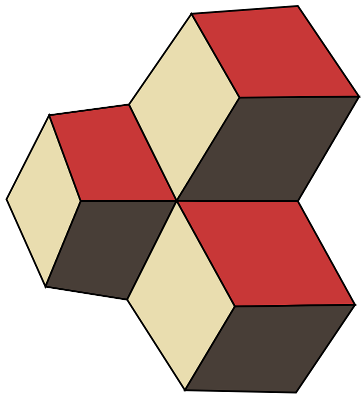
WST



Dr. D. Adams
Institut für Mathematik und Naturwissenschaften (IMN)
donat.adams@fhnw.ch
Büro: 5.1C01

Windisch, 14. Dezember 2021

Inhaltsverzeichnis

KAPITEL 1

Beschreibende Statistik

Lernziele 1.1 Beschreibende Statistik

- Die Studierenden kennen die Inhalte der Stochastik und der Statistik
- Sie können ein statistisches Experiment planen.
- Sie können Lageparameter einer Stichprobe: Mittelwert, Standard-Abweichung, Median und Quartile.
- Sie können diskrete, metrische und qualtative Grössen unterscheiden.
- Sie können Datensätze darstellen als Häufigkeitstabellen, Histogramme, Kreisdiagramme oder Boxplot.
- Sie kennen verschiedene kennen Möglichkeiten um in einer graphischen Darstellung Besonderheiten hervorzuhaben.

1.1 Stochastik vs. Statistik

Definition 1.1 Statistik

Analyse von Daten, die durch Zufall beeinflusst sind

Definition 1.2 Stochastik

Beschreibung und Untersuchung von Ereignissen, die vom Zufall beeinflusst werden.
Wahrscheinlichkeitstheorie + Statistik

Infobox 1.1 Einsatzgebiete Stochastik

- Technik, Physik
- Meteorologie
- Ökonomie

Infobox 1.2 Arbeitsweise Statistik

A) Formulierung Problem

B) Planung Experiment

C) Ausführung Experiment

D) Beschreibung experimentelle Daten

E) Schluss von Stichprobe auf Grundgesamtheit

Definition 1.3 Mittelwert $\overline{x}$, Standard-Abweichung s

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

Stichprobenumfang: n

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

Definition 1.4 Grundgesamtheit, Stichprobe

- Grundgesamtheit: Menge der Elemente, die untersucht werden soll (μ, σ)
- Stichprobe: Teilmenge der Grundgesamtheit, die untersucht wird ($\overline{x}$, s)

Beispiel 1.1 Beispiel Neonröhren **5R3XXH**

Wie gross ist die Lebensdauer der Neonröhren, die an FHNW verwendet werden?

Lösung:

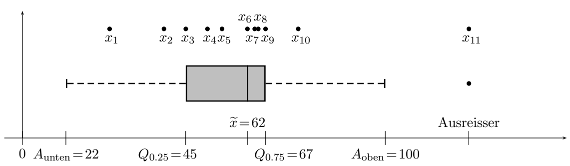
A) Formulierung Problem:
Wie gross ist Lebensdauer der Neonröhren, die an FHNW verwendet werden?

B) Planung Experiment:
Test einer Röhre genügt nicht. Alle können nicht getestet werden. Wir testen 11 Röhren.

C) Ausführung Experiment: siehe oben

D) Beschreibung experimentelle Daten:
Lageparameter: Durchschnitt und Standardabweichung $\overline{x} = 61 \pm 25.25$

E) Schluss von Stichprobe auf Grundgesamtheit:
Durchschnittliche Lebensdauer 61 Monate
Daten können verbessert werden, wenn wir Stichprobe vergrössern.



Infobox 1.3 Quantitative Grössen, qualitative Grössen

Quantitative Merkmale unterteilt in **diskrete** und **metrische**

- **diskrete** Merkmale werden meistens von mehreren Merkmalsträgern angenommen
- bei **diskreten** Merkmalen ist es sinnvoll zu zählen wie of eine Merkmals-Ausprägung angenommen wird
- bei **metrischen** Grössen gibt es zu fast jedem Merkmalsträger eine, von anderen verschiedene Merkmal-Ausprägung (sogar bei grossen Stichproben)

Für **qualitative** Grössen ist die Ordnung meist willkürlich: rot>grün; Zürich< Zug < Aarau

Beispiel 1.2 Metrische (stetige) Grössen (m), diskrete Grössen (d), qualitative Merkmale **H3ITET**

Ordnen sie zu:

a) Windgeschwindigkeit auf dem Arbeitsweg

b) Sonnenschein-Dauer am letzten Tag im Monat

c) Anzahl Regentage im April

d) Luftdruck während 24 Stunden

e) Stau-Stunden am Gotthard

f) Anzahl Lastwagen durch Belchentunnnel pro Stunde

g) Zivilstand der Studierenden in Klasse

h) steuerbares Einkommen der Studierenden in Klasse

i) abgeschlossene Diplome der Studierenden in Klasse

Lösung:

a) Windgeschwindigkeit (m)

b) Sonnenschein-Dauer am letzten Tag im Monat (m)

c) Anzahl Regentage im April (d)

d) Luftdruck (m)

e) Stau-Stunden am Gotthard (m)

f) Anzahl Lastwagen durch Belchentunnnel (d/m)

g) Zivilstand (q)

h) steuerbares Einkommen (m)

i) abgeschlossene Diplome (q)

Infobox 1.4 Graphische Darstellung

- Häufigkeitstabellen (Tabelle, m/d)
- Histogramme (Plot, m/d)
- Kreisdiagramme (q)

Infobox 1.5 Häufigkeitstabellen

- Anzahl Klassen (Richtwert) $k \approx \sqrt{n}$
- Klassenbreite $d \approx \frac{x_{\max} - x_{\min}}{k}$
- Intervalle $[a_i, a_{i+1}[$ (Daten auf Grenzen konsistent zu rechten Klasse gezählt)

Beispiel 1.3 Häufigkeitsverteilungen **C5EMXI**

i	0	1	2	3
h_i	48	38	10	4

Visualisieren Sie die Anzahl Betriebsstörungen an Baumaschinen. **Lösung:**

Beispiel 1.4 Zugfestigkeit Walzdraht **SLZQZI**

a) Erstellen Sie eine Häufigkeitstabelle und ein Histogramm zu den Daten der Zugfestigkeit (Skirpt Steiner, S. 6)

Lösung:

1.2 Lageparameter

Definition 1.5 Median $\tilde{x}$, Quartile $Q_{0.25}$ $Q_{0.75}$, Ausreissergrenzen A_{unten}

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & n \text{ ungerade} \\ 1/2 \cdot \left[x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right] & n \text{ gerade} \end{cases}$$

- Oberhalb von $Q_{0.75}$, liegt 1/4 der Messungen, unterhalb von $Q_{0.25}$ liegt 1/4 der Messungen
- Ausreissergrenzen $A_{\text{unten}} = Q_{0.25} - 1.5 \cdot d_Q$ und $A_{\text{oben}} = Q_{0.75} + 1.5 \cdot d_Q$
- Quartilsweite: $d_Q = Q_{0.75} - Q_{0.25}$

Infobox 1.6 Lageparameter (praktisch)

Liste der Länge n

- Ordne die Liste und berechne den Median
- Teile die Werte in zwei Listen auf
  - n ungerade: Der Median wird aus beiden Listen ausgeschlossen.
  - n gerade: Teile die Liste in exakt gleich lange Listen auf
- $Q_{0.25}$ ist der Median der ersten Liste, $Q_{0.75}$ ist der Median der zweiten Liste

Beispiel 1.5 Median/Quartile/Ausreisser **B41441**

$$\begin{pmatrix}5.9 & 4.5 & 5.4 & 5.1 & 3.5 & 2.9 & 2.3 & 4.7 & 2.7\end{pmatrix}$$

Berechnen Sie die Lageparameter
Lösung:
Ordnen:

$$\begin{pmatrix}2.3 & 2.7 & 2.9 & 3.5 & 4.5 & 4.7 & 5.1 & 5.4 & 5.9\end{pmatrix}$$

Mittelwert: $\overline{x} = 4.1$
Median: $\tilde{x} = 4.5$
Quartile: $Q_{0.25} = \frac{2.7+2.9}{2} = 2.8$ und $Q_{0.75} = \frac{5.1+5.4}{2} = 5.25$
Aussreissergrenzen: $A_{\text{unten}} = Q_{0.25} - 1.5 \cdot d_Q = -0.875$ und $A_{\text{oben}} = Q_{0.75} + 1.5 \cdot d_Q = 8.925$, d.h. es gibt keine Ausreisser

Beispiel 1.6 Median/Quartile/Ausreisser **UHC9IV**

5.3 3.8 4.0 19.5 5.0 4.9 2.2 4.1 3.1 5.5

Berechnen Sie die Lageparameter **Lösung:**

Ordnen:

2.2 3.1 3.8 4.0 4.1 4.9 5.0 5.3 5.5 19.5

Mittelwert: $\overline{x} = 5.74$
Median: $\tilde{x} = \frac{1}{2}(4.1 + 4.9) = 4.5$
Quartile: $Q_{0.25} = 3.8$ und $Q_{0.75} = 5.3$
Aussreissergrenzen: $A_{\text{unten}} = Q_{0.25} - 1.5 \cdot d_Q = 1.55$ und $A_{\text{oben}} = Q_{0.75} + 1.5 \cdot d_Q = 7.55$
$\Rightarrow$ 19.5 ist Ausreisser
Erstelle Box- und Whiskperplot
Ohne Ausreisser (Mittelwert/Median):
Mittelwert: $\overline{x} = 4.2$
Median: $\tilde{x} = \frac{1}{2}(4.1 + 4.9) = 4.1$

Infobox 1.7 Mittelwert vs. Median

Median ist stabiler gegenüber Ausreissern als Mittelwert

Definition 1.6 Alternative Lageparameter

- Arithmetisches Mittel
$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
- Harmonisches Mittel
$$\frac{1}{\overline{x}_\mathrm{h}} = \frac{1}{n}\cdot\sum_{i=1}^{n}\frac{1}{x_i}$$
- Geometrisches Mittel
$$\overline{x}_\mathrm{g} = (x_1 \cdot x_2 \, \ldots \, x_n)^{1/n} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

Beispiel 1.7 Mittlere Geschwindigkeit **V81PBS**

Strecke [km]	20	20	20
Geschwindigkeit [km/h]	40	120	80

Lösung:

$$\tilde{v} = \frac{\sum_i^n s_i}{\sum_i^n t_i} = \frac{\sum_i^n s_i}{\sum_i^n \frac{s_i}{v_i}}$$
$$\overline{v}_\mathrm{h} = 65.45 \text{ km/h}$$

Beispiel 1.8 Mittlere Verzinsung **W3LJRG**

Jahr	1	2	3
Zins %	8.5	12.2	-4.5

Wenn jedes Jahr der selbe Zins gelten würde, wie gross müsste dieser sein, damit am Schluss das selbe Kapital entsteht, wie bei der welchselnden Verzinsung? **Lösung:**

$$K = K_0 \cdot (1 + r_1) \cdot (1 + r_2) \cdot (1 + r_3)$$
$$\overline{v}_\mathrm{g} = 5.15 \text{ \%}$$

1.3 Das Summenzeichen

Das Summenzeichen $\sum$ erspart uns viel Schreibarbeit. Wir benutzen dazu das Σ (Sigma) aus dem griechischen Alphabet, weil der Buchstaben gleich klingt wie das ‘S’ im Wort Summe.

Definition 1.7 Summenzeichen, Bezeichnungen

$$\sum_{i=1}^{18} a_i$$

- a: Summationsvariable
- i: Laufindex. Sie nimmt hier die Werte von 1 bis 18 in aufsteigender Folge an
- 1 : untere Grenze
- 18 : obere Grenze

Die Wahl der Bezeichnung des Laufindex ist nicht von Bedeutung. Gewöhnlich werden hierfür i, k oder j verwendet.

Beispiel 1.9 Summenzeichen

Schreiben Sie die Summe mit dem Summenzeichen

a)
$$V = 1 + 2 + 3 + 4 + 5 + \ \ldots \ + 10 + 11 + 12$$

b)
$$W = 4^0 + 4^2 + 4^3 + \ \ldots \ + 4^{10} + 4^{11} + 4^{15}$$

c)
$$X = 7 + 7 + 7 + 7 + 7$$

d)
$$S = 2 + 4 + 6 + 8 + 10 + \ \ldots \ + 32 + 34 + 36$$

e)
$$P = 1 + 5 \ + 2 + 10 \ + 3 + 15 \ + \ \ldots \ + 8 + 40$$

Lösung:

a)
$$V = \sum_{i=1}^{12} i$$

b)
$$W = \sum_{i=0}^{15} 4^i$$

c) Es sind 5 Summanden. Es lohnt sich zu merken, dass die obere Grenze die Anzahl der Summanden angibt, falls die untere Grenze $i = 1$ ist.

$$X = \sum_{i=1}^{5} 7$$

d) Jeder Summand ist hier durch 2 teilbar. Wir schreiben also

$$S = 2 \cdot 1 + 2 \cdot 2 + 2 \cdot 3 + \ \ldots \ + 2 \cdot 18$$

Wir können also schreiben $S = \sum_{i=1}^{18} 2 \cdot i$ Da die 2 immer auftaucht, ist es eine Konstante in der Summe. Wir klammern sie aus

$$S = 2 \cdot (1 + 2 + 3 + \ \ldots \ + 18)$$

d.h. $S = 2 \cdot \sum_{i=1}^{18} i$

e) Wir gruppieren die Summanden. Wir sehen, dass in den Gruppen jeweils die zweiten Summanden Vielfache von 5 sind:

$$P = 1 + 5 \cdot 1 \ + 2 + 5 \cdot 2 \ + 3 + 5 \cdot 3 \ + \ \ldots \ + 8 + 5 \cdot 8$$

Wir können also schreiben

$$P = \sum_{i=1}^{8} (i + 5 \cdot i)$$

Wir können die Summer aber auch umordnen:

$$P = 1 + 2 + 3 \ldots 8 \quad + 5 \cdot 1 + 5 \cdot 2 + 5 \cdot 3 + \ \ldots \ + 5 \cdot 8$$

also

$$P = \sum_{i=1}^{8} i + \sum_{i=1}^{8} 5 \cdot i$$

Infobox 1.8 Gesetzte für das Summenzeichen

- Ausklammern (Multiplikation mit einer Konstanten):

$$\sum_{i=1}^{n} c \cdot a_i = c \cdot \sum_{i=1}^{n} a_i$$

- Umordnen (Zerlegen in Teilsummen):

$$\sum_{i=1}^{n} (a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$$

Achtung: Aus der Summe darf nur ausgeklammert werden, was nicht vom Laufindex abhängt.

Beispiel 1.10 Schreiben Sie die Summen aus **W315NQ**

a) $\sum_{i=1}^{5} i^3$

b) $\sum_{i=2}^{7} (2 - i)$

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

d) $\sum_{i=3}^{8} (10 - \frac{i}{2})$

e) $\sum_{i=2}^{6} (-i)^i$

f) $\sum_{i=0}^{5} \frac{i}{i+2}$

Lösung:

a) $1 + 8 + 27 + 64 + 125$

b) $0 - 1 - 2 - 3 - 4 - 5$

c) $1 + 4 + 9 + 16 + 25$

d)

$$\frac{17}{2} + 8 + \frac{15}{2} + 7 + \frac{13}{2} + 6$$

e) $4 - 27 + 256 - 3125 + 46656$

f)

$$0 + \frac{1}{3} + \frac{1}{2} + \frac{3}{5} + \frac{2}{3} + \frac{5}{7}$$

Beispiel 1.11 Ermitteln Sie die Grenzen **N3J50P**

a)

$$\sum \frac{1}{i} = \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ \ldots \ + \frac{1}{10}$$

b)

$$\sum \frac{2^k}{3} = \frac{1}{3} + \frac{2}{3} + \frac{4}{3} + \ \ldots \ + \frac{4\,096}{3}$$

c)

$$\sum (-1)^i \cdot i^2 = (-9) + 16 - 25 + \ \ldots \ + 324$$

d)

$$\sum (4j^2 - j) = 14 + 33 + 60 + 96 + \ \ldots \ + 1\,580$$

Lösung:

a)

$$\sum_{i=3}^{10} \frac{1}{i}$$

b)

$$\sum_{k=0}^{12} \frac{2^k}{3}$$

c)

$$\sum_{i=3}^{18} (-1)^i \cdot i^2$$

d)

$$\sum_{j=2}^{20} (4j^2 - j)$$

Satz 1.1 Summen Zusammenstellung

- Summe der Einsen
$$\sum_{i=1}^{n} 1 = n$$
- Summe der Indices
$$\sum_{i=1}^{n} i = \frac{n \cdot (n+1)}{2}$$
- Summe von Quadraten
$$\sum_{i=1}^{n} i^2 = \frac{n \cdot (n+1) \cdot (2n+1)}{6}$$
- Summe von Kuben
$$\sum_{i=1}^{n} i^3 = \frac{n^2 \cdot (n+1)^2}{2^2}$$
- Exponentialsumme
$$\sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r}$$

Falls die obere Grenze gegen ∞ geht, sprechen wir von Reihen. Summen der Form $\sum_{i=1}^{n} i$ sind dann algebraische Reihen und Summen der Form $\sum_{k=0}^{n} r^k$ sind dann geometrische Reihen. Ihre Konvergenz wird in der Analysis besprochen.

