cap F_2)}{P(F_1 \cup F_2)} = \frac{0,02}{0,1} = 0,2 \end{aligned}$$

d) Gesucht: $P(F_2 | \overline{F}_1)$

$$P(F_2 | \overline{F}_1) = \frac{P(F_2 \cap \overline{F}_1)}{P(\overline{F}_1)} = \frac{0,03}{0,93} = 0,032258$$

e) Gesucht: $P(F_1 \cap \overline{F}_2 | F_1)$

$$\begin{aligned} P(F_1 \cap \overline{F}_2 | F_1) &= \frac{P(F_1 \cap \overline{F}_2 \cap F_1)}{P(F_1)} = \frac{P(F_1 \cap \overline{F}_2)}{P(F_1)} \\ &= \frac{0,05}{0,07} = 0,714286 \end{aligned}$$

Aufgabe 4.15

Ein medizinischer Test für eine bestimmte Krankheit liefert mit einer Wahrscheinlichkeit von 99% das richtige Ergebnis, d.h. wenn der Patient krank ist, ist der Test mit 99%-iger Wahrscheinlichkeit positiv; und wenn der Patient gesund ist, ist der Test mit 99%-iger Wahrscheinlichkeit negativ. Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person krank ist, beträgt 0,001 (ein Tausendstel !). Eine zufällig ausgewählte Person wird nun dem Test unterzogen. Der Test fällt positiv aus. Wie groß ist die Wahrscheinlichkeit, dass die Person wirklich an der Krankheit leidet?

Lösung:

Setze
$$\begin{aligned} P &:= \{\text{Test positiv}\} \\ \overline{P} &:= \{\text{Test negativ}\} \\ K &:= \{\text{Person krank}\} \\ \overline{K} &:= \{\text{Person nicht krank}\} \end{aligned}$$

Gegeben:
$$\begin{aligned} P(P|K) &= 0,99 \\ P(\overline{P}|\overline{K}) &= 0,99 \\ P(K) &= 0,001 \end{aligned}$$

Gesucht: $P(K|P)$.

Anwendung der Formel von Bayes (wobei K und $\overline{K}$ eine disjunkte Zerlegung von Ω bilden)

$$\begin{aligned} P(K|P) &= \frac{P(K) \cdot P(P|K)}{P(K) \cdot P(P|K) + P(\overline{K}) \cdot P(P|\overline{K})} \\ &= \frac{P(K) \cdot P(P|K)}{P(K) \cdot P(P|K) + (1 - P(K))(1 - P(\overline{P}|\overline{K}))} \\ &= \frac{0,001 \cdot 0,99}{0,001 \cdot 0,99 + 0,999 \cdot 0,01} \\ &= \frac{0,00099}{0,01098} = 0,0902 \end{aligned}$$

Aufgabe 4.16

In einer Schraubenfabrik wird an drei Maschinen A, B und C produziert. Maschine A produziert 25%, B produziert 35% und C produziert 40% der Schrauben. Die Ausschussanteile betragen 5% an Maschine A, 4% an Maschine B und 2% an Maschine C. Eine Schraube der Gesamtproduktion wird zufällig ausgewählt. Sie ist kaputt. Wie groß ist die Wahrscheinlichkeit, dass sie an Maschine A (B, C) produziert wurde?

Lösung:

Setze $K :=$ ´Schraube ist kaputt´

Gegeben:
$$\begin{array}{ll} P(A) = 0,25 & P(K|A) = 0,05 \\ P(B) = 0,35 & P(K|B) = 0,04 \\ P(C) = 0,4 & P(K|C) = 0,02 \end{array}$$

Gesucht: $P(A|K), P(B|K), P(C|K)$; (Berechnung über Bayes–Theorem)

$$\begin{aligned} P(A|K) &= \frac{P(A) \cdot P(K|A)}{P(A) \cdot P(K|A) + P(B) \cdot P(K|B) + P(C) \cdot P(K|C)} \\ &= \frac{0,25 \cdot 0,05}{0,25 \cdot 0,05 + 0,35 \cdot 0,04 + 0,4 \cdot 0,02} \\ &= \frac{0,0125}{0,0345} = 0,362319 \end{aligned}$$

$$P(B|K) = \frac{0,35 \cdot 0,04}{0,0345} = 0,405797$$

$$P(C|K) = \frac{0,4 \cdot 0,02}{0,0345} = 0,231884$$

Aufgabe 4.17

Eine unverfälschte Münze (mit $K =$ „Kopf“ und $Z =$ „Zahl“) wird dreimal geworfen. Betrachtet werden die folgenden Ereignisse:

A : In den ersten zwei Würfen liegt jeweils die gleiche Seite oben.
B : In den letzten zwei Würfen liegt jeweils die gleiche Seite oben.
C : Im ersten und im letzten Wurf liegt die gleiche Seite oben.

Prüfen Sie, ob die Ereignisse A, B und C paarweise unabhängig sind. Sind die drei Ereignisse global unabhängig?

Lösung:

Es ist Ω die Menge aller 3er-Tupel, bestehend aus den Werten K und Z. Somit ist $|\Omega| = 2^3 = 8$.
Man hat zunächst:

$$\begin{aligned}
A = &\ \text{„in den ersten zwei Würfen liegt die gleiche Seite oben“} \\
= &\ \{KKZ, KKK, ZZZ, ZZK\} \\
B = &\ \text{„in den letzten zwei Würfen liegt die gleiche Seite oben“} \\
= &\ \{KKK, ZKK, KZZ, ZZZ\} \\
C = &\ \text{„im ersten und letzten Wurf liegt die gleiche Seite oben“} \\
= &\ \{ZZZ, ZKZ, KZK, KKK\}
\end{aligned}$$

mit

$$P(A) = P(B) = P(C) = \frac{4}{8} = \frac{1}{2}.$$

Es ist

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{2}{8} = \frac{1}{4}.$$

Also sind die drei Ereignisse paarweise unabhängig.
Nun ist aber

$$P(A \cap B \cap C) = \frac{2}{8} = \frac{1}{4} \neq \frac{1}{8} = \left(\frac{1}{2}\right)^3 = P(A) \cdot P(B) \cdot P(C)$$

somit sind die drei Ereignisse A, B und C nicht global unabhängig.

Aufgabe 4.18

Ein Industrieprodukt bestehe aus drei Teilen. In der Fertigung bestehe Unabhängigkeit zwischen den Teilen. Die Wahrscheinlichkeit, dass ein Teil Ausschuss ist, betrage für die drei Teile jeweils 3%, 5% und 6%.
Wie groß ist die Wahrscheinlichkeit, dass das Endprodukt völlig einwandfrei ist?

Lösung:

Man berechnet

$$P = (1 - 0,03) \cdot (1 - 0,05) \cdot (1 - 0,06) = 0,8662$$

Aufgabe 4.19

In einer Reifenfabrik werden 40% der Reifen während der Frühschicht, 40% der Reifen während der Spätschicht und die restlichen 20% der Reifen während der Nachtschicht produziert. Es ist bekannt, dass der Anteil fehlerhafter Reifen für die Frühschicht 2%, für die Spätschicht 3% und für die Nachtschicht 8% beträgt.

a) Wie hoch ist die Wahrscheinlichkeit, dass ein in dieser Fabrik produzierter Reifen fehlerhaft ist?

b) Wie groß ist die Wahrscheinlichkeit, dass ein Reifen, von dem bekannt ist, dass er fehlerhaft ist, während der Nachtschicht (der Frühschicht, der Spätschicht) produziert wurde?

Lösung:

Wir definieren die Ereignisse

A_1: „Reifen wird während der Frühschicht produziert."
A_2: „Reifen wird während der Spätschicht produziert."
A_3: „Reifen wird während der Nachtschicht produziert."
B: „Reifen ist fehlerhaft."

Gegeben sind die Wahrscheinlichkeiten

$$\begin{aligned} P(A_1) &= 0.4, \quad P(B|A_1) = 0.02, \\ P(A_2) &= 0.4, \quad P(B|A_2) = 0.03, \\ P(A_3) &= 0.2, \quad P(B|A_3) = 0.08. \end{aligned}$$

a) Gesucht ist $P(B)$.
Mit dem Satz von der totalen Wahrscheinlichkeit erhält man

$$\begin{aligned} P(B) &= \sum_{i=1}^{3} P(B|A_i)P(A_i) \\ &= 0.02 \cdot 0.4 + 0.03 \cdot 0.4 + 0.08 \cdot 0.2 \\ &= 0.036 \end{aligned}$$

b) Gesucht ist $P(A_3|B)$.
Mit dem Satz von Bayes erhält man

$$P(A_3|B) = \frac{P(B|A_3)P(A_3)}{\sum_{i=1}^{3} P(B|A_i)P(A_i)} = \frac{0.08 \cdot 0.2}{0.036} = 0.4444\,.$$

Entsprechend erhält man

$$P(A_1|B) = \frac{0.02 \cdot 0.4}{0.036} = 0.2222 \quad \text{bzw.} \quad P(A_2|B) = \frac{0.03 \cdot 0.4}{0.036} = 0.3333\,.$$

Aufgabe 4.20

Wie oft muss ein fairer Würfel mindestens geworfen werden, damit die Wahrscheinlichkeit, mindestens einmal eine gerade Augenzahl zu werfen, mindestens $0,9$ ist?

Lösung:

Sei A_i das Ereignis, dass im i-ten Wurf eine gerade Augenzahl geworfen wird. Die Ereignisse A_i sind stochastisch unabhängig, und es ist (weil es ein fairer Würfel ist) $P(A_i) = \frac{1}{2}$ für alle $i = 1, ..., n$.

Nun ist n derart gesucht, dass

$$
\begin{aligned}
P(A_1 \cup ... \cup A_n) &\geq 0,9 \\
1 - P(\overline{A_1 \cup ... \cup A_n}) &\geq 0,9 \\
1 - P(\overline{A_1} \cap ... \cap \overline{A_n}) &\geq 0,9 \\
P(\overline{A_1} \cap ... \cap \overline{A_n}) &\leq 0,1 \\
\left[P(\overline{A_1})\right]^n &\leq 0,1 \qquad \text{weil die } A_i \text{ unabhängig sind} \\
\left(\frac{1}{2}\right)^n &\leq 0,1 \\
n &\geq \frac{\ln 0,1}{\ln 0,5} = 3,3219
\end{aligned}
$$

Also muss der Würfel mindestens viermal geworfen werden.

Aufgabe 4.21

Ein Unternehmen lässt ein bestimmtes Bauteil in den Betrieben A und B herstellen, und zwar 70 % in A. Bei der Produktion können zwei Fehler auftreten, nämlich F1 und F2. Von Betrieb A ist bekannt, dass

- 90 % der gefertigten Bauteile fehlerfrei sind,
- 7 % den Fehler F1 aufweisen,
- 5 % den Fehler F2 aufweisen.

Von Betrieb B ist bekannt, dass

- 4 % der gefertigten Bauteile den Fehler F1 aufweisen,
- alle Bauteile, die den Fehler F1 aufweisen, ebenfalls den Fehler F2 aufweisen,
- 95 % der gefertigten Bauteile fehlerfrei sind.

Wie groß ist die Wahrscheinlichkeit, dass ein aus der Gesamtproduktion des Unternehmens zufällig ausgewähltes Bauteil

a) beide Fehler aufweist?

b) aus A stammt, falls es den Fehler F1 aufweist?

c) den Fehler F1 nicht aufweist, falls es den Fehler F2 aufweist?

d) Sind die Ereignisse ,,Fehler F1 tritt auf" und ,,Fehler F2 tritt auf" stochastisch unabhängig?

Ein Tipp zur Bearbeitung: Erstellen Sie zuerst die Vier-Felder-Tafeln für die beiden Betriebe getrennt.

Lösung:

Es ist $\overline{F}_1 \cap \overline{F}_2$ = Zustand "fehlerfrei", dann sind laut Aufgabenstellung folgende Wahrscheinlichkeiten für Produktionsstätte A gegeben:

$$P(F_1 \mid A) = 0,07 \quad , \quad P(\overline{F}_1 \cap \overline{F}_2 \mid A) = 0,9 \quad ,$$
$$P(F_2 \mid A) = 0,05 \quad , \quad P(A) = 0,7$$

Für Produktionsstätte B hat man:

$$P(F_1 \mid B) = 0,04 \qquad , \qquad P(\overline{F}_1 \cap \overline{F}_2 \mid B) = 0,95 \qquad ,$$

$$P(F_2 \mid F_1 \cap B) = \frac{P(F_1 \cap F_2 \mid B)}{P(F_1 \mid B)} = 1 \qquad , \qquad P(B) = 0,3$$

Für jede Produktionsstätte kann man eine Vierfelder-Tafel aufstellen
(die kursiven Werte ergeben sich aus der Aufgabenstellung).

Für A:

$P_A(.\cap.)$	F_1	$\overline{F}_1$	
F_2	$0,02$	$0,03$	*0,05*
$\overline{F}_2$	$0,05$	*0,9*	$0,95$
	0,07	$0,93$	*1*

und für B:

$P_B(.\cap.)$	F_1	$\overline{F}_1$	
F_2	*0,04*	$0,01$	$0,05$
$\overline{F}_2$	0	*0,95*	$0,95$
	0,04	$0,96$	*1*

a)

$$\begin{aligned} P(F_1 \cap F_2) &= P(F_1 \cap F_2 \mid A) \cdot P(A) + P(F_1 \cap F_2 \mid B) \cdot P(B) \\ &= 0,026 \end{aligned}$$

b)

$$P(A \mid F_1) = \frac{P(F_1 \mid A) \cdot P(A)}{P(F_1 \mid A) \cdot P(A) + P(F_1 \mid B) \cdot P(B)} = 0,803$$

c)

$$\begin{aligned} P(\overline{F}_1 \mid F_2) &= \frac{P(\overline{F}_1 \cap F_2)}{P(F_2)} \\ &= \frac{P(\overline{F}_1 \cap F_2 \mid A) \cdot P(A) + P(\overline{F}_1 \cap F_2 \mid B) \cdot P(B)}{P(F_2 \mid A) \cdot P(A) + P(F_2 \mid B) \cdot P(B)} = 0,48 \end{aligned}$$

d) Da $P(F_1) \cdot P(F_2) = 0,061 \cdot 0,05 = 0,003 \neq 0,026 = P(F_1 \cap F_2)$, sind die Ereignisse (stochastisch) abhängig.
$P(F_1) = P(F_1 \mid A) \cdot P(A) + P(F_1 \mid B) \cdot P(B) = 0,07 \cdot 0,7 + 0,04 \cdot 0,3 = 0,061$
$P(F_2) = P(F_2 \mid A) \cdot P(A) + P(F_2 \mid B) \cdot P(B) = 0,05 \cdot 0,7 + 0,05 \cdot 0,3 = 0,05$

5 Zufallsgrößen und Verteilungen

Aufgabe 5.1

Ein Würfel wird zweimal geworfen. Der Wert der Zufallsvariablen X sei die kleinere der beiden geworfenen Augenzahlen bzw. die geworfene Augenzahl, wenn beide Würfel die gleiche Seite zeigen. Geben Sie die Wahrscheinlichkeitsfunktion und die Verteilungsfunktion von X an.

Lösung:

Sei X das Minimum der beiden geworfenen Augenzahlen. Der Ergebnisraum ist $\Omega = \{(i,j) \text{ mit } i,j \in \{1,2,3,4,5,6\}\}$. Es ist $|\Omega| = 6^2 = 36$.
Die Wahrscheinlichkeitsfunktion lautet

$$f_X(x) = P(X = x) = \begin{cases} \frac{11}{36} & \text{für } x = 1, \\ \frac{9}{36} & \text{für } x = 2, \\ \frac{7}{36} & \text{für } x = 3, \\ \frac{5}{36} & \text{für } x = 4, \\ \frac{3}{36} & \text{für } x = 5, \\ \frac{1}{36} & \text{für } x = 6, \end{cases}$$

und man erhält die zugehörige Verteilungsfunktion

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{für } x < 1, \\ \frac{11}{36} & \text{für } 1 \leq x < 2, \\ \frac{20}{36} & \text{für } 2 \leq x < 3, \\ \frac{27}{36} & \text{für } 3 \leq x < 4, \\ \frac{32}{36} & \text{für } 4 \leq x < 5, \\ \frac{35}{36} & \text{für } 5 \leq x < 6, \\ 1 & \text{für } 6 \leq x. \end{cases}$$

Aufgabe 5.2

Sei X eine diskrete Zufallsvariable mit $T_X = \{1, 2, 3, 4, 5, 6\}$ und

$$P(X = x) = \begin{cases} 0,1 & \text{für } x = 1, 2, 3, 4, 5, \\ 0,5 & \text{für } x = 6. \end{cases}$$

a) Geben Sie die Verteilungsfunktion von X an.

b) Berechnen Sie $E[X]$ und $V[X]$.

c) Geben Sie die Quantilfunktion von X an.

Lösung:

a) Da die ZV. X diskret ist, ist die Verteilungsfunktion eine Treppenfunktion. Sie lautet

$$F_X(x) = \begin{cases} 0 & \text{für} \quad x < 1, \\ 0,1 & \text{für} \quad 1 \leq x < 2, \\ 0,2 & \text{für} \quad 2 \leq x < 3, \\ 0,3 & \text{für} \quad 3 \leq x < 4, \\ 0,4 & \text{für} \quad 4 \leq x < 5, \\ 0,5 & \text{für} \quad 5 \leq x < 6, \\ 1 & \text{für} \quad 6 \leq x. \end{cases}$$

b) Für Erwartungswert und Varianz erhält man

$$\begin{aligned} E[X] &= \sum_{x=1}^{6} x \cdot P(X = x) = 4,5 \\ E[X^2] &= \sum_{x=1}^{6} x^2 \cdot P(X = x) = 23,5 \\ V[X] &= E[X^2] - E[X]^2 = 23,5 - 20,25 = 3,25 \end{aligned}$$

c) Die Quantilfunktion lautet

$$Q_X(p) = \begin{cases} 1 & \text{für} \quad 0 < p \leq 0,1, \\ 2 & \text{für} \quad 0,1 < p \leq 0,2, \\ 3 & \text{für} \quad 0,2 < p \leq 0,3, \\ 4 & \text{für} \quad 0,3 < p \leq 0,4, \\ 5 & \text{für} \quad 0,4 < p \leq 0,5, \\ 6 & \text{für} \quad 0,5 < p < 1. \end{cases}$$

Aufgabe 5.3

Eine (diskrete) Zufallsvariable X habe die Wahrscheinlichkeitsfunktion

$$f(x) = \begin{cases} \frac{1}{6}k & \text{für } x = -1, \\ \frac{5}{12}k & \text{für } x = 0, \\ \frac{1}{12}k & \text{für } x = 1, \\ 0 & \text{sonst.} \end{cases}$$

a) Bestimmen Sie den Wert der Konstanten k.

b) Geben Sie die Verteilungsfunktion von X an.

c) Berechnen Sie den Erwartungswert und die Varianz dieser Zufallsvariablen.

d) Berechnen Sie die Wahrscheinlichkeit für die folgenden Ereignisse:

(i) $A = \{X > 0\}$
(ii) $B = \{X \leq 0,5\}$
(iii) $C = \{-0,8 \leq X < 1\}$
(iv) $D = \{0,3 < X \leq 2\}$.

Lösung:

a) Für eine diskrete Zufallsvariable muss gelten:

$$\sum_{i=1}^{\infty} f(x_i) = 1$$

hier:

$$\begin{aligned} \sum_{i=1}^{3} f(x_i) &= 1 \\ \frac{1}{6}k + \frac{5}{12}k + \frac{1}{12}k &= 1 \\ \frac{8}{12}k &= 1 \\ k &= \frac{3}{2} = 1,5 \end{aligned}$$

b) Es ist

$$f(x) = \begin{cases} \frac{2}{8} & \text{für } x = -1, \\ \frac{5}{8} & \text{für } x = 0, \\ \frac{1}{8} & \text{für } x = 1, \\ 0 & \text{sonst.} \end{cases}$$

und

$$F(x) = \begin{cases} 0 & \text{für } x < -1, \\ \frac{2}{8} & \text{für } -1 \leq x < 0, \\ \frac{7}{8} & \text{für } 0 \leq x < 1, \\ 1 & \text{für } 1 \leq x. \end{cases}$$

c) X diskret:

$$E[X] = \sum_{i=1}^{3} x_i f(x_i) = -1 \cdot \frac{2}{8} + 0 + 1 \cdot \frac{1}{8} = -\frac{1}{8}$$

$$E[X^2] = \sum_{i=1}^{3} x_i^2 f(x_i) = (-1)^2 \cdot \frac{2}{8} + 0 + 1^2 \cdot \frac{1}{8} = \frac{3}{8}$$

$$V[X] = E[X^2] - E[X]^2 = \frac{3}{8} - \left(-\frac{1}{8}\right)^2 = \frac{23}{64}$$

d) Die Wahrscheinlichkeiten lassen sich über die Verteilungsfunktion und/oder Wahrscheinlichkeitsfunktion berechnen als

(i)

$$P(A) = P(X > 0) = f(1) = 1 - F(0) = 1 - \frac{7}{8} = \frac{1}{8}$$

(ii)

$$P(B) = P(X \leq 0,5) = f(-1) + f(0) = F(0,5) = \frac{7}{8}$$

(iii)

$$P(C) = P(-0,8 \leq X < 1) = f(0) = \frac{5}{8}$$

(iv)

$$P(D) = P(0,3 < X \leq 2) = f(1) = F(2) - F(0,3) = 1 - \frac{7}{8} = \frac{1}{8}$$

Aufgabe 5.4

Gegeben sei die Verteilungsfunktion

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0, \\ \frac{1}{4}x^2 & \text{für } 0 \leq x \leq 2, \\ 1 & \text{für } x > 2. \end{cases}$$

a) Bestimmen Sie die Dichte von X und geben Sie den Träger von X an.

b) Geben Sie die zugehörige Quantilfunktion Q_X an.

c) Berechnen Sie den Median, den Erwartungswert und die Varianz von X.

d) Berechnen Sie die Wahrscheinlichkeit für das Ereignis $\{X < 0,5\}$.