Beispiel 1.12 Rechenregeln für Summen **R4K83X**

a) $\sum_{i=1}^{15} 4i - 5$

b) $\sum_{k=0}^{8} \frac{3^k}{4}$

Lösung:

a)
$$\sum_{i=1}^{15} 4i - 5 \quad = \quad 4 \cdot \sum_{i=1}^{15} i - 5 \sum_{i=1}^{15} 1 = 4 \cdot \frac{15 \cdot 16}{2} - 5 \cdot 15 = 405$$

b)
$$\sum_{k=0}^{8} \frac{3^k}{4} \quad = \quad \frac{1}{4} \cdot \sum_{k=0}^{8} 3^k = \frac{1}{4} \cdot \frac{1 - 3^9}{1 - 3} = \frac{1}{4} \cdot 9841 = 2460,25$$

Beispiel 1.13 Rechenregeln für Summen **R7B3FX**

a) $\sum_{i=1}^{10} (3i)^2$

b) $\sum_{i=0}^{9} 2^{i+1}$

c) $\sum_{i=1}^{15} i^2 - 5$

d) $\sum_{i=1}^{11} (3i - 1) \cdot (3i + 1)$

Lösung:

a)

$$\sum_{i=1}^{10}(3i)^2 = 9\cdot\sum_{i=1}^{10} i^2 = 9\frac{10\cdot 11\cdot 21}{6} = 3465$$

b)

$$\sum_{i=0}^{9} 2^{i+1} = 2\sum_{i=0}^{9} 2^i = 2\cdot\frac{1-2^{10}}{1-2} = 2046$$

c)

$$\begin{aligned}\sum_{i=1}^{15} i^2 - 5 &= \sum_{i=1}^{15} i^2 - 5\sum_{i=1}^{15} 1\\ &= \frac{15\cdot 16\cdot 31}{6} - 5\cdot 15 = 1165\end{aligned}$$

d)

$$\begin{aligned}\sum_{i=1}^{11}(3i-1)\cdot(3i+1) &= \sum_{i=1}^{11} 9i^2 + 3i - 3i - 1\\ &= 9\sum_{i=1}^{11} i^2 - \sum_{i=1}^{11} 1\\ &= 9\cdot\frac{11\cdot 12\cdot 23}{6} - 11 = 4543\end{aligned}$$

Beispiel 1.14 Indexverschiebung **2ZXHEU**

$$\sum_{i=6}^{10}(i-1)$$

Lösung:

Um die Rechengesetzte anwenden zu können, führen wir eine Indexverschiebung durch. Das Ziel ist, dass die untere Grenze zu 1 wird. Also

$$i = 6 = \underbrace{k}_{=1} + 5$$

Es ergibt sich das ‘Übersetzungschema’

$$\left|\begin{array}{rrcl} & i & = & k+5\\ \text{oder auch:} & k & = & i-5\end{array}\right|$$

Wir ersetzen damit alle i

$$\sum_{k=1}^{10-5}\big((k+5)-1\big) = \sum_{k=1}^{5}\big(k+4\big) = 35$$

Beispiel 1.15 Indexverschiebung **VUMPOV**

Bestimmen Sie die Werte der Platzhalter $\square$

a) $\sum_{i=4}^{10}(i+4) = \sum_{k=1}^{\square}(\square)$

b) $\sum_{i=2}^{8}(3-i) = \sum_{k=1}^{\square}(\square)$

c) $\sum_{i=5}^{20}(i^2+2i) = \sum_{k=\square}^{17}(\square)$

d) $\sum_{i=6}^{10}(25-10i+i^2) = \sum_{k=1}^{\square}(\square)$

e) $\sum_{j=6}^{10} \frac{1}{4} \cdot 3^j = \sum_{k=0}^{\square}(\square)$

f) $\sum_{j=2}^{80} \left(1+\frac{1}{4}\right)^j = \sum_{k=\square}^{78}(\square)$

Lösung:

a) Wir finden $4 = i = \underbrace{k}_{=1} + 3$ also $i = k+3$ und $k = i-3$

$$\sum_{i=4}^{10}(i+4) = \sum_{k=1}^{7}(i+1)$$

b) Wir finden $2 = i = \underbrace{k}_{=1} + 1$ also $i = k+1$ und $k = i-1$

$$\sum_{i=2}^{8}(3-i) = \sum_{k=1}^{7}(3-(k+1)) = \sum_{k=1}^{7}(2-k)$$

c) Wir finden $20 = i = \underbrace{k}_{=17} + 3$ also $i = k+3$ und $k = i-3$

$$\sum_{i=5}^{20}(i^2+2i) = \sum_{k=2}^{17}((k+3)^2+2(k+3)) \sum_{k=2}^{17}(k^2+8k+15)$$

d) Wir finden $6 = i = \underbrace{k}_{=1} + 5$ also $i = k+5$ und $k = i-5$

$$\sum_{i=6}^{10}(25-10i+i^2) = \sum_{k=1}^{5}\left(25-10\cdot(k+5)+(k+5)^2\right) = \sum_{k=1}^{5} k^2$$

e) Wir finden $6 = j = \underbrace{k}_{=0} + 6$ also $j = k+6$ und $k = j-6$

$$\sum_{j=6}^{10} \frac{1}{4} \cdot 3^j = \sum_{k=0}^{4}(\frac{1}{4}3^{k+6}) = \sum_{k=0}^{4}(\frac{729}{4}3^k)$$

f) Wir finden $80 = j = \underbrace{k}_{=78} + 2$ also $j = k+2$ und $k = j-2$

$$\sum_{j=2}^{80}\left(1+\frac{1}{4}\right)^j = \sum_{k=0}^{78}(1+\frac{1}{4})^{k+2}$$

Beispiel 1.16 Rechenregeln + Indexverschiebung für Summen **2W2RMM**

Benutzen Sie die Regeln unten um die Summen zu berechnen.

a) $\sum_{i=0}^{15} 5+3i$

b) $\sum_{i=0}^{9} 50-5i$

c) $\sum_{i=5}^{10} \frac{i\cdot(i+1)}{2}$

d) $\sum_{i=0}^{12} 1+(i+3)^2$

e) $\sum_{j=1}^{1} 83^{j-10}$

f) $\sum_{i=2}^{8}(-5)^i$

Lösung:

a)

$$\begin{aligned}\sum_{i=0}^{15} 5+3i &= \sum_{i=1}^{16} 5+3\cdot(i-1)\\ &= \sum_{i=1}^{16} 2+3\sum_{i=1}^{16} i\\ &= 2\cdot 16+3\cdot\frac{16\cdot 17}{2}=440\end{aligned}$$

b)

$$\begin{aligned}\sum_{i=0}^{9} 50-5i &= \underbrace{50}_{i=0}+50\sum_{i=1}^{9} 1-\underbrace{5\cdot 0}_{i=0}-5\cdot\sum_{i=1}^{9} i\\ &= 50+50\cdot 9-5\cdot\frac{9\cdot 10}{2}=275\end{aligned}$$

c)

$$\begin{aligned}\sum_{i=5}^{10}\frac{i\cdot(i+1)}{2} &= \frac{1}{2}\cdot\left\{\sum_{i=5}^{10} i^2+i\right\}\\ &= \frac{1}{2}\cdot\left\{\sum_{i=1}^{10} i^2-\sum_{i=1}^{4} i^2+\sum_{i=1}^{10} i-\sum_{i=1}^{4} i\right\}\\ &= \frac{1}{2}\cdot\left\{\frac{10\cdot 11\cdot 21}{6}-\frac{4\cdot 5\cdot 9}{6}+\frac{10\cdot 11}{2}-\frac{4\cdot 5}{2}\right\}=200\end{aligned}$$

d)

$$\begin{aligned}\sum_{i=0}^{12} 1+(i+3)^2 &= \sum_{i=1}^{13} 1+(i+2)^2=\sum_{i=1}^{13} i^2+4i+5\\ &= \sum_{i=1}^{13} i^2+4\cdot\sum_{i=1}^{13} i+5\sum_{i=1}^{13} 1\\ &= \frac{13\cdot 14\cdot 27}{6}+4\cdot\frac{13\cdot 14}{2}+5\cdot 13=1248\end{aligned}$$

e)

$$\begin{aligned}\sum_{j=1}^{1} 83^{j-10} &= \sum_{j=0}^{1} 73^{j-9} = \frac{1}{3^9} \sum_{j=0}^{1} 73^{j} \\ &= \frac{1}{3^9} \cdot \frac{1-3^{17+1}}{1-3} = 9841.5\end{aligned}$$

f)

$$\begin{aligned}\sum_{i=2}^{8} (-5)^i &= \sum_{k=0}^{7} (-5)^{k+1} = (-5) \cdot \sum_{k=0}^{7} (-5)^k \\ &= (-5) \cdot \sum_{k=0}^{7} (-5)^k = (-5) \cdot \frac{1-(-5)^{7+1}}{1-(-5)} = 325\,520\end{aligned}$$

KAPITEL 3

Kombinatorik

Lernziele 3.1 Kombinatorik

- Die Studierenden Kennen die Grundexperimente der Kombinatorik
  - i) Kombination ohne Wiederholung
  - ii) Kombination mit Wiederholung
  - iii) Variation ohne Wiederholung
  - iv) Variation mit Wiederholung
- Sie können in Textaufgaben die Grundexperimente wiedererkennen und die Anzahl Möglichkeiten berechnen.
- Sie kennen die Fakultät $n!$ und den Binomialkoeffizienten $\binom{n}{k}$ und können algebraische Umformungen mit diesen Ausdrücken vornehmen.
- Mit Hilfe von Ereignisbäumen können die Studierenden die Anzahl Möglichkeiten in Fällen berechnen die keinem Grundexperiment entspricht.

Beispiel 3.1 Bingo **18X2JW**

Wir haben eine Schachtel mit den Karten, die die Zahlen 1-9 tragen. Daraus ziehen wir eine Karte, notieren die gezogene Zahl, legen die Karte zurück in die Schachten und ziehen erneut eine Karte. Auf wie viele Arten kann man eine Summe von 8 erhalten?
Lösung:

1. Karte	1	2	3	4	5	6	7
2. Karte	7	6	5	4	3	2	1

Beispiel 3.2 Bingo **GSGQ9F**

Aus der Schachtel mit den Karten, die die Zahlen 1-9 tragen, ziehen wir eine Karte und dann eine zweite Karte. Auf wie viele Arten kann man eine Summe von 8 erhalten?

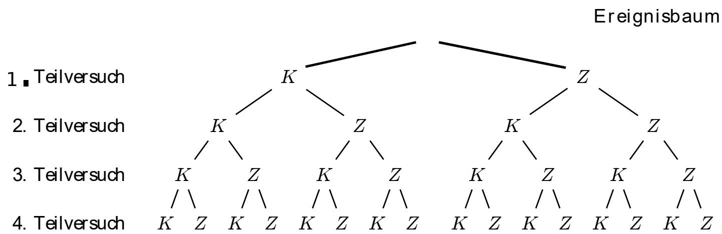
Lösung:

1. Karte	1	2	3	5	6	7
2. Karte	7	6	5	3	2	1

Die beiden Beispiele zeigen, was mir unter 'mit Zurücklegen' und 'ohne Zurücklegen' meinen: Beim ersten Versuch, kann die 4 zweimal verwendet werden, während das zweite Experiment, die Wiederverwendung von der 4 ausschließt.

Beispiel 3.3 Sandwich **W6P8CP**

Wir machen Sie ein Sandwich aus einer von drei Brotsorten und einer von 4 Käsesorten, mit oder ohne Gurken. Wie viele verschiedene Arten von Sandwich können gemacht werden?
Lösung:
Mit jeder Brotsorte können wir 4 Sandwiches belegen. So ergeben sich 12 mögliche Sandwiches. Diese sind ohne Gurke; wenn wir Sandwiches mit oder ohne Gurken wollen, dann haben wir 24 mögliche Sandwiches. Das heißt, es gibt $3 \times 4 \times 2 = 24$ mögliche Sandwiches.



Satz 3.1 Produktregel

Besteht ein zusammengesetzter Versuch aus m unabhängigen Teilversuchen mit jeweils $[n_1, n_2, n_3, n_4, \ldots, n_m$ Ausfallsmöglichkeiten, dann besitzt der Versuch

$$n_1 \cdot n_2 \ \ldots \ \cdot n_m \text{ Ausfälle.}$$

Beispiel 3.4 Dreistellige gerade Zahlen **5QDH2E**

Wie viele dreistellige gerade Zahlen gibt es? Zugelassen sind die Ziffern 0-9. Ausgeschlossen sind Zahlen die mit der Ziffer 0 beginnen.
Lösung:

Die erste Ziffer kann nicht Null sein, da die Zahl dreistellig sein muss - Es gibt also 9 Möglichkeiten, die Hunderterstelle zu belegen. Es gibt keine Bedingung dafür, was die Zehnerstelle sein soll, also haben wir dort 10 Möglichkeiten. Um eine gerade Zahl erhalten muss die Zahl mit 0, 2, 4, 6 oder 8 enden. Wir haben also

$$9 \cdot 10 \cdot 5 = 450$$

dreistellige gerade Zahlen.

Beispiel 3.5 Dreistellige Zahlen: Variation **HAT89D**

Wieviele Zahlen mit den Ziffern 1 bis 9 gibt es? Beispiele:

$$(1, 1, 1),\ (1, 2, 3),\ (9, 1, 1)$$

a) Geben Sie zwei weitere Zahlen an.

b) Erfinden sie ein Verfahren an, das mit Sicherheit alle Variationen erzeugt.

c) Zählen Sie nun alle Variationen.

Lösung:

a) Geben Sie zwei weitere Zahlen an.

$$(2, 2, 2),\ (3, 2, 1),\ (1, 1, 9)$$

b) Wir gehen die Möglichkeiten systematisch durch, wie bei einem Zahlenschloss: Wir stellen die erste Ziffer auf 1, die zweite auf 1 und 'drehen am hintersten Rädchen' alle Möglichkeiten von 1 bis 9; Nun stellen wir die zweite Ziffer auf 2 und 'drehen am hintersten Rädchen' alle Möglichkeiten von 1 bis 9; etc.

c) Zählen Sie nun alle Variationen.

$$V(9; 3) = 9 \cdot 9 \cdot 9 = 9^3 = 729$$

Fachterm: Variation 3-ter Ordnung mit Zurücklegen.

Permutationen

Beispiel 3.6 Bücherregal **QXG629**

Sie haben 5 Bücher, die Sie in ein Regal stellen wollen: Mathematik (M), Physik (P), Englisch (E), Biologie (B) und Geschichte (G). Auf wie viele verschiedene Arten können Sie dies tun?
Gehen Sie systematisch vor: Nummerieren Sie die Positionen, auf die die Bücher platzieren werden können

$$\begin{array}{ccccc} \sqcup & \sqcup & \sqcup & \sqcup & \sqcup \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

Belegen Sie die Positionen in dieser Reihenfolge. Wieviele Möglichkeiten gibt es für die erste Position? Für die zweite? etc.
Lösung:
Wir besetzen zuerst die erste Position

$$\begin{array}{ccccc} \text{G} & & & & \\ \sqcup & \sqcup & \sqcup & \sqcup & \sqcup \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

dann die zweite

$$\begin{array}{ccccc} \text{G} & \text{P} & & & \\ \sqcup & \sqcup & \sqcup & \sqcup & \sqcup \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

die dritte

$$\begin{array}{ccccc} G & P & M & & \\ \sqcup & \sqcup & \sqcup & \sqcup & \sqcup \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

und die vierte

$$\begin{array}{ccccc} G & P & M & B & \\ \sqcup & \sqcup & \sqcup & \sqcup & \sqcup \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

und schliesslich die letzte

$$\begin{array}{ccccc} G & P & M & B & E \\ \sqcup & \sqcup & \sqcup & \sqcup & \sqcup \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

Wir sehen z.B. wenn bereits drei Bücher im Regal stehen, gibt es zwei Bücher für die vierte Position und nur eine Möglichkeit, das fünfte Buch zu platzieren. Also ist die Anzahl der Möglichkeiten alle 5 Bücher zu platzieren

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120 = 5!$$

Beispiel 3.7 Fakultät **DUD65E**

Wir werden immer wieder Ausdrücken begegen, wo nacheinanderfolgende ganze Zahlen miteinander multipliziert werden

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$$

a) Berechnen Sie $4!,\ 5!,\ 9!$

b) Benutzen Sie die Fakultät um folgende Produkte auszurechnen

   i) $9 \cdot 8 \cdot 7 \cdot 6 \cdot 5$
   ii) $11 \cdot 10 \cdot 9 \cdot 8$
   iii) $8 \cdot 7 \cdot 6 \cdot 5 \cdot 4$

c) Was könnte $0!$ bedeuten?

d) Wie geben Sie allgemein das folgende Produkt mit Faktultäten an

$$n \cdot (n-1) \cdot (n-2) \ldots (n-k)$$

wobei $k < n$ und $n, k \in \mathbb{N}$

Lösung:

a) Wir erhalten 24, 120 und 362880.

b) Benutzen Sie die Fakultät um folgende Produkte auszurechnen

   i) $\frac{9!}{4!} = 15120$
   ii) $\frac{11!}{7!} = 7920$
   iii) $\frac{8!}{3!} = 6720$

c) $0! = 1$, das ist so definiert

d) Wir schreiben

$$n \cdot (n-1) \cdot (n-2) \dots (n-k) = \frac{n!}{(n-k-1)!}$$

Beispiel 3.8 Permutationen **KUPTEG**

Wieviele Permutationen der Ziffern 1 bis 5 gibt es? Beispiele:

$$(1,2,3,4,5),\ (2,1,3,4,5),\ (5,4,3,2,1)$$

a) Geben Sie zwei weitere Permutationen an.

b) Erfinden Sie ein Verfahren an, das mit Sicherheit alle Permutationen erzeugt.

c) Zählen Sie nun alle Permutationen

Lösung:

a) Weitere Permutationen, z.B.

$$(2,1,4,3,5),\ (2,1,3,5,4)$$

b) Wir betrachten die leere Liste $(\ ,\ ,\ ,\ ,\ )$ und plazieren darin die Ziffern $1,2,3,4$ und 5. Zuerst setzen wir die 1 auf den ersten Platz, dann die 2 auf einen der leeren Plätze, dann ... . Nun setzen wir die 1 auf den zweiten Platz, dann die 2 auf einen der leeren Plätze, dann ... Dies gibt jedes Mal eine neue Permutation. Und wir finden so alle möglichen Permutationen.

c) Um die 1 zu plazieren haben wir 5 Möglichkeiten; um die 2 zu plazieren, haben wir 4 Möglichkeiten, etc. Möglichkeiten zählen:

$$P(5) = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 720$$

Beispiel 3.9 Kombination **F1VF9G**

Wieviele fünfstellige Kombinationen der Ziffern 1 bis 9 gibt es? Beispiele:

$$(9,8,3,7,5),\ (2,1,3,9,8),\ (9,8,7,6,5)$$

a) Geben Sie zwei weitere Kombinationen an.

b) Erfinden Sie ein Verfahren an, das mit Sicherheit alle Kombinationen erzeugt.

c) Zählen Sie nun alle Permutationen

Lösung:

a) Weitere Permutationen, z.B.

$$(1,2,3,4,5),\ (9,8,3,1,2)$$

b) Wir betrachten die Ziffern 1 bis 9. In der Liste benennen wir zunächst die Plätze: (A, B, C, D, E). Nun ordnen wir die Plätze den 9 Ziffern zu: Zuerst setzen wir die A auf die Ziffer 1, dann die B auf einen der leeren Plätze, dann … . Nun setzen wir die A auf die Ziffer 2, dann die B auf einen der leeren Plätze, dann … Dies gibt jedes Mal eine neue Kombinationen. Und wir finden so alle möglichen Permutationen.

c) Um die A zu plazieren haben wir 9 Möglichkeiten; um die B zu plazieren, haben wir 8 Möglichkeiten, etc. Möglichkeiten zählen:

$$9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = \frac{9!}{4!} = 15\,120$$

Beispiel 3.10 Anzahl mögliche vierstelligen Zahlen? **A8T7BK**

Wieviele vierstellige Zahlen mit den Ziffern 1 bis 9 gibt es? Jede Ziffer soll nur ein Mal verwenden. (Ziffern 1 bis 9: keine Probleme mit führenden 0)
Lösung:

$$V(9;4) = \frac{9!}{(9-4)!} = 7 \cdot 8 \cdot 9 \cdot 8 = 4032$$

Variation 3-ter Ordnung ohne Zurücklegen.

Beispiel 3.11 Pferdetoto **2KWVHJ**

Wir schauen uns ein Pferderennen an mit 10 Pferden. Wieviele Möglichkeiten gibt es für den Zieleinlauf der ersten 3 Pferde? **Lösung:**

$$V(10;3) = \frac{10!}{(10-3)!} = 720$$

Variation 3-ter Ordnung ohne Wiederholungen.

Kombination

Definition 3.1 Kombination

Eine **Kombination** ist eine Auswahl von Objekten. Die Auswahl besteht aus einer festen Anzahl von Objekten, die Reihenfolge der Objekte in der Zielmenge spielt keine Rolle.

Beispiel 3.12 Permutation vs. Kombination **BGFEH1**

ABC	ABD	ACB	ACD
ADB	ADC	BAC	BAD
BCA	BCD	BDA	BDC
CAB	CAD	CBA	CBD
CDA	CDB	DAB	DAC
DBA	DBC	DCA	DCB

Hier sind die 24 Permutationen der Länge 3 der Buchstaben ABCD. Wie viele

Kombinationen gibt es? Für die Kombination spielt die Reihenfolge keine Rolle, wir haben z.B.

$$\text{ABC} = \text{CBA}$$

Gehen Sie wie folgt vor:

a) Berechnen Sie zunächst die Anzahl Möglichkeiten aus den Symbolen ABCD 3 Symbole auszuwählen (mit Berücksichtigung der Reihenfolge bei der Wahl).

b) Wir haben nun 3 Symbole gewählt, z.B. ABC. Wieviele Möglichkeiten gibt es, diese anzuordnen?

c) Wie können Sie die Zahlen oben benutzen, um die Anzahl Kombinationen zu berechnen?

d) Identifizieren Sie in der Liste oben die äquivalenten Einträge und zählen Sie die möglichen Kombinationen.

Lösung:

a) Es gibt $4 \cdot 3 \cdot 2 = 24$ Möglichkeiten für die Wahl von Symbolen

b) Es gibt $3 \cdot 2 \cdot 1 = 6$ Möglichkeiten die 3 Symbole anzuordnen.

c) Wir Zahlen zunächst, als würde die Reihenfolge eine Rolle spielen, dann teilen wir durch die Anzahl der möglichen Anordnungen.

$$N = \frac{24}{6} = \frac{\frac{4!}{1!}}{3!} = \frac{4!}{1! \cdot 3!}$$

d) Viele Einträge sind gleich. Indem wir die Einträge alphabetisch ordnen, finden wir dass es nur 4 Kombinationen gibt

ABC ABD ACD BCD

Definition 3.2 Binomialkoeffizient

Für $n > k$ und $n, k \in \mathbb{N}$ ist

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Beispiel 3.13 Projektteam **AVYE9M**

Aus einer Klasse mit 12 Studierenden soll ein Projektteam mit 5 Mitgliedern zusammengestellt werden.

a) Im Team gibt es 5 Funktionen (A: ChefIn, B: UnterchefIn, C: Kassier, D: ProtokollführerIn, E: AngestellteR). Wieviele mögliche Teams gibt es.

b) Wir haben die 5 Personen ausgewählt. Wie viele Möglichkeiten gibt es die Funktionen diesen Personen zuzuordnen?

c) Wir haben ein Team ohne Funktionen (jedeR ist gleichgestellt). Wieviele mögliche Teams gibt es jetzt? Benutzen Sie die Zahlen von oben.

d) Wie unterscheidet sich die letzte Unter-Aufgabe von den vorigen?

e) Versuchen Sie Ihre Resultate zu Verallgemeinern: Wieviele Möglichkeiten gibt es, aus n Personen ein Projektteam zusammenzustellen mit k Mitgliedern, die keine definierte Funktion haben?

Lösung:

a) Anzahl $\frac{12!}{(12-5)!} = 95\,040$

b) Anzahl Möglichkeiten Funktionen zuzuordnen: $5! = 120$

c) Wir rechnen zunächst, als ob die Funktionen unterscheidbar wären und teilen durch die Anzahl Möglichkeiten, die dabei zu viel waren

$$N = \frac{95\,040}{120} = 792$$

also $N = frac12!(12-5)! \cdot 5!$

d) Die Reihenfolge im Team spielt keine Rolle.

e) Wähle k Personen aus einer Gruppe von n ohne Berücksichtigung der Reihenfolge: $n!/(n-k)!$ ist Anzahl der Möglichkeit mit beachtung der Reihenfolge und also

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

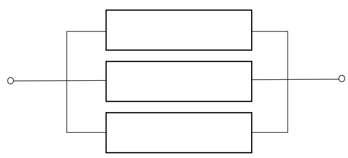
ist die Anzahl ohne Berücksichtigung der Reihenfolge.

Satz 3.2 Kombination ohne Zurücklegen

Wir wollen aus einer Menge mit n Elementen k auswählen. In der Stichprobe gibt es keine Reihenfolge. Dann gibt es N Möglichkeiten

$$N = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Beispiel 3.14 Anzahl mögliche Schaltungen **G3U4RB**



Wir wollen 5 verschiedene Widerstände $R_1, \ldots, R_5$ in die folgende Schaltung einbauen. Jeder Widerstand kann nur ein Mal verwenden werden. Wieviel Möglichkeiten gibt es für die Schaltung? **Lösung:**

Wir belegen die Plätze nacheinander mit einem Widerstand:

$$N' = 5 \cdot 4 \cdot 3$$

Viele der Konfigurationen, die so entstehen, sind gleich. Deshalb teilen wir durch die Anzahl Möglichkeiten drei Widerstände auf drei Plätze zu verteilen:

$$N = \frac{N}{3 \cdot 2 \cdot 1} = 10$$

Oder mit der Formel

$$C(5;3) = \binom{5}{3} = \frac{5!}{3! \cdot 2!} = 10$$

Kombination 3-ter Ordnung ohne Zurücklegen.

Beispiel 3.15 Anzahl mögliche ungeordneter Stichproben einer Lieferung von Batterien? **5AIURL**

Gelieferte Batterien (unterscheibar): 100, Stichprobe: 10; Position in der Stichprobe spielt keine Rolle.

Lösung:

$$C(100;10) = \binom{100}{10} = \frac{100!}{10! \cdot 90!} = 17\,310\,309\,456\,440$$

Kombination 10-ter Ordnung ohne Zurücklegen.

Beispiel 3.16 Äpfel verteilen **RC8SB8**

Wir wollen 8 Äpfel an 4 Kinder (Adam, Bert, Caroline und Dan). Die Äpfel betrachten wir als ununterscheibar. Ein Kind kann auch mehrere (alle) Äpfel erhalten. Wie viel Möglichkeiten gibt es die Äpfel zu verteilen?

a) Geben Sie zwei mögliche Verteilungen an, z.B. Adam 3, Bert 0, Caroline 5 und Dan 0

b) Geben Sie die Verteilungen graphisch an, z.B.

$$(\underbrace{000}_{A}|\underbrace{\quad}_{B}|\underbrace{00000}_{C}|\underbrace{\quad}_{D})$$

c) Wie können Sie mit der oben eingeführten Codierung die Anzahl Verteilungen zählen?

d) Wie viele mögliche Verteilungen gibt es?

e) Verallgemeinern Sie Ihre Resultate auf k ununterscheidbare Äpfel, die wir an n Kinder verteilen, wobei jedes Kind *mehrere Äpfel* erhalten kann.

Lösung:

a) Verteilung ii): Adam 1, Bert 2, Caroline 5, Dan 0 ; Verteilung iii): Adam 8, Bert 0, Caroline 0 und Dan 0

b) Verteilung ii):

$$(\underbrace{0}_{A}|\underbrace{00}_{B}|\underbrace{00000}_{C}|\underbrace{\quad}_{D})$$

Verteilung iii):

$$(\underbrace{00000000}_{A}|\underbrace{\qquad}_{B}|\underbrace{\qquad}_{C}|\underbrace{\qquad}_{D})$$

c) Die Codierung führt zu einem Strichcode, bestehend aus 0 und 1 (wenn wir die $|$ als 1 denken. Verteilung i):

$$(000||00000|) \;\rightarrow\; 00011000001$$

Verteilung ii):

$$(0|00|00000|) \;\rightarrow\; 01001000001$$

Verteilung iii):

$$(00000000|||) \;\rightarrow\; 00000000111$$

Er besteht aus 8 Nullen und 3 Seperatoren, also hat er $8+(4-1)=11$ Stellen.

d) Beim Abzählen gehen wir wie folgt vor: Wir stellen die Sepratoren nacheinander auf. Das gibt $11\cdot 10\cdot 9$ Möglichkeiten. Den Rest füllen wir mir mit Nullen: 1 Möglichkeit. Es gibt also

$$N = \frac{11\cdot 10\cdot 9}{3\cdot 2\cdot 1}\cdot 1 = 165$$

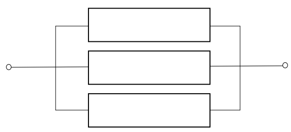
Möglichkeiten.

e) Verallgemeinerung:

$$N = \frac{(n+k-1)!}{(k)!\cdot(n-1)!} = \binom{n+k-1}{k}$$

dabei ist n die Anzahl unterscheidbare Urnen (hier die Kinder), und k die Anzahl der ununterscheibaren Objekte (hier die Äpfel).

Beispiel 3.17 Anzahl mögliche Schaltungen **P6DR9M**



5 verschiedene Widerstände $R_1,\ldots,R_5$
Jeder Widerstand beliebig oft verwendbar. **Lösung:**

Hier sind die Widerständ (1,2,3,4,5) unterscheidbar, die Plätze sind ununterscheidbar.
Eine Konfiguration entsteht, indem wir den 5 Widerständen die 3 Plätze zuordnen, z.B. $(000|||)$ bedeutet, dass wir alle drei Pätze mit dem Widerstand 1 besetzen.
Wir setzen $n=5$ und $k=3$, also

$$C_w(5;3) = \binom{5+3-1}{3} = \frac{7!}{4!\cdot 3!} = 35$$

Kombination 3-ter Ordnung mit Zurücklegen.

	ohne Wiederholung	mit Wiederholung	
Kombination k-ter Ordnung	$C(n;k) = \binom{n}{k}$	$C_w(n;k) = \binom{n+k-1}{k}$	ungeordnete Stichprobe
Variation k-ter Ordnung	$V(n;k) = \frac{n!}{(n-k)!}$	$V_w(n;k) = n^k$	geordnete Stichprobe
	Ziehung ohne Zurücklegen	Ziehung mit Zurücklegen	

- **Kombination ohne Wiederholung:** k ununterscheidbare Äpfel an n Kinder verteilen, wobei jedes Kind *nur einen Apfel* erhalten kann ($k < n$); k-er Gruppen aus n Personen bilden; aus einer Menge von n Objekte eine ungeordnete Stichprobe mit k Objekten machen

- **Kombination mit Wiederholung:** k ununterscheidbare Äpfel an n Kinder verteilen, wobei jedes Kind *mehrere Äpfel* erhalten kann

- **Variation ohne Wiederholung:** Aus n unterscheidbaren Büchern k aufstellen (jedes Buch kann nur einmal aufgestellt werden, $k < n$)

- **Variation mit Wiederholung:** Zahlenschloss, k Rädchen mit den Zahlen 1 bis n; Passwort der Länge k mit n möglichen Symbolen an jedem Platz.

KAPITEL 4

Wahrscheinlichkeit

Lernziele 4.1 Wahrscheinlichkeit

- Die Studierenden kennen die Begriffe Zufallsexperiment, Stichproblenraum und Ereignis
- Die Studierenden können Wahrscheinlichkeiten einfache Zufallsexperimente berechnen
- Die Studierenden können Wahrscheinlichkeiten für verknüpfte Zufallsexperimente berechnen

Definition 4.1 Zufallsexperiment und Stichprobenraum

Zufallsexperiment verstehen wir einen Vorgang,

- der gedanklich beliebig oft wiederholbar und
- dessen Ausgang innerhalb einer Menge möglicher Ergebnisse ungewiss (zufällig) ist.

Die Menge S der sämtlichen Ausfallsmöglichkeiten heisst **Stichprobenraum**.

[Steiner, 2015, S. 17]

Definition 4.2 Ereignis

Ereignis eine Teilmenge des Stichprobenraums S.
Alternativ: Ein Ereignis ist ein Resultat oder eine Menge von Resultaten eines Zufallsexperiments.

[Steiner, 2015, S. 18]

Beispiel 4.1 Stichprobenraum **MSUX6L**

Bestimmen Sie für folgende Zufallsexperimente den Stichprobenraum S und die Anzahl Elemente in $n(S)$

a) Wir werfen eine Münze zwei Mal (geordnete Stichproben).

b) Wir werfen eine Münze zwei Mal (ungeordnete Stichproben).

c) Wir werfen einen Würfel zwei Mal (geordnete Stichproben).

d) Wir werfen einen Würfel zwei Mal (ungeordnete Stichproben).

Lösung:

a) Wir werfen eine Münze zwei Mal (geordnete Stichproben).

$$S = \{KK, KZ, ZK, ZZ\}, \quad n(S) = 4$$

b) Wir werfen eine Münze zwei Mal (ungeordnete Stichproben).

$$S = \{KK, KZ, ZZ\}, \quad n(S) = 3$$

c) Wir werfen einen Würfel zwei Mal (geordnete Stichproben).

$$S = \begin{array}{cccccc} \{(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6)\} \end{array}$$

und $n(S) = 36$

d) Wir werfen einen Würfel zwei Mal (ungeordnete Stichproben).

$$S = \begin{array}{cccccc} \{(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ & & (3,3), & (3,4), & (3,5), & (3,6), \\ & & & (4,4), & (4,5), & (4,6), \\ & & & & (5,5), & (5,6), \\ & & & & & (6,6)\} \end{array}$$

und $n(S) = 21$

Beispiel 4.2 Stichprobenraum **1TIXNY**

Bestimmen Sie für folgende Zufallsexperimente den Stichprobenraum S und die Anzahl Elemente in $n(S)$

a) Wir werfen einen Würfel 5 Mal und zählen die Anzahl Sechser.

b) Wir werfen eine Münze 5 Mal und zählen die Anzahl Köpfe.

Lösung:

a) Wir werfen einen Würfel 5 Mal und zählen die Anzahl Sechser.

$$S = \{0, 1, 2, 3, 4, 5\}$$

Anzahl Elemente 6.

b) Wir werfen eine Münze 5 Mal und zählen die Anzahl Köpfe.

$$S = \{0, 1, 2, 3, 4, 5\}$$

Anzahl Elemente 6.

Beispiel 4.3 Zufällige Karte **2R1AN4**

In den französichen Karten gibt es die Farben Kreuz $[\clubsuit K]$, Pik $[\spadesuit K]$, Herz $[\heartsuit K]$ und Karo $[\diamondsuit K]$ (Trèfle, Pique, Cœur, Carreau) z.B. mit den Werten

$$[\spadesuit 2], [\spadesuit 3] \ldots [\spadesuit 10], [\spadesuit J], [\spadesuit Q], [\spadesuit K], [\spadesuit A]$$

Wir ziehen eine zufällige Karte. Geben Sie folgende Ereignisse explizit an:

a) Es wird ein König gezogen.

b) Es wird eine Herz-Karte gezogen.

c) Es wird ein Jack oder eine bessere Karte gezogen.

Lösung:

a) $K = \{[\heartsuit K][\spadesuit K][\clubsuit K][\diamondsuit K]\}$

b)

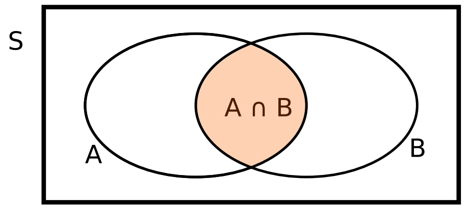
$$K = \{[\heartsuit 2], [\heartsuit 3] \ldots [\heartsuit 10], [\heartsuit J], [\heartsuit Q], [\heartsuit K], [\heartsuit A]\}$$

c)

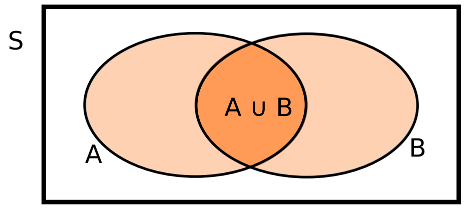
$$P = \{\begin{matrix} [\heartsuit J], & [\heartsuit Q], & [\heartsuit K], & [\heartsuit A], \\ [\spadesuit J], & [\spadesuit Q], & [\spadesuit K], & [\spadesuit A], \\ [\clubsuit J], & [\clubsuit Q], & [\clubsuit K], & [\clubsuit A], \\ [\diamondsuit J], & [\diamondsuit Q], & [\diamondsuit K], & [\diamondsuit A] \end{matrix}\}$$

Infobox 4.1 Verknüpfungen

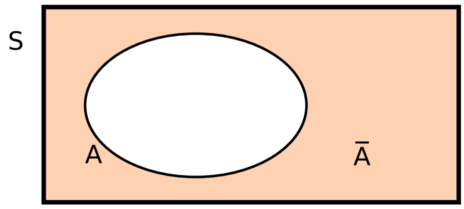
- $\cup$: Oder-Verknüpfung. $A \cup B = C$: C enthält alle Elemente die in A liegen und zusätzlich die, die in B liegen aber nicht in A.



- $\cap$: Und-Verküpfung. $A \cap B = C$: C enthält nur Elemente die in A und gleichzeitig die Bedingungen für B erfüllen.



- A' (oder auch $\overline{A}$): Komplement. A' enthält alle Elemente, die nicht in A liegen.



Satz 4.1 Morgansche Regeln

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$

Beispiel 4.4 Ereignis **W5BQ2P**

Wir würfeln mit einem Würfel (Punkte 1 bis 6). Beschreiben Sie folgende Ereignisse explizit:

a) A: Wir erhalten eine ungerade Zahl.

b) B: Wir erhalten Zahl kleiner als 5.

c) $A \cup B$

d) $A \cap B$

e) A'

f) B'

Lösung:

a) $A = \{1, 3, 5\}$

b) $B = \{1, 2, 3, 4\}$

c) $A \cup B = \{1, 2, 3, 4, 5\}$

d) $A \cap B = \{1, 3\}$

e) $A' = \{2, 4, 6\}$

f) $B' = \{5, 6\}$

Beispiel 4.5 Mengen **D7SI7B**

Erkläre mit Hilfe eines Venn-Diagrams folgende Gesetze der Mengenlehre

a) $A \cup B = B \cup A$

b) $(A')' = A$

c) $A \cap S = A$

d) $A \cup S = S$

e) $A \cap A' = \{\}$

f) $A \cup A' = S$

Lösung:

a) Beweis `https://www.youtube.com/watch?v=6neDspDoR0U`

b) Wir betrachten zuerst den inneren Ausdruck: A', d.h. alle Elemente, die nicht in A sind. Wenn wir jetzt den äusseren Ausdruck $(\ )'$ finden wollen, interessieren wir uns also für das, was in A liegt.

c) A geschnitten mit dem Stichprobenraum ist A.

d) A vereinigt mit dem Stichprobenraum ist S. Es können in der Vereinigung keine weiteren Elemente dazukommen, da es keine Elemente ausserhalb von S gibt.

e) Ein Element kann nicht gleichzeigtig in A und ausserhalb von A liegen. Deshalb ist die Schnittmenge von A und A' die leere Menge $\{\}$.

f) $A \cup A' = S$

Beispiel 4.6 Trinkverhalten **E8TJ8C**

Im Rahmen einer Studie in Soziologie werden folgenden Merkmale unterschieden

- Geschlecht: f (G_1),m (G_2)
- Trinkverhalten: abstinent (K_1), gelegentlicher Alkoholkonsum (K_2), regelmässiger Alkoholkonsum (K_3)
- Zivilstand: ledig (M_1), verheiratet (M_2), geschieden (M_3), verwitwet (M_4)

a) Gib den Stichprobenraum an

b) Gib folgende Ereignisse explizit an (Liste):

- A : die Person ist männlich
- B : die Person trinkt
- C : die Person ist unverheiratet

c) Beschreibe folgende Mengen in Worten

$$A \cup B,\ A \cap C,\ C',\ A \cap B \cap C,\ A' \cap B$$

Lösung:

a)

$$S = \{G_i K_j M_l;\ i \in \{1,2\}, i \in \{1,2,3\}, i \in \{1,2,3,4\}, \}$$

b) Gib folgende Ereignisse explizit an (Liste):

- $A = \{G_2 K_j M_l\}$
- $B = \{G_i K_2 M_l,\ G_i K_3 M_l\}$
- $C = \{G_i K_j M_1,\ G_i K_j M_3,\ G_i K_j M_4\}$

c) $A \cup B$: Personen, die männlich sind oder die trinken (eine der Bedingungen muss mindestens zutreffen). $A \cap C$: Personen, die männlich und unverheiratet sind (beide der Bedingungen müssen zutreffen). C': Verheiratete Personen. $A \cap B \cap C$: Männlich und trinkt und unverheiratet (alle drei Bedingungen müssen zutreffen). $A' \cap B$: Frauen, die trinken.