Lösung:

a) Durch Differentiation der Verteilungsfunktion erhält man die Dichtefunktion:

$$f(x) = F'(x) = \begin{cases} \frac{1}{2}x & \text{für } 0 \leq x \leq 2, \\ 0 & \text{sonst.} \end{cases}$$

Der Träger ist folglich das Intervall $[0; 2]$.

b) Verteilungsfunktion:
$F(x) = p =$ Wahrscheinlichkeit, dass X einen Wert $\leq x$ annimmt.
Quantilfunktion:
$Q(p) = x_p = \min\{x | F(x) \geq p\}$.
Die folgende Beziehung muss nach x_p aufgelöst werden:

$$\begin{aligned} F(x_p) &= p \\ \frac{1}{4}x_p^2 &= p \\ x_p &= 2\sqrt{p} \end{aligned}$$

Somit ist

$$Q(p) = 2\sqrt{p} \qquad \text{für } 0 < p < 1.$$

(Die negative Lösung $-2\sqrt{p}$ beim Wurzelziehen entfällt, da der Träger $T_X = [0; 2]$ ist.)

c) Median $x_{0,5}$ wird mit Wahrscheinlichkeit $p = 0,5$ erreicht oder unterschritten. Einsetzen von $p = 0,5$ in obige Beziehung ergibt:

$$\begin{aligned} F(x_{0,5}) &= 0,5 \\ x_{0,5} &= 2\sqrt{0,5} = \sqrt{2} \approx 1,4142 \end{aligned}$$

Erwartungswert:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{2} \frac{1}{2} x^2 dx = \left[\frac{1}{6} x^3\right]_0^2 = \frac{8}{6} = \frac{4}{3}$$

Varianz:

$$\begin{aligned} E\left[X^2\right] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{0}^{2} \frac{1}{2} x^3 dx = \left[\frac{1}{8} x^4\right]_0^2 = \frac{16}{8} = 2 \\ V[X] &= E\left[X^2\right] - E\left[X\right]^2 = 2 - \frac{16}{9} = \frac{2}{9} \end{aligned}$$

d) Gesucht:

$$P(X < 0,5) = F(0,5) = \frac{1}{4}(0,5)^2 = \frac{1}{16} = 0,0625$$

Aufgabe 5.5

Gegeben sei die Funktion

$$f(x) = \begin{cases} 2x - c & \text{für } 1 \leq x \leq 2, \\ 0 & \text{sonst.} \end{cases}$$

a) Bestimmen Sie c so, dass f eine Dichtefunktion ist.

b) Berechnen Sie die zugehörige Verteilungsfunktion und die Quantilfunktion.

c) Berechnen Sie das p-Quantil von X für die Werte $p = 0,2$ und $p = 0,9$.

d) Wie groß ist die Wahrscheinlichkeit dafür, dass X vom Erwartungswert um weniger als $\frac{1}{3}$ abweicht?

Lösung:

a) Es müssen zwei Bedingungen erfüllt sein, damit $f(x)$ eine Dichtefunktion ist:

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

$$\int_1^2 f(x)dx = \int_1^2 (2 \cdot x - c)\ dx = \left[x^2 - c \cdot x\right]_1^2$$
$$= (4 - 2c) - (1 - c) = 3 - c = 1$$
$$\Longrightarrow c = 2$$

- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
  $f(x) = 0 \geq 0$ für $x < 1$ und $x > 2$.
  $f(x) = 2x - 2 \geq 0$ für $1 \leq x \leq 2$.

Für $c = 2$ ist $f(x) := \begin{cases} 2x - 2 & \text{für } 1 \leq x \leq 2, \\ 0 & \text{sonst,} \end{cases}$ eine Dichtefunktion.

b) Verteilungsfunktion:

$$F(x) = \begin{cases} 0 & \text{für } x < 1, \\ \int\limits_1^x f(z)\ dz & \text{für } 1 \leq x \leq 2, \\ 1 & \text{für } x > 2. \end{cases}$$

Dabei ist:

$$\int\limits_1^x f(z)\ dz = \int\limits_1^x (2z - 2)\ dz = \left[z^2 - 2z\right]_1^x = x^2 - 2x + 1$$

Quantilfunktion: Auflösen von $p = F(x_p) = x_p^2 - 2x_p + 1$ nach x_p ergibt

$$\begin{aligned} x_p^2 - 2x_p + 1 &= p \\ x_p^2 - 2x_p + 1 - p &= 0 \end{aligned}$$

Lösen der Gleichung über p, q–Formel (quadratische Ergänzung) liefert:

$$x_{p;1,2} = 1 \pm \sqrt{1 - (1 - p)}$$

Da bei der Umkehrfunktionsbildung hier der Bereich $1 \leq x \leq 2$ relevant ist, muss die Lösung $x_1 = 1 + \sqrt{p}$ richtig sein, da nur hier $x-$Werte größer Eins herauskommen. Somit:

$$Q(p) = 1 + \sqrt{p} \qquad \text{für } 0 < p < 1.$$

c) 0,2-Quantil:

$$\begin{aligned} x_{0,2} &= Q(0,2) = 1 + \sqrt{0,2} \\ &= 1,4472 \end{aligned}$$

0,9-Quantil:

$$\begin{aligned} x_{0,9} &= Q(0,9) = 1 + \sqrt{0,9} \\ &= 1,9487 \end{aligned}$$

d) Erwartungswert:

$$E[X] = \int_1^2 x f(x)\ dx = \int_1^2 \left(2x^2 - 2x\right)\ dx = \left[\frac{2}{3}x^3 - x^2\right]_1^2$$
$$= \frac{16}{3} - 4 - \frac{2}{3} + 1$$
$$= \frac{5}{3}$$

Gesuchte Wahrscheinlichkeit:

$$P\left(|X - E[X]| < \frac{1}{3}\right) = P\left(E[X] - \frac{1}{3} < X < E[X] + \frac{1}{3}\right)$$
$$= P\left(\frac{4}{3} < X < 2\right) = F(2) - F\left(\frac{4}{3}\right)$$
$$= 1 - \left[\left(\frac{4}{3}\right)^2 - \frac{8}{3} + 1\right]$$
$$= 1 - \frac{1}{9} = \frac{8}{9}$$
$$= 0,8889$$

Zum Vergleich: Abschätzung mittels Tschebyscheff-Ungleichung

$$E[X^2] = \int_1^2 x^2 f(x)\ dx = \left[\frac{2}{4}x^4 - \frac{2}{3}x^3\right]_1^2 = \frac{17}{6}$$
$$V[X] = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}$$
$$P\left(|X - E[X]| < \frac{1}{3}\right) \geq_{\varepsilon = \frac{1}{3}} 1 - \frac{\frac{1}{18}}{\left(\frac{1}{3}\right)^2} = 0,5$$

Aufgabe 5.6

Die Zufallsvariable X habe Erwartungswert -3 und Varianz 25.

a) Bestimmen Sie Erwartungswert und Standardabweichung der Zufallsvariablen $Y = -4X + \frac{1}{2}$.

b) Schätzen Sie die Wahrscheinlichkeit ab, dass die Zufallsvariable X einen Wert annimmt, der sich vom Erwartungswert um mindestens zehn unterscheidet.

Lösung:

Man hat $E[X] = -3$ und $V[X] = 25$.

a) Für die lineare Transformation einer Zufallsvariablen gilt:

$$E[aX + b] = aE[X] + b$$
$$V[aX + b] = a^2 V[X]$$

Damit ist

$$E[Y] = E\left[-4X + \frac{1}{2}\right] = -4 \cdot E[X] + \frac{1}{2} = 12,5$$
$$V[Y] = V\left[-4X + \frac{1}{2}\right] = (-4)^2 \cdot V[X] = 400$$
$$\sqrt{V[Y]} = 20$$

b) Unter Anwendung der Tschebyscheff´schen Ungleichung berechnet man

$$P\left(|X - E[X]| \geq 10\right)$$
$$= P\left(|X - E[X]| \geq 2 \cdot \sqrt{V[X]}\right)$$
$$\leq \frac{1}{2^2} = 0,25$$

Aufgabe 5.7

Sei X die Wertveränderung (in %) der Aktie A am morgigen Tag. Die Verteilungsfunktion von X sei

$$F(x) = \frac{e^x}{1+e^x}$$

a) Bestimmen Sie die Quantilfunktion von X.

b) Bestimmen Sie den Median und den Quartilsabstand von X.

c) In der Finanzierung werden Quantile x_p auch als "value at risk" (VaR) bezeichnet, da sie angeben, welche Wertminderung (Wertsteigerung) mit einer Wahrscheinlichkeit von $(1-p)$ nicht überschritten (unterschritten) wird. Bestimmen Sie den VaR für die Aktie A für eine Wahrscheinlichkeit von 95%.

Lösung:

a) Die Quantilfunktion $Q(p)$ ist definiert als

$$Q(p) = \min(x|F(x) \geq p) = x_p\,, \qquad p \in [0,1]\,,$$

d. h. $Q(p)$ ist der Wert, der mit Wahrscheinlichkeit p angenommen oder unterschritten wird, also $F(x_p) = p$. Durch Auflösen der folgenden Gleichung nach x_p erhält man die Quantilfunktion Q :

$$\begin{aligned}
\frac{e^{x_p}}{1+e^{x_p}} &= p \\
e^{x_p} &= p(1+e^{x_p}) \\
e^{x_p} &= p + pe^{x_p} \\
e^{x_p} - pe^{x_p} &= p \\
e^{x_p}(1-p) &= p \\
e^{x_p} &= \frac{p}{1-p} \\
x_p &= \ln\left(\frac{p}{1-p}\right)\,. \\
Q(p) &= \ln\left(\frac{p}{1-p}\right)\,.
\end{aligned}$$

b) Median:

$$\begin{aligned}x_{0,5} &= Q(0,5)\\ &= \ln\left(\frac{0,5}{1-0,5}\right)\\ &= \ln\left(\frac{0,5}{0,5}\right)\\ &= \ln 1\\ &= 0\,.\end{aligned}$$

Quartilsabstand:

$$\begin{aligned}&x_{0,75}-x_{0,25}\\ =&Q(0,75)-Q(0,25) = \ln\left(\frac{0,75}{0,25}\right) - \ln\left(\frac{0,25}{0,75}\right)\\ =&\ln(3) - \ln\left(\frac{1}{3}\right) = \ln\left(\frac{3}{\frac{1}{3}}\right) = \ln 9\\ =&2,1972\end{aligned}$$

c) Gesucht ist der VaR für eine Wahrscheinlichkeit von $0,95 = 1-p$, d. h. $p = 0,05$. Also entspricht der VaR dem 0,05–Quantil (VaR $= x_{0,05}$).

$$\begin{aligned}x_{0,05} &= Q(0,05)\\ &= \ln\left(\frac{0,05}{0,95}\right)\\ &= -2,9444\end{aligned}$$

Somit entspricht der VaR für eine Wahrscheinlichkeit von 95% einer Wertminderung von -2,9444%.

Aufgabe 5.8

Ein fairer Würfel wird zweimal unabhängig geworfen. Wie groß ist die Wahrscheinlichkeit dafür, dass die Augenzahl beim zweiten Wurf größer ist als beim ersten Wurf?

Lösung:

Der Ergebnisraum ist $\Omega = \{(i,j) \text{ mit } i,j \in \{1,2,3,4,5,6\}\}$.
Es ist $|\Omega| = 6^2 = 36$.
Sei X die Differenz zwischen der Augenzahl beim zweiten Wurf und der Augenzahl beim ersten Wurf, d.h. $X(\omega) = j - i \in \{-5,...,5\}$ mit $\omega = (i,j) \in \Omega$.
Die Wahrscheinlichkeitsfunktion lautet

$$f_X(x) = P(X = x) = \begin{cases} \frac{1}{36} & \text{für } x = -5, \\ \frac{2}{36} & \text{für } x = -4, \\ \frac{3}{36} & \text{für } x = -3, \\ \frac{4}{36} & \text{für } x = -2, \\ \frac{5}{36} & \text{für } x = -1, \\ \frac{6}{36} & \text{für } x = 0, \\ \frac{5}{36} & \text{für } x = 1, \\ \frac{4}{36} & \text{für } x = 2, \\ \frac{3}{36} & \text{für } x = 3, \\ \frac{2}{36} & \text{für } x = 4, \\ \frac{1}{36} & \text{für } x = 5. \end{cases}$$

Die zugehörige Verteilungsfunktion lautet

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{für } x < -5, \\ \frac{1}{36} & \text{für } -5 \leq x < -4, \\ \frac{3}{36} & \text{für } -4 \leq x < -3, \\ \frac{6}{36} & \text{für } -3 \leq x < -2, \\ \frac{10}{36} & \text{für } -2 \leq x < -1, \\ \frac{15}{36} & \text{für } -1 \leq x < 0, \\ \frac{21}{36} & \text{für } 0 \leq x < 1, \\ \frac{26}{36} & \text{für } 1 \leq x < 2, \\ \frac{30}{36} & \text{für } 2 \leq x < 3, \\ \frac{33}{36} & \text{für } 3 \leq x < 4, \\ \frac{35}{36} & \text{für } 4 \leq x < 5, \\ 1 & \text{für } 5 \leq x. \end{cases}$$

Somit ist

$$P(X > 0) = 1 - P(X \leq 0) = 1 - F(0) = 1 - \frac{21}{36} = \frac{15}{36} = 0,4167.$$

Aufgabe 5.9

Eine (faire) Münze werde so lange geworfen bis entweder Zahl (Z) oder fünfmal hintereinander Wappen (W) erscheint. Die Zufallsvariable X sei die Anzahl der durchgeführten Würfe.

a) Geben Sie eine geeignete Ergebnismenge an.

b) Bestimmen Sie die Wahrscheinlichkeitsfunktion und die Verteilungsfunktion der Zufallsvariablen X.

c) Bestimmen Sie die Quantilfunktion. Welchen Wert haben die Quartile von X ?

d) Berechnen Sie den Erwartungswert und die Varianz von X.

Lösung:

a) Die Ergebnismenge ist

$$\Omega = \{(Z), (W, Z), (W, W, Z), (W, W, W, Z), (W, W, W, W, Z), (W, W, W, W, W)\}\,.$$

b) Für die Wahrscheinlichkeitsfunktion erhält man (wegen der Unabhängigkeit der einzelnen Würfe):

$$\begin{aligned}
P(X = 1) &= \frac{1}{2} \\
P(X = 2) &= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\
P(X = 3) &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \\
P(X = 4) &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16} \\
P(X = 5) &= 1 - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}\right) = \frac{1}{16}\,.
\end{aligned}$$

Die Verteilungsfunktion ist somit

$$F(x) = \begin{cases} 0, & \text{falls } x < 1, \\ \frac{1}{2}, & \text{falls } 1 \leq x < 2, \\ \frac{3}{4}, & \text{falls } 2 \leq x < 3, \\ \frac{7}{8}, & \text{falls } 3 \leq x < 4, \\ \frac{15}{16}, & \text{falls } 4 \leq x < 5, \\ 1, & \text{falls } x \geq 5. \end{cases}$$

c) Allgemein ist

$$Q(p) = \min\{x \in IR \mid F(x) \geq p\}\,.$$

Mit Verteilungsfunktion erhalten wir somit:

$$Q(p) = \begin{cases} 1, & \text{falls } 0 < p \leq \frac{1}{2}, \\ 2, & \text{falls } \frac{1}{2} < p \leq \frac{3}{4}, \\ 3, & \text{falls } \frac{3}{4} < p \leq \frac{7}{8}, \\ 4, & \text{falls } \frac{7}{8} < p \leq \frac{15}{16}, \\ 5, & \text{falls } \frac{15}{16} < p < 1. \end{cases}$$

Der Median hat den Wert $x_{0.5} = 1$. Für das erste und dritte Quartil erhält man $x_{0.25} = 1$ bzw. $x_{0.75} = 2$.

d) Es ist

$$E[X] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + 5 \cdot \frac{1}{16} = \frac{31}{16} = 1.9375\,.$$

Die Varianz berechnen wir mit der Formel $V[X] = E[X^2] - E[X]^2$. Für $E[X^2]$ ergibt sich

$$\begin{aligned} E[X^2] &= 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{1}{8} + 4^2 \cdot \frac{1}{16} + 5^2 \cdot \frac{1}{16} \\ &= 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 9 \cdot \frac{1}{8} + 16 \cdot \frac{1}{16} + 25 \cdot \frac{1}{16} \\ &= \frac{83}{16} \\ &= 5.1875\,. \end{aligned}$$

Für die Varianz ergibt sich somit

$$V[X] = 5.1875 - 1.9375^2 = 1.4336\,.$$

Aufgabe 5.10

Die Zufallsvariable X sei stetig verteilt mit der Dichte

$$f(x) = \begin{cases} 24x^{-4}, & \text{falls } x \geq c, \\ 0, & \text{sonst.} \end{cases}$$

a) Bestimmen Sie den Wert des Parameters c.

b) Berechnen Sie die Verteilungsfunktion und die Quantilfunktion der Zufallsvariablen X.

c) Berechnen Sie Erwartungswert, Varianz und Median der Zufallsvariablen X.

d) Bestimmen Sie die Wahrscheinlichkeiten

$$P(X \leq 3), P(2,5 \leq X \leq 3,5), P(X = 3) \text{ und } P(X \geq 4).$$

Lösung:

a) Damit f eine Dichte ist, muss $f(x) \geq 0$ für alle $x \in \mathbb{R}$ sein und das Integral über die Dichte muss den Wert 1 haben.

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\,dx &= \int_{c}^{\infty} 24x^{-4} dx \\ &= \left[-8x^{-3}\right]_c^{\infty} \\ &= 0 - \left(-8\frac{1}{c^3}\right) \\ &= \frac{8}{c^3}\,. \end{aligned}$$

Damit $8/c^3 = 1$ ist, muss c offensichtlich den Wert $c = 2$ haben. Die Bedingung, dass die Dichte überall größer oder gleich Null ist, ist dann ebenfalls erfüllt.

b) Für $x \geq 2$ ist

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\,dt = \int_{2}^{x} 24t^{-4} dt = \left[-8t^{-3}\right]_2^x \\ &= \left(-8\frac{1}{x^3}\right) - \left(-8\frac{1}{2^3}\right) \\ &= 1 - \frac{8}{x^3}\,. \end{aligned}$$

Also ist

$$F(x) = \begin{cases} 1 - \frac{8}{x^3}, & \text{falls } x \geq 2, \\ 0 & \text{sonst} \end{cases}$$

Die Quantilfunktion erhält man bei stetigen Zufallsvariablen durch Auflösen der Gleichung $F(x) = p$ nach x:

$$1-\frac{8}{x^3} = p \quad \Longleftrightarrow \quad \frac{8}{x^3} = 1-p \quad \Longleftrightarrow \quad x^3 = \frac{8}{1-p} \quad \Longleftrightarrow \quad x = \sqrt[3]{\frac{8}{1-p}}\,.$$

Somit erhält man

$$Q(p) = \sqrt[3]{\frac{8}{1-p}}, \quad 0 < p < 1\,.$$

c) Zunächst der Erwartungswert:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x)\, dx = \int_{2}^{\infty} x \cdot 24x^{-4}\, dx \\ &= \int_{2}^{\infty} 24x^{-3}\, dx = \left[-12x^{-2}\right]_2^{\infty} \\ &= 0 - \left(-12\frac{1}{4}\right) \\ &= 3\,. \end{aligned}$$

Die Varianz berechnen wir mit der Formel $V[X] = E[X^2] - E[X]^2$:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_{2}^{\infty} x^2 \cdot 24x^{-4}\, dx \\ &= \int_{2}^{\infty} 24x^{-2}\, dx = \left[-24x^{-1}\right]_2^{\infty} \\ &= 0 - \left(-24\frac{1}{2}\right) \\ &= 12\,. \end{aligned}$$

Für die Varianz ergibt sich somit

$$V[X] = 12 - 3^2 = 3\,.$$

Der Median ergibt sich aus

$$x_{0,5} = Q(0,5) = \sqrt[3]{\frac{8}{1-0,5}} \approx 2.5198\,.$$

d) Wir berechnen die gesuchten Wahrscheinlichkeiten mit der in b) bestimmten Verteilungsfunktion. Hierbei ist zu beachten, dass bei *stetig verteilten* Zufallsvariablen $P(X \leq x) = P(X < x)$ gilt!

$$\begin{aligned}
P(X \leq 3) &= F(3) = 1 - \frac{8}{3^3} = \frac{19}{27} = 0.7037\,, \\
P(2.5 \leq X \leq 3.5) &= F(3.5) - F(2.5) = \left[1 - \frac{8}{3.5^3}\right] - \left[1 - \frac{8}{2.5^3}\right] = 0.3254\,, \\
P(X = 3) &= [F(3) - F(3) =]\ 0\,, \\
P(X \geq 4) &= 1 - F(4) = 1 - \left[1 - \frac{8}{4^3}\right] = \frac{1}{8} = 0.125\,.
\end{aligned}$$

Aufgabe 5.11

X und Y seien zwei voneinander unabhängige Zufallsvariablen mit $E[X] = 27$, $E[Y] = 13$, $V[X] = 10$ und $V[Y] = 12$. Berechnen Sie für die Zufallsvariablen

a) $Z_1 = 3X + 7Y$

b) $Z_2 = 3X - 7Y$

c) $Z = Z_1 + Z_2$

jeweils den Erwartungswert und die Varianz.