4.1 Wahrscheinlichkeitsberechnung

Definition 4.3 Wahrscheinlichkeit (Laplace)

Die Wahrscheinlichkeit für das Eintreten eines Ereignisses E ist

$$P(E) = \frac{E}{n}$$

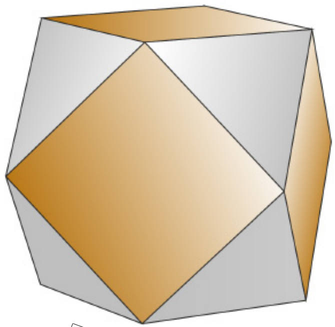
mit

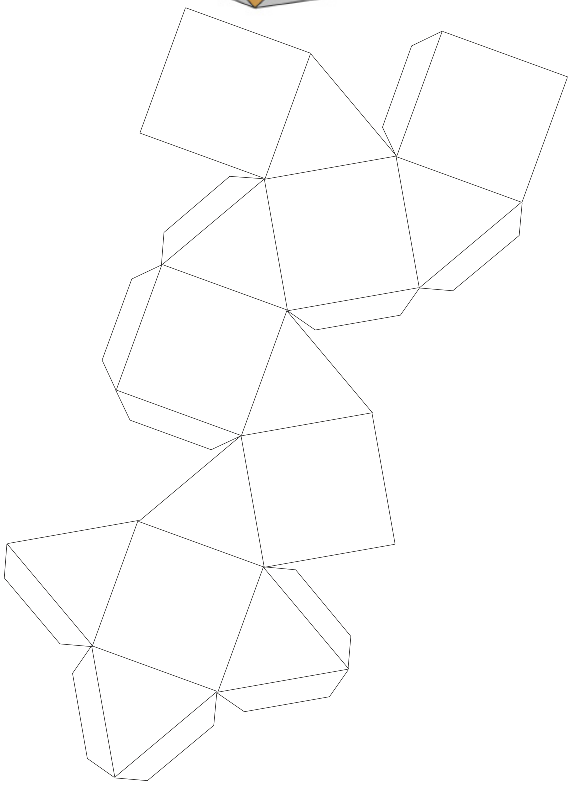
- E die Anzahl der günstigen Ergebnisse
- n die Anzahl der möglichen Ergebnisse

Wir unterscheiden die experimentelle Wahrscheinlichkeit und die theoretische Wahrscheinlichkeit.

Beispiel 4.7 Kuboktaeder **EIUWX9**

Sie sehen abgebildet einen Kuboktaeder und sein Netz.





a) Nummerieren und zählen Sie die Seiten des Kuboktaeders. Tipp: Zuerst Vierecke nummerieren, dann fortlaufend die Dreiecke.

b) Sie sollen 10 CHF setzen, auf Quadrate oder auf Dreieck. Bei einem richtigen Tipp wird ihr Einsatz verdoppelt. Was tippen Sie?

c) Werfen Sie den Kuboktaeder 10 Mal. Wie viel mal erhalten Sie oben ein Quadrate?

d) Wiederholen Sie das Experiment 100 Mal. Notieren sie bei jedem 10er Schritt wie viele Quadrate Sie insgesamt erhalten haben.

e) Berechnen sie die Wahrscheinlichkeit ein Quadrate zu würfeln. Ist die Wahrscheinlichkeit nach 10 Würfen gleich wie nach 100 Würfen? Welche Zahlen würden Sie bei 100 000 Würfen erwarten?

Lösung:

a) Wir erhalten 1-6 für die Quadrate, und 7-14 für die Dreieck.

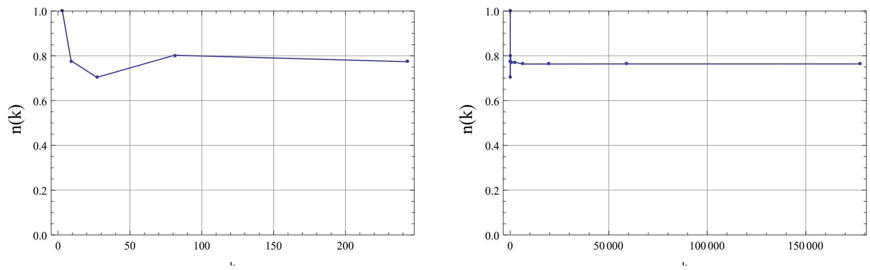
b) Alle Antworten korrekt. Eine Möglichkeit wäre: “Dreiecke hat es 8, also tippe ich auf die Dreiecke.”

c) Wir erhalten Zahlen zwischen 1 bis 10.

d) Siehe Tabelle unten.

e) Die Wahrscheinlichkeit konvergiert, wenn wir oft Würfeln. Wir würden etwas erwarten im Bereich $0.73 \pm |0.73 - 0.82| = 0.73 \pm |0.11|$

Anz. Würfe k	10	20	30	40	50	60	70	80	90	100
Total Anz. $\square$	7	14	26	34	37	44	47	53	74	73
$P(\square)$	0.7	0.7	0.87	0.85	0.74	0.73	0.67	0.66	0.82	0.73



Beispiel 4.8 Oktaeder **S3HRYC**

Wir werfen einen Oktoeder, mit 8 gleichen Seiten. Bestimmen Sie die Wahrscheinlichkeit für folgende Ereignisse

a) Wir würfeln eine gerade Zahl

b) Wir würfeln ein Vielfaches von 3

c) Wir würfeln ein Vielfaches von 4

d) Wir würfeln kein Vielfaches von 4

e) Wir würfeln weniger als 4

f) Wir würfeln eine 9

Lösung:

a) $P(\text{gerade}) = 5/10 = 1/2$

b) $P(\text{vielf.}) = 2/8 = 1/4$

c) $P(\text{vielf.}) = 2/8 = 1/4$

d) $P(\text{n.vielf.}) = 6/8 = 3/4$

e) $P(X < 4) = 3/8$

f) $P(X = 9) = 0/8$

Beispiel 4.9 Münze werfen **DU9UUW**

Wir werfen eine Münze drei Mal und notieren **ungeordnet** die Ergebnisse.

a) Geben Sie den Stichprobenraum an.

b) Was ist die Wahrscheinlichkeit, dass genau zwei Mal Kopf geworfen wird (A)?

c) Was ist die Wahrscheinlichkeit, dass genau zwei Mal Zahl geworfen wird (B)?

d) Was ist die Wahrscheinlichkeit, dass mindestens ein Mal Kopf geworfen wird (C)?

e) Was ist die Wahrscheinlichkeit, dass mindestens ein Mal Zahl geworfen wird (D)?

Lösung:
Die Aufgaben können entweder durch Auszählen im Stichprobenraum oder mit den Gesetzen für die Wahrscheinlichkeit berechnet werden.

a) $S = \{\text{KKK}, \text{KKZ}, \text{KZZ}, \text{ZZZ}\}$

b) $P(A) = \frac{1}{4}$

c) $P(B) = \frac{1}{4}$

d) $P(C) = \frac{3}{4}$

e) $P(D) = \frac{3}{4}$

Beispiel 4.10 Münze werfen **6QW1XT**

Wir werfen eine Münze drei Mal und notieren **geordnet** die Ergebnisse.

a) Geben Sie den Stichprobenraum an.

b) Was ist die Wahrscheinlichkeit, dass genau zwei Mal Kopf geworfen wird (A)?

c) Was ist die Wahrscheinlichkeit, dass genau zwei Mal Zahl geworfen wird (B)?

d) Was ist die Wahrscheinlichkeit, dass mindestens ein Mal Kopf geworfen wird (C)?

e) Was ist die Wahrscheinlichkeit, dass mindestens ein Mal Zahl geworfen wird (D)?

Lösung:
Die Aufgaben können entweder durch Auszählen im Stichprobenraum oder mit den Gesetzen für die Wahrscheinlichkeit berechnet werden.

a) $S = \{\text{KKK, KKZ, KZK, KZZ, ZKK, ZKZ, ZZK, ZZZ}\}$

b) $P(A) = \frac{3}{8}$

c) $P(B) = \frac{3}{8}$

d) $P(C) = \frac{7}{8}$

e) $P(D) = \frac{3}{4}$

Beispiel 4.11 'STATISTIK' **91PBAQ**

Wir schreiben die Buchstaben des Wortes 'STATISTIK' auf Karten. Wir ziehen dann zufällig aus diesen Karten. Was ist die Wahrscheinlichkeit

a) dass wir ein K erwischen?

b) dass einen Vokal erwischen?

c) dass ein P erwischen?

Lösung:

a) $P(\mathrm{K}) = 1/9$

b) $P(\mathrm{Voc.}) = 3/9 = 1/3$

c) $P(\mathrm{P}) = 0/9 = 0$

4.2 Verknüpfte Zufallsexperimente

Beispiel 4.12 Spiele **LPHB76**

In einer Klasse mit 38 StudentInnen spielen 29 Computer-Spiele (C), 10 Gesellschaftsspiele (G) und 9 spielen beides.

a) Zeichnen Sie ein Venn-Diagramm, das die Situation beschreibt.

b) Wir wählen einen Studierenden zufällig aus. Wie gross ist die Wahrscheinlichkeit, dass er/sie weder in C noch in G ist?

c) Berechnen Sie $P(C)$ und $P(G)$.

d) Wie gross ist $P(C \cup G)$?

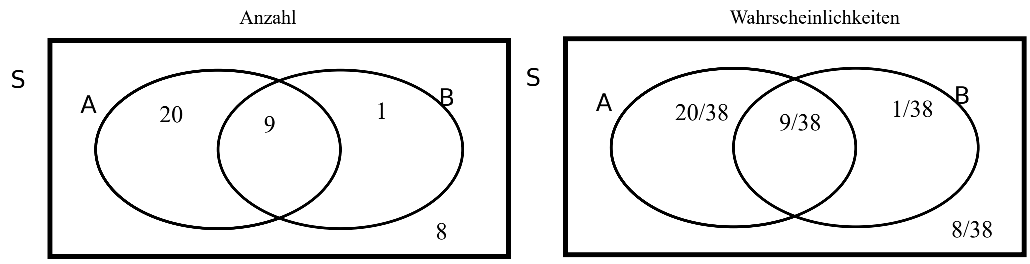
Lösung:

a) Venn-Diagramm unten

b) Wahrscheinlichkeit weder in C noch in G $P = 8/38$

c) $P(C) = 29/38$ und $P(G) = 10/38$.

d) $P(C \cup G) = (29/38 - 9/38) + 1/38 = 29/38 + 1/38 - 9/38$



Beispiel 4.13 Allgemeine Gesetze **F9GT6F**

Betrachten Sie noch einmal die vorherige Aufgabe

a) Wie kann allgemein $P(A \cup B)$ berechnet werden, falls $P(A)$ und $P(B)$ bekannt sind?

b) Es ergeben sich oben vier Teilmengen. Wie gross ist die Wahrscheinlichkeit aller Teilmengen zusammen?

Lösung:

a) Es gilt $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, d.h. es muss zusätzlich $P(A \cap B)$ bekannt sein. Schliessen sich A und B aus, d.h. $A \cap B = \{\}$, dann ist $P(A \cap B) = 0$ und $P(A \cup B) = P(A) + P(B)$.

b) Wahrscheinlichkeit aller Teilmengen

$$\frac{20}{38} + \frac{9}{38} + \frac{1}{38} + \frac{8}{38} = \frac{38}{38} = 1$$

Die Summe wird immer $\frac{N}{N} = 1$ sein, denn auf dem Bruchstrich addieren wir alle Möglichkeiten und unter dem Bruchstrich steht die Zahl aller Möglichkeiten.

Satz 4.2 Additionssatz

Für die Ereignisse A und B ist die Wahrscheinlichkeit, dass entweder A oder B eintritt

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

Oft schliessen sich A und B aus, d.h. $A \cap B = \{\}$, dann ist $P(A \cap B) = 0$ und $P(A \cup B) = P(A) + P(B)$.

Beispiel 4.14 Sportclub **NNM3F1**

In einem Sportclub gibt es 40 Mitglieder die Badminton spielen. 37 spielen Squash, 21 spielen beides und 7 spielen weder das eine noch das andere.

a) Stellen Sie die Angaben in einem Venn-Diagramm dar.

b) Wie viele Mitglieder hat der Club?

c) Was ist die Wahrscheinlichkeit, dass ein Miglied

i) Badminton spielt
ii) beides spielt
iii) weder das eine noch das andere spielt
iv) mindestens eine Sportart macht

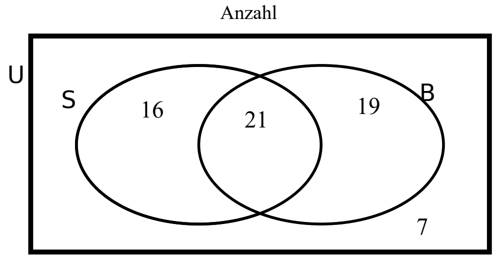
Lösung:

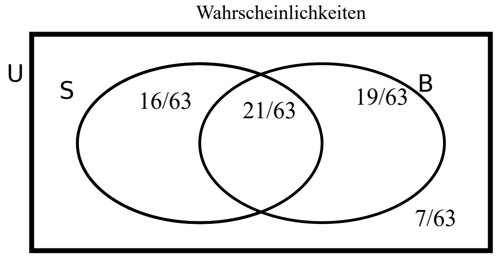
a) Venn-Diagramm unten

b) 63

c) Wahrscheinlichkeiten

i) $P(B) = \frac{40}{63}$
ii) $P(B \cap S) = \frac{21}{63}$
iii) $P((B \cup S)') = \frac{7}{63}$
iv) $P(B \cup S) = 1 - \frac{7}{63} = \frac{56}{63}$





Beispiel 4.15 Sandwich **V7Q9DP**

Robi stellt Sandwiches zufällig zusammen, aus einer von drei Brotsorten (V,K,Z) und einer von 4 Käsesorten (E,G,S,P), mit oder ohne Gurken.

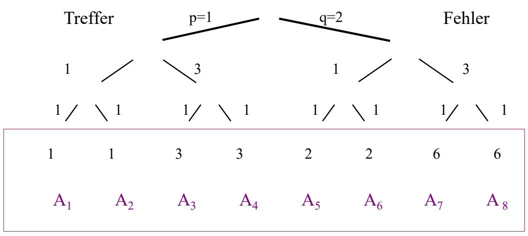
a) Betrachten Sie die Baumdiagramme. Was stellen Sie dar? Was könnte man mit ihnen berechnen?

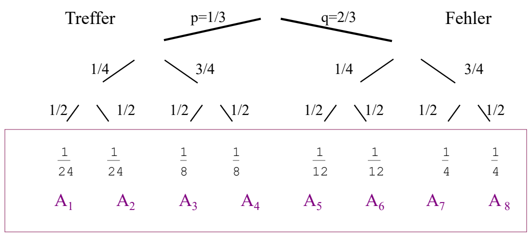
b) Jemand bestellt $[\mathrm{V}, \mathrm{G}, \mathrm{m}]$. Wie gross ist die Chance, dass Robi das richtig Sandwich liefert?

c) Wir bestellen $[\mathrm{V}, \mathrm{G}, \mathrm{m}]$. Wie gross ist die Chance, dass bei Robis Sandwich das Brot und die Gurke stimmen, nicht aber der Käse?

d) Wir bestellen $[\mathrm{K}, \mathrm{P}, \mathrm{o}]$. Wie gross ist die Chance, dass bei Robis Sandwich das Brot und Käse stimmen nicht aber die Gurke?

e) Wir bestellen $[\mathrm{K}, \mathrm{P}, \mathrm{o}]$. Wie gross ist die Chance, dass zumindest das Brot und Käse stimmen?





Lösung:

a) A) beschreibt die Anzahl Möglichkeiten beim Tippen. Jede Zutat ist auf einer Stufe. Z.B. tippen wir V für das Brot, dann gibt es eine Möglichkeit, dass wir richtig liegen und 2, dass wir falsch liegen. Multiplikation entlang eines Astes ergibt die Anzahl Möglichkeiten in diesem Ast. B) beschreibt die Wahrscheinlichkeiten beim Tippen. Multiplikation entlang eines Astes ergibt die Anzahl Wahrscheinlichkeit für diesen Ast.

b) $P = 1/24$

c) Weg über die Anzahl (A): Es gibt $1 \cdot 3 \cdot 1 = 3$ Möglichkeiten, dass Robis Sandwich korrekt ist. $P = 3/24 = \frac{1}{8}$
Weg über die Wahrscheinlichkeit: $P = \frac{1}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} = \frac{1}{8}$

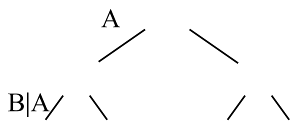
d) Die Wahrscheinlichkeit ist $P = \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{24}$

e) Die Wahrscheinlichkeit ist $P = \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{12}$

Satz 4.3 Multiplikationssatz

Für die Ereignisse A und B ist die Wahrscheinlichkeit, dass beide gleichzeitig eintreten

$$P(A \cap B) = P(B|A) \cdot P(A)$$



Oft hat der Ausgang von A keinen Einfluss auf den Ausgang von B. Wir nennen dann A und B unabhängig und schreiben

$$P(A \cap B) = P(A) \cdot P(B)$$

Beispiel 4.16 Verkehrsregelverstösse **EXRYUV**

Wir betrachten die Wahrscheinlichkeit für Verkehrsregelverstösse

Altersgruppe	18-20	21-29	30-39	≥ 40	Total
Wahrscheinlichkeit	0.06	0.47	0.29	0.18	1

Was sind die Wahrscheinlichkeiten, dass gegen die Regeln verstossen wird

a) in der jüngsten Altersgruppe

b) zwischen 21 und 40

c) vor 40

Lösung:

a) 6%

b) 0.76

c) 1-0.18=0.82

Beispiel 4.17 Führende Einsen **XU3CAM**

Es ist zwar überraschend, aber es ist eine Tatsache, dass in Geheimzahlen (z.B. zum Entsprerren des Handys) für die erste Ziffer folgende Wahrscheinlichkeitsverteilung befolgt.[a]
Bestimmen Sie folgende Wahrscheinlichkeiten für

a) $A = \{$die erste Ziffer ist eine 1$\}$

b) $B = \{$die erste Ziffer grösser als 5$\}$

c) $C = \{$die erste Ziffer ist eine ungerade Zahl$\}$

d) $A \cap B$

e) A'

f) $C \cap B$

Lösung:

a) $P(A) = 0.3$

b) $P(B) = P(6) + P(7) + P(8) + P(9) = 0.221$

c) $P(C) = P(1) + P(3) + P(5) + P(7) + P(9) = 0.603$

d) $P(A \cap B) = P(A) \cdot P(B|A) = 0.3 \cdot 0$

e) $P(A') = 1 - P(A) = 0.7$

f) Da wir weder $P(C \cap B)$ noch $P(C|B)$ kennen, berechnen wir direkt

$$P(C \cap B) = P(\{7, 9\}) = P(7) + P(9) = 0.103$$

[a] Noch überraschender ist, dass auch Zahlen in Buchaltung und Naturwissenschaftlichen Berechnungen eine ähnliche Verteilung befolgen. Mehr dazu unter Benfordsches Gesetz.

Beispiel 4.18 Führende Einsen, Gleichverteilung **R96ZRG**

Aufgabe wie oben.
Wenn einzelne Wahrscheinlichkeiten gleich sind, was passiert bei der oder-Verknüpfung?
Fassen Sie anschliessend Ihre Erkenntnisse in einem (Lehr)-Satz zusammen.

Lösung:

a) $P(A) = 0.1$

b) $P(B) = P(6) + P(7) + P(8) + P(9) = 4 \cdot 0.1 = 0.4$

c) $P(C) = P(1) + P(3) + P(5) + P(7) + P(9) = 0.5$

d) $P(A \cap B) = P(A) \cdot P(B|A) = 0.1 \cdot 0$

e) $P(A') = 1 - P(A) = 0.9$

f) Da wir weder $P(C \cap B)$ noch $P(C|B)$ kennen, berechnen wir direkt

$$P(C \cap B) = P(\{7, 9\}) = 0.2$$

Sind die Wahrscheinlichkeiten gleichverteilt, können die Wahrscheinlichkeit berechnen über

$$P = k \cdot p$$

wobei k die Anzahl Elemente sind und p die Wahrscheinlichkeit, dass ein einzelnes Ereignis eintritt.