Lösung:

$$\begin{array}{lll} X: & E[X] = 27\ ; & V[X] = 10 \\ Y: & E[Y] = 13\ ; & V[Y] = 12 \end{array}$$

X und Y unabhängig

a)

$$\begin{aligned} Z_1 &= 3 \cdot X + 7 \cdot Y \\ E[Z_1] &= E[3 \cdot X + 7 \cdot Y] = 3 \cdot E[X] + 7 \cdot E[Y] = 81 + 91 = 172 \\ V[Z_1] &= V[3 \cdot X + 7 \cdot Y] = 3^2 \cdot V[X] + 7^2 \cdot V[Y] = 90 + 588 = 678 \end{aligned}$$

b)

$$\begin{aligned} Z_2 &= 3 \cdot X - 7 \cdot Y \\ E[Z_2] &= E[3 \cdot X - 7 \cdot Y] = 3 \cdot E[X] - 7 \cdot E[Y] = 81 - 91 = -10 \\ V[Z_2] &= V[3 \cdot X - 7 \cdot Y] = 3^2 \cdot V[X] + 7^2 \cdot V[Y] = 90 + 588 = 678 \end{aligned}$$

c) Im Gegensatz zu X und Y sind Z_1 und Z_2 nicht unabhängig voneinander, denn $Z_2 = Z_1 - 14 \cdot Y$. Daher empfiehlt sich die Berechnung mittels der Zufallsvariablen X und Y.

$$\begin{aligned} Z &= Z_1 + Z_2 = 6 \cdot X \\ E[Z] &= E[6 \cdot X] = 6 \cdot E[X] = 162 \\ V[Z] &= V[6 \cdot X] = 6^2 \cdot V[X] = 360 \end{aligned}$$

Aufgabe 5.12

Einem Verband gehören 5000 Selbständige an. Für einen zufällig ausgewählten Selbständigen sei
X = Einkommen aus selbständiger Arbeit,
Y = Einkommen aus Vermietung und Verpachtung,
Z = Einkommen aus Kapitalerträgen.
Gehen Sie davon aus, dass gilt:

$$\begin{array}{rrrr} E[X] = & 200\,000 \text{ DM} & \sqrt{V[X]} = & 100\,000 \text{ DM} \\ E[Y] = & 40\,000 \text{ DM} & \sqrt{V[Y]} = & 10\,000 \text{ DM} \\ E[Z] = & 60\,000 \text{ DM} & \sqrt{V[Z]} = & 15\,000 \text{ DM} \end{array}$$

Errechnen Sie Erwartungswert und Standardabweichung des Gesamteinkommens eines zufällig ausgewählten Selbständigen

a) unter der Annahme, dass X, Y und Z unabhängig sind.

b) unter der Annahme, dass $\rho_{XY} = \rho_{YZ} = \rho_{XZ} = 0,2$ ist.

Lösung:

Sei G das Gesamteinkommen eines zufällig ausgewählten Selbständigen, d.h. $G = X + Y + Z$.

a) Sind X, Y und Z unabhängig, so ist

$$E[G] = E[X + Y + Z] = E[X] + E[Y] + E[Z] = 300000 \text{ DM}$$
$$\sqrt{V[G]} = \sqrt{V[X + Y + Z]} = \sqrt{V[X] + V[Y] + V[Z]} = 101612 \text{ DM}$$

b) Sind X, Y, Z korreliert mit $\rho_{XY} = \rho_{YZ} = \rho_{XZ} = 0,2$, so ist

$$\begin{aligned} E[G] &= E[X + Y + Z] = E[X] + E[Y] + E[Z] = 300000 \text{ DM} \\ \sqrt{V[G]} &= \sqrt{V[X + Y + Z]} \\ &= \sqrt{V[X] + V[Y] + V[Z] + 2\,(Cov[X, Y] + Cov[Y, Z] + Cov[X, Z])} \\ &= \sqrt{V[X] + V[Y] + V[Z] + 2 \cdot 0,2 \cdot (\sigma_X\sigma_Y + \sigma_Y\sigma_Z + \sigma_X\sigma_Z)} \\ &= 106701 \text{ DM} \end{aligned}$$

Aufgabe 5.13

Gegeben seien zwei exponentialverteilte Zufallsvariablen $X \sim Exp(\lambda_X = 0,01)$ und $Y \sim Exp(\lambda_Y = 0,02)$. Die Zufallsvariablen seien voneinander unabhängig.
Berechnen Sie $E[X+Y]$, $V[X+Y]$, $Cov(X,Y)$, $\rho(X,Y)$ sowie $E[X \cdot Y]$.

Lösung:

Man hat $X \sim Exp(0,01)$ und $Y \sim Exp(0,02)$ und X, Y unabhängig. Damit ist

$$
\begin{aligned}
E[X+Y] &= E[X] + E[Y] = \frac{1}{0,01} + \frac{1}{0,02} = 150 \\
V[X+Y] &= V[X] + V[Y] = \frac{1}{(0,01)^2} + \frac{1}{(0,02)^2} = 12500 \\
Cov(X,Y) &= 0 \\
\rho(X,Y) &= \frac{Cov(X,Y)}{\sqrt{V[X]V[Y]}} = 0 \\
E[X \cdot Y] &= E[X] \cdot E[Y] = \frac{1}{0,01} \cdot \frac{1}{0,02} = 5000
\end{aligned}
$$

6 Spezielle Verteilungen

Aufgabe 6.1

Ein Würfel wird fünfmal geworfen. Sei X die Anzahl der ,,Sechsen".

a) Bestimmen Sie die Verteilung von X.

b) Wie groß ist die Wahrscheinlichkeit, dass genau zwei ,,Sechsen" geworfen werden?

c) Wie groß ist die Wahrscheinlichkeit, dass mehr als eine ,,Sechs" geworden wird?

Lösung:

X = Anzahl der ,,Sechsen" bei fünfmaligem Werfen des Würfels.

a) Verteilung:

$$X \sim B\left(n = 5; \pi = \frac{1}{6}\right)$$

b) Gesucht:

$$P(X = 2) = \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = 0,16075$$

c) Gesucht:

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - \binom{5}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5 - \binom{5}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4 \\ &= 1 - 2 \cdot \left(\frac{5}{6}\right)^5 \\ &= 0,19624 \end{aligned}$$

Aufgabe 6.2

Bei einer Prüfung wird einem Kandidaten ein "Multiple-Choice"-Fragebogen vorgelegt. Dabei stehen unter jeder der neun Fragen in zufälliger Reihenfolge die richtige und zwei falsche Antworten. Zum Bestehen der Prüfung müssen mindestens 50% Antworten richtig angekreuzt werden. Ein Kandidat kreuzt bei jeder Frage eine der Antworten zufällig an.

a) Mit welcher Wahrscheinlichkeit besteht er die Prüfung?

b) Man bestimme den Erwartungswert und die Streuung der Zufallsvariablen X, der Anzahl der richtigen Antworten, die man durch zufälliges Ankreuzen erreicht.

Lösung:

X = Anzahl richtiger Antworten bei neun Fragen $\sim B(n = 9; \pi = \frac{1}{3})$

a) Wahrscheinlichkeit, dass der Kandidat die Prüfung besteht

$$\begin{aligned}P(X \geq 5) &= 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} \binom{9}{x} \left(\frac{1}{3}\right)^{x} \left(\frac{2}{3}\right)^{9-x} \\ &= 1 - 0,02601 - 0,11706 - 0,23411 - 0,27313 - 0,20485 \\ &= 0,14484\end{aligned}$$

b)

$$\begin{aligned}E\left[X\right] &= n \cdot \pi = 9 \cdot \frac{1}{3} = 3 \\ V\left[X\right] &= n \cdot \pi \cdot (1 - \pi) = 9 \cdot \frac{1}{3} \cdot \frac{2}{3} = 2\end{aligned}$$

Aufgabe 6.3

Die Wahrscheinlichkeit, dass ein deutscher Urlauber sich in einem bestimmten afrikanischen Land eine seltene Infektionskrankheit zuzieht, sei $p = 0,0001$. Nehmen Sie an, in einem Jahr verbringen 20 000 Deutsche ihren Urlaub in diesem Land.

a) Wie groß ist die Wahrscheinlichkeit, dass sich mindestens ein deutscher Urlauber die Infektionskrankheit zuzieht? Welche Annahmen treffen Sie zur Berechnung der Wahrscheinlichkeit?

b) Berechnen Sie Erwartungswert und Standardabweichung der Anzahl deutscher Urlauber, die sich mit der Krankheit infizieren.

Lösung:

X = Anzahl infizierter Urlauber

a) Unter der Annahme, dass sich die Urlauber unabhängig voneinander infizieren, ist $X \sim B(n = 20000, \pi = 0,0001)$. Die gesuchte Wahrscheinlichkeit berechnet sich als

$$
\begin{aligned}
P(X \geq 1) = 1 - P(X = 0) &= 1 - \binom{20000}{0} \cdot 0,0001^0 \cdot 0,9999^{20000} \\
&= 1 - 0,1353 \\
&= 0,8647.
\end{aligned}
$$

Möglich ist hier auch die approximative Berechnung der gesuchten Wahrscheinlichkeit mittels der Poisson-Verteilung (Bedingungen sind erfüllt), also $B(n = 20000, \pi = 0,0001) \simeq Po(\mu = 2)$, d.h.

$$
P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-2}\frac{2^0}{0!} = 1 - 0,1353 = 0,8647.
$$

b) Es ist

$$
\begin{aligned}
E[X] &= 20000 \cdot 0,0001 = 2 \\
V[X] &= 20000 \cdot 0,0001 \cdot 0,9999 = 1,9998 \\
\sqrt{V[X]} &= \sqrt{1,9998} = 1,4141
\end{aligned}
$$

Aufgabe 6.4

Eine Urne enthält 50 Kugeln, von denen fünf rot sind. Wie groß ist die Wahrscheinlichkeit, dass Sie mindestens zwei rote Kugeln ziehen, wenn Sie vier Kugeln ohne Zurücklegen aus der Urne ziehen? Wie groß wäre die entsprechende Wahrscheinlichkeit, wenn Sie mit Zurücklegen ziehen würden?

Lösung:

Zum Berechnen der Wahrscheinlichkeit beim Ziehen ohne Zurücklegen benutzt man die hypergeometrische Verteilung mit den Parametern $N = 50$, $M = 5$, $n = 4$. Sei X die Anzahl der gezogenen roten Kugeln. Man erhält

$$\begin{aligned} P(X \geq 2) = 1 - P(X \leq 1) &= 1 - \sum_{x=0}^{1} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \\ &= 1 - \sum_{x=0}^{1} \frac{\binom{5}{x}\binom{45}{4-x}}{\binom{50}{4}} \\ &= 1 - (0.647 + 0.3081) \\ &= 1 - 0.9551 = 0.0449 \end{aligned}$$

Würde man mit Zurücklegen ziehen, so wäre X binomialverteilt mit Parametern $n = 4$ und $\pi = 0,1$. Die Wahrscheinlichkeiten $P(X = 0)$ und $P(X = 1)$ sind tabelliert. Man erhält

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &= 1 - (0.6561 + 0.2916) \\ &= 1 - 0.9447 = 0.0523 \end{aligned}$$

Aufgabe 6.5

Ein Veterinär soll eine (kleine) Rinderherde auf BSE untersuchen. Die Herde besteht aus 20 Tieren. Vier davon sind infiziert, aber noch nicht erkrankt; die übrigen Tiere sind nicht infiziert. Da man zum Nachweis einer Infektion das Tier töten muss, werden zufällig fünf Rinder aus der Herde ausgewählt und geschlachtet. Wie groß ist die Wahrscheinlichkeit, dass auf diese Weise die Erkrankung der Herde nachgewiesen wird?

Lösung:

Ziehen ohne Zurücklegen (weil Tiere getötet werden).

X = Anzahl infizierter Tiere unter den fünf geschlachteten Tieren

$X \sim H(n = 5; N = 20; M = 4)$

Die Krankheit wird erkannt, wenn mindestens ein geschlachtetes Tier infiziert war, also wenn $X \geq 1$.

Die Wahrscheinlichkeit, dass die Krankheit erkannt wird, ist folglich

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{4}{0}\binom{16}{5}}{\binom{20}{5}} = 1 - \frac{4368}{15504} = 0,7183$$

Aufgabe 6.6

Ein regionaler Verkehrsbetrieb untersucht die Verluste durch Schwarzfahrten im Kurzstreckenbereich. Es wird festgestellt, dass die Wahrscheinlichkeit dafür, kontrolliert zu werden, 5% beträgt.

a) Wie groß ist die Wahrscheinlichkeit, dass ein Schwarzfahrer das erste Mal bei der 10. Fahrt kontrolliert wird?

b) Wie groß ist die Wahrscheinlichkeit, dass ein Schwarzfahrer frühestens bei der 10. Fahrt kontrolliert wird?

c) Wie hoch müsste das Bußgeld für Schwarzfahrer mindestens sein, damit sich Schwarzfahren nicht lohnt? (Gehen Sie von Fahrkartenpreisen von 3 DM aus.)

Lösung:

Erfolg = „Schwarzfahrer fährt und wird kontrolliert"
Misserfolg = „Schwarzfahrer fährt und wird nicht kontrolliert"
X = Anzahl Schwarzfahrten (ohne Kontrolle) bis zur ersten Kontrolle
$X \sim G(\pi = 0,05)$ (X zählt die Anzahl Fahrten inkl. der Fahrt, wo der Schwazrfahrer erwischt wird)

a) Gesucht:

$$P(X = 10) = 0,05 \cdot 0,95^{10-1} = 0,0315$$

b) Gesucht:

$$\begin{aligned} P(X \geq 10) &= 1 - P(X \leq 9) \\ &= 1 - \sum_{i=1}^{9} 0,05 \cdot 0,95^{i-1} \underset{\text{FS S. 67}}{=} 1 - 0,05 \cdot \frac{0,95^9 - 1}{0,95 - 1} \\ &= 0,95^9 = 0,6302 \end{aligned}$$

Zur Verwendung der FS:
Es handelt sich um die geometrische Reihe mit $a_i = a_1 \cdot q^{i-1}$

$$\sum_{i=1}^{n} a_i = a_1 \cdot \frac{q^n - 1}{q - 1}$$

Hier ist also $a_1 = 0,05$ und $q = 0,95$.

c) Das Bußgeld müsste mindestens $3 \cdot E[X] = 3 \cdot \frac{1}{0{,}05} = 60$ DM betragen.
Im Mittel wird man nämlich alle $E[X] = \frac{1}{\pi} = 20$ Fahrten kontrolliert. Folglich muss die Strafe so hoch bemessen sein, dass man durch dauerhaftes Schwarzfahren keine Ersparnis hat, die bisherigen Schwarzfahrten also (mindestens) nachzahlen muss.

Aufgabe 6.7

In dem Call-Center einer Direktbank treffen erfahrungsgemäß werktags in der Zeit von 17.00 Uhr bis 19.00 Uhr durchschnittlich 24 Anrufe ein. Die Anzahl Anrufe kann als Poisson-verteilt angesehen werden. Wie groß ist die Wahrscheinlichkeit, dass an einem Werktag

a) zwischen 17.00 Uhr und 17.20 Uhr keine Anrufe eintreffen?

b) in der Zeit zwischen 18.00 Uhr und 18.30 Uhr mehr als vier Anrufe eingehen?

c) zwischen 17.15 Uhr und 17.45 Uhr maximal fünf Anrufe eingehen, wenn bekannt ist, dass zwischen 17.15 Uhr und 18.15 Uhr genau zehn Anrufe eingegangen sind? (Nehmen Sie an, dass die Anzahl der Anrufe zwischen 17.15 Uhr und 17.45 Uhr unabhängig von der Anzahl der Anrufe zwischen 17.45 Uhr und 18.15 Uhr ist.)

Lösung:

Laut Aufgabenstellung erfolgen 24 Anrufe pro 120 Minuten, d.h. $\frac{1}{5}$ Anruf pro Minute.

a) In 20 Minuten erfolgen im Schnitt $\frac{1}{5} \cdot 20 = 4$ Anrufe.
X = Anzahl der Anrufe pro 20 Minuten, also $X \sim Po(\mu = 4)$

$$P(X = 0) = e^{-4}\frac{4^0}{0!} = 0,0183$$

b) In einer halben Stunde erfolgen im Schnitt $\frac{1}{5} \cdot 30 = 6$ Anrufe
Y = Anzahl der Anrufe pro halbe Stunde, also $Y \sim Po(\mu = 6)$

$$\begin{aligned} P(Y > 4) &= 1 - P(Y \leq 4) = 1 - F(4) \\ &= 1 - 0,2851 = 0,7149 \end{aligned}$$

c) X = Anzahl Anrufe zwischen 17.15 Uhr und 17.45 Uhr $\sim Po(\mu = 6)$
Y = Anzahl Anrufe zwischen 17.45 Uhr und 18.15 Uhr $\sim Po(\mu = 6)$
Z = Anzahl Anrufe zwischen 17.15 Uhr und 18.15 Uhr $\sim Po(\mu = 12)$

Somit ist $Z = X + Y$ (bzw. $Y = Z - X$).
Gesucht:

$$
\begin{aligned}
P(X \leq 5 \mid Z = 10) &= \frac{P(X \leq 5; Z = 10)}{P(Z = 10)} = \sum_{x=0}^{5} \frac{P(X = x; Z = 10)}{P(Z = 10)} \\
&= \sum_{x=0}^{5} \frac{P(X = x; Y = 10 - x)}{P(Z = 10)} = \sum_{x=0}^{5} \frac{P(X = x) \cdot P(Y = 10 - x)}{P(Z = 10)} \\
&= \sum_{x=0}^{5} \frac{e^{-6} \frac{6^x}{x!} \cdot e^{-6} \frac{6^{10-x}}{(10-x)!}}{e^{-12} \frac{12^{10}}{10!}} = \sum_{x=0}^{5} \frac{10!}{x!(10-x)!} \frac{6^{10}}{12^{10}} \\
&= \sum_{x=0}^{5} \binom{10}{x} \left(\frac{1}{2}\right)^{10} = 0,6230
\end{aligned}
$$

Man sieht, dass $X \mid Z = z$ binomialverteilt ist.

Aufgabe 6.8

Ein Billiganbieter stellt Mikrochips her. Ein von diesem Anbieter hergestellter Chip ist mit einer Wahrscheinlichkeit von 5% defekt. Wie groß ist die Wahrscheinlichkeit, beim Kauf von fünf dieser Chips

a) höchstens zwei Chips,

b) weniger als zwei Chips,

c) mindestens drei Chips,

d) mindestens einen Chip

mit einem Defekt zu erhalten?