Beispiel 4.19 Tasche mit Zurücklegen EJRTYR

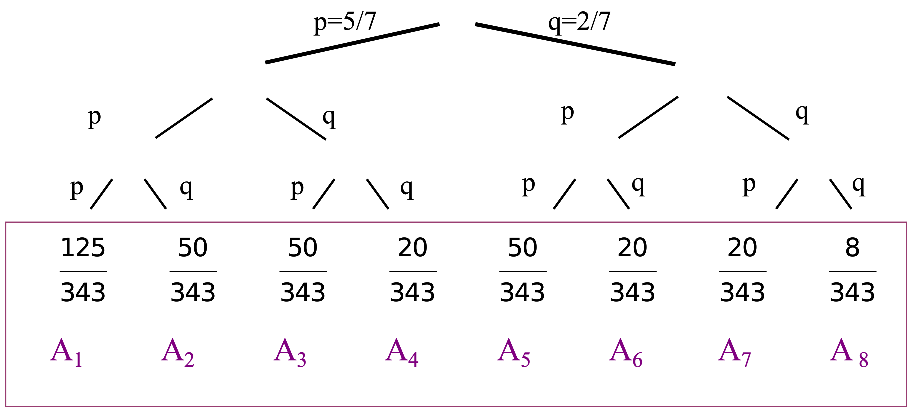
Eine Tasche enthält 5 blaue Bälle (b) und 2 grüne Bälle (g). Wir ziehen einen Ball aus der Taschen notieren die Farbe und legen ihn dann zurück. Wir wiederholen dies 3 Mal. Zeichnen Sie einen Ereignisbaum.

a) Wie gross ist die Wahrscheinlichkeit für (bbb) und (bbg), geordnete Stichproben.

b) Wie gross ist die Wahrscheinlichkeit für (bbb) und (bbg), ungeordnete Stichproben.

c) Für ungeordnete Stichproben: Auf welches Ergebnis sollten sie wetten, (bbb), (bbg), (bgg) oder (ggg)?

Lösung:



a) (bbb) geordnet, d.h. A_1:

$$P = \frac{125}{343}$$

(bbg) geordnet, d.h. A_2:

$$P = \frac{50}{343}$$

b) (bbb) ungeordnet, d.h. A_1 :

$$P = \frac{125}{343} \approx 0.364431$$

(bbg) ungeordnet, d.h. A_2, A_3 und A_5:

$$P = 3 \cdot \frac{50}{343} \approx 0.437318$$

c) Wir berechnen noch (bgg) ungeordnet, d.h. A_4, A_6 und A_7:

$$P = 3 \cdot \frac{20}{343} \approx 0.174927$$

(ggg) ungeordnet, d.h. A_8:

$$P = 3 \cdot \frac{8}{343} \approx 0.0233236$$

Am wahrscheinlichsten ist also (bbg).

Beispiel 4.20 Tasche ohne Zurücklegen **1WGNRL**

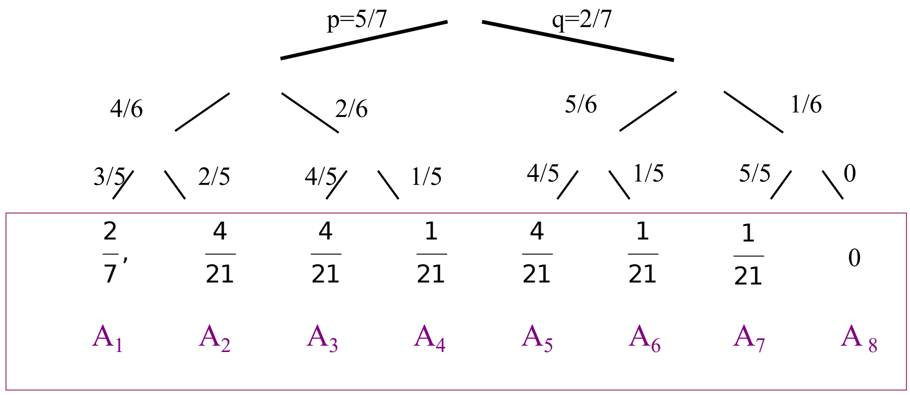
Eine Tasche enthält 5 blaue Bälle (b) und 2 grüne Bälle (g). Wir nehmen nacheinander drei Bälle aus der Tasche. Vervollständigen Sie den Ereignisbaum.

a) Wie gross ist die Wahrscheinlichkeit für (bbb) und (bbg), geordnete Stichproben.

b) Wie gross ist die Wahrscheinlichkeit für (bbb) und (bbg), ungeordnete Stichproben.

c) Für ungeordnete Stichproben: Auf welches Ergebnis sollten sie wetten, (bbb), (bbg), (bgg) oder (ggg)?

Lösung:



a) (bbb) geordnet, d.h. A_1:

$$P = \frac{2}{7}$$

(bbg) geordnet, d.h. A_2:

$$P = \frac{4}{21}$$

b) (bbb) ungeordnet, d.h. A_1 :

$$P = \frac{2}{7} \approx 0.285714$$

(bbg) ungeordnet, d.h. A_2, A_3 und A_5:

$$P = 3 \cdot \frac{4}{21} \approx 0.571429$$

c) Wir berechnen noch (bgg) ungeordnet, d.h. A_4, A_6 und A_7:

$$P = 3 \cdot \frac{1}{21} \approx 0.142857$$

(ggg) ungeordnet, d.h. A_8:

$$P = 0$$

Am wahrscheinlichsten ist also (bbg).

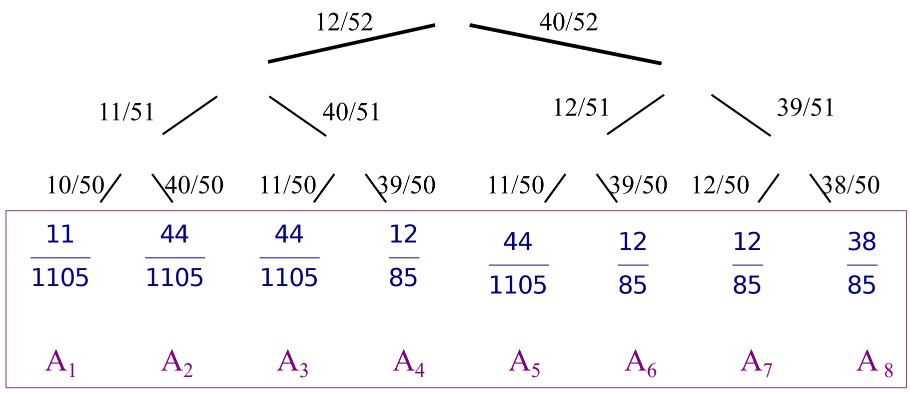
Beispiel 4.21 Kartensatz **71T1FW**

Aus einem Stapel mit 52 Karten — davon 12 mit Abbildungen drauf — werden 3 Karten gezogen. Zeichnen Sie einen Ereignisbaum und bestimmen Sie die Wahrscheinlichkeit

a) dass wir drei Karten mit Abbildungen ziehen.

b) dass wir (genau) zwei Karten mit Abbildungen ziehen.

Lösung:



a) drei Karten mit Abbildungen erhalten wir nur über A_1:

$$P = \frac{11}{1105} \approx 0.00995475$$

b) genau zwei Karten mit Abbildungen erhalten wir über A_2 A_3 und A_5:

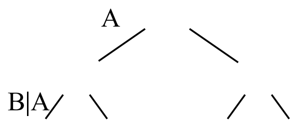
$$P = \frac{44}{1105} + \frac{44}{1105} + \frac{44}{1105} = \frac{132}{1105} \approx 0.119457$$

4.3 Bedingte Wahrscheinlichkeit

Infobox 4.2 Bedingte Wahrscheinlichkeit

Der Produktsatz lässt sich wie folgt auflösen

$$P(A \cap B) = P(B|A) \cdot P(A) \;\Rightarrow\; P(B|A) = \frac{P(A \cap B)}{P(A)}$$



D.h. aus $P(A \cap B)$ (Wahrscheinlichkeit, dass beides gleichzeitig eintritt) und aus $P(A)$, lässt sich berechnen, dass B eintritt, wenn A schon eingetreten ist. Alternativ: A ist schon eingetreten. Das heisst uns interessieren uns nur noch die Anzahl der Fälle, wo A Eingetrenten ist, d.h.

$$P(B|A) = \frac{N(A \cap B)}{N(A)} = \frac{N(A \cap B)/N_{\text{tot}}}{N(A)/N_{\text{tot}}} = \frac{P(A \cap B)}{P(A)}$$

Beispiel 4.22 Stifte **L3Z6XI**

In einer Schachtel gibt es 9 Stifte, die funktionieren und 5, die kaputt sind. Mark und Peter nehmen noch einander zwei Stifte aus der Schachtel.

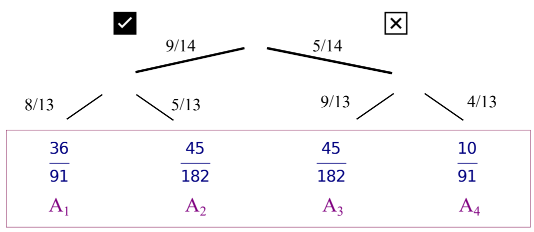
a) Zeichnen Sie einen Ereignisbaum.

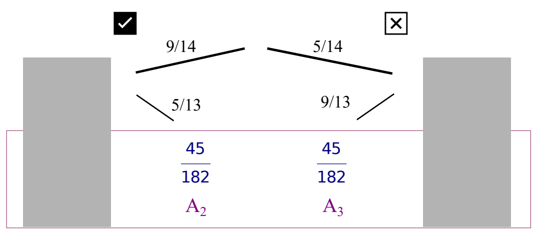
b) Was ist die Wahrscheinlichkeit, dass beide einen kaputten Stift erwischen?

c) Was ist die Wahrscheinlichkeit, dass mindestens ein kaputter Stift gezogen wird?

d) Es wurde genau ein kaputter Stift gezogen. Was ist die Wahrscheinlichkeit, dass Peter ihn zog.

Lösung:





a) Ereignisbaum unten

b) beide Stift kaputt, d.h. A_4

$$P = \frac{10}{91} \approx 0.10989$$

c) mindestens ein kaputter Stift, d.h. A_2, A_3 und A_4:

$$P = \frac{45}{182} + \frac{45}{182} + \frac{10}{91} = \frac{55}{91} \approx 0.604396$$

d) Wir wissen, dass genau ein kaputter Stift gezogen, d.h. A_1 und A_4 wurden also nicht beschritten, also decken wir sie ab. Jetzt lässt sich die Wahrscheinlichkeit so berechnen:

$$P = \frac{P_{\text{Peter}}}{P_{\text{tot}}} = \frac{\frac{45}{182}}{\frac{45}{182} + \frac{45}{182}} = \frac{1}{2} = 0.5$$

Beispiel 4.23 Team **B1USEU**

Ein Team besteht aus 7 Elektro- und 3 Maschinen-Ingenieuren.

a) Die Leitung wird zufällig bestimmt. Was ist die Wahrscheinlichkeit, dass ein Maschineningenieur gewählt wird?

b) Zwei Teammitglieder sollen eine Schaltung konzipieren. Was ist die Wahrscheinlichkeit dafür, dass zufällig ein Elektro- und ein Maschineningenieur gewährt werden?

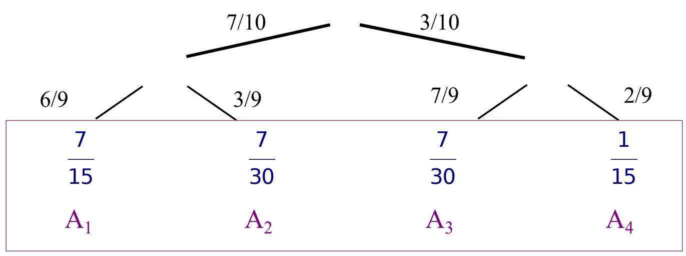
Lösung:

a) Leitung

$$P = \frac{3}{10} = 0.3$$

b) Elektro+Maschinening. d.h. A_2 und A_3

$$P = \frac{7}{30} + \frac{7}{30} = \frac{7}{15} = 0.466667$$



Beispiel 4.24 Luca und Heiri **EI2SE2**

Durchschnittlich löst Luca 5 von 7 korrekt. Heiri beantwortet 5 von 9 korrekt. Beide bearbeiten nun die selbe Aufgabe.

a) Zeichnen Sie einen Ereignisbaum.

b) Was ist die Wahrscheinlichkeit, dass mindestens einer der beiden die Aufgabe korrekt löst?

c) Was ist die Wahrscheinlichkeit, dass *Luca* richtig geantwortet hat, wenn wir wissen, dass die Frage richtig beantwortet wurde.

d) Was ist die Wahrscheinlichkeit, das *Heiri* richtig geantwortet hat, wenn wir wissen, dass die Frage richtig beantwortet wurde.

e) Was ist die Wahrscheinlichkeit, das beide richtig geantwortet haben, wenn wir wissen, dass die Frage richtig beantwortet wurde.

Lösung:

a) Ereignisbaum unten

b) P mindestens einer der beiden hat die Aufgabe korrekt gelöst, d.h. . d.h. A_1, A_2 und A_3

$$P = \frac{25}{63} + \frac{20}{63} + \frac{10}{63} = \frac{55}{63} = 0.818182$$

c) Wir wissen, dass die Frage richtig beantwortet wurde, also decken wir den Ast ab, der nicht beschritten wurde. Luca hat richtig geantwortet in A_1 und A_2. Jetzt lässt sich die Wahrscheinlichkeit so berechnen:

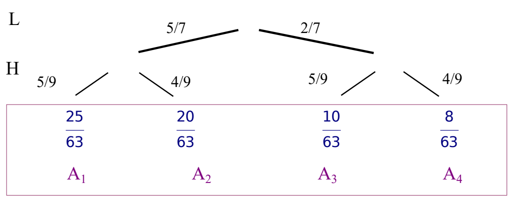
$$P = \frac{P_{\text{Luca}}}{P_{\text{tot}}} = \frac{\frac{25}{63} + \frac{20}{63}}{\frac{25}{63} + \frac{20}{63} + \frac{10}{63}} = \frac{9}{11} \approx 0.818182$$

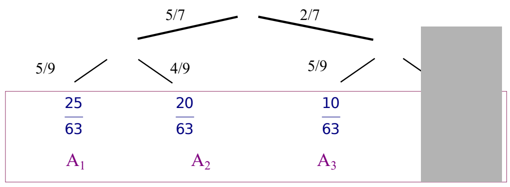
d) Wir wissen, dass die Frage richtig beantwortet wurde, also decken wir den Ast ab, der nicht beschritten wurde. Heiri hat richtig geantwortet in A_1 und A_3. Jetzt lässt sich die Wahrscheinlichkeit so berechnen:

$$P = \frac{P_{\text{Heiri}}}{P_{\text{tot}}} = \frac{\frac{25}{63} + \frac{10}{63}}{\frac{25}{63} + \frac{20}{63} + \frac{10}{63}} = \frac{7}{11} \approx 0.636364$$

e) Beide haben richtig geantwortet in A_1. Jetzt lässt sich die Wahrscheinlichkeit so berechnen:

$$P = \frac{P_{\text{beide}}}{P_{\text{tot}}} = \frac{\frac{25}{63}}{\frac{25}{63} + \frac{20}{63} + \frac{10}{63}} = \frac{5}{11} \approx 0.454545$$





Beispiel 4.25 Kate und Jane spielen Tennis JF9KBJ

Kate und Jane haben folgendes Spiel erfunden: Sie servieren abwechselnd nacheinander, und die erste, die ihren Service verliert, zahlt ein Stück Schokolade. Beide gewinnen ihren Service mit einer Wahrscheinlichkeit von $p = 0.6$. Jane beginnt mit dem Service.
Berechne die Wahrscheinlichkeit, dass

a) Jane nach dem zweiten Serive zahlen muss (dafür muss 3 Mal gespielt werden!)

b) Jane die Schokolade (überhaupt) zahlen muss

c) Kate die Schokolade bezahlt.

Lösung:

a) $P = 0.4 \cdot 0.6 \cdot 0.6$

b) Um die Schokolade (überhaupt) bezahlen zu müssen, muss eines der folgenden Fälle eintreffen:

$$P = 0.4 + 0.4 \cdot 0.6^2 + 0.4 \cdot 0.6^4 + 0.4 \cdot 0.6^6 \ldots = \frac{0.4}{1-(0.6)^2} = 0.625$$

c) Wir rechnen mit der Gegenwahrscheinlichkeit $P = 1 - 0.625 = 0.375$

Beispiel 4.26 RaucherInnen **GV3VND**

Verhalten von 768 Studierenden an FH.

	Raucher (R)	**Gelegentlich (G)**	**Nicht-Raucher (NR)**	**Gesamt**
Männlich (m)	127	73	214	414
Weiblich (w)	99	66	189	354
Gesamt	226	139	403	768

Wir wählen zufällig einen Studierenden. Berechnen Sie:

a) Wahrscheinlichkeit, dass es eine Frau ist

b) Wahrscheinlichkeit, dass es ein männlicher Raucher ist

c) Wahrscheinlichkeit, dass es ein Nicht-Raucher ist

d) Wahrscheinlichkeit, dass es eine Nicht-Raucherin ist, wenn wir schon wissen, dass wir eine Frau erwischt haben

Lösung:

a) $P(\mathrm{w}) = \frac{354}{768} = 0.461$

b) $P(\mathrm{m} \cap \mathrm{R}) = \frac{127}{768} = 0.165$

c) $P(\mathrm{NR}) = \frac{403}{768} = 0.525$

d) $P(\mathrm{NR}|\mathrm{w}) = \frac{189}{354} = 0.534$

Beispiel 4.27 Extremfälle **R4VXZF**

Wir betrachten die Fälle

- keine der Frauen raucht
- alle Frauen rauchen

a) Zeichnen Sie für die Fälle je ein Venn-Diagramm

b) Geben Sie die Wahrscheinlichkeit $P(\mathrm{NR}|\mathrm{w})$ für diese Fälle an.

Lösung:

- keine der Frauen raucht, dann gilt $P(\mathrm{NR}|\mathrm{w}) = 1$

- alle Frauen rauchen, dann gilt $P(\mathrm{NR}|\mathrm{w}) = 0$

Beispiel 4.28 RaucherInnen II **E9AX8J**

	Raucher (R)	Gelegentlich (G)	Nicht-Raucher (NR)
Männlich	0.165	0.095	0.279
Weiblich	0.129	0.086	0.246

Wir wissen aus der Aufgabe oben auch, dass $P(\mathrm{NR}|\mathrm{w}) = 0.534$. Lesen Sie aus oder berechnen Sie und erklären Sie das Resultat.

a) $P(\mathrm{NR} \cap \mathrm{w})$

b) $P(\mathrm{w})$

c) $P(\mathrm{NR}|\mathrm{w}) \cdot P(\mathrm{w})$

d) Vergleichen Sie $P(\mathrm{NR} \cap \mathrm{w})$ mit $P(\mathrm{NR}|\mathrm{w}) \cdot P(\mathrm{w})$.