Lösung:

X sei die Anzahl defekter Chips beim Kauf von fünf dieser Chips. Dann ist $X \sim B(5, 0.05)$.

a)

$$P(X \leq 2) = \sum_{k=0}^{2} \binom{5}{k} \cdot 0.05^k \cdot 0.95^{5-k} = 0.9988$$

b)

$$P(X < 2) = P(X \leq 1) = \sum_{k=0}^{1} \binom{5}{k} \cdot 0.05^k \cdot 0.95^{5-k} = 0.9774$$

c)

$$P(X \geq 3) = 1 - P(X \leq 2) = 0.0012$$

d)

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.95^5 = 0.2262$$

Aufgabe 6.9

Die Firma Invinoveritas lagert in einem ihrer Weinkeller 80 Rotwein- und 40 Weißwein-Fässer. Dieser Weinkeller wurde beim letzten Hochwasser überflutet, so dass die Fässer nun völlig ungeordnet im Keller liegen. Der Inhaber lässt zunächst sechs Fässer aus dem Keller holen. Wie groß ist die Wahrscheinlichkeit, dass sich darunter

a) genau drei Weißwein- und drei Rotwein-Fässer,

b) mindestens ein Rotwein-Fass,

c) höchstens vier Weißwein-Fässer

befinden?

Lösung:

X sei die Anzahl gezogener Weißwein-Fässer. Dann ist X $H(6, 120, 40)$-verteilt. Da der Auswahlsatz $\frac{n}{N} = \frac{6}{120} = 0.05 \leq 0.05$ ist, kann die hypergeometrische Verteilung durch die Binomialverteilung mit Parametern $n = 6$ und $\pi = \frac{M}{N} = \frac{1}{3}$ approximiert werden.

a)

$$P(X = 3) = \binom{6}{3} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^{6-3} = 20 \cdot \frac{8}{729} = 0.2195\,.$$

b) $Y = 6 - X$ ist Anzahl entnommener Rotwein-Fässer.

$$\begin{aligned} P(Y \geq 1) &= P(X \leq 5) \\ &= 1 - P(X = 6) \\ &= 1 - \binom{6}{6} \left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^{6-6} \\ &= 1 - \frac{1}{729} \\ &= 0.9986\,. \end{aligned}$$

c)

$$\begin{aligned} P(X \leq 4) &= 1 - (P(X = 5) + P(X = 6)) \\ &= 1 - \left[\binom{6}{5} \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^{6-5} + \binom{6}{6} \left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^{6-6}\right] \\ &= 1 - [0.01646 + 0.00137] \\ &= 0.9822 \end{aligned}$$

Aufgabe 6.10

Die Versuche eines bestimmten Fahrschülers, die Führerscheinprüfung zu bestehen, können als unabhängige Versuche aufgefasst werden. Die Wahrscheinlichkeit, dass er die Prüfung besteht, sei konstant $0,6$.

a) Bestimmen Sie die durchschnittliche Anzahl Versuche (bzw. Prüfungen) bis zum Erhalt des Führerscheins.

b) Berechnen Sie die Wahrscheinlichkeit dafür, dass der Fahrschüler mehr als vier weitere Versuche benötigt, wenn er bereits bei zwei Prüfungen durchgefallen ist.

Lösung:

Erfolg = „Bestehen der Fahrprüfung"
X = Anzahl Versuche bis zum ersten Erfolg $\sim G(\pi = 0,6)$

a)
$$E[X] = \frac{1}{0,6} = 1,6667$$

b) Die Verteilungsfunktion von X lautet (mit geometrischer Reihe)

$$P(X \leq x) = \sum_{k=1}^{x} \pi(1-\pi)^{k-1} = \pi \cdot \frac{(1-\pi)^x - 1}{(1-\pi) - 1} = 1 - (1-\pi)^x$$

Damit berechnet man

$$\begin{aligned} P(X > 6 \mid X > 2) &= \frac{P(X > 6)}{P(X > 2)} = \frac{1 - P(X \leq 6)}{1 - P(X \leq 2)} \\ &= \frac{(1-0,6)^6}{(1-0,6)^2} = (1-0,6)^4 = 0,4^4 = 0,0256 \end{aligned}$$

Aufgrund der „Gedächtnislosigkeit" der geometrischen Verteilung hätte man auch (kürzer) rechnen können:

$$\begin{aligned} P(X > 6 \mid X > 2) &= \frac{P(X > 6)}{P(X > 2)} = P(X > 4) \\ &= (1-0,6)^4 = 0,4^4 = 0,0256 \end{aligned}$$

Aufgabe 6.11

An einem bestimmten Drive-In-Restaurant halten im Schnitt neun Motorroller pro Stunde. Gehen Sie davon aus, dass die Anzahl der Motorroller pro Stunde Poisson-verteilt ist.

a) Wie viele Motorroller kommen durchschnittlich innerhalb einer halben Stunde?

b) Wie groß ist die Wahrscheinlichkeit, dass innerhalb von 20 Minuten mehr als vier Motorroller kommen?

Lösung:

Sei X_1 = Anzahl Motorroller pro Stunde $\sim Po(\mu_1 = 9)$

a) Sei X_2 = Anzahl Motorroller pro halbe Stunde $\sim Po(\mu_2 = \frac{9}{2} = 4,5)$. Dann ist

$$E[X_2] = 4,5$$

b) Sei X_3 = Anzahl Motorroller pro 20 Minuten $\sim Po(\mu_3 = \frac{9}{3} = 3)$. Damit ist

$$\begin{aligned} P(X_3 > 4) &= 1 - P(X_3 \leq 4) \\ &= 1 - (0,0498 + 0,1494 + 0,2240 + 0,2240 + 0,1680) \\ &= 1 - 0,8152 \\ &= 0,1848 \end{aligned}$$

Aufgabe 6.12

Jemand wartet auf einen Anruf. Es ist abgemacht, dass der Anruf irgendwann zwischen 14.00 Uhr und 16.00 Uhr erfolgen soll. Der Zeitpunkt des Anrufs, X , sei rechteckverteilt (zwischen 14.00 Uhr und 16.00 Uhr).

a) Wie groß ist die Wahrscheinlichkeit dafür, dass der Anruf zwischen 15.30 Uhr und 16.00 Uhr erfolgt?

b) Es ist nun 15.15 Uhr, und das Telefon hat noch nicht geklingelt. Wie groß ist jetzt die Wahrscheinlichkeit, dass der Anruf zwischen 15.30 Uhr und 16.00 Uhr erfolgt?

Lösung:

X = Zeitpunkt des Anrufs $\sim R_{[14;16]}$, so dass

$$f(x) = \begin{cases} \frac{1}{16-14} = \frac{1}{2} & 14 \leq x \leq 16 \\ 0 & \text{sonst} \end{cases}$$

$$\text{und} \quad F(x) = \begin{cases} 0 & x < 14 \\ \int\limits_{14}^{x} \frac{1}{2} dt = \frac{1}{2}x - 7 & 14 \leq x \leq 16 \\ 1 & 16 < x \end{cases}$$

a)

$$\begin{aligned} P(15,5 \leq X \leq 16) &= F(16) - F(15,5) \\ &= 1 - 0,75 = 0,25 \end{aligned}$$

b)

$$\begin{aligned} P(15,5 \leq X \leq 16 \mid X \geq 15,25) &= \frac{F(16) - F(15,5)}{1 - F(15,25)} \\ &= \frac{0,25}{0,375} = \frac{2}{3} = 0,6667 \end{aligned}$$

Teil b) kann auch derart gelöst werden, dass eine neue Rechteckverteilung betrachtet wird, nämlich $Y \sim R_{[15,25;16]}$, und man von dieser dann $P(15,5 \leq Y \leq 16)$ berechnet.

Aufgabe 6.13

Die Zufallsvariable X beschreibe die Lebensdauer eines Fernsehgeräts (in h). Angenommen, X ist exponentialverteilt. Bestimmen Sie den Erwartungswert und den Median dieser Zufallsvariablen, wenn bekannt ist, dass $P(X \geq 10\,000) = 0,7$ gilt.

Lösung:

X = Lebensdauer des Fernsehgeräts $\sim Exp\,(\lambda)$

$$\begin{aligned}
P(X \geq 10000) &= 0,7 \\
&\Longrightarrow P(X < 10000) = 0,3 \\
&\Longrightarrow F_X(10000) = 1 - e^{-\lambda \cdot 10000} = 0,3 \\
&\Longrightarrow 0,7 = e^{-\lambda \cdot 10000} \\
&\Longrightarrow \lambda = \frac{\ln 0,7}{-10000} = 0,00003567
\end{aligned}$$

$$E\,[X] = \frac{1}{\lambda} = \frac{1}{0,00003567} = 28036,7325 \text{ h } \approx 3,2 \text{ Jahre}$$

$$\begin{aligned}
Q_X(p) &= -\frac{\ln(1-p)}{\lambda} \\
&\text{denn :} \\
1 - e^{-\lambda x} &= p \\
1 - p &= e^{-\lambda x} \\
\ln(1-p) &= -\lambda x \\
x &= -\frac{\ln(1-p)}{\lambda}
\end{aligned}$$

$$x_{0,5} = Q_X(0,5) = -\frac{\ln 0,5}{0,00003567} = 19432,217 \text{ h } \approx 2,2 \text{ Jahre}$$

D.h. 50% der Fernseher halten höchstens 2,2 Jahre.

Aufgabe 6.14

Eine Zufallsvariable X sei exponentialverteilt mit Parameter $\lambda = 0,01$.

a) Berechnen Sie $P(X > 100)$.

b) Berechnen Sie $P(X > 1000 \mid X > 900)$.

c) Halten Sie es für angemessen, die Lebensdauer von technischen Geräten durch eine exponentialverteilte Zufallsvariable zu beschreiben? Kurze Begründung. (Hinweis: Machen Sie sich die Ergebnisse aus den Aufgabenteilen a) und b) zunutze.)

Lösung:

Die Zufallsvariable $X \sim Exp(\lambda = 0,01)$.

a) Es ist

$$\begin{aligned} P(X > 100) &= 1 - P(X \leq 100) \\ &= 1 - (1 - e^{-0,01 \cdot 100}) \\ &= e^{-1} = 0,3679 \end{aligned}$$

b) Es ist

$$\begin{aligned} P(X > 1000 \mid X > 900) &= \frac{P(\{X > 1000\} \cap \{X > 900\})}{P(X > 900)} \\ &= \frac{P(X > 1000)}{P(X > 900)} = \frac{1 - P(X \leq 1000)}{1 - P(X \leq 900)} \\ &= \frac{1 - (1 - e^{-0,01 \cdot 1000})}{1 - (1 - e^{-0,01 \cdot 900})} = \frac{e^{-10}}{e^{-9}} = e^{-1} = 0,3679 \end{aligned}$$

c) Die Beschreibung der Lebensdauer durch eine Exponentialverteilung erscheint nicht angemessen. Die Ergebnisse aus a) und b) zeigen, dass es theoretisch keinen Einfluss hat, wie lange z.B. das technische Gerät bereits im Einsatz ist. Praktisch jedoch treten Abnutzungserscheinungen auf, die auf diese Weise unberücksichtigt bleiben.
Diese besondere Eigenschaft der "Gedächtnislosigkeit" findet sich auch bei der Geometrischen Verteilung, z.B. $P(X > 5) = P(X > 7 \mid X > 2)$.

Aufgabe 6.15

In einer Fernseh-Spielshow wird von einem Mitspieler ein Glücksrad mit stetiger 360°-Einteilung in Bewegung gesetzt und der Stillstand abgewartet. Der angezeigte Winkel in Grad wird als Zufallsvariable X aufgefasst. Nehmen Sie an, dass $X \sim R(0; 360)$ gilt.

a) Wie groß ist die Wahrscheinlichkeit, dass das Glücksrad einen Winkel zwischen 100 und 200 Grad anzeigt?
Bestimmen Sie den Median von X.

b) Wie groß ist die Wahrscheinlichkeit, dass das Glücksrad einen Winkel zwischen $10 \cdot n$ und $10 \cdot (n+1)$ Grad anzeigt ($n \in \{0, 1, ..., 35\}$)?

c) Der Spieler erhält einen Gewinn in Höhe von n Euro, wenn das Glücksrad einen Winkel zwischen $10 \cdot n$ und $10 \cdot (n+1)$ Grad anzeigt.
Fassen Sie den Gewinn als Zufallsvariable Y auf und geben Sie deren Verteilung sowie Erwartungswert und Standardabweichung an.
Wie groß ist die Wahrscheinlichkeit, weniger als 15 Euro als Gewinn zu erhalten?
Berechnen Sie den Median von Y.

Lösung:

a) Die Verteilungsfunktion von $X \sim R(0; 360)$ lautet für $0 \leq x \leq 360$

$$F_X(x) = \frac{x}{360}$$

Die gesuchte Wahrscheinlichkeit berechnet man als

$$\begin{aligned} P(100 \leq X \leq 200) &= F_X(200) - F_X(100) \\ &= \frac{200 - 100}{360} = \frac{100}{360} \\ &= 0,2778 \end{aligned}$$

Den Median von X erhält man durch Auflösen der Gleichung

$$\begin{aligned} F_X(x) &= 0,5 \\ x_{0,5} &= 180 \end{aligned}$$

b) Die gesuchte Wahrscheinlichkeit berechnet man als

$$\begin{aligned}
P(10n \leq X \leq 10(n+1)) &= F_X(10(n+1)) - F_X(10n) \\
&= \frac{10(n+1) - 10n}{360} = \frac{10}{360} \\
&= 0,0278
\end{aligned}$$

c) Die Zufallsvariable Y hat die Wahrscheinlichkeitsfunktion

$$f_Y(y) = \begin{cases} 0,0278 & \text{für } y = 0, ..., 35 \\ 0 & \text{sonst.} \end{cases}$$

Der Erwartungswert ist

$$\begin{aligned}
\mu_Y = E[Y] &= \sum_{y=0}^{35} y \cdot f(y) \\
&= 0,0278 \cdot \sum_{y=0}^{35} y = 0,0278 \cdot \frac{35 \cdot 36}{2} \\
&= 17,5140
\end{aligned}$$

und die Standardabweichung

$$\begin{aligned}
\sigma_Y &= \sqrt{V[Y]} = \sqrt{E[Y^2] - \mu_Y^2} \\
&= \sqrt{\sum_{y=0}^{35} y^2 \cdot f(y) - 17,5140^2} = \sqrt{0,0278 \cdot \sum_{y=0}^{35} y^2 - 17,5140^2} \\
&= \sqrt{0,0278 \cdot \frac{35 \cdot 36 \cdot 71}{6} - 17,5140^2} = \sqrt{107,7578} \\
&= 10,38
\end{aligned}$$

Die gesuchte Wahrscheinlichkeit berechnet man als

$$\begin{aligned}
P(Y < 15) &= P(Y \leq 14) \\
&= \sum_{y=0}^{14} f(y) = 15 \cdot 0,0278 \\
&= 0,4170
\end{aligned}$$

Der Median ist der kleinste Wert y, so dass

$$\begin{aligned} F_Y(y) = \sum_{k=0}^{y} f(k) &\geq 0,5 \\ (y+1) \cdot 0,0278 &\geq 0,5 \\ y &\geq \frac{0,5 - 0,0278}{0,0278} = 16,9856 \end{aligned}$$

Also ist der Median $y_{0,5} = 17$.

Aufgabe 6.16

Die Lebensdauer X eines elektonischen Bauteils sei exponentialverteilt mit Parameter λ. Für die Aufgabenteile a) bis c) sei $\lambda = \frac{1}{500}$.

a) Wie groß ist die Wahrscheinlichkeit, dass das Bauteil vor dem Zeitpunkt $t_0 = 200$ nicht ausfällt?

b) Wie groß ist die Wahrscheinlichkeit, dass das Bauteil zwischen den Zeitpunkten $t_1 = 200$ und $t_2 = 300$ ausfällt?

c) Welche Zeitpunkte überlebt das Bauteil mit mindestens 90% Wahrscheinlichkeit?

d) Welchen Wert müsste der Parameter λ besitzen, damit mit Wahrscheinlichkeit 0,9 die Lebensdauer des Bauteils mindestens 50 Stunden beträgt?

Lösung:

Die Verteilungsfunktion einer $Exp(\lambda)$-verteilten Zufallsvariablen X erhält man durch Integration der Dichte zu:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

a)
$$\begin{aligned} P(X > 200) &= 1 - P(X \leq 200) \\ &= 1 - F(200) \\ &= e^{-\frac{1}{500} \cdot 200} \\ &= 0.6703 \end{aligned}$$

b)
$$\begin{aligned} P(200 \leq X \leq 300) &= F(300) - F(200) \\ &= \left[1 - e^{-\frac{3}{5}}\right] - \left[1 - e^{-\frac{2}{5}}\right] \\ &= e^{-\frac{2}{5}} - e^{-\frac{3}{5}} \\ &= 0.1215 \end{aligned}$$

c)

$$\begin{aligned}
P(X > t_0) &\geq 0.9 \\
1 - F(t_0) &\geq 0.9 \\
F(t_0) &\leq 0.1 \\
1 - e^{-\frac{1}{500}t_0} &\leq 0.1 \\
-\frac{1}{500}t_0 &\geq \ln 0.9 \\
t_0 &\leq -500 \cdot \ln 0.9 \\
t_0 &\leq 52.68
\end{aligned}$$

Mit mindestens 90% Wahrscheinlichkeit überlebt das Bauteil also alle Zeitpunkte vor $t_0 = 52.68$.

d)

$$\begin{aligned}
P(X > 50) &= 0.9 \\
1 - F(50) &= 0.9 \\
e^{-\lambda \cdot 50} &= 0.9 \\
-\lambda \cdot 50 &= \ln 0.9 \\
\lambda &= -\frac{1}{50} \cdot \ln 0.9 \\
\lambda &= 0.002107
\end{aligned}$$

7 Normalverteilung und zentraler Grenzwertsatz

Aufgabe 7.1

Gegeben sei die normalverteilte Zufallsvariable X. Es ist bekannt, dass in 15,87% aller Fälle der Wert von X kleiner ist als 4 und in 2,5% aller Fälle größer als 21,76. Bestimmen Sie Erwartungswert und Varianz der Zufallsvariablen.

Lösung:

Es gilt zunächst:

$$\begin{aligned}
P(X \leq 4) &= 0,1587 \\
\Phi\left(\frac{4-\mu_X}{\sigma_X}\right) &= 0,1587 \\
\frac{4-\mu_X}{\sigma_X} &= u_{0,1587} \\
4-\mu_X &= u_{0,1587}\sigma_X
\end{aligned}$$

Dabei ist $u_{0,1587} = -u_{1-0,1587} = -u_{0,8413} = -1$. Weiter kann man berechnen:

$$\begin{aligned}
P(X > 21,76) &= 0,025 \\
P(X \leq 21,76) &= 1-0,025 \\
\Phi\left(\frac{21,76-\mu_X}{\sigma_X}\right) &= 0,975 \\
\frac{21,76-\mu_X}{\sigma_X} &= u_{0,975} \\
21,76-\mu_X &= 1,96\sigma_X
\end{aligned}$$

Man hat zwei Gleichungen mit zwei Unbekannten:

$$\begin{aligned}
4-\mu_X &= -\sigma_X \\
21,76-\mu_X &= 1,96\sigma_X
\end{aligned}$$

Zieht man nun die erste Gleichung von der zweiten ab, so erhält man:

$$\begin{aligned}
4-\mu_X &= -\sigma_X \\
17,76 &= 2,96\sigma_X
\end{aligned}$$

Damit ergibt sich: 1) $\sigma_X = \frac{17,76}{2,96} = 6$.

2) $4-\mu_X = -6 \quad \Longrightarrow \mu_X = 10$.

Aufgabe 7.2

In einer Fabrik werden Konservendosen abgefüllt, deren Nettofüllgewicht als normalverteilt angesehen werden kann. Das durchschnittliche Nettofüllgewicht der Dosen beträgt 249,2 Gramm, die Standardabweichung 11,37 Gramm. Das Label der Dosen gibt ein Nettofüllgewicht von 240 Gramm an. Berechnen Sie den Anteil Dosen, die untergewichtig sind.

Lösung:

$X =$ Nettofüllgewicht einer Dose $\sim N\big(\mu = 249,2; \sigma^2 = (11,37)^2\big)$

$$\begin{aligned} P(X \leq 240) = F(240) &= \Phi\left(\frac{240 - 249,2}{11,37}\right) \\ &= \Phi(-0,81) = 1 - \Phi(0,81) = 1 - 0,791 = 0,209 \end{aligned}$$

D.h. $20,9\%$ der Dosen sind untergewichtig.

Aufgabe 7.3

Die Zufallsvariable X sei standardnormalverteilt. Berechnen Sie die Wahrscheinlichkeit dafür, dass die Zufallsvariable $Y = 25 \cdot X - 6$

a) höchstens den Wert 30 annimmt.

b) mindestens den Wert 10 annimmt.

c) einen Wert von mindestens 20, aber kleiner als 24 annimmt.

d) den Wert 25 annimmt.

Lösung:

Man hat

$$X \sim N(0;1) \Longrightarrow Y = (25 \cdot X - 6) \sim N(-6; 25^2)$$

a)

$$P(Y \leq 30) = \Phi\left(\frac{30-(-6)}{25}\right) = \Phi\,(1,44) = 0,9251$$

b)

$$\begin{aligned} P(Y \geq 10) = 1 - P(Y \leq 10) &= 1 - \Phi\left(\frac{10-(-6)}{25}\right) \\ &= 1 - \Phi\,(0,64) = 1 - 0,7389 = 0,2611 \end{aligned}$$

c)

$$\begin{aligned} P(20 \leq Y \leq 24) &= \Phi\left(\frac{24-(-6)}{25}\right) - \Phi\left(\frac{20-(-6)}{25}\right) \\ &= \Phi\,(1,2) - \Phi\,(1,04) = 0,8849 - 0,8508 = 0,0341 \end{aligned}$$

d)

$$P(Y = 25) = 0 \qquad \text{, da } X \text{ stetige ZV}$$

Aufgabe 7.4

Die tatsächliche Flugzeit X [in Minuten] von Köln nach Washington kann als normalverteilte Zufallsvariable mit Erwartungswert μ und Varianz σ^2 aufgefasst werden. Es sei bekannt, dass $P(X \leq 470) = 0,9332$ und $P(X > 426) = 0,7580$.

a) Bestimmen Sie aus diesen Angaben μ und σ^2.

b) Welche Flugzeit wird mit einer Wahrscheinlichkeit von 95% nicht überschritten?

c) Mit welcher Wahrscheinlichkeit liegt die tatsächliche Flugzeit zwischen 420 und 460 Minuten?

Lösung:

a) Es ist

$$P(X \leq 470) = \Phi\left(\frac{470-\mu}{\sigma}\right) = 0.9332\,.$$

Einer Tabelle der Verteilungsfunktion der Standardnormalverteilung entnimmt man, dass $\Phi(1.5) = 0.9332$, d.h.

$$\frac{470-\mu}{\sigma} = 1.5 \qquad \Longleftrightarrow \qquad \mu = 470 - 1.5\sigma\,.$$

Ferner ist

$$P(X > 426) = 1 - \Phi\left(\frac{426-\mu}{\sigma}\right) = \Phi\left(-\frac{426-\mu}{\sigma}\right) = 0.7580\,.$$

Wie oben sieht man, dass $\Phi(0.7) = 0.7580$. Somit ist

$$-\frac{426-\mu}{\sigma} = 0.7 \qquad \Longleftrightarrow \qquad \mu = 426 + 0.7\sigma\,.$$

Wir erhalten also das lineare Gleichungssystem

$$\begin{aligned} \mu &= 470 - 1.5\sigma \\ \mu &= 426 + 0.7\sigma\,. \end{aligned}$$

Subtraktion der Gleichungen liefert

$$0 = 44 - 2.2\sigma \qquad \Longleftrightarrow \qquad \sigma = 20\,.$$

Einsetzen von $\sigma = 20$ in die erste Gleichung liefert

$$\mu = 470 - 1.5 \cdot 20 = 440\,.$$

Die Flugzeit ist also $N(440, 20^2)$-verteilt.

b) Die Flugzeit x ist so zu bestimmen, dass $P(X \leq x) = 0.95$ ist. Gesucht ist also das 0.95-Quantil der $N(440, 20)$-Verteilung. Allgemein ist das p-Quantil einer $N(\mu, \sigma^2)$-Verteilung gegeben durch $\tilde{x}_p = \mu + \sigma u_p$. Man erhält also

$$\tilde{x}_{0.95} = 440 + 20u_{0.95} = 440 + 20 \cdot 1.6449 = 472.898\,.$$

Kennt man die Formel für das Quantil nicht, rechnet man so: Es ist

$$P(X \leq x) = \Phi\left(\frac{x - 440}{20}\right) = 0.95$$

genau dann, wenn

$$\frac{x - 440}{20} = u_{0.95}\,.$$

Also ist

$$x = 440 + 20u_{0.95} = 440 + 20 \cdot 1.6449 = 472.898\,.$$

Ein Flug dauert also mit einer Wahrscheinlichkeit von 95% nicht länger als 472.898 Minuten.

c) Gesucht ist

$$\begin{aligned} P(420 \leq X \leq 460) &= \Phi\left(\frac{460 - 440}{20}\right) - \Phi\left(\frac{420 - 440}{20}\right) \\ &= \Phi(1) - \Phi(-1) \\ &= 2\Phi(1) - 1 \\ &= 2 \cdot 0.8413 - 1 \\ &= 0.6826\,. \end{aligned}$$

Aufgabe 7.5

Sei X eine normalverteilte Zufallsvariable. Sie kennen zwei Quantile der Verteilung, nämlich $x_{0,4} = 0,7467$ und $x_{0,7} = 1,5244$.

a) Bestimmen Sie den Erwartungswert und die Varianz von X.

b) Wie groß ist die Wahrscheinlichkeit, dass X negativ wird?

Lösung:

a) Wegen $x_p = \sigma u_p + \mu$ und $u_{0,4} = -0,25335$ sowie $u_{0,7} = 0,5244$ ergibt sich das lineare Gleichungssystem

$$\begin{aligned} 0,7467 &= \sigma \cdot (-0,25335) + \mu \\ 1,5244 &= \sigma \cdot 0,5244 + \mu \end{aligned}$$

Subtrahiert man die erste Gleichung von der zweiten, so erhält man

$$\begin{aligned} 0,7467 &= -0,25335 \cdot \sigma + \mu \\ 0,7777 &= 0,7777 \cdot \sigma \end{aligned}$$

Also

$$\begin{aligned} \sigma &= 1 \\ 0,7467 = -0,25335 + \mu &\Longrightarrow \mu = 1 \end{aligned}$$

D.h. die Lösung dieses Systems ist $\mu = 1$ und $\sigma^2 = 1$.

b) Die Wahrscheinlichkeit, dass eine normalverteilte Zufallsvariable X mit $E[X] = V[X] = 1$ einen negativen Wert annimmt, ist

$$\begin{aligned} P(X < 0) &= F(0) \\ &= \Phi\left(\frac{0-\mu}{\sigma}\right) \\ &= \Phi(-1) \\ &= 0,1587 \end{aligned}$$

Aufgabe 7.6

Eine große Hausverwaltung, die einen Bestand von 10 000 Wohnungen betreut, möchte etwas über die Zufriedenheit der Mieter mit ihren Wohnungen wissen. Dazu werden aus dem Bestand 100 Wohnungen zufällig und ohne Zurücklegen ausgewählt und die Mieter befragt. Gehen Sie im Folgenden davon aus, dass in der Grundgesamtheit der 10 000 Wohnungen 70% der Mieter mit der Wohnung im Allgemeinen zufrieden sind. 5% der Mieter sind jedoch sehr unzufrieden.

a) Wie groß ist die Wahrscheinlichkeit dafür, dass unter den 100 ausgewählten Mietern weniger als 50 mit der Wohnung im Allgemeinen zufrieden sind?
Geben Sie zunächst eine Formel für die gesuchte Wahrscheinlichkeit an und berechnen Sie die Wahrscheinlichkeit dann näherungsweise durch geeignete Approximationen.

b) Wie groß ist die Wahrscheinlichkeit dafür, dass unter den 100 ausgewählten Mietern mindestens zwei mit der Wohnung sehr unzufrieden sind?

Lösung:

a) X = Anzahl Mieter, die i.Allg. zufrieden sind $\sim H(n = 100; N = 10000; M = 0,7 \cdot 10000 = 7000)$
Gesucht ist also:

$$P(X < 50) = \sum_{x=0}^{49} \frac{\binom{7000}{x} \cdot \binom{3000}{100-x}}{\binom{10000}{100}}.$$

Die Wahrscheinlichkeit kann approximativ berechnet werden, denn:

$$\begin{aligned} &H(n = 100; N = 10000; M = 7000) \\ &\quad \simeq B\left(n = 100; \pi = \frac{M}{N} = 0,7\right) \quad \text{denn: } \frac{n}{N} = 0,01 \leq 0,05 \\ &\quad \simeq N\left(\mu = n\pi = 70; \sigma^2 = 21\right) \quad \text{denn: } \sigma^2 = n\pi(1-\pi) = 21 > 9. \end{aligned}$$

Also ist (unter Berücksichtigung einer Stetigkeitskorrektur):

$$P(X < 50) \approx \Phi\left(\frac{50 - 0,5 - 70}{\sqrt{21}}\right) = \Phi(-4,473) \approx 0.$$

b) Y = Anzahl Mieter, die sehr unzufrieden sind
$Y \sim H(n = 100; N = 10000; M = 0,05 \cdot 10000 = 500)$
Gesucht ist also:

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - \sum_{x=0}^{1} \frac{\binom{500}{x} \cdot \binom{9500}{100-x}}{\binom{10000}{100}}.$$

Die Wahrscheinlichkeit kann approximativ berechnet werden, denn:

$$\begin{aligned} &H(n = 100; N = 10000; M = 500) \\ &\quad \simeq B\left(n = 100; \pi = \frac{M}{N} = 0,05\right) \quad \text{denn: } \frac{n}{N} = 0,01 \leq 0,05 \\ &\quad \simeq Po\left(\mu = n\pi\right) \qquad \text{denn: } n = 100 \geq 50; \mu = n\pi = 5 \leq 9; \pi = 0,05 \leq 0,1. \end{aligned}$$

Also ist:

$$\begin{aligned} P(Y \geq 2) &\approx 1 - 0,0067 - 0,0337 \\ &= 0,9596. \end{aligned}$$

Eine Approximation über die Normalverteilung ist nicht möglich wegen $\sigma^2 = 5 \cdot 0,95 = 4,75 < 9$.
Die Wahrscheinlichkeit kann auch exakt berechnet werden. Man erhält

$$\begin{aligned} P(Y \geq 2) = 1 - P(Y \leq 1) &= 1 - \frac{\binom{500}{0} \cdot \binom{9500}{100}}{\binom{10000}{100}} - \frac{\binom{500}{1} \cdot \binom{9500}{99}}{\binom{10000}{100}} \\ &= 1 - 0,00577 - 0,03067 = 0,96356. \end{aligned}$$

Über die Binomialverteilung erhält man $P\left(Y \geq 2\right) = 0,9629$.

Aufgabe 7.7

Der Ausschuss-Prozentsatz in einer Produktionsserie von Lautsprecherboxen betrage 6%. Wie groß ist die Wahrscheinlichkeit, bei einer Palette von 50 Stück aus dieser Produktion höchstens drei Ausschussstücke zu finden?

Lösung:

X = Anzahl Ausschussstücke in der Palette

$X \sim H(n = 50, N = \infty, M = 0,06 \cdot N = 0,06 \cdot \infty)$

$\simeq B(n = 50; \pi = 0,06)$ (denn: $\frac{n}{N} \approx 0 \leq 0,05$ und $\pi = \frac{M}{N} = 0,06$)

$\simeq Po(\mu = 3)$ (denn: $n \geq 50$, $\mu = 3 \leq 9$ und $\pi = 0,06 \leq 0,1$)

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0,0498 + 0,1494 + 0,2240 + 0,2240 \\ &= 0,6472 \end{aligned}$$

Aufgabe 7.8

In einer Lieferung von 400 Computernetzgeräten sind 40 defekte Geräte enthalten. Wie groß ist die Wahrscheinlichkeit, dass sich bei einer Überprüfung von 20 zufällig ausgewählten Geräten mehr als vier Geräte als defekt erweisen?

Lösung:

$X =$ Anzahl der defekten Geräte $\sim H(n = 20; N = 400; M = 40)$

$\simeq B\left(n = 20; \pi = \frac{40}{400}\right) = B\left(n = 20; \pi = 0,1\right)$, denn der Auswahlsatz $\frac{n}{N} = \frac{20}{400} = 0,05 \leq 0,05$ ist hinreichend klein

$$\begin{aligned} P\left(X > 4\right) &= 1 - P\left(X \leq 4\right) \\ &= 1 - (0,1216 + 0,2702 + 0,2852 + 0,1901 + 0,0898) \\ &= 1 - 0,9569 \\ &= 0,0431 \end{aligned}$$

Zum Vergleich die exakt berechnete Wahrscheinlichkeit:

$$P\left(X > 4\right) = 1 - P\left(X \leq 4\right) = 0,0389.$$

Aufgabe 7.9

In einer Urne befinden sich zehn Kugeln, welche mit den Ziffern $0, 1, 2, ..., 8, 9$ durchnummeriert sind.

a) Berechnen Sie Erwartungswert und Varianz der gezogenen Ziffer bei einmaligem zufälligen Ziehen aus der Urne.

b) Aus der Urne wird 50 Mal mit Zurücklegen gezogen. Berechnen Sie Erwartungswert und Varianz der Summe der gezogenen Ziffern.

c) Wie ist die Summe der gezogenen Ziffern näherungsweise verteilt?

Lösung:

a) X = gezogene Ziffer bei einmaligem Ziehen
(unter Zuhilfenahme der Summenformeln aus der Formelsammlung)

$$\begin{aligned} E[X] &= \frac{1}{10}(0 + 1 + ... + 8 + 9) = \frac{45}{10} = 4,5 \\ E[X^2] &= \frac{1}{10}\left(0^2 + 1^2 + ... + 8^2 + 9^2\right) = \frac{285}{10} = 28,5 \\ V[X] &= E[X^2] - E[X]^2 = 8,25 \end{aligned}$$

b) $Z = \sum_{i=1}^{50} X_i$ mit X_i wie in Teil a)

$$\begin{aligned} E[Z] &= 50 \cdot E[X_1] = 225 \\ V[Z] &= 50 \cdot V[X_1] = 412,5 \qquad \text{da die } X_i \text{ unabhängig wegen ZmZ} \end{aligned}$$

c) Nach dem Zentralen Grenzwertsatz ist Z annähernd normalverteilt mit Erwartungswert und Varianz wie in Teil b), also

$$Z \simeq N(225; 412,5).$$

Aufgabe 7.10

Zwei Würfel werden 100mal gleichzeitig geworfen. Wie groß ist die Wahrscheinlichkeit, dass mindestens dreimal zwei Sechsen oben liegen?

Lösung:

Sei X = Anzahl Doppelsechser bei 100 Würfen $\sim B(n = 100; \pi = \frac{1}{36})$.
Damit ist

$$\begin{aligned}
P(X \geq 3) &= 1 - P(X \leq 2) \\
&= 1 - \left[\binom{100}{0}\left(\frac{1}{36}\right)^0\left(\frac{35}{36}\right)^{100} + \binom{100}{1}\left(\frac{1}{36}\right)^1\left(\frac{35}{36}\right)^{99} + \binom{100}{2}\left(\frac{1}{36}\right)^2\left(\frac{35}{36}\right)^{98}\right] \\
&= 1 - [0,0598 + 0,1708 + 0,2416] \\
&= 1 - 0,4722 \\
&= 0,5278
\end{aligned}$$

Mit der Poisson-Approximation $X \simeq Po\left(\mu = \frac{100}{36}\right)$ (denn: $n = 100 \geq 50$, $\pi = \frac{1}{36} \leq 0,1$ und $n\pi = 2,78 \leq 9$) erhält man:

$$\begin{aligned}
P(X \geq 3) &= 1 - P(X \leq 2) \\
&\approx 1 - e^{-\frac{100}{36}}\left[\frac{1}{0!}\left(\frac{100}{36}\right)^0 + \frac{1}{1!}\left(\frac{100}{36}\right)^1 + \frac{1}{2!}\left(\frac{100}{36}\right)^2\right] \\
&= 1 - [0,0622 + 0,1727 + 0,2399] \\
&= 1 - 0,4748 \\
&= 0,5252
\end{aligned}$$

Zum Vergleich: Die Normalapproximation ist hier nicht anwendbar, da sie ein zu stark abweichendes Ergebnis (0,5675 = 7,5% Abweichung) liefert. Außerdem ist die Faustregel nicht erfüllt.

8 Schließende Statistik – Schätzen

Aufgabe 8.1

Aus der Grundgesamtheit der Wähler eines Bundeslandes werden $n = 1100$ Personen zufällig und mit Zurücklegen ausgewählt. Sei

$$X_i = \begin{cases} 1 & i\text{-te Person wählt die A-Partei,} \\ 0 & \text{sonst.} \end{cases}$$

a) Wie ist $\sum_{i=1}^{1100} X_i$ zu interpretieren?

b) Wie ist $\sum_{i=1}^{1100} X_i$ verteilt?

c) Nehmen Sie an, genau 40% der Wähler des Bundeslandes wählen die A-Partei. Wie groß ist die Wahrscheinlichkeit dafür, dass sich in der Stichprobe mehr als 50% (also mehr als 550) A-Wähler befinden?

Lösung:

a) Der Ausdruck bzw. die Zufallsvariable $\sum_{i=1}^{1100} X_i$ zählt die Summe/Anzahl aller $n = 1100$ Wähler, welche die Partei A gewählt haben.

b) Sei π die Wahrscheinlichkeit, dass die i-te Person die Partei A wählt (für alle $i = 1, ..., 1100$).
Dann ist X_i Bernoulli-verteilt und $E[X_i] = \pi$ sowie $V[X_i] = \pi(1 - \pi)$.
Als Summe Bernoulli-verteilter Zufallsvariablen ist $\sum_{i=1}^{1100} X_i$ binomialverteilt mit Erwartungswert 1100π und Varianz $1100\pi(1 - \pi)$, also

$$\sum_{i=1}^{1100} X_i \sim B(1100; \pi) \simeq N(1100\pi; 1100\pi(1 - \pi))$$

(nach dem Zentralen Grenzwertsatz).

c) Nun: $\pi = 0,4$.
Gesucht:

$$\begin{aligned}
P\left(\sum_{i=1}^{1100} X_i > 550\right) &= 1 - P\left(\sum_{i=1}^{1100} X_i \leq 550\right) \\
&\approx 1 - \Phi\left(\frac{550 + 0,5 - 1100 \cdot 0,4}{\sqrt{1100 \cdot 0,4 \cdot 0,6}}\right) \\
&= 1 - \Phi\left(\frac{110,5}{\sqrt{264}}\right) \\
&= 1 - \Phi\left(6,80\right) \\
&\approx 0
\end{aligned}$$

Aufgabe 8.2

Die Brenndauer von Glühlampen (in 1000 Stunden) eines bestimmten Typs sei $Exp(\lambda)$-verteilt. $X_1, X_2, ..., X_{50}$ sei eine einfache Stichprobe aus $X \sim Exp(\lambda)$.

a) Wie sind $\sum_{i=1}^{50} X_i$ und $\frac{1}{50}\sum_{i=1}^{50} X_i$ näherungsweise verteilt, wenn $\lambda = 0,2$ ist?

b) Betrachten Sie die von X_i abhängenden Zufallsvariablen

$$Y_i = \begin{cases} 1 & X_i \geq 3, \\ 0 & X_i < 3. \end{cases}$$

(i) Wie sind $\sum_{i=1}^{50} Y_i$ und $\frac{1}{50}\sum_{i=1}^{50} Y_i$ zu interpretieren?

(ii) Wie ist $\sum_{i=1}^{50} Y_i$ verteilt, wenn $\lambda = 0,2$ ist?

Lösung:

a) Aufgrund des Zentralen Grenzwertsatzes ist

$$\sum_{i=1}^{50} X_i \simeq N\left(\frac{50}{\lambda} = 250; \frac{50}{\lambda^2} = 1250\right)$$

$$\frac{1}{50}\sum_{i=1}^{50} X_i \simeq N\left(\frac{1}{50}\cdot\frac{50}{\lambda} = 5; \frac{1}{50^2}\cdot\frac{50}{\lambda^2} = 0,5\right)$$

b) Die ZVen Y_i sind Bernoulli-verteilt mit $\pi = P(Y_i = 1) = P(X_i \geq 3) = 1 - P(X_i \leq 3) = 1 - (1 - e^{-3\lambda}) = e^{-3\lambda}$.

(i) $\sum_{i=1}^{50} Y_i$ ist die Anzahl Glühlampen, deren Brenndauer mindestens 3000 Stunden beträgt; $\frac{1}{50}\sum_{i=1}^{50} Y_i$ ist der Anteil Glühlampen, deren Brenndauer mindestens 3000 Stunden beträgt.

(ii) Als Summe Bernoulli-verteilter ZVen ist $\sum_{i=1}^{50} Y_i$ binomialverteilt mit $n = 50$ und (mit $\lambda = 0,2$) $\pi = e^{-3\cdot 0,2} = e^{-0,6} = 0,5488$.

Die Summe ist approximativ normalverteilt, denn $\sum_{i=1}^{50} Y_i \sim B(n = 50; \pi = 0,5488) \simeq N(\mu = 27,44; \sigma^2 = 12,381)$.