Lösung:

a) $P(\mathrm{NR} \cap \mathrm{w}) = 0.246$ auslesen

b) $P(\mathrm{w}) = 0.129 + 0.086 + 0.246 = 0.461$

c) $P(\mathrm{NR}|\mathrm{w}) \cdot P(\mathrm{w}) = 0.246$

d) Wir stellen also fest

$$P(\mathrm{NR} \cap \mathrm{w}) = P(\mathrm{NR}|\mathrm{w}) \cdot P(\mathrm{w})$$

Beispiel 4.29 Farbblindheit **VZPWTP**

	Männlich (m)	Weiblich (w)	Total
Farbenblind (FB)	40	2	42
Nicht FB (NFB)	470	488	958

Was ist die Wahrscheinlichkeit, dass jemand farbenblind ist, gegeben es ist eine Frau? **Lösung:**

Da wir wissen, dass es sich um eine weibliche Person handelt, decken wir die anderen Spalten der Tabelle ab und berechnen:

$$P(\mathrm{FB}|\mathrm{w}) = \frac{n(\mathrm{FB} \cap \mathrm{w})}{n(\mathrm{w})} = \frac{2}{2 + 488}$$

Beispiel 4.30 Pünklichkeit **HLKBLS**

Die Fluggesellschaft AUA ist bekannt für ihre Pünklichkeit. Ihre Wahrscheinlichkeit für einen pünklichen Abflug ist $P(D) = 0.83$, die für eine pünkliche Ankunft ist $P(A) = 0.92$ und die Wahrscheinlichkeit dass beides zusammen auftritt ist $P(D \cap A) = 0.78$.
Wir bezeichnen einen unpünktlichen Abflug mit D' und eine unpünkliche Ankunft mit A'. Erstellen Sie eine Tabelle mit den Wahrscheinlichkeiten für alle möglichen Ereignissen und berechnen Sie die Wahrscheinlichkeit für

a) eine pünkliche Ankunft vorausgesetzt, dass der Abflug pünklich sein wird

b) einen pünklichen Ablug vorausgesetzt, dass die Ankunft pünklich war

Lösung:
Wir ergänzen die Tabelle Schritt für Schritt. Geben waren

	A	A'	total
D	0.78		0.83
D'			
total	0.92		

Also können wir zuerst ergänzen

	A	A'	total
D	0.78	0.05	0.83
D'	0.14		
total	0.92		

und schliesslich mit $\sum P_i = 1$

	A	A'	total
D	0.78	0.05	0.83
D'	0.14	0.03	
total	0.92		

1. $P(A|D) = \frac{P(A\cap D)}{P(D)} = \frac{0.78}{0.83} = 0.94$

2. $P(D|A) = \frac{P(A\cap D)}{P(A)} = \frac{0.78}{0.92} = 0.848$

4.4 Unabhängigkeit

Definition 4.4 Unabhängigkeit

Zwei Ereignisse sind unabhängig, falls ihre Wahrscheinlichkeit

$$P(A \cap B) = P(A) \cdot P(B)$$

d.h. auch

$$P(A|B) = P(A)$$

[Wazir et al., 2012, S. 545]

Beispiel 4.31 Unabhängigkeit **E81EU9**

Ist die Pünklichkeit der Abflüge der Fluggesellschaft AUA unabhängig von der Pünklichkeit der Ankünfte?

Lösung:
Nein, die Pünlichkeit von Ankunft und Abflug sind abhängig.

a) $P(A) = 0.92$ und $P(A|D) = 0.94$

b) Alternativ: $P(A \cap D) = 0.78$ und $P(A) \cdot P(D) = 0.92 \cdot 0.76 = 0.76 \neq P(A \cap D)$

Beispiel 4.32 Blutprobe RCCB5T

Die Polizei hält verdächtige Autofahrer an. Die Fahrer machen dann einen Atemtest ($P(A) = 81\%$), oder einen Bluttest ($P(B)$ =40%) oder beides ($P(A \cap B)$ =25%). Erstellen Sie eine Tabelle mit den Wahrscheinlichkeiten für alle möglichen Ereignisse und geben Sie folgende Wahrscheinlichkeiten an:

a) Der Fahrer macht keinen Test.

b) Der Fahrer macht einen Test.

c) Der Fahrer macht genau einen Test.

d) Sind A und B unabhängig voneinander?

Lösung:
Die grünen Einträge sind in der Aufgabenstellung gegeben:

	A	A'	total
B	0.25	0.15	0.4
B'	0.56	0.04	
total	0.81		

a) $P(A' \cap B') = 0.04$

b) $P(A \cup B) = 1 - P((A \cup B)') = 1 - P(A' \cap B') = 1 - 0.04 = 0.96$

c) $P(A \cap B') + P(A' \cap B) = 0.56 + 0.15 = 0.71$

d) Falls A und B unabhängig sind, müsste gelten

$$P(A) = P(A|B)$$

hier ist aber $P(A) = 0.81$ und $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.25}{0.4} = 0.625$ also sind A und B abhängig voneinander.

4.5 Screening

Definition 4.5 Sensitivität, Spezifität

Wir bezeichnen Personen, die eine gewisse Krankheit haben als ‘Merkmalsträger’.

- Sensitivität: Fähigkeit eines Tests, Merkmalsträger vollständig aus einer Gruppe herauszufiltern. Wir die Sensitivität ausschliesslich mit einer Gruppe von Kranken bestimmt, bedeutet eine Sensitivität von 100%, dass *alle* Merkmalsträger ein richtig-positives Testergebnis haben.

- Spezifität: Fähigkeit eines Tests, ausschliesslich Merkmalsträger zu erfassen. Wird die Spezifität mit einer Gruppe von Gesunden bestimmt, bedeutet eine Spezifität von 100%, dass alle Nicht-Merkmalsträger ein richtig-negatives Testergebnis haben.

Beispiel 4.33 Sensitivität und Spezifiät symbolisch IIT6Y1

Wir bezeichnen das Ereignis 'positiver Test' mit T und 'Person hat das Merkmal' mit M.

a) Geben Sie Sensitivität und Spezifiät symbolisch an.

b) Wie kann die Sensitivität mit Hilfe von $P(T \cap M)$ und $P(M)$ angegeben werden?

c) Wie kann die Spezifiät angegeben werden?

Lösung:

a) Sensitivität: $P(T|M)$ Spezifiät: $P(T'|M')$

b) Sensitivität
$$P(T|M) = \frac{P(T \cap M)}{P(M)}$$

c) Spezifiät
$$P(T'|M') = \frac{P(T' \cap M')}{P(M')}$$

Beispiel 4.34 Prostatakrebs NSQ9ZG

Prostatakrebs gehört mit 25% zu den häufigen Krankheiten bei Männern. Er kann mit Hilfe der PSA (Prostate-Specific Antigen) im Blut diagnostiziert werden. Der Test hat eine Sensitivität von 53.2% und eine Spezifizität von 83.6%.

a) Beschriften Sie die Zellen mit Symbolen

b) Berechnen Sie $P(T \cap M)$.

c) Berechnen Sie $P(T' \cap M')$ und füllen Sie die Tabelle auf

d) Wir testen 100 Personen. Wie viele werden positiv getestet und sind wirklich krank?

e) Wie viele werden positiv getestet, obwohl sie nicht krank sind?

f) Ist es eine gute Idee die gesamte männliche Bevölkerung zu screenen? Wie könnte der Test gewinnbringend eingesetzt werden?

Lösung:

		Test		
		pos (T)	neg (T')	Total
Krankheit	ja (M)	$P(T \cap M)$	$P(T' \cap M)$	$P(M)$
	nein (M')	$P(T \cap M')$	$P(T' \cap M')$	$P(M')$
	Total	$P(T)$	$P(T')$	

a) Beschriftung oben

b) $P(T \cap M) = P(T|M) \cdot P(M) = 0.532 \cdot 0.25 = 0.133$

c) $P(T' \cap M') = P(T'|M') \cdot P(M') = 0.836 \cdot 0.75 = 0.627$. Wir ergänzen die Tabelle:

$$0.133 + x = 0.25 \;\Rightarrow\; x = 0.117$$

und

$$0.627 + x = 0.75 \;\Rightarrow\; x = 0.123$$

d) Von 100 werden rund 13 positiv getestet und sind wirklich krank.

e) Es werden werden 12 positiv getestet, obwohl sie nicht krank sind.

f) Screening ist keine gute Idee. Der Test muss auf eine Gruppe angewandt werden (z.B. Männer ab 50 Jahren), in der Prostatakrebs häufig ist. Hier das Beispiel mit dem gleichen Test für eine Gruppe, in der 85% erkrankt sind:

	T	T'
M	0.4522	0.3978
M'	0.0246	0.1254

Auf 100 Personen werden also 45 korrekt positiv getestet und nur 2 falsch positiv.

		Test		
		pos (T)	neg (T')	Total
Krankheit	ja (M)	0.133	0.117	0.25
	nein (M')	0.123	0.627	0.75
	Total	0.256	0.744	1

Beispiel gemässt “So lügt man mit Statistik” Krämer [2015].

Beispiel 4.35 HIV **VC8BDB**

Der HIV Virus ist in der Schweiz nicht sehr verbreitet und betrifft aktuell in der Schweiz 6.4 pro $100\,000$ Einwohner. Die heutigen Tests sind sehr gut entwickelt und zeigen eine Sensitivität von 99% und eine Spezifizität von mindestens 99%.

a) Füllen Sie die Tablle auf

b) Wir testen alle Personen im Kanton Bern, d.h. rund 1 Mio. Personen. Wie viele werden positiv getestet und sind wirklich krank?

c) Wie viele werden positiv getestet obwohl sie nicht krank sind?

d) Ist es eine gute Idee die gesamte Bevölkerung zu screenen? Wo liegt der Unterschied zu den Krebs-Diagonose-Tests?

e) Wie kann man die HIV-Tests trotzdem gewinnbringend anwenden?

Lösung:

		Test		
		pos (T)	neg (T')	Total
Krankheit	ja (M)	.00006336	$6.4 \cdot 10^{-7}$	0.000064
	nein (M')	0.00999936	0.989937	0.999936
	Total	0.0100627	0.989937	1

a) Tabelle

b) Auf 1 Mio. Personen, werden 63 positiv getestet und sind wirklich krank.

c) Es werden 9999 positiv getestet, obwohl sie nicht krank sind.

d) Screening ist keine gute Option.

e) Gewinnbringend wird der Test angewendet, wenn er gezielt bei Personen angewandt wird, die einem Ansteckungsrisiko ausgesetzt waren.

Beispiel 4.36 Sally Clark **IPIMEX**

Sally Clarks beide Kinder sind Kindstod gestorben. Der Staatsanwalt gibt an, dass die Wahrscheinlichkeit für einen doppelten natürlichen plötzlichen Kindstod bei

$$p = \frac{1}{8\,543} \cdot \frac{1}{8\,543} \approx \frac{1}{70\,\text{Mio.}}$$

liegen würde. Sie wird anschliessend als schuldig verurteilt. Wir bezeichnen das Ereignis 'Eltern schuldig' mit S und 'doppelter natürlichen plötzlichen Kindstod' mit K.

a) Drücken Sie die Angaben des Staatsanwaltes symbolisch aus. Tragen Sie die Angaben in die Tabelle unten ein.

b) Für welche Grösse interessiert sich der Staatsanwalt? Wie könnte er sie berechnen?

c) Mit welchem Argument befreien Sie Sally Clark aus dem Gefängnis?

Lösung:

	(K)	(K')	Total
S			
S'			
Total	$P(K) = \left(\frac{1}{8\,543}\right)^2$		

a) Die Angaben des Staatsanwaltes sind $p(S' \cap K)$.

b) Der Staatsanwalt interessiert sich für

$$P(S'|K) = \frac{P(S' \cap K)}{P(K)}$$

c) Argument: Der doppelte Kindstod hat stattgefunden, deshalb müsste sich der Staatsanwaltes für die bedingten Wahrscheinlichkeiten $P(S'|K) = \frac{P(S' \cap K)}{P(K)}$ und $P(S|K) = \frac{P(S \cap K)}{P(K)}$ interessieren. Sie hängen von $P(S' \cap K)$ und $P(S \cap K)$ ab und werden mit $P(K)$ normalisiert, d.h. man könnte auch einfach $P(S' \cap K)$ und $P(S \cap K)$ vergleichen. Über diese Wahrscheinlichkeiten, d.h. über die Verknüpfung von Doppeltem Kindstod K mit S oder mit S' ist aber nichts bekannt, deshalb sind alle Zahlen oben nicht relevant für den Fall.
Ausserdem: Nach geltendem Recht kann jemand nicht aufgrund von Wahrscheinlichkeiten verurteilt werden sondern nur augrund von Tatbeweisen.

KAPITEL 6

Diskrete Zufallsgrössen

Lernziele 6.1 Diskrete Zufallsgrössen

- Die Studierenden können Wahrscheinlichkeitsverteilung darstellen u.a. als Verteilfunktion.
- Sie können für Zufallsgrössen die Wahrscheinlichkeit bestimmen anhand einer Wahrscheinlichkeitsverteilung oder anhand der Verteilfunktion.
- Sie können anhand der Wahrscheinlichkeitsverteilung den Erwartungswert und die Varianz der Verteilung berechnen.
- Sie die Binomialverteilung in Excel oder in Matlab erzeugen und Wahrscheinlichkeiten für Ereignisse ($P(X = x)$, $P(X \geq x)$ und $P(X \leq x)$) berechnen
- Sie können für die Binomialverteilung Erwartungswert und die Varianz berechnen.
- Sie Simulation durchführen mit Matlab und experimentelle Wahrscheinlichkeiten bestimmen.

6.1 Wahrscheinlichkeitsverteilungen

Beispiel 6.1 Wahrscheinlichkeitsverteilungen: Münze **9CZTV6**

Versuch: zwei unabhängige Münzwürfe.
Erstellen Sie eine Tabelle (Resultat ungeordnet notiert):

Lösung:

Ausgang Experiment	(K,K)	(K,Z)	(Z,Z)
Wahrscheinlichkeit	0.25	0.5	0.25
Aufsummieren	0.25	0.75	1

Definition 6.1 Zufallsgrösse

$$\begin{aligned} X : S &\to \mathbb{R} \\ s_i &\mapsto X(s_i) = x_i \end{aligned}$$

$S = \{s_1, s_2, \ldots\}$: Stichprobenraum Zufallsgrösse X kann Werte x_i annehmen.

Definition 6.2 Wahrscheinlichkeitsverteilung:

Jedem Ausfall s_i wird eine Wahrscheinlichkeit zugeordnet: $P(x_i) = p_i$.

Definition 6.3 Verteilfunktion

Summe der Wahrscheinlichkeiten von links $P(X \leq x_i) = \sum_{x_i}^{k} P(X = x_i)$ Synonym: **Summenfunktion**
Synonym: **Wahrscheinlichkeitsverteilung kumulativ**

Beispiel 6.2 Wahrscheinlichkeitsverteilungen: Münze **M4E6X3**

Versuch: Zwei unabhängige Münzwürfe. Resultat **geordnet** notiert.

a) Erstellen Sie eine Tabelle mit dem Stichprobenraum, der Wahrscheinlichkeitsverteilung und Verteilfunktion des Stichprobenraums.

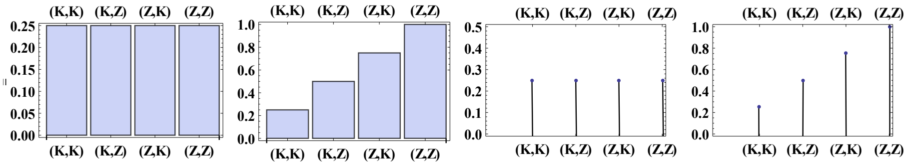
b) Visualisieren Sie für die Beispiele Wahrscheinlichkeitsverteilung und die Verteilfunktion.

Lösung:

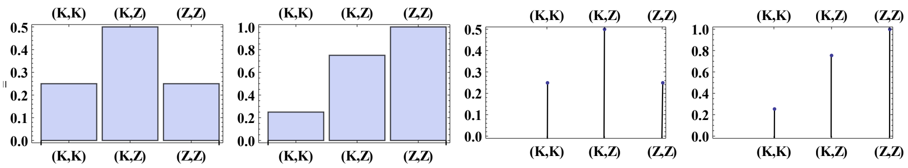
Stichprobenraum	(K,K)	(K,Z)	(Z,K)	(Z,Z)
Wahrscheinlichkeitv.	0.25	0.25	0.25	0.25
Verteilf.	0.25	0.5	0.75	1

Visualisierung

Geordnet:



Ungeordnet:



6.2 Erwartungswert und Varianz

Beispiel 6.3 Tetraheder **UA4CU3**

Wir wollen berechnen, wieviel Gewinn wir durchschnittlich bei einem Wurf mit dem Tetraeder erwarten können. Gehen Sie dafür wie folgt vor:

a) Wir betrachten zunächst $N = 100$ Würfe. Wieviel Mal dürfen wir erwarten, dass wir eine 1, 2, 3 oder eine 4 würfeln?

b) Bei $N = 100$ Würfen: Wieviel Geld können wir insgesamt erwarten? Wieviel Einnahmen erwarten wir also pro Wurf?

c) Wie ändern sich die erwarteten Einnahmen, bei $N = 25$, $N = 200$ und $N = 1000$ Würfen?

d) Berechnen Sie nun für N Würfe die erwarteten Einnahmen pro Wurf, rechnen Sie allgemein mit den Wahrscheinlichkeiten p_1, p_2, p_3, p_4 und dem Gewinn x_1, x_2, x_3, x_4.

Lösung:

a) Tabelle

b) Wir erwarten CHF 2520, also pro Wurf CHF 25.2.

c) Die erwarteten Einnahmen pro Wurf ändern sich nicht. Wir erhalten immer CHF 25.2.

d) $x_{\text{tot}} = p_1 \cdot N \cdot x_1 + p_2 \cdot N \cdot x_2 + p_3 \cdot N \cdot x_3 + p_4 \cdot N \cdot x_4$ also pro Wurf

$$\bar{x} = \frac{x_{\text{tot}}}{N} = p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 + p_4 \cdot x_4$$

Seite	1	2	3	4	total.
p_i	0.26	0.23	0.24	0.27	1
Anzahl erwartet	26	23	24	27	100
Gewinn (CHF)	10	20	30	40	
Gewinn total (CHF)	260	460	720	1080	2520

Definition 6.4 Erwartungswert

$$\mu = E(X) = \sum_{i=1}^{n} p_i \cdot x_i$$

Definition 6.5 Varianz

$$\sigma^2 = Var(X) = E((X-\mu)^2) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2$$

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

Beispiel 6.4 Erwartungswert Münze **S6VG73**

Max wirft einen Würfel und gewinnt Geld, falls er mehr als 3 würfelt.

a) Wieviel Geld gewinnt er pro Wurf durchschnittlich?

b) Varianz des Gewinns?