Aufgabe 8.3

Eine Zufallsvariable X sei normalverteilt mit Erwartungswert 100 und Varianz 36. Es wird eine einfache Zufallsstichprobe $X_1, X_2, ..., X_{10}$ vom Umfang $n = 10$ gezogen und der Stichprobenmittelwert $\overline{X}$ berechnet.

a) Bestimmen Sie den Erwartungswert des Stichprobenmittelwerts.

b) Bestimmen Sie die Varianz des Stichprobenmittelwerts.

c) Standardisieren Sie den Stichprobenmittelwert und geben Sie die Verteilung der neuen Zufallsvariable an.

d) Wie groß ist die Wahrscheinlichkeit, dass $\overline{X}$ größer als 102 ist?

e) Berechnen Sie das 0,1-Quantil und das 0,9-Quantil von $\overline{X}$. Was sagen sie aus?

Lösung:

Die Stichprobenvariablen X_i sind $N(\mu_X = 100; \sigma_X^2 = 36)$ -verteilt. Der Umfang der einfachen Zufallsstichprobe ist $n = 10$. Somit ist $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ auch wieder normalverteilt.

a)

$$E\left[\overline{X}\right] = \mu_{\overline{X}} = \mu_X = 100$$

b)

$$V\left[\overline{X}\right] = \sigma_{\overline{X}}^2 = \frac{1}{n}\sigma_X^2 = 3,6$$

c) Aus den Teilen a) und b) folgt, dass $\overline{X} \sim N(100; 3,6)$.
Damit ist

$$\frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - 100}{\sqrt{3,6}} \sim N(0;1)$$

d)

$$\begin{aligned} P(\overline{X} > 102) &= 1 - P(\overline{X} \leq 102) = 1 - \Phi\left(\frac{102 - 100}{\sqrt{3,6}}\right) \\ &= 1 - \Phi(1,0541) = 1 - 0,8531 = 0,1469 \end{aligned}$$

e) 0,1-Quantil von $\overline{X}$ = Wert $x_{0,1}$, so dass $P(\overline{X} \leq x_{0,1}) = 0,1$ (0,9-Quantil entsprechend).
Das p-Quantil ist derjenige Wert, der mit Wahrscheinlichkeit p höchstens erreicht wird.
Es ist $x_{0,9} = \mu_{\overline{X}} + \sigma_{\overline{X}} \cdot u_{0,9}$ und $x_{0,1} = \mu_{\overline{X}} + \sigma_{\overline{X}} \cdot u_{0,1}$ (vgl. Formelsammlung), wobei u_p das entsprechende Quantil der Standardnormalverteilung bezeichnet.
Mit $u_{0,9} = 1,2816$ und $u_{0,1} = -u_{0,9}$ erhält man

$$x_{0,9} = 102,43 \quad \text{und} \quad x_{0,1} = 97,57$$

(Da die Normalverteilung symmetrisch um den Erwartungswert ist, hätte man den zweiten Wert auch durch Symmetrieüberlegungen erhalten.)

Aufgabe 8.4

Sei $X_1, X_2, ..., X_n$ eine einfache Stichprobe aus X mit $E[X] = \mu$ und $V[X] = \sigma^2$. Zur Schätzung von μ werden die folgenden Schätzer vorgeschlagen:

$$\begin{aligned} T_1 &= \overline{X} \\ T_2 &= \frac{1}{2}(X_2 + X_n) \\ T_3 &= \frac{4}{5}\overline{X} \end{aligned}$$

a) Prüfen Sie die Schätzer auf Erwartungstreue.

b) Bestimmen Sie die Varianzen der Schätzer.

c) Welchen der Schätzer würden Sie vorziehen?

Lösung:

a)

$$\begin{aligned} T_1 &= \overline{X} \\ E[T_1] &= E\left[\overline{X}\right] = E\left[\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \cdot E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n} \cdot n \cdot \mu = \mu \\ &\Rightarrow T_1 = \overline{X} \text{ ist erwartungstreu für } \mu \\ T_2 &= \frac{1}{2}X_2 + \frac{1}{2}X_n \\ E[T_2] &= E\left[\frac{1}{2}X_2 + \frac{1}{2}X_n\right] = \frac{1}{2} \cdot (E[X_2] + E[X_n]) = \frac{1}{2} \cdot 2 \cdot \mu = \mu \\ &\Rightarrow T_2 = \frac{1}{2}(X_2 + X_n) \text{ ist erwartungstreu für } \mu \\ T_3 &= \frac{4}{5}\overline{X} \\ E[T_3] &= E\left[\frac{4}{5}\overline{X}\right] = \frac{4}{5} \cdot E\left[\overline{X}\right] = \frac{4}{5} \cdot \frac{1}{n} \cdot n \cdot \mu = \frac{4}{5}\mu \\ &\Rightarrow T_3 = \frac{4}{5}\overline{X} \text{ ist nicht erwartungstreu für } \mu \end{aligned}$$

b)

$$
\begin{aligned}
T_1 &= \overline{X} \\
V[T_1] &= V\left[\overline{X}\right] = V\left[\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \cdot \sum_{i=1}^{n} V[X_i] = \frac{1}{n^2} \cdot n \cdot V[X] = \frac{V[X]}{n} \\
T_2 &= \frac{1}{2}X_2 + \frac{1}{2}X_n \\
V[T_2] &= V\left[\frac{1}{2}X_2 + \frac{1}{2}X_n\right] = \frac{1}{4} \cdot \left(V\left[X_2\right] + V\left[X_n\right]\right) = \frac{1}{4} \cdot 2 \cdot V[X] = \frac{1}{2}V[X] \\
T_3 &= \frac{4}{5}\overline{X} \\
V[T_3] &= V\left[\frac{4}{5}\overline{X}\right] = \frac{16}{25} \cdot V\left[\overline{X}\right] = \frac{16 \cdot V[X]}{25 \cdot n}
\end{aligned}
$$

c) Man wählt denjenigen Schätzer, der unter den erwartungstreuen Schätzern (asymptotisch) die geringste Varianz hat. $T_3 = \frac{4}{5}\overline{X}$ fällt weg, da dieser Schätzer nicht erwartungstreu ist. Als bester Schätzer stellt sich $T_1 = \overline{X}$ heraus, da seine Varianz mit $\frac{1}{n} \cdot V\left[X\right]$ für $n > 2$ immer kleiner ist als die Varianz des Schätzers $T_2 = \frac{1}{2}\left(X_2 + X_n\right)$.

Aufgabe 8.5

Auf einer Maschine werden Werkstücke hergestellt, deren Länge eine Varianz von $(2,8 \text{ mm})^2$ aufweist. Eine Zufallsstichprobe von zehn Stück ergab folgende Werte in mm:

41,6	37,1	42,4	39,3	40,2
36,1	37,6	43,5	36,7	37,2

Geben Sie ein Konfidenzintervall für den unbekannten Mittelwert der Werkstücklänge zu einem Niveau von mindestens $1 - \alpha = 0,95$ an.

Lösung:

X = Länge des Werkstücks ; X ist beliebig verteilt

$V[X] = \sigma^2 = (2,8 \text{ mm})^2$

$n = 10$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 39,17$$

Tschebyscheff-Konfidenzintervall für μ zum Niveau von mindestens $1 - \alpha = 0,95$:

$$\begin{aligned}
\widehat{\theta}_{u,o} &= \overline{X} \mp \frac{\sigma}{\sqrt{n\alpha}} \\
&= 39,17 \mp \frac{2,8}{\sqrt{10 \cdot 0,05}} \\
&= 39,17 \mp 3,96 \\
&= [35,21; 43,13]
\end{aligned}$$

Aufgabe 8.6

Das Einzelgewicht X von Bananen einer bestimmten Sorte sei als normalverteilte Zufallsvariable anzusehen. Eine einfache Stichprobe vom Umfang $n = 9$ erbrachte ein Gesamtgewicht von $1686,60$ (Gramm) und eine Stichprobenvarianz von $s^2 = 6,74^2$ (Gramm^2).

a) Geben Sie eine Punktschätzung für das durchschnittliche Gewicht einer einzelnen Banane dieser Sorte an.

b) Geben Sie ein 0,95-Konfidenzintervall für den unbekannten Mittelwert μ an.

Lösung:

a) Punktschätzung

$$\widehat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{n=9}{=} \frac{1686,60}{9} = 187,4 \text{ (Gramm)}$$

b) Standardabweichung bzw. Varianz unbekannt, Normalverteilung =>Konfidenzintervall

$$\begin{aligned}
\widehat{\theta}_{u,o} &= \overline{X} \mp t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}} \\
&= 187,4 \mp t_{8,0,975} \frac{6,74}{\sqrt{8}} \\
&= 187,4 \mp 2,3060 \cdot \frac{6,74}{\sqrt{8}} \\
&= 187,4 \mp 5,495 \\
&= [181,905;\ 192,895]
\end{aligned}$$

Aufgabe 8.7

In der Mühle eines Freilichtmuseums wird nach alter Tradition Mehl gemahlen und in Säcke verpackt. Das Einzelgewicht X eines Mehlsacks kann als normalverteilte Größe angesehen werden. Die Varianz des Gewichts sei aus langjähriger Erfahrung bekannt und betrage (15 Gramm)^2. Eine einfache Stichprobe vom Umfang $n = 16$ erbrachte ein Gesamtgewicht von 7936 Gramm.

a) Geben Sie eine Punktschätzung für das durchschnittliche Gewicht eines einzelnen Mehlsacks an.

b) Bestimmen Sie ein konkretes 0,90-Konfidenzintervall für den Erwartungswert μ an.

c) Wie groß muss der Stichprobenumfang n mindestens sein, wenn das 90%-Konfidenzintervall höchstens eine Breite von 6 Gramm haben soll?

Lösung:

a) Punktschätzung

$$\widehat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{n=16}{=} \frac{7936}{16} = 496 \text{ (Gramm)}$$

b) Varianz bekannt, Normalverteilung => 90%–Konfidenzintervall

$$\begin{aligned}\widehat{\theta}_{u,o} &= \overline{X} \mp u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\ &= 496 \mp u_{0,95} \frac{15}{\sqrt{16}} \\ &= 496 \mp 1,6449 \cdot \frac{15}{\sqrt{16}} \\ &= 496 \mp 6,1684 \\ &= [489,8316;\ \ 502,1684]\end{aligned}$$

c) Man hat $2b = 6$, also $b = 3$.
Der Mindeststichprobenumfang (Varianz bekannt) ergibt sich aus der Formel

$$\begin{aligned}n_{\min} &= \frac{u^2_{1-\frac{\alpha}{2}}\sigma^2}{b^2} \\ &= \frac{1,6449^2 \cdot 15^2}{3^2} \\ &= 67,6424\end{aligned}$$

Also beträgt der Mindeststichprobenumfang 68.

Aufgabe 8.8

Die Vorschriften für die Produktion einer bestimmten Aluminiumlegierung verlangen einen durchschnittlichen Natriumgehalt von mindestens $0,03\%$ und höchstens $0,1\%$. Aus Erfahrung weiß man, dass der Natriumgehalt einer Legierung normalverteilt mit dem Erwartungswert μ und der Standardabweichung $\sigma = 0,015\%$ ist.

Eine Stichprobe der Länge 9 ergab die folgenden Werte für den Natriumgehalt (in %):

$$0,0701 \ ; \ 0,0697 \ ; \ 0,0693 \ ; \ 0,0698 \ ; \ 0,0703 \ ;$$
$$0,0700 \ ; \ 0,0702 \ ; \ 0,0701 \ ; \ 0,0705 \ .$$

a) Geben Sie ein Konfidenzintervall für μ zum Konfidenzniveau $1 - \alpha = 0,99$ an.

b) Wie groß muss die Stichprobenlänge mindestens sein, damit ein Konfidenzintervall für μ zum Konfidenzniveau $1 - \alpha = 0,99$ eine Breite von höchstens $0,01$ hat?

Lösung:

a) Bei den angegebenen Daten ist

$$n = 9 \quad \text{und} \quad \sum_{i=1}^{9} x_i = 0.63 \, .$$

Für $\overline{x}$ erhält man also

$$\overline{x} = \frac{1}{9} 0.63 = 0.07 \, .$$

Das Konfidenzintervall für μ einer Normalverteilung bei bekannter Varianz ist gegeben durch

$$T_{1,2} = \overline{x} \mp \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \, .$$

Mit $u_{0.995} = 2.5758$ erhält man

$$T_{1,2} = 0.07 \mp \frac{0.015}{\sqrt{9}} \cdot 2.5758 = 0.07 \mp 0.012879 \, .$$

Das Konfidenzintervall für μ ist somit

$$[0.057121, 0.082879] \, .$$

b) Der nötige Stichprobenumfang für ein Konfidenzintervall der Breite $2b$ ist

$$n_{min} = \frac{u_{1-\frac{\alpha}{2}}^2 \sigma^2}{b^2} \,.$$

Mit $u_{1-\frac{\alpha}{2}} = 2.5758$, $\sigma = 0.015$ und $2b = 0.01$ erhält man

$$n_{min} = \frac{2.5758^2 \cdot 0.015^2}{0.005^2} = 59.71 \,.$$

Der Stichprobenumfang muss also mindestens $n = 60$ sein.

9 Schließende Statistik – Testen

Aufgabe 9.1

Eine Ladenkette fordert von den Erzeugern für Eisbergsalat ein mittleres Kopfgewicht von mindestens 1000 Gramm. Das Gewicht kann als normalverteilte Größe angesehen werden. Aus einer Lieferung wird eine Stichprobe vom Umfang $n = 7$ gezogen. Man erhält $\overline{x} = 960,7$ (Gramm) und $s = 46,5$ (Gramm).
Prüfen Sie auf einem Signifikanzniveau von $\alpha = 0,05$, ob das mittlere Kopfgewicht der Forderung entspricht.

Lösung:

Test für Mittelwert μ bei unbekanntem σ, einfache Zufallsstichprobe aus Normalverteilung $\Longrightarrow$ t-Test

- $H_0 : \mu \geq 1000$ gegen $H_1 : \mu < 1000$
- Prüffunktion:
$$T = \frac{\overline{X} - \mu_0}{S}\sqrt{n-1} \sim t_{n-1} \text{ unter } H_0$$
- Wert der Prüffunktion:
$$t = \frac{960,7 - 1000}{46,5}\sqrt{6} = -2,07$$
- Linksseitiger Test mit $\alpha = 0,05$. Kritischer Bereich:
$$K_{0,05} = \{t | t < -t_{n-1;1-\alpha}\} = \{t | t < -t_{6;0,95} = -1,9432\}$$
- Entscheidung:
$$t = -2,07 \in K_{0,05}$$
$$\Longrightarrow H_0 \text{ ablehnen}$$

  D.h. zum Signifikanzniveau von 5% ist das mittlere Kopfgewicht geringer als 1000 Gramm.

Aufgabe 9.2

Ein Hersteller von Ventilatoren für PC's gibt für die Ventilatoren eine mittlere Lebensdauer von mindestens 3000 Stunden an. Ein Verbraucherinstitut behauptet, die Lebensdauer sei geringer, und testet 50 Ventilatoren. Für diese ergibt sich eine mittlere Lebensdauer von 2900 Stunden bei einer Stichprobenvarianz von (160 Stunden)^2.

Ist die Behauptung des Herstellers mit dem Stichprobenergebnis vereinbar ($\alpha = 0,05$)?

Lösung:

Test für Mittelwert μ bei unbekanntem σ, einfache Zufallsstichprobe, keine Normalverteilung, aber $n = 50 > 40 \Longrightarrow$ approximativer t-Test

- $H_0 : \mu \geq 3000 \qquad \text{gegen} \qquad H_1 : \mu < 3000$
- Prüffunktion:
$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n-1}}} \sim t_{n-1} \text{ unter } H_0$$
- Wert der Prüffunktion:
$$t = \frac{2900 - 3000}{\frac{160}{\sqrt{49}}} = -4,375$$
- Linksseitiger Test mit $\alpha = 0,05$. Kritischer Bereich:
$$K_{0,05} = \{t | t < -t_{49;0,95} \approx -u_{0,95}\} = \{t | t < -1,6449\}$$
  Approximation der Quantile der t-Verteilung durch die Quantile der Standardnormalverteilung (da $\nu > 40$)
- Entscheidung:
$$t = -4,375 \in K_{0,05}$$
$$\Longrightarrow H_0 \text{ ablehnen}$$
  D.h. die mittlere Lebensdauer ist signifikant kleiner als 3000 Stunden.

Zusatz:

Nehmen Sie an, die wahre durchschnittliche Lebensdauer betrage nur 2950 Stunden. Wie groß ist die Wahrscheinlichkeit, Nullhypothese fälschlicherweise beizubehalten?

Lösung: Gesucht ist die Wahrscheinlichkeit für einen Fehler 2. Art

$$
\begin{aligned}
P(T \notin K_{0,05} \mid \mu = 2950) &= P\left(\frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n-1}}} \geq -1,6449 \mid \mu = 2950\right) \\
&= P\left(\frac{\overline{X} - 2950 + 2950 - 3000}{\frac{160}{\sqrt{49}}} \geq -1,6449 \mid \mu = 2950\right) \\
&= P\left(\frac{\overline{X} - 2950}{\frac{160}{\sqrt{49}}} \geq -1,6449 - \frac{2950 - 3000}{\frac{160}{\sqrt{49}}}\right) \\
&= 1 - \Phi\,(0,5426) \\
&= 1 - 0,7063 \qquad \text{(mit linearer Interpolation)} \\
&= 0,2937
\end{aligned}
$$

Zusatz (nicht veröffentlicht):

Ein Statistiker hat für den β-Fehler bei dem oben angegebenen Test einen Wert von 0,8 berechnet. Welcher Wert für μ liegt seiner Berechnung zugrunde?

Lösung: Wahrscheinlichkeit für einen Fehler 2. Art

$$
\begin{aligned}
P(T \notin K_{0,05} \mid \mu = \mu_1) &= P\left(\frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n-1}}} \geq -1,6449 \mid \mu = \mu_1\right) \\
&= P\left(\frac{\overline{X} - \mu_1 + \mu_1 - \mu_0}{\frac{S}{\sqrt{n-1}}} \geq -1,6449 \mid \mu = \mu_1\right) \\
&= P\left(\frac{\overline{X} - \mu_1}{\frac{S}{\sqrt{n-1}}} \geq -1,6449 - \frac{\mu_1 - \mu_0}{\frac{S}{\sqrt{n-1}}} \mid \mu = \mu_1\right) = 0,8 \\
&\Longrightarrow u_{0,2} = -u_{0,8} = -1,6449 - \frac{\mu_1 - \mu_0}{\frac{S}{\sqrt{n-1}}} \\
&\Longrightarrow u_{0,8} = 1,6449 + \frac{\mu_1 - 3000}{\frac{160}{\sqrt{49}}} \\
&\Longrightarrow 0,8416 = 1,6449 + \frac{\mu_1 - 3000}{\frac{160}{\sqrt{49}}} \\
&\Longrightarrow \mu_1 = 3000 + (0,8416 - 1,6449) \cdot \frac{160}{7} = 2981,64
\end{aligned}
$$

Aufgabe 9.3

Die Zufallsvariablen $X_1, X_2, ..., X_n$ seien unabhängig und identisch normalverteilt mit Standardabweichung $\sigma = 5$ und unbekanntem Erwartungswert.
Eine einfache Zufallsstichprobe vom Umfang $n = 81$ ergibt einen Stichprobenmittelwert von $\overline{x} = 37$.

a) Überprüfen Sie mittels eines geeigneten Testverfahrens die Nullhypothese, dass der Erwartungswert μ größer oder gleich 38, gegen die Alternative, dass der Erwartungswert kleiner als 38 ist, bei einem α von 0,05.

b) Nehmen Sie an, dass der Erwartungswert tatsächlich 37 ist. Wie groß ist die Wahrscheinlichkeit für einen Fehler 2. Art?

Lösung:

Die Stichprobenvariablen X_i sind $N(\mu; \sigma^2 = 5^2)$ -verteilt. Der Umfang der einfachen Zufallsstichprobe ist $n = 81$.

a) Gegeben sind $\overline{x} = 37$ und $\alpha = 0,05$.
Test für unbekannten Mittelwert μ bei bekanntem σ. Einfache Zufallsstichprobe aus Normalverteilung $\Longrightarrow$ Gauß-Test

- $H_0 : \mu_X \geq 38 \qquad H_1 : \mu_X < 38$
- Prüffunktion:

$$T = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0;1) \text{ unter } H_0$$

- Wert der Prüffunktion:

$$t = \frac{37 - 38}{\frac{5}{9}} = -1,8$$

- Linksseitiger Test mit $\alpha = 0,05$. Kritischer Bereich:

$$K_\alpha = K_{0,05} = \{t \mid t < -u_{1-0,05} = -u_{0,95} = -1,6449\}$$

- Entscheidung:

$$t = -1,8 \in K_{0,05}$$
$$=> H_0 \text{ ablehnen}$$

D.h. der Mittelwert ist signifikant kleiner als 38.

b) Sei $\mu = 37$. Die Fehlerwahrscheinlichkeit 2.Art ist (bei $\alpha = 0,05$) und mit Interpolation

$$\beta = P(T \notin K_{0,05} \mid \mu = 37) = P\left(\frac{\overline{X} - 38}{\frac{5}{9}} \geq -1,6449 \mid \overline{X} \sim N\left(37; \frac{25}{81}\right)\right)$$
$$= P\left(\frac{\overline{X} - 37}{\frac{5}{9}} + \frac{37 - 38}{\frac{5}{9}} \geq -1,6449\right)$$
$$= P\left(\frac{\overline{X} - 37}{\frac{5}{9}} \geq -1,6449 - \frac{37 - 38}{\frac{5}{9}}\right)$$
$$= P\left(\frac{\overline{X} - 37}{\frac{5}{9}} \geq -1,6449 + 1,8\right)$$
$$= 1 - \Phi(0,1551)$$
$$= 0,4384$$

Aufgabe 9.4

Von der Abfüllanlage einer Brauerei werden Flaschen gefüllt, wobei die Füllmenge X pro Flasche gewissen Schwankungen unterliegt und als normalverteilte Zufallsvariable mit bekannter Standardabweichung $\sigma = 1,5[cm^3]$ angesehen werden kann. Die Hypothese H_0, dass der Erwartungswert μ dieser Normalverteilung gleich dem Sollwert $\mu_0 = 330[cm^3]$ ist, soll anhand einer Stichprobe der Länge $n = 30$ überprüft werden.

Aufgrund der Interessenlage derjenigen Personen, die die Untersuchung vornehmen, unterscheiden wir drei Fälle, nämlich: Die Überprüfung geschieht durch:

a) (i) eine Eichkommission, die an einer Abweichung vom Sollwert $\mu_0 = 330$ sowohl nach unten als auch nach oben interessiert ist,

(ii) eine Verbraucherorganisation, deren Interesse nur der Frage gilt, ob der wahre Erwartungswert μ kleiner als der Sollwert μ_0 ist,

(iii) den Brauereibesitzer, von dem wir hier annehmen, dass er lediglich wissen will, ob im Mittel zu viel abgefüllt wird.

b) Formulieren Sie für jeden der drei Fälle das entsprechende Testproblem (d.h. H_0 und H_1) und bestimmen Sie dafür jeweils einen Test. Das Niveau der Tests sei jeweils $\alpha = 0,01$. Für welche Werte von $\overline{x}$ wird die Nullhypothese bei den drei Tests abgelehnt?

c) Als Stichprobenmittel für die Füllmenge ergab sich der Wert $\overline{x} = 329,33[cm^3]$. Wie entscheiden Sie sich in den drei Fällen?

Lösung:

a) Die Testgröße für alle drei Tests ist

$$T = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} = \frac{\overline{x} - 330}{1.5}\sqrt{30}\,.$$

Nullhypothese, Gegenhypothese und kritischer Bereich K sind wie folgt zu wählen:

(i) $H_0 : \mu = 330$, $H_1 : \mu \neq 330$, H_0 ablehnen, falls $|T| > 2.5758$,

(ii) $H_0 : \mu \geq 330$, $H_1 : \mu < 330$, H_0 ablehnen, falls $T < -2.3263$,

(iii) $H_0 : \mu \leq 330$, $H_1 : \mu > 330$, H_0 ablehnen, falls $T > 2.3263$.

b) Im Fall 1. wird H_0 abgelehnt, wenn

$$|T| > 2.5758 \quad \Longleftrightarrow \quad \left|\frac{\overline{x} - \mu_0}{\sigma}\sqrt{n}\right| > 2.5758 \quad \Longleftrightarrow \quad \overline{x} < 329.29 \text{ oder } \overline{x} > 330.71\,.$$

Entsprechend wird im Fall 2. H_0 abgelehnt, wenn

$$\overline{x} < 329.36$$

und im Fall 3., wenn

$$\overline{x} > 330.64\,.$$

c) Es ist

$$t = \frac{329.33 - 330}{1.5}\sqrt{30} = -2.4465\,.$$

Somit wird im Fall 2. H_0 abgelehnt, in den Fällen 1. und 3. jedoch nicht.

10 Regressionsanalyse

Aufgabe 10.1

Bei $n = 12$ Personen wurde die Körpergröße x_i [in cm] und das Gewicht y_i [in kg] ermittelt. Dabei ergaben sich die folgenden zusammengefassten Werte:

$$\sum_{i=1}^{n} x_i = 2088, \qquad \sum_{i=1}^{n} y_i = 888$$

$$\sum_{i=1}^{n} x_i^2 = 364464, \qquad \sum_{i=1}^{n} y_i^2 = 67330, \qquad \sum_{i=1}^{n} x_i y_i = 155784$$

Nehmen Sie an, dass die Werte $y_1, ..., y_{12}$ Realisierungen stochastisch unabhängiger Zufallsvariablen $Y_1, ..., Y_{12}$ sind und dass für $i = 1, ..., 12$ gilt:

$$Y_i = a + bx_i + U_i \qquad \text{mit } U_i \sim N\left(0, \sigma^2\right).$$

Ferner seien $U_1, ..., U_{12}$ stochastisch unabhängig.

a) Berechnen Sie Schätzwerte für a, b und σ^2.

b) Schätzen Sie die Standardabweichungen der Schätzer für a und b.

c) Geben Sie ein Konfidenzintervall für a zum Konfidenzniveau $1 - \alpha = 0,95$ an.

d) Testen Sie die Nullhypothese $H_0 : b = 1$ gegen die Alternative $H_1 : b \neq 1$ zu einem Signifikanzniveau $\alpha = 0,05$.

Lösung:

Aus den angegebenen Werten erhält man

$$\overline{x} = \frac{1}{12}2088 = 174, \qquad \overline{y} = \frac{1}{12}888 = 74,$$

$$s_X^2 = \frac{1}{12}364464 - 174^2 = 96, \qquad s_Y^2 = \frac{1}{12}67330 - 74^2 = 134.8333$$

$$s_{XY} = \frac{1}{12}155784 - 174 \cdot 74 = 106.$$

a) Mit den obigen Werten erhält man

$$\hat{b} = \frac{s_{XY}}{s_X^2} = \frac{106}{96} = 1.1042\,,$$

und

$$\widehat{a} = \overline{y} - \hat{b}\overline{x} = 74 - 1.1042 \cdot 174 = -118.1250\,.$$

Der Schätzer für σ^2 ist

$$\hat{\sigma}^2 = \frac{n}{n-2}\left(s_Y^2 - \frac{s_{XY}^2}{s_X^2}\right) = \frac{12}{10}\left(134.8333 - \frac{106^2}{96}\right) = 21.35\,.$$

b) Die Schätzwerte für die Standardabweichungen der Schätzer für a und b erhält man gem

$$\widehat{\sigma_{\hat{b}}} = \frac{\hat{\sigma}}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{\hat{\sigma}^2}{n s_X^2}} = \sqrt{\frac{21.35}{12 \cdot 96}} = 0.1361$$

bzw.

$$\widehat{\sigma_{\hat{a}}} = \widehat{\sigma_{\hat{b}}} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = 0.1361 \cdot \sqrt{\frac{1}{12}364464} = 23.7252\,.$$

c) Das Konfidenzintervall für a ist gegeben durch:

$$\widehat{a} \mp t_{n-2,1-\frac{\alpha}{2}}\widehat{\sigma_{\hat{a}}}$$

Das 0.975-Quantil der t-Verteilung mit 10 Freiheitsgraden ist $t_{10,0.975} = 2.2281$. Damit ergibt sich der Wert des Konfidenzintervalls für a zu

$$-118.1250 \mp 2.2281 \cdot 23.7252 = -118.1250 \mp 52.8621$$

bzw.

$$[-170.9871, -65.2629]\,.$$

d) Der Test für b verwendet die Testgröße

$$T = \frac{\hat{b} - b_0}{\widehat{\sigma_{\hat{b}}}}\,.$$

Die Nullhypothese $H_0 : b = b_0$ wird zugunsten der Alternative $H_1 : b \neq b_0$ abgelehnt, wenn $|T| \geq t_{n-2,1-\frac{\alpha}{2}}$ ist.

Das benötigte t-Quantil wurde bereits in Aufgabenteil c) bestimmt: $t_{10,0.975} = 2.2281$.

Mit $b_0 = 1$ erhält man den Wert der Testgröße zu

$$t = \frac{1.1042 - 1}{0.1361} = 0.7656\,.$$

Wegen $|t| = 0.7656 \leq 2.2281 = t_{10,0.975}$ wird $H_0 : b = 1$ nicht abgelehnt.

Aufgabe 10.2

Eine Unternehmensabteilung ist ausschließlich mit der Herstellung eines Produkts beschäftigt. Für 10 Perioden wurden folgende Produktionsmengen x_i und Gesamtkosten y_i registriert.

i	1	2	3	4	5	6	7	8	9	10
x_i	9	12	14	12	12	13	10	11	12	15
y_i	1216	1300	1356	1288	1276	1292	1260	1244	1288	1360

Nehmen Sie an, dass die Werte $y_1, \ldots, y_{10}$ Realisierungen stochastisch unabhängiger Zufallsvariablen $Y_1, \ldots, Y_{10}$ sind und dass für $i = 1, \ldots, 10$ gilt:

$$Y_i = a + bx_i + U_i \quad \text{mit} \quad U_i \sim N(0, \sigma_U^2)\,.$$

Ferner seien $U_1, \ldots, U_{10}$ stochastisch unabhängig.

a) Berechnen Sie Schätzwerte für a, b und σ^2.

b) Schätzen Sie die Standardabweichungen der Schätzer $\hat{a}$ und $\hat{b}$.

c) Geben Sie ein Konfidenzintervall für b zum Konfidenzniveau $1 - \alpha = 0,95$ an.

d) Testen Sie die Nullhypothese $H_0 : a = 900$ gegen die Alternative $H_1 : a \neq 900$ zu einem Signifikanzniveau $\alpha = 0,05$.

Lösung:

Für die angegebenen Daten ist

$$n = 10, \quad \sum_{i=1}^{10} x_i = 120, \quad \sum_{i=1}^{10} y_i = 12880,$$

$$\sum_{i=1}^{10} x_i^2 = 1468, \quad \sum_{i=1}^{10} y_i^2 = 16607456, \quad \sum_{i=1}^{10} x_i y_i = 155232.$$

Damit ist

$$\overline{x} = \frac{1}{10}120 = 12, \qquad \overline{y} = \frac{1}{10}12880 = 1288,$$

$$s_x^2 = \frac{1}{10}1468 - 12^2 = 2.8, \qquad s_y^2 = \frac{1}{10}16607456 - 1288^2 = 1801.6,$$

$$s_{xy} = \frac{1}{10}155232 - 12 \cdot 1288 = 67.2.$$

a) Die Punktschätzer für a und b sind

$$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{67.2}{2.8} = 24$$

und

$$\hat{a} = \overline{y} - \hat{b}\overline{x} = 1288 - 24 \cdot 12 = 1000\,.$$

Der Schätzer für σ^2 ist

$$\hat{\sigma}^2 = \frac{n}{n-2}\left(s_y^2 - \hat{b}^2 s_x^2\right) = 1,25\left(1801,6 - 24^2 \cdot 2,8\right) = 236\,.$$

b) Es ist

$$\widehat{\sigma_{\hat{b}}} = \frac{\hat{\sigma}}{\sqrt{n s_x^2}} = \frac{\sqrt{236}}{\sqrt{10 \cdot 2,8}} = 2,9032$$

sowie

$$\widehat{\sigma_{\hat{a}}} = \widehat{\sigma_{\hat{b}}} \cdot \sqrt{s_x^2 + \overline{x}^2} = 2,9032 \cdot \sqrt{2,8 + 12^2} = 35,1755\,.$$

c) Das Konfidenzintervall für b ist

$$\hat{b} \mp t_{n-2,1-\frac{\alpha}{2}} \cdot \widehat{\sigma_{\hat{b}}}\,.$$

Das $0,975$-Quantil der t-Verteilung mit 8 Freiheitsgraden ist $t_{8;0,975} = 2,3060$. Damit ergibt sich das Konfidenzintervall für b zu

$$24 \mp 2,3060 \cdot 2,9032 = 24 \mp 6,6948 \quad \text{bzw.} \quad [17,3052; 30,6948]\,.$$

d) Die Testgröße des Tests auf a ist

$$T = \frac{\hat{a} - a_0}{\widehat{\sigma_{\hat{a}}}}\,.$$

Die Nullhypothese $H_0 : a = a_0$ wird zugunsten der Alternative $H_1 : a \neq a_0$ abgelehnt, wenn $|T| \geq t_{n-2,1-\frac{\alpha}{2}}$ ist.

Mit $a_0 = 900$ erhält man den Wert der Testgröße zu

$$T = \frac{1000 - 900}{35,1755} = 2.8429\,.$$

Wegen $|T| = 2,8429 > 2,3060 = t_{8;0,975}$ wird $H_0 : a = 900$ abgelehnt.

Aufgabe 10.3

Bei 50 Neugeborenen wurden jeweils die Größe (in cm) und das Gewicht (in kg) gemessen. Dabei ergab sich eine durchschnittliche Größe von 51,77cm bei einer Standardabweichung von 3,04cm. Das mittlere Gewicht der Kinder betrug 3,38kg bei einer Standardabweichung von 0,63kg. Die Kovarianz von Größe und Gewicht hatte den Wert 1,65.

Nehmen Sie an, dass die Annahmen des Modells der linearen Einfachregression mit normalverteilten Störgrößen erfüllt sind.

a) Schätzen Sie die Koeffizienten einer linearen Regression vom Gewicht auf die Größe.

b) Geben Sie ein Konfidenzintervall für den Koeffizienten b zum Konfidenzniveau 0,95 an.

c) Testen Sie zum Niveau $\alpha = 0,05$ die Nullhypothese, dass der Koeffizient b höchstens den Wert 0,15 hat.

Lösung:

a) Die Punktschätzer der Koeffizienten sind

$$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{1,65}{3,04^2} = 0,1785$$

und

$$\hat{a} = \overline{y} - \hat{b}\overline{x} = 3,38 - 0,1785 \cdot 51,77 = -5,8609\,.$$

b) Wir berechnen zunächst den Schätzer für die Standardabweichung von $\hat{b}$. Dazu berechnen wir als erstes der Schätzer für σ^2

$$\hat{\sigma}^2 = \frac{n}{n-2}\left(s_y^2 - \hat{b}^2 s_x^2\right) = \frac{50}{48}\left(0,63^2 - 0,1785^2 \cdot 3,04^2\right) = 0,1067\,.$$

Damit erhält man

$$\widehat{\sigma_{\hat{b}}} = \frac{\hat{\sigma}}{\sqrt{n s_x^2}} = \frac{\sqrt{0,1067}}{\sqrt{50 \cdot 3,04^2}} = 0,01519\,.$$

Das Konfidenzintervall für b ist

$$\hat{b} \mp t_{n-2,1-\frac{\alpha}{2}} \cdot \widehat{\sigma_{\hat{b}}}\,.$$

Das $0,975$-Quantil der t-Verteilung mit 48 Freiheitsgraden ist $t_{48;0,975} \approx 1,96$. Damit erhält man das Konfidenzintervall

$$0,1785 \mp 1,96 \cdot 0,01519 = 0,1785 \mp 0,02977 \quad \text{bzw.} \quad [0,1488; 0,2083]\,.$$

c) Zu testen ist die Hypothese $H_0 : b \leq 0,15$ gegen $H_1 : b > 0,15$. Die Testgröße ist

$$T = \frac{\hat{b} - b_0}{\widehat{\sigma_{\hat{b}}}} = \frac{0,1785 - 0,15}{0,01519} = 1,8793\,.$$

Die Nullhypothese wird zugunsten der Alternative abgelehnt, wenn $T \geq t_{n-2,1-\alpha}$ ist. Wegen $T = 1,8793 > 1,6449 = t_{48;0,95}$ wird die Nullhypothese $H_0 : b \leq 0,15$ abgelehnt.

Aufgabe 10.4

In einer Studie der britischen Regierung über das Konsumverhalten von Haushalten in verschiedenen Regionen des Landes wurden unter anderem die Anteile der Ausgaben für Tabak X und Alkohol Y an den Gesamtausgaben der Haushalte ermittelt.

Region	X	Y
North	6.47	4.03
Yorkshire	6.13	3.76
Northeast	6.19	3.77
East Midlands	4.89	3.34
West Midlands	5.63	3.47
East Anglia	4.52	2.92
Southeast	5.89	3.20
Southwest	4.79	2.71
Wales	5.27	3.53
Scotland	6.08	4.51
Northern Ireland	4.02	4.56

a) Berechnen Sie den Korrelationskoeffizienten r_{XY}. Wie würden Sie den Zusammenhang zwischen den Ausgabenanteilen für Tabak und Alkohol charakterisieren?
Es gilt

$$\sum_{i=1}^{11} x_i = 59.88\,, \qquad \sum_{i=1}^{11} x_i^2 = 332.3292\,,$$

$$\sum_{i=1}^{11} y_i = 39.80\,, \qquad \sum_{i=1}^{11} y_i^2 = 147.4930\,,$$

$$\sum_{i=1}^{11} x_i y_i = 217.7103\,.$$

b) Berechnen Sie die Koeffizienten der linearen Regression von Alkohol auf Tabak sowie den Anteil der durch die Regression erklärten Varianz.

c) Führen Sie a) und b) erneut durch, jedoch diesmal ohne die Werte für Northern Ireland. Was fällt Ihnen auf?

Lösung:

a) $r_{XY} = \frac{s_{XY}}{s_X s_Y}$ mit $\overline{x} = \frac{59{,}88}{11} = 5,44364$ und $\overline{y} = \frac{39{,}8}{11} = 3,61818$

$$\begin{aligned}
s_{XY} &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy} \\
&= \frac{1}{11}217,7103 - \frac{1}{11}59,88 \cdot \frac{1}{11}39,8 = 0,09578 \\
s_X &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2} \\
&= \sqrt{\frac{1}{11}332,3292 - \left(\frac{1}{11}59,88\right)^2} = 0,760637 \\
s_Y &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \overline{y}^2} \\
&= \sqrt{\frac{1}{11}147,493 - \left(\frac{1}{11}39,8\right)^2} = 0,563218 \\
r_{XY} &= \frac{0,09578}{0,760637 \cdot 0,563218} = 0,2236
\end{aligned}$$

Der Korrelationskoeffizient hat den Wert 0.2236. Es besteht eine schwache positive Korrelation zwischen den Ausgaben für Alkohol und Tabak.

b)

$$\begin{aligned}
b &= \frac{s_{XY}}{s_X^2} = r_{XY}\frac{s_Y}{s_X} = \frac{0,09578}{0,57857} = 0,16555 \\
a &= \overline{y} - b\overline{x} = 3,61818 - 0,16555 \cdot 5,44364 = 2,71699 \\
R^2 &= r_{XY}^2 = 0,22357^2 = 0,04998
\end{aligned}$$

Sei X = Anteil von Tabak an den Gesamtausgaben und Y = Anteil von Alkohol an den Gesamtausgaben. Die Regressionsgerade lautet:

$$y = 2,71699 + 0.16555 \cdot x$$

Der Anteil der durch die Regression erklärten Streuung beträgt 5%.

c) Man hat

$$\begin{aligned}
s_{XY} &= \frac{1}{10}199,3791 - 5,586 \cdot 3,524 = 0,252846 \\
s_X &= \sqrt{\frac{1}{10}316,1688 - 5,586^2} = 0,64303 \\
s_Y &= \sqrt{\frac{1}{10}126,6994 - 3,524^2} = 0,50136 \\
r_{XY} &= \frac{0,252846}{0,64303 \cdot 0,50136} = 0,7843
\end{aligned}$$

und damit

$$\begin{aligned}
b &= \frac{0,252846}{0,413484} = 0,6115 \\
a &= 3,524 - 0,6115 \cdot 5,586 = 0,108161 \\
R^2 &= 0,78429^2 = 0,61511
\end{aligned}$$

Ignoriert man die Werte für Northern Ireland, so erhält man einen Korrelationskoeffizienten von 0.7843, also eine recht hohe Korrelation. Für die Regressionsgerade ergibt sich die Gleichung:

$$y = 0,108161 + 0,6115 \cdot x.$$

Der Anteil der durch die Regression erklärten Streuung beträgt jetzt 61.51%. Die Regressionskoeffizienten haben sich durch den Wegfall Northern Irelands grundlegend geändert. Gleichzeitig ist das Bestimmtheitsmaß stark angestiegen. Dies liegt daran, dass Northern Ireland im Vergleich zu den anderen Regionen ein deutlich anderes Konsumverhalten in Bezug auf Alkohol und Tabak aufweist. Die Ausgaben für Tabak sind im Vergleich zu den Ausgaben für Alkohol sehr hoch.