Resultat	1	2	3	4	5	6
Gewinn (CHF)	0	0	0	4	5	6

Lösung:

a) Erwartungswert

$$\mu = E(X) = \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 2.5$$

b) Varianz

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^{n} p_i (x_i - \mu)^2 = (-2.5)^2 \cdot \frac{1}{6} + (-2.5)^2 \cdot \frac{1}{6} + (-2.5)^2 \\
& +1.5^2 \cdot \frac{1}{6} + 2.5^2 \cdot \frac{1}{6} + 3.5^2 \cdot \frac{1}{6} \\
&\approx 6.58
\end{aligned}$$

Beispiel 6.5 Beweis Varianz **1C34Y5**

Zeigen Sie dass gilt

$$\sum_{i=1}^{n} p_i (x_i - \mu)^2 = -\mu^2 + \sum_{i=1}^{n} p_i (x_i)^2$$

a) Ausmultiplizieren

b) Definition μ verwenden

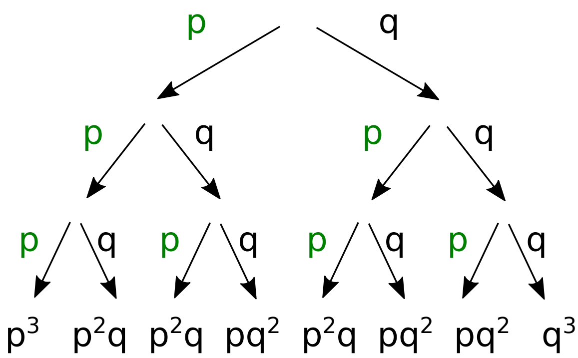
Lösung:

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^{n} p_i \left[(x_i)^2 - 2x_i\mu + \mu^2 \right] = \sum_{i=1}^{n} p_i (x_i)^2 - 2x_i p_i \mu + p_i \mu^2 \\
&= \sum_{i=1}^{n} p_i (x_i)^2 - 2\mu \sum_{i=1}^{n} x_i \cdot p_i + \mu^2 \sum_{i=1}^{n} p_i \\
&= \sum_{i=1}^{n} p_i (x_i)^2 - 2\mu\mu + \mu^2 \cdot 1
\end{aligned}$$

Infobox 6.1 Praktische Berechnung der Varianz

$$\sigma^2 = -\mu^2 + \sum_{i=1}^{n} p_i (x_i)^2$$

6.3 Binomialverteilung



Definition 6.6 Binomialverteilung

Bei mehrfacher Ausführung (n Mal) eines Zufallsexperiments mit der Gewinnchance p ist die Wahrscheinlichkeit für x Gewinne $P(X = x)$ mit

$$P(X = x) = \binom{n}{x} p^x \cdot (1 - p)^{n-x}$$

Oft auch $q := (1 - p)$

Satz 6.1 Binomialverteilung: Erwartungswert und Varianz

$$\mu = n \cdot p$$
$$\sigma^2 = n \cdot p \cdot q$$

Beispiel 6.6 Anwendung Erwartungswert und Varianz Binomialverteilung GM6KIF

Berechnen Sie Erwartungswert und Varianz für die Anzahl gewürfelter Fünfen bei 10 Würfen mit einem Würfel. **Lösung:**

$$\mu = 10 \cdot 1/6 = 1.667$$
$$\sigma^2 = 3 \cdot 0.25 = 1.3889$$

Beispiel 6.7 Kuboktaeder **2I8Z7X**

Der Kuboktaeder fällt mit einer Wahrscheinlichkeit von $p = 0.73$ auf eine quadratische Fläche (□), sonst auf eine dreieckige Fläche. Wir betrachten $n = 12$ Würfe mit dem Kuboktaeder.

a) Berechnen Sie die Wahrscheinlichkeit, dass genau 5 mal □ gewürfel wird.

b) Berechnen Sie die Wahrscheinlichkeit, dass genau 5 mal ein Dreieck gewürfelt wird.

c) Berechnen Sie die Wahrscheinlichkeit, dass mindestens 5 mal □ gewürfel wird.

d) Berechnen Sie die Wahrscheinlichkeit, dass mindestens 5 mal ein Dreieck gewürfel wird.

e) Jemand wettet, dass er bei 12 Würfen mit dem Kuboktaeder genau 4 mal $\square$ erziele. Wie gross ist die Gewinnwahrscheinlichkeit?

f) Was müsst er tippen um die höchsten Gewinnchancen zu erzielen?

Lösung:
Wir betrachtetn die Gewinnwahrscheinlichkeit für die Anzahl Zahl gemäss der Binomialverteilung:

x_i	0	1	2	3	4	5	6
p_i	$1.5 \cdot^{-7}$	$4.87 \cdot 10^{-6}$	0.0000724	0.000653	0.00397	0.0172	0.0542

x_i	7	8	9	10	11	12
p_i	0.126	0.212	0.255	0.207	0.102	0.0229

a) $P(X = 5) = 0.0171746$

b) $P(X = 7) = 0.125546$

c) $P(X \geq 5) = 0.9953$ (wir zählen die Felder in der Tabelle zusammen)

d) $P(X \leq 7) = 0.201595$

e) $P(X = 4) = 0.00397015$

f) Die höchsten Gewinnchancen ist bei $P(X = 9)$ und beträgt 0.255. Alternative, wir rechnen mit dem Erwartungswert:

$$\mu = 8.76 \text{ also } P(X = 9) = 0.225 \ .$$

Beispiel 6.8 Glücksrad (Erwartungswert, Varianz) **2Y8ZUF**

Berechnen Sie Erwartungswert und Varianz für die Anzahl (total) der richtigen Tips beim Glückrad für $n = 10, \ \ldots, 100$ Spiele. Trefferwahrscheinlichkeit $p = 1/3$.

Lösung:

n	10	20	50	100
μ	3.33	6.67	16.67	33.33
σ^2	2.1	4.2	10.5	21

Herleitung μ und σ Binomialverteilung

Beispiel 6.9 Wichtige Ausdrücke **P1XQ8T**

Vereinfache die Ausdrücke für die Binomialverteilung $P(X = x)$

$$\begin{aligned}
P(X=x) &= ? \\
\sum_{x=1}^{n} x \cdot P(X=x) &= ? \\
\sum_{x=1}^{n} (x)^2\, P(X=x) &= ? \\
q = 1-p \;\Rightarrow p+q &= ? \\
g(t) = \sum_{k=1}^{n} 4k \cdot t^k \;\Rightarrow\; g'(t) &= ? \\
\sum_{x=0}^{n} \binom{n}{x} a^{n-x} \cdot b^x &= ?
\end{aligned}$$

Lösung:

$$\begin{aligned}
P(X=x) &= \binom{n}{x} q^{n-x} \cdot p^x \\
\mu &= \sum_{x=1}^{n} x \cdot P(X=x) \\
\sum_{x=1}^{n} (x)^2\, P(X=x) &= \sigma^2 + \mu^2 \\
p+q &= 1 \\
g'(t) &= \sum_{k=1}^{n} 4k^2 \cdot t^{k-1} \\
\sum_{x=0}^{n} \binom{n}{x} a^{n-x} \cdot b^x &= (a+b)^n
\end{aligned}$$

Beispiel 6.10 Herleitung Erwartungswert μ bei der Binomialverteilung 7NKW-LE

$$\begin{aligned}
f(t) = (q + p \cdot t)^n &= \sum_{x=0}^{n} \binom{n}{x} q^{n-x} \cdot p^x \cdot t^x \\
f'(t) &= \\
f'(1) &=
\end{aligned}$$

Ableitung zuerst links $(q + p \cdot t)^n$, dann rechts $(\sum_{x=0}^{n} \ldots)$ durchführen.

Lösung:

$$
\begin{aligned}
f(t) = (q + p \cdot t)^n &= \sum_{x=0}^{n} \binom{n}{x} q^{n-x} \cdot p^x \cdot t^x \\
f'(t) = n \cdot (q + p \cdot t)^{n-1} \cdot p &= \sum_{x=0}^{n} \binom{n}{x} q^{n-x} \cdot p^x \cdot x \cdot t^{x-1} \\
f'(1) = n \cdot (q + p)^{n-1} \cdot p &= \sum_{x=0}^{n} \binom{n}{x} q^{n-x} \cdot p^x \cdot x \\
n \cdot \underbrace{(q + p)^{n-1}}_{=1} \cdot p &= \underbrace{\sum_{x=0}^{n} P(X = x) \cdot x}_{=\mu} \\
\mu &= n \cdot p
\end{aligned}
$$

Beispiel 6.11 Herleitung Varianz σ^2 bei der Binomialverteilung **8IVN74**

$$
f(t) = (q + p \cdot t)^n = \sum_{x=0}^{n} \binom{n}{x} q^{n-x} \cdot p^x \cdot t^x
$$

$$
\begin{aligned}
f''(t) &= \\
f''(1) &=
\end{aligned}
$$

Ableitung zuerst links $(q + p \cdot t)^n$, dann rechts $(\sum_{x=0}^{n} \ldots)$ durchführen.

Lösung:

$$
f''(t) = n \cdot (n-1) \cdot (q + p \cdot t)^{n-1} \cdot p^2 = \sum_{x=0}^{n} \binom{n}{x} q^{n-x} \cdot p^x \cdot x \cdot (x-1) \cdot t^{x-2}
$$

$$
\begin{aligned}
f''(1) = n \cdot (n-1) \cdot p^2 &= \sum_{x=0}^{n} \binom{n}{x} q^{n-x} \cdot p^x \cdot x \cdot (x-1) \\
\underbrace{n^2 \cdot p^2}_{=\mu^2} - \underbrace{n \cdot p^2}_{=\mu \cdot p} &= \underbrace{\sum_{x=0}^{n} P(X = x) \cdot x^2}_{=\underbrace{\sigma^2 + \mu^2}} - \underbrace{\sum_{x=0}^{n} P(X = x) \cdot x}_{=\mu} \\
\mu^2 - \mu \cdot p &= \sigma^2 + \mu^2 - \mu \\
\underbrace{-\mu \cdot p + \mu}_{\mu \cdot (-p+1)} &= \sigma^2
\end{aligned}
$$

6.4 Simulation der Binomialverteilung

In den folgenden Beispielen wird gezeigt, wie in Simulationen experimentelle Wahrscheinlichkeiten bestimmt werden. Dabei wird gleichzeitig die praktische Konsequenz der obigen Gesetze für die Binomialverteilung gezeigt.

Beispiel 6.12 Simulation Glücksrad **84J2YS**

Führen Sie mit `matlab` folgenden Versuch durch.

a) Am Glücksrad werden 10 zufällige Zahlen zwischen 1 und 3 erzeugt. Ein Hellseher gibt vor der Ziehung jeweils seinen Tip ab. Zählen Sie die richtigen Tips.
Befehle: `1:10, randi([1 5],1,6), sum`

b) Schreiben Sie eine Häufigkeitsverteilung. Die Position in der Liste sagt, wieviele richtige Tipps abgegeben wurden. Z.B bedeutet die Liste

$$(0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)$$

dass ein Versucht durchgeführt wurde und dabei 2 richtige Tipps waren. Z.B.

$$(2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0)$$

bedeutet, dass 3 Versuche gemacht wurden, davon 2 mit 0 richtigen Tipps und 1 mit 3 richtigen Tips.
Führen Sie dann den Versuch 1000 mal durch (jedes Mal Anzahl richtige Tips in Liste hinzufügen!). Erstellen Sie so eine Häufigkeitsverteilung.
Befehle: `for i=1:mmax ... end`

Lösung:

```
n=10
r= randi([1 3],1,n) % Zahlen vom Rad
t= randi([1 3],1,n) % Tips von Hellseher
inx=sum(r==t)       % Anz richtige Tips

n=10 ; mmax=10^3;
sta=(1:n+1)*0; %
for i=1:mmax
 r= randi([1 3],1,n);
 t= randi([1 3],1,n);
 inx=sum(r==t);
 sta(inx+1)=sta(inx+1)+1;
end
```

Beispiel 6.13 Glücksrad: Wahrsch.-verteilung und Summenfunktion **R7B3FX**

- Berechnen Sie die Summenfunktion.
- Notieren Sie die Wahrscheinlichkeitsverteilung und die Summenfunktion in einer Tabelle,
- und stellen Sie die Daten graphisch dar.

Befehle: `sum,cumsum, plot, hold on/off`

Lösung:

```
xi=(1:n+1)-1;    % Legende für  sta: Anzahl der Treffer
```

```
pp=sta/sum(sta); % probabbility distribution from experiment
plot(xi,pp)
ppv=cumsum(pp)
hold on
plot(xi,ppv)
hold off
```

Treffer	0	1	2	3	4	5	6	7	8	9	10
W.verteil.	0.014	0.076	0.210	0.265	0.216	0.141	0.054	0.021	0.003	0	0
Summenf.	0.014	0.090	0.300	0.565	0.781	0.922	0.976	0.997	1.0	1.0	1.0

Beispiel 6.14 Glücksrad: $\mu(n)$ und $\sigma^2(n)$ **3XDGE6**

a) Benutzen Sie das Model für das Glücksrad. Berechnen Sie den Erwartungswert der Anzahl Treffer und die Varianz.
Befehle: `*` `'`

b) Ändern Sie nun das Spiel: Am Glücksrad werden `n` zufällige Zahlen zwischen 1 und 3 erzeugt (mit n=10,20,50,100). Machen Sie wieder 1000 Durchläufe und notieren Sie Erwartungswert und Varianz der Treffer.

c) Stellen Sie die Tabelle graphisch dar.

d) Formulieren Sie eine Vermutung für $\mu(n)$ und $\sigma^2(n)$.

Lösung:

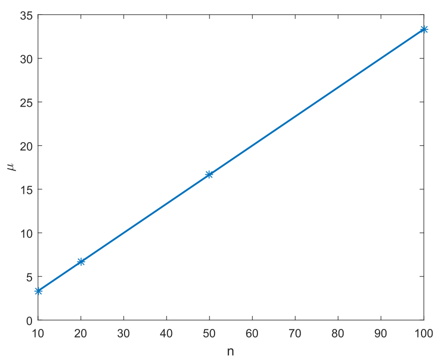
a) Erwartungswert der Anzahl Treffer und die Varianz

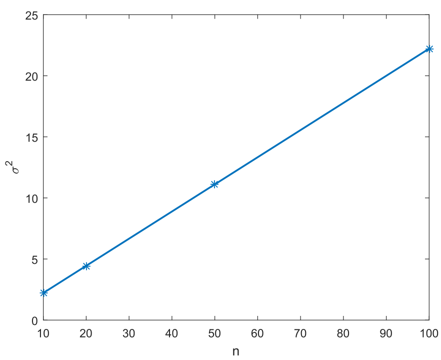
```
mu=xi*pp'              % mean
%mu =       3.3120
 sig=(xi-mu).^2*pp' % varianz
%sig  =      2.2707
```

b) 1000 Durchläufe

n	10	20	50	100
μ	3.3120	6.7150	16.5310	33.3400
σ^2	2.2707	4.3518	11.2490	23.9944

c) Graphische Darstellung





d) Vermutung

$$\mu(n) \propto n \text{ und } \sigma^2(n) \propto n$$

Beispiel 6.15 Glücksrad $\mu(p)$ und $\sigma^2(p)$ **ERRPHC**

a) Ändern Sie nun das Spiel: Am Glücksrad werden 10 zufällige Zahlen zwischen 1 und r erzeugt (mit *r=2,3,5,10,20*). Machen Sie wieder 1000 Durchläufe und notieren Sie Erwartungswert und Varianz der Treffer.

b) Stellen Sie die Tabelle grafisch dar

c) Formulieren Sie eine Vermutung für die Abhängigkeit von $\mu(p)$ von p und für die Abhängigkeit von $\sigma^2(p)$ von

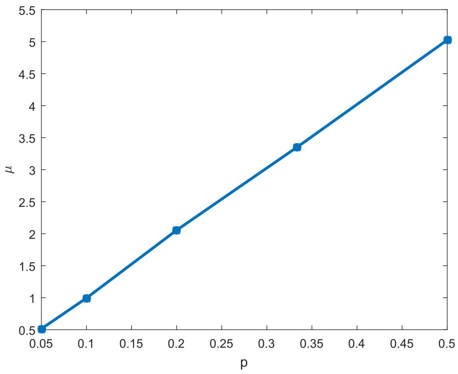
$$p \cdot (1-p) \; .$$

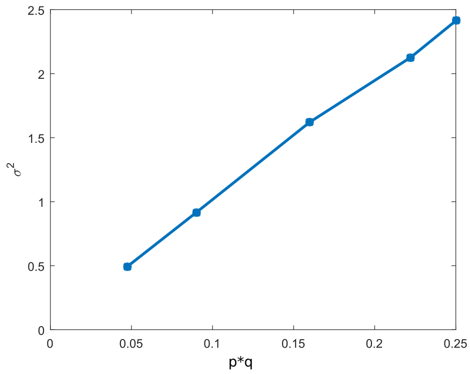
Lösung:

a) Erwartungswert und Varianz der Treffer

r	2	3	5	10	20
μ	2.4133	2.1275	1.6220	0.9160	0.4938
σ^2	5.0260	3.3500	2.0550	0.9940	0.5140
p	0.5	0.33	0.2	0.1	0.02

b) Graphische Darstellung der Tabelle





c) Vermutungen:

$$\mu(n,p) \propto n \cdot p \text{ und } \sigma^2(n,p) \propto n \cdot p \cdot \underbrace{(1-p)}_{q}$$

Beispiel 6.16 Glücksrad (Erwartungswert, Varianz) **LH6GTH**

Berechnen Sie die Wahrscheinlichkeitsverteilung und die Summenfunktion für das Glücksrad ($p = 1/3$, $n = 10$) und vergleichen Sie mit den Resultaten aus der Simulation.

Befehle: cumsum, nchoosek(n,x) .* ./ .^

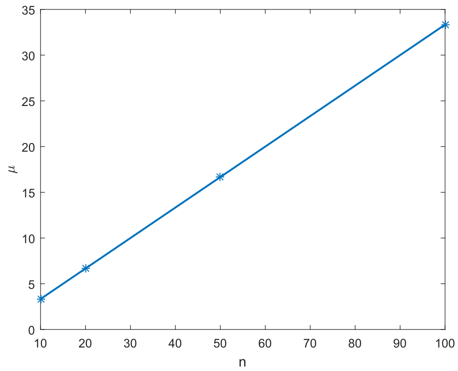
Lösung:
Simulation:

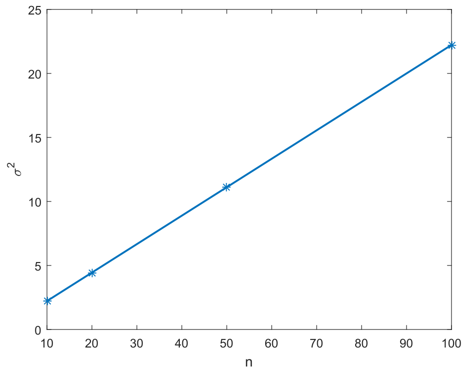
Treffer	0	1	2	3	4	5	6	7	8	9	10
W.verteil.	0.014	0.076	0.210	0.265	0.216	0.141	0.054	0.021	0.003	0	0
Summenf.	0.014	0.090	0.300	0.565	0.781	0.922	0.976	0.997	1.0	1.0	1.0

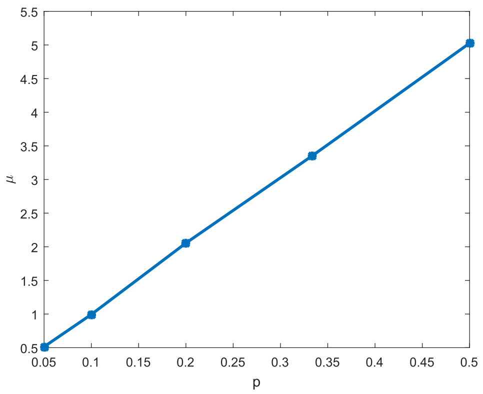
Theorie:

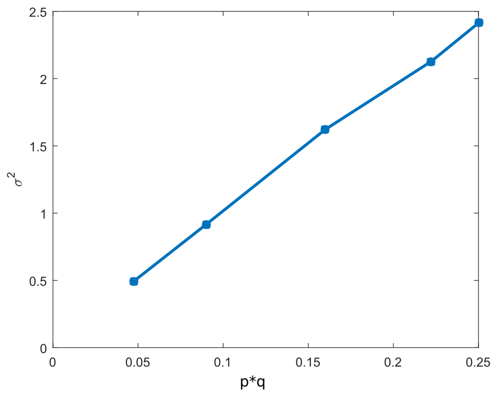
Treffer	0	1	2	3	4	5	6	7	8
W.verteil.	0.0173	0.0867	0.1951	0.2601	0.2276	0.1366	0.0569	0.0163	0.0030
Summenf.	0.0173	0.1040	0.2991	0.5593	0.7869	0.9234	0.9803	0.9966	0.9996

Zusammenstellung der Resultate









Beispiel 6.17 Glücksrad (Erwartungswert, Varianz) **2Y8ZUF**

n	10	20	50	100
μ	3.33	6.67	16.67	33.33
σ^2	2.1	4.2	10.5	21

Berechnen Sie Erwartungswert und Varianz das Glücksrad ($p = 1/3$) für ihre Simulation. Zum Vergleich finden Sie die Angaben der Binomialverteilung in der Tabelle.

Lösung:
Simulation:

n	10	20	50	100
μ	3.31	6.71	16.53	33.34
σ^2	2.27	4.35	11.25	23.99

KAPITEL 7

Wahrscheinlichkeitsverteilungen

Lernziele 7.1 Wahrscheinlichkeitsverteilungen

- Die Studierenden kennen die Binomial- und die Poissonverteilung. Mit Hilfe von Excel oder Matlab können sie Wahrscheinlichkeiten für die Verteilungen angeben.
- Die Studierenden kennen Wahrscheinlichkeitsverteilungen und Summenfunktionen. Sie können Summenfunktionen benutzen um Wahrscheinlichkeiten für Intervalle ($P(x > 5)$) zu berechnen.
- Die Studierenden können die Bionomialverteilung mit der Poissonverteilung nähern und kennen die Bedingung, dass die Näherung sinnvoll ist ($n \geq 100$ und $p < 0.1$).
- Die Studierenden kennen stetige Verteilfunktionen. Sie können normierbare Funktionen normieren und so stetige Verteilfunktionen generieren.
- Die Studierenden können anhand von stetigen Verteilfunktionen ϕ und deren Summenfunktion Φ Wahrscheinlichkeiten auf Intervallen berechnen.
- Die Studierenden kennen die Normalverteilung. Mit Hilfe von Tabellen, Excel und Matlab können sie Wahrscheinlichkeiten und Quantile für die Verteilungen angeben.
- Die Studierenden können Angaben zwischen Realraum x und z-Raum (Argument der standardisierten Normalverteilung) in beide Richtungen transformieren.
$$z = \frac{x - \mu}{\sigma}$$
- Die Studierenden kennen den Unterschied zwischen einseitigen und zweiseitigen Quantilen. Sie können über $p = 1-\alpha$ einseitige und über $p = 1-\alpha/2$ zweiseitige Quantile in Tafeln auslesen.
- Die Studierenden können die Bionomialverteilung mit der Normalverteilung nähern und kennen die Bedingung, dass die Näherung sinnvoll ist ($\sigma^2 > 9$, Moivre und Laplace).

Beispiel 7.1 Glücksrad, Binomial-Verteilung **84J2YS**

Bei einem Glücksrad (3 unterscheidbare Felder am Rad) errät jemand von bei 10 mal Drehen 7 Mal den Ausfall richtig.
Binomialverteilung mit $n = 10$, $p = 1/3$

Anz. Richtige	0	1	2	3	4	5	6	7	8	9	10
p_i	0.0173	0.0867	0.1951	0.2601	0.2276	0.1366	0.0569	0.0163	0.0030	0.0003	0.0000
$\sum p_i$	0.0173	0.1040	0.2991	0.5593	0.7869	0.9234	0.9803	0.9966	0.9996	1.0000	1.0000

a) Was ist die Wahrscheinlichkeit genau 7 Resultate richtig vorauszusagen?

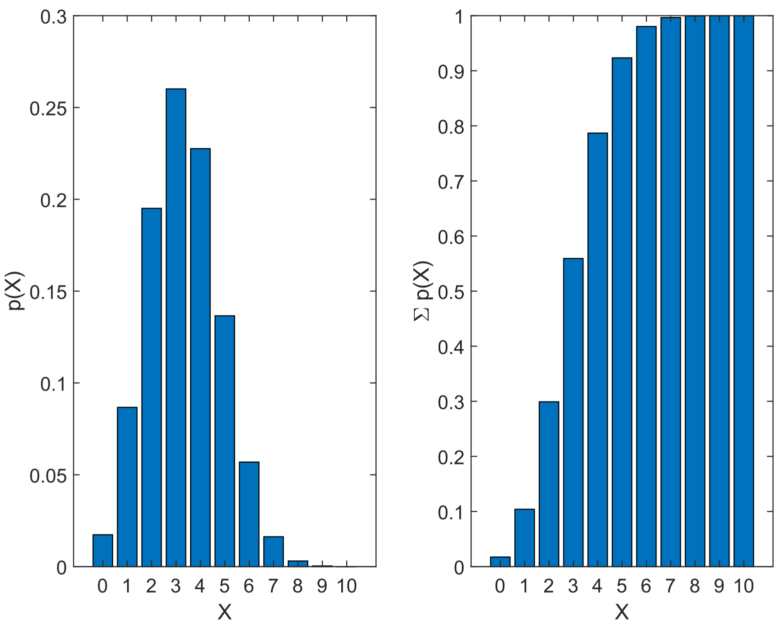
b) Was ist die Wahrscheinlichkeit genau 5 Resultate richtig vorauszusagen?

c) Was ist die Wahrscheinlichkeit höchstens 5 Resultate richtig vorauszusagen?

d) Was ist die Wahrscheinlichkeit mindestens 7 Resultate richtig vorauszusagen?

e) Wieviele richtige Voraussagen erwarten wir? (Mittelwert und Standard-Abweichung)

Lösung:



Die Wahrscheinlichkeitsverteilung und die Summenfunktion der Binomialverteilung mit $n = 10$, $p = 1/3$.

a) $P(X = 7) = 0.0163$

b) $P(X = 5) = 0.1366$

c) $P(X \leq 5) = 0.9234$

d) $P(X \geq 7) = 1 - \sum P(X \leq 6) = 0.0197$

e) $\mu = n \cdot p = 3.33$, $\sigma^2 = n \cdot p \cdot (1-p) = 2.22$, $\sigma = 1.50$

7.1 Näherung der Binomialverteilung mit der Poissonverteilung

Infobox 7.1 Anwendung der Poisson-Verteilung

Für $n \geq 100$ und $p \leq 0.1$ kann statt der Binomial-Verteilung auch die Poisson-Verteilung benutzt werden.
Für die Poisson-Verteilung gilt

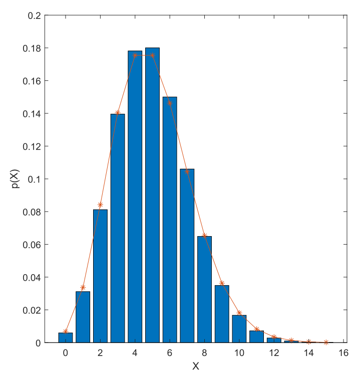
$$\mu = n \cdot p \text{ und } \sigma^2 = n \cdot p$$

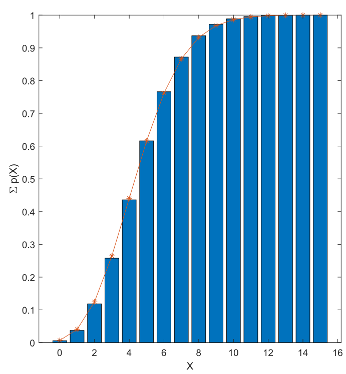
Beispiel 7.2 Glücksrad, Poisson-Verteilung **R7B3FX**

Bei einem Glücksrad (20 unterscheidbare Felder am Rad) errät jemand von bei 100 mal Drehen 11 Mal den Ausfall richtig.
Wir nähern mit der Poisson-Verteilung mit $n = 100$ und $p = 1/20$.

Anz. Richtige	0	1	2	3	4	5	6
p_i	0.0067	0.0337	0.0842	0.1404	0.1755	0.1755	0.1462
$\sum p_i$	0.0067	0.0404	0.1247	0.2650	0.4405	0.6160	0.7622

Anz. Richtige	7	8	9	10	11	12
p_i	0.1044	0.0653	0.0363	0.0181	0.0082	0.0034
$\sum p_i$	0.8666	0.9319	0.9682	0.9863	0.9945	0.9980





a) Was ist die Wahrscheinlichkeit genau 11 Resultate richtig vorauszusagen?

b) Was ist die Wahrscheinlichkeit genau 5 Resultate richtig vorauszusagen?

c) Was ist die Wahrscheinlichkeit höchstens 5 Resultate richtig vorauszusagen?

d) Was ist die Wahrscheinlichkeit mindestens 11 Resultate richtig vorauszusagen?

e) Wieviele richtige Voraussagen erwarten wir? (Mittelwert und Standard-Abweichung)

Lösung:
Graphik oben: Vergleich der exakten Rechnung (blaue Balken für die Binomial-Verteilung) mit der Poissonverteilung (rote Linie). Wahrscheinlichkeitsverteilung und die Summenfunktion mit $n = 100$, $p = 1/20$.

a) $P(X = 11) = 0.0082$, exakt mit der Binomial-Verteilung

$$P(X = 11) = 0.0072$$

b) $P(X = 5) = 0.1755$, exakt mit der Binomial-Verteilung

$$P(X = 5) = 0.1800$$

c) $P(X \leq 5) = 0.6160$, exakt mit der Binomial-Verteilung

$$P(X \leq 5) = 0.6160$$

d) $P(X \geq 11) = 1 - \sum P(X \leq 10) = 0.0137$, exakt mit der Binomial-Verteilung

$$P(X \geq 11) = 1 - \sum P(X \leq 10) = 0.0115$$

e) $\mu = n \cdot p = 5$, $\sigma^2 = n \cdot p = 5$, $\sigma = 2.24$

Beispiel 7.3 Bremsen für LKW (mit Excel, poisson_lkw_aufg.xlc) **8FB1HI**

Eine Firma stellt Bremsen für LKWs in grosser Zahl her. Wir wissen, dass im Mittel 0.5% Fehler aufweisen. Wie gross ist die Wahrscheinlichkeit, dass die Lieferung von 100 Bremsen genau 2 fehlerhafte Stücke enthält?

a) Berechnen Sie die Wahrscheinlichkeit mit der Binomial-Verteilung (Spalten G-H in poisson_lkw_aufg).

b) Berechnen Sie die Wahrscheinlichkeit mit der Poisson-Verteilung. Benutzen Sie dazu ausnahmsweise

$$P(X = x) = \frac{\mu^x}{x!} \cdot e^{-\mu} \text{ mit } \mu = n \cdot p$$

und n die Grösse der Stichprobe, p die Trefferwahrscheinlichkeit. (Spalten O-P in poisson_lkw_aufg).
In den anderen Aufgaben werden wir in Excel stets POISSON.VERT für die Poisson-Verteilung benutzen. Sie ist numerisch robuster, z.B. bei grossen x.

c) Berechnen Sie die Wahrscheinlichkeitsverteilung gemäss Binomial-Verteilung (Spalten G-H) und Poisson-Verteilung (Spalten O-P) für 0 bis 100 fehlerhafte Stücke. Visualisieren Sie die beiden Verteilungen und vergleichen Sie die Graphen. Was stellen Sie fest?

d) Berechnen Sie für die Binomial-Verteilung Erwartungswert und Varianz über

$$\mathrm{E}(X) = \sum_{x=0}^{n} x \cdot p(X = x) \text{ und } \mathrm{Var}(X) = \sum_{x=0}^{n} (x - \mu)^2 \cdot p(X = x)$$

(Spalten I-J). Übertragen Sie die Rechnung auf die Poisson-Verteilung (Spalten Q-R). Vergleichen Sie mit den Ausdrücken $\mu = n \cdot p$ und $\sigma^2 = n \cdot p \cdot q$. Wo sind die Unterschiede, wo die Gemeinsamkeiten der Verteilungen?

e) Formulieren Sie eine Hypothese für E(X) und Var(X) für die Poisson-Verteilung.

f) Überprüfen Sie, ob die Kriterien für die Anwendung der Poisson-Verteilung erfüllt sind:
Für $p \leq 0.1$ und $n \geq 100$ kann statt der Binomialverteilung auch die Poissonverteilung verwendet werden.

Lösung:

a) Wir lesen aus $P(X = 2) = 0.3044$

b) Wir lesen aus $P(X = 2) = 0.3032$

c) Visualisierung unten. Die Graphen lassen sich von Auge nicht unterscheiden

d) Wir erhalten mit der Bionomial-Verteilung

$$\mu = 0.5 \text{ und } \sigma^2 = 0.4975$$

Mit der Poisson-Verteilung ergibt sich

$$\mu = 0.5 \text{ und } \sigma^2 = 0.5$$

Mit den Ausdrücken $\mu = n \cdot p$ und $\sigma^2 = n \cdot p \cdot q$ ergibt sich hingegen

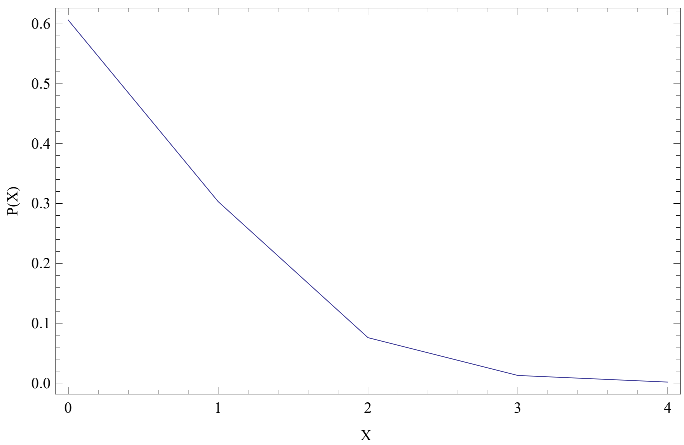
$$\mu = 0.5 \text{ und } \sigma^2 = 0.4975$$

Wir stellen fest, dass sich ein kleiner Unterschied zwischen den Verteilungen bei der Streuung ergibt, nicht aber beim Mittelwert.

e) $\mathrm{E(X)}$ und $\mathrm{Var(X)}$ für die Poisson-Verteilung:

$$\mu = n \cdot p \text{ und } \sigma^2 = n \cdot p$$

f) Wir finden $0.005 \leq 0.1$ und $100 \geq 100$, d.h. die Poissonverteilung kann verwendet werden.



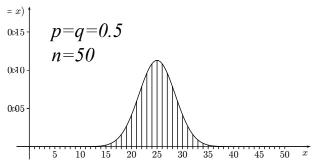
7.2 Näherung der Binomialverteilung mit der Normalverteilung

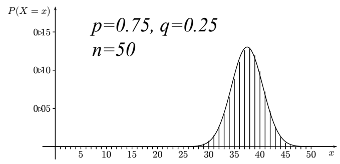
Satz 7.1 Grenzwertsatz von de Moivre und Laplace

Eine binomialverteilte Zufallsgrösse X mit Erwartungswert $E(X) = np$ und Varianz $Var(X) = np(1-p)$ ist näherungsweise normalverteilt mit den Parametern $\mu = n \cdot p$ und $\sigma^2 = n \cdot p \cdot (1-p)$ Annäherung nur zulässig für

$$\sigma^2 = n \cdot p \cdot (1-p) > 9$$

Beispiel 7.4 Näherung der Binomialverteilung





Ist die Näherung durch die Normalverteilung zulässig? **Lösung:**

$$n \cdot p \cdot (1-p) = 12.5 > 9$$

$$n \cdot p \cdot (1-p) = 9.375 > 9$$

Beispiel 7.5 Glückrad, Normalverteilung **3XDGE6**

Bei einem Glücksrad (20 unterscheidbare Felder am Rad) errät jemand von bei 200 mal Drehen 9 Mal den Ausfall richtig.
Wir nähern mit der Normal-Verteilung mit $n = 200$, $p = 1/20$.

Anz. Richtige	0	1	2	3	4	5
p_i	0.0004	0.0012	0.0030	0.0068	0.0142	0.0266
$\sum p_i$	0.0006	0.0018	0.0047	0.0116	0.0258	0.0524

Anz. Richtige	6	7	8	9	10
p_i	0.0448	0.0680	0.0930	0.1146	0.1272
$\sum p_i$	0.0972	0.1652	0.2582	0.3728	0.5000

a) Überprüfen Sie, ob die Anwendung der Normalverteilung gerechtfertigt ist.

b) Wieviele richtige Voraussagen erwarten wir? (Mittelwert und Standard-Abweichung)

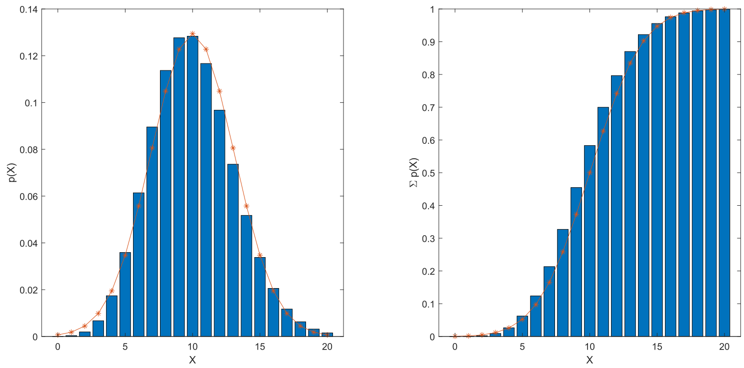
c) Was ist die Wahrscheinlichkeit genau 9 Resultate richtig vorauszusagen?

d) Was ist die Wahrscheinlichkeit genau 7 Resultate richtig vorauszusagen?

e) Was ist die Wahrscheinlichkeit höchstens 7 Resultate richtig vorauszusagen?

f) Was ist die Wahrscheinlichkeit mindestens 9 Resultate richtig vorauszusagen?

Lösung:



Vergleich der exakten Rechnung (blaue Balken für die Binomial-Verteilung) mit der Normalverteilung (rote Linie). Wahrscheinlichkeitsverteilung und die Summenfunktion mit $n = 200$, $p = 1/20$.

a) Gerechtfertigt, denn

$$200 > \frac{9}{1/20 \cdot (1 - 1/20)} \approx 190$$

b) $\mu = 10$, $\sigma^2 = 9.5$, $\sigma = 3.08$

c) $P(X = 9) = 0.1146$

d) $P(X = 7) = 0.068$

e) $P(X \leq 7) = 0.1652$

f) $P(X \geq 9) = 1 - \sum P(X \leq 8) = 0.7418$

Infobox 7.2 Notation Normalverteilung

Wir schreiben $X \sim \mathcal{N}(2, 4)$ um auszudrücken, dass die Zufallsvariable X normalverteilt ist, mit Mittelwert $\mu = 2$ und Varianz $\sigma^2 = 4$.
Allgemein also $X \sim \mathcal{N}(\mu, \sigma^2)$

Infobox 7.3 Wahrscheinlichkeitsverteilungen in Excel

Wir benutzen folgende Ausdrücke:

- x Anzahl Treffer
- n Anzahl Versuche
- p Trefferwahrscheinlichkeit

Für die Binomial und für die Normalverteilung ergibt sich

$$\mu = n \cdot p \text{ und } \sigma^2 = n \cdot p \cdot (1 - p)$$

Für die Poisson-Verteilung

$$\mu = n \cdot p \text{ und } \sigma^2 = n \cdot p$$

In Excel können die Verteilungen wie folgt generiert werden werden:

- Binomialverteilung
  `BINOM.DIST`(x; n;p; `cumulativ`)
- Poisson-Verteilung
  `POISSON.DIST`(x; μ; `cumulativ`)
- Normalverteilung
  `NORM.DIST`(x; μ; σ;`cumulativ`)

Infobox 7.4 Unterschiede der W.verteilungen in Excel

- Beachte, dass die Verteilungen unterschiedliche Input-Parameter benutzen, z.B. n und p bei `BINOM.DIST`(x; n;p; `cumulativ`) aber μ und σ bei der Normalverteilung `NORM.DIST`(x; μ; σ;`cumulativ`).
- Bei der Normalverteilung sind μ und σ manchmal direkt gegeben; ansonsten berechnen wir diese Grössen durch

$$\mu = n \cdot p \text{ und } \sigma^2 = n \cdot p \cdot (1 - p)$$

- Beachte auch, dass in den Aufgabenstellungen mit $X \sim \mathcal{N}(\mu, \sigma^2)$ die Varianz gegeben ist und nicht der Input-Parameter σ.

Beispiel 7.6 Wahrscheinlichkeitsverteilungen in Excel **E13RR8**

Erzeugen Sie folgende Wahrscheinlichkeitsverteilungen in Excel und lesen die gefragten Werte aus. Vergleichen Sie mit den Wahrscheinlichkeitsverteilungen oben.

a) Binomialverteilungen, $n = 10$, $p = 1/3$;

$$P(X = 7),\ P(X = 5),\ P(X \leq