Aufgabe 10.5

Für fünf ausgewählte private Haushalte wurden jeweils Monatsdurchschnitte für die Höhe des Nettoeinkommens X und die Höhe der Telefonrechnung Y ermittelt:

Haushalt	Nettoeinkommen X (in 1 000 Euro)	Telefonrechnung Y (in 100 Euro)
1	2.6	1.2
2	2.1	0.7
3	1.4	0.5
4	3.5	2.0
5	1.7	0.6

a) Berechnen Sie die Koeffizienten der linearen Einfachregression von Y auf X.

b) Prognostizieren Sie unter Verwendung der Ergebnisse von Aufgabe a) die Höhe der Telefonrechnung für einen Haushalt mit einem durchschnittlichen Nettoeinkommen von 2 200 Euro im Monat.

c) Um wie viel veränderten sich die durchschnittlichen Ausgaben für das Telefon pro Haushalt, wenn das monatliche Nettoeinkommen um 200 Euro steigt?

d) Berechnen Sie das Bestimmtheitsmaß und interpretieren Sie es. Welchen Wert hat der Korrelationskoeffizient?

Lösung:

a) Aus den angegebenen Daten erhält man:

i	x_i	y_i	x_iy_i	x_i^2
1	$2,6$	$1,2$	$3,12$	$6,76$
2	$2,1$	$0,7$	$1,47$	$4,41$
3	$1,4$	$0,5$	$0,7$	$1,96$
4	$3,5$	$2,0$	7	$12,25$
5	$1,7$	$0,6$	$1,02$	$2,89$
$\sum$	$11,3$	5	$13,31$	$28,27$

X = Einkommen, Y = Telefonrechnung
Lineare Regression von Y auf X :

$$\begin{aligned} \overline{x} &= \frac{1}{5} \cdot 11,3 = 2,26 \mathrel{\widehat{=}} 2260 \text{ Euro} \\ \overline{y} &= \frac{1}{5} \cdot 5 = 1 \mathrel{\widehat{=}} 100 \text{ Euro} \end{aligned}$$

Damit

$$
\begin{aligned}
b &= \frac{s_{XY}}{s_{X^2}} = \frac{\sum x_i y_i - n\overline{xy}}{\sum x_i^2 - n \cdot \overline{x}^2} \\
&= \frac{13,31 - 5 \cdot 2,26 \cdot 1}{28,27 - 5 \cdot (2,26)^2} = \frac{2,01}{2,732} = 0,7357
\end{aligned}
$$

und

$$a = \overline{y} - b\overline{x} = -0,6627$$

Also

$$\widehat{y} = -0,6627 + 0,7357 \cdot x$$

b) $x = 2.2 : \widehat{y}_{2.2} = -0.6627 + 0,7357 \cdot 2.2 = 0.9558$ [100 Euro] $\approx 95,58$ [Euro]

c) Interpretation von b :
Bei einer Steigerung des Nettoeinkommens um 1000 Euro steigt die Telefonrechnung um $b \cdot 100$ Euro.

$$
\begin{aligned}
\widehat{y}_{x+0,2} &= a + b \cdot (x_{alt} + 0,2) \\
&= [a + bx_{alt}] + \underbrace{b \cdot 0,2}_{=0,2 \cdot 0,7357 = 0,1471 \quad [100 Euro]}
\end{aligned}
$$

Die durchschnittlichen Ausgaben steigen um $b \cdot 0,2 = 0.1471$ [100 Euro] = 14,71 Euro

d) Es ist

$$b = r_{XY} \cdot \frac{s_Y}{s_X} \Leftrightarrow r_{XY} = b \cdot \frac{s_X}{s_Y} = 0.73572 \cdot \frac{\sqrt{0.5464}}{\sqrt{0.308}} = 0.9799$$

$R^2 = r_{XY}^2 = 0.96026$
D.h., ca. 96% der Gesamtvarianz wird durch die lineare Regression erklärt.

Aufgabe 10.6

Der Output Y einer Volkswirtschaft sei in Abhängigkeit der eingesetzten Produktionsfaktoren Arbeit L und Kapital K betrachtet. Im Folgenden wird unterstellt, dass diese Abhängigkeit durch eine Cobb-Douglas-Produktionsfunktion der Form $Y = \alpha K^{b_1} L^{b_2}$ beschrieben werden kann. Für die Jahre 2005 bis 2010 liegen folgende Beobachtungen des Outputs und der Produktionsfaktoren vor:

i	1	2	3	4	5	6
y_i (Mrd Euro)	2270	2300	2400	2470	2500	2570
k_i (Mrd Euro)	16000	17000	19000	22000	22000	24000
l_i (1000 Besch.)	22000	23000	24000	24000	23500	25000

a) Wie muss die Produktionsfunktion transformiert werden, damit die partiellen Elastizitäten b_1 und b_2 mittels linearer Regression geschätzt werden können?

b) Schätzen und interpretieren Sie die Koeffizienten α, b_1 und b_2.

c) Berechnen Sie das Bestimmtheitsmaß der Regression.

Lösung:

a) Da die Produktionsfunktion nicht linear in den Parametern b_1 und b_2 ist, kann sie nicht mittels linearer Regression geschätzt werden. Logarithmieren der Produktionsfunktion liefert jedoch $\ln(Y) = \ln \alpha + b_1 \ln K + b_2 \ln L$. Daraus folgt der Regressionsansatz

$$\ln y_i = a + b_1 \ln k_i + b_2 \ln l_i + u_i .$$

b) Für den Fall zweier Regressoren können die Formeln für die Schätzung der Koeffizienten explizit angegeben werden:

$$\begin{aligned} b_1 &= \frac{s^2_{X_2} s_{YX_1} - s_{X_1X_2} s_{YX_2}}{s^2_{X_1} s^2_{X_2} - (s_{X_1X_2})^2} , \\ b_2 &= \frac{s^2_{X_1} s_{YX_2} - s_{X_1X_2} s_{YX_1}}{s^2_{X_1} s^2_{X_2} - (s_{X_1X_2})^2} , \\ a &= \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 . \end{aligned}$$

Die logarithmierten Werte sind der folgenden Tabelle zu entnehmen.

i	1	2	3	4	5	6
$\ln y_i$ (Mrd Euro)	7.7275	7.7407	7.7832	7.8112	7.8240	7.8517
$\ln k_i$ (Mrd Euro)	9.6803	9.7410	9.8522	9.9988	9.9988	10.0858
$\ln l_i$ (1000 Besch.)	9.9988	10.0432	10.0858	10.0858	10.0648	10.1266

Mit

$$
\begin{aligned}
\overline{\ln k} &= \frac{1}{n}\sum_{i=1}^{n} \ln k_i = \frac{1}{6}59.3569 = 9.8928\,,\\
\overline{\ln l} &= \frac{1}{n}\sum_{i=1}^{n} \ln l_i = \frac{1}{6}60.4050 = 10.0675\,,\\
\overline{\ln y} &= \frac{1}{n}\sum_{i=1}^{n} \ln y_i = \frac{1}{6}46.7383 = 7.7897\,,\\
s^2_{\ln k} &= \frac{1}{n}\sum_{i=1}^{n} \ln k_i{}^2 - \overline{\ln k}^2 = 97.8894 - 9.8928^2 = 0.02191\,,\\
s^2_{\ln l} &= \frac{1}{n}\sum_{i=1}^{n} \ln l_i{}^2 - \overline{\ln l}^2 = 101.3563 - 10.0675^2 = 0.001544\,,\\
s_{\ln y \ln k} &= \frac{1}{n}\sum_{i=1}^{n} \ln y_i \ln k_i - \overline{\ln y}\,\overline{\ln k} = 77.0687 - 7.7897 \cdot 9.8928 = 0.006756\,,\\
s_{\ln y \ln l} &= \frac{1}{n}\sum_{i=1}^{n} \ln y_i \ln l_i - \overline{\ln y}\,\overline{\ln l} = 78.4245 - 7.7897 \cdot 10.0675 = 0.001695\,,\\
s_{\ln k \ln l} &= \frac{1}{n}\sum_{i=1}^{n} \ln k_i \ln l_i - \overline{\ln k}\,\overline{\ln l} = 99.6011 - 9.8928 \cdot 10.0675 = 0.005236\,,
\end{aligned}
$$

folgt

$$
\begin{aligned}
b_1 &= 0.2427\,,\\
b_2 &= 0.2749\,,\\
\alpha &= \exp a = \exp 2.6216 = 13.7577\,.
\end{aligned}
$$

Das bedeutet, dass die partielle Kapitalelastizität der Produktion gleich 0.2427 und die partielle Arbeitselastizität gleich 0.2749 ist. Ein Anstieg des Kapitalstocks um 1% würde also einen Anstieg des Outputs um 0.2427% zur Folge haben. (Ohne Rundungen: $b = (0.2844, 0.0634, 76.5051)^T$.)

c) Die Regressionsfehler können nach der Formel $u_i = \ln y_i - \ln \hat{y}_i = \ln y_i - (a + b_1 \ln x_{i1} + b_2 \ln x_{i2})$ berechnet werden.

i	1	2	3	4	5	6
$\ln y_i$	7.7275	7.7407	7.7832	7.8112	7.8240	7.8517
$\ln \hat{y}_i$	7.7192	7.7462	7.7849	7.8205	7.8147	7.8528
u_i	0.0083	-0.0055	-0.0017	-0.0093	0.0093	-0.0011

Damit folgt

$$R^2 = 1 - \frac{s_U^2}{s_Y^2} = 1 - \frac{0.00004604}{0.001966} = 0.9766\,.$$

Aufgabe 10.7

Aufgrund einer Erhebung für zwei metrische Merkmale bei $n = 8$ Einheiten einer Grundgesamtheit sind folgende Werte bekannt:

$$\sum_{i=1}^{n} x_i y_i = 297\,, \quad \sum_{i=1}^{n} x_i = 36\,, \quad \sum_{i=1}^{n} y_i = 57\,, \quad \sum_{i=1}^{n} {x_i}^2 = 188\,, \quad \sum_{i=1}^{n} {y_i}^2 = 517\,.$$

Berechnen Sie

a) eine Maßzahl für die Stärke des linearen Zusammenhangs zwischen X und Y,

b) die Koeffizienten der beiden Regressionsgeraden $y = a + bx$ und $x = c + dy$.

Lösung:

Maßzahl:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y}$$

Man berechnet:

$$\begin{aligned}
s_X^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{8}188 - \left(\frac{1}{8}36\right)^2 = 23.5 - 20.25 \\
&= 3.25 \\
s_Y^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2 = \frac{1}{8}517 - \left(\frac{1}{8}57\right)^2 = 64.625 - 50.765 \\
&= 13.8594 \\
s_{XY} &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{8}297 - \frac{1}{8}36\frac{1}{8}57 = 37.125 - 32.0625 \\
&= 5.0625
\end{aligned}$$

a)

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{5.0625}{\sqrt{3.25}\sqrt{13.8594}} = 0.7543$$

b)

$$
\begin{aligned}
b &= \frac{s_{XY}}{s_X^2} = 1.5577, \qquad a = \bar{y} - b\bar{x} = 0.1154 \\
y &= 0,1154 + 1,5577x
\end{aligned}
$$

$$
\begin{aligned}
d &= \frac{s_{XY}}{s_Y^2} = 0.3653, \qquad c = \bar{x} - d\bar{y} = 1.8974 \\
x &= 1,8974 + 0,3653y \\
(y &= -5,1941 + 2,7375x)
\end{aligned}
$$

Letztere Gleichung ist übrigens nicht die Umkehrfunktion der ersten Regressionsgerade. Beide Regressionsgeraden verlaufen durch den Punkt $(\overline{x}/\overline{y}) = (4,5$ / $7,125)$.

Aufgabe 10.8

Berechnen Sie zu der folgenden Kontingenztabelle die Regressionsgerade $y = a + bx$.

X	Y			
	$\eta_1 = 1$	$\eta_2 = 2$	$\eta_3 = 3$	$\eta_4 = 4$
$\xi_1 = 2$	2	5		
$\xi_2 = 4$	4	8	2	
$\xi_3 = 6$	3	3	7	
$\xi_4 = 8$		3	2	5
$\xi_5 = 10$			3	3

Lösung:

$\xi_j \backslash \eta_k$	1	2	3	4	$n_{j\cdot}$	ξ_j^2
2	2	5			7	4
4	4	8	2		14	16
6	3	3	7		13	36
8		3	2	5	10	64
10			3	3	6	100
$n_{\cdot k}$	9	19	14	8	50	
η_k^2	1	4	9	16		

$$\bar{x} = 5.7600, \qquad s_X^2 = 6.0224$$
$$\bar{y} = 2.4200, \qquad s_Y^2 = 0.9236$$

$$s_{XY} = 1.5408$$
$$b = 0.2558 \qquad a = 0.9463$$

Aufgabe 10.9

Die Weltbevölkerung betrug in den Jahren 1960 bis 2000:

Jahr	Bevölkerung (in Mrd.)
1960	3.02
1970	3.70
1980	4.45
1990	5.30
2000	6.06

a) Stellen Sie das Wachstum der Weltbevölkerung graphisch dar.

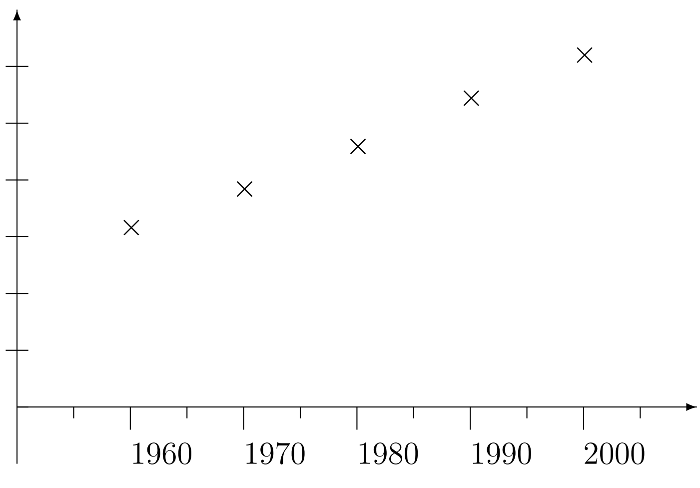
b) Berechnen Sie den linearen Trend sowie den exponentiellen Trend.

c) Schätzen Sie mit Hilfe der beiden Ansätze unter b) jeweils den Wert für 2006 und vergleichen ihn mit dem tatsächlichen Wert von 6.55 Milliarden.

d) Interpretieren Sie jeweils den Koeffizienten b.

Lösung:

a) Bildchen:



b) Linearer Trend:
Arbeitstabelle:

i	$t_i - 1980$	y_i	t_i^2	$t_i \cdot y_i$
1	-20	3,02	400	-60,4
2	-10	3,70	100	-37,0
3	0	4,45	0	0,0
4	10	5,30	100	53,0
5	20	6,06	400	121,2
$\sum$	0	22,53	1000	76,8
$\oslash$	0	4,5060	—	—

Man hat:

$$g_i = a + b \cdot (t_i - 1980)$$

$$b = \frac{76,8 - 5 \cdot 0 \cdot 4,506}{1000 - 5 \cdot 0^2} = \frac{76,8}{1000} = 0,0768$$

$$a = 4,5060 - 0,0768 \cdot 0 = 4,5060$$

$$g_i = 4,5060 + 0,0768 \cdot (t_i - 1980)$$

Exponentieller Trend:
Arbeitstabelle:

i	$t_i - 1980$	$\log y_i$	t_i^2	$t_i \cdot \log y_i$
1	-20	0,4800	400	-9,6000
2	-10	0,5682	100	-5,6820
3	0	0,6484	0	0,0000
4	10	0,7243	100	7,2430
5	20	0,7825	400	15,6500
$\sum$	0	3,2034	1000	7,6110
$\oslash$	0	0,6407	—	—

Man hat:

$$
\begin{aligned}
g_i &= a \cdot b^{(t_i - 1980)} \\
\\
B &= \frac{7,6110 - 0 \cdot 3,2033}{1000 - 5 \cdot 0} = \frac{7,6110}{1000} = 0,007611 \\
b &= 1,0177 \\
A &= 0,6407 - 0,007611 \cdot 0 = 0,6407 \\
a &= 4,3722 \\
\\
g_i &= 4,3722 \cdot 1,0177^{(t_i - 1980)}
\end{aligned}
$$

c) Mit der linearen Trendgleichung ergibt sich ein Schätzwert von 6,5028 Mrd. für 2006, mit der exponentiellen Trendgleichung ein Schätzwert von 6,8994 Mrd., denn

$$
\begin{aligned}
g_{2006} &= 4,5060 + 0,0768 \cdot (2006 - 1980) = 6,5028 \qquad \text{(linear)} \\
g_{2006} &= 4,3722 \cdot 1,0177^{(2006-1980)} = 6,8994 \quad \text{(exponentiell)}
\end{aligned}
$$

d) Linearer Trend:
Pro Jahr wächst die Weltbevölkerung um ca. 76 Mio.
Exponentieller Trend: Pro Jahr wächst die Weltbevölkerung um ca. $1,8\%$

Aufgabe 10.10

Der schweizerische "Landesindex der Konsumentenpreise" (Basis Mai 2000 = 100) betrug in den Jahren 1990 bis 2005:

Jahr	1990	1991	1992	1993	1994	1995	1996	1997
Index	82.8	87.6	91.2	94.2	95.0	96.7	97.5	98.0

Jahr	1998	1999	2000	2001	2002	2003	2004	2005
Index	98.0	98.8	100.3	101.3	102.0	102.6	103.4	104.7

a) Berechnen Sie die durchschnittliche Wachstumsrate dieses Indexes für 1990 bis 2005 sowie 2000 bis 2005.

b) Bestimmen Sie den linearen und den exponentiellen Trend. Vergleichen Sie das jeweilige Ergebnis mit a).

Lösung:

a)

$$\bar{w} = \sqrt[T]{\frac{x_T}{x_0}} - 1$$

$$\begin{aligned}
\bar{w}_{1990,2005} &= \sqrt[15]{\frac{104.7}{82.8}} - 1 = 0.015768 = 1.5768\% \\
\bar{w}_{2000,2005} &= \sqrt[5]{\frac{104.7}{100.3}} - 1 = 0.008624 = 0.8624\%
\end{aligned}$$

b) Gesucht ist eine lineare Trendgleichung der Form

$$y_i = a + b \cdot t_i$$

mit

$$b = \frac{\sum_{i=1}^{n} t_i\, y_i - n\, \bar{y}\, \bar{t}}{\sum_{i=1}^{n} t_i^2 - n\, \bar{t}^2}$$

und

$$a = \bar{y} - b\, \bar{t}.$$

Arbeitstabelle:

i	Jahr	t_i	y_i	$t_i\, y_i$	t_i^2
1	1990	0	82.8	0.0	0
2	1991	1	87.6	87.6	1
3	1992	2	91.2	182.4	4
4	1993	3	94.2	282.6	9
5	1994	4	95.0	380.0	16
6	1995	5	96.7	483.5	25
7	1996	6	97.5	585.0	36
8	1997	7	98.0	686.0	49
9	1998	8	98.0	784.0	64
10	1999	9	98.8	889.2	81
11	2000	10	100.3	1003.0	100
12	2001	11	101.3	1114.3	121
13	2002	12	102.0	1224.0	144
14	2003	13	102.6	1333.8	169
15	2004	14	103.4	1447.6	196
16	2005	15	104.7	1570.5	225
$\sum$	-	120	1554.1	12053.5	1240

Berechnung:

$$\begin{aligned}
\bar{t} &= \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{120}{16} = 7.5 \\
\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1554.1}{16} = 97.1313
\end{aligned}$$

$$\begin{aligned}
b &= \frac{12053.5 - 16 \cdot 7.5 \cdot 97.1313}{1240 - 16 \cdot 7.5^2} = 1.1698 \\
a &= 97.1313 - 1.1698 \cdot 7.5 = 88.3578
\end{aligned}$$

Daraus ergibt sich die lineare Trendgleichung $g(t) = 88.3578 + 1.1698 \cdot (t - 1990)$.

Zur Berechnung der durchschnittlichen Zuwachsraten berechnet man $g(0)$, $g(10)$ und $g(15)$

$$\begin{aligned}
g(0) &= 88.3578 \\
g(10) &= 100.0558 \\
g(15) &= 105.9048
\end{aligned}$$

Daraus ergeben sich

$$\begin{aligned}
\bar{w}_{1990,2005} &= \sqrt[15]{\frac{105.9048}{88.3578}} - 1 = 0.012150 = 1.2150\% \\
\bar{w}_{2000,2005} &= \sqrt[5]{\frac{105.9048}{100.0558}} - 1 = 0.011427 = 1.1427\%
\end{aligned}$$

Die exponentielle Trendgleichung lautet $g = 88.4149 \cdot 1.012369^{(t-1990)}$. Hieraus ergibt sich immer eine durchschnittliche Zuwachsrate von 1.2369%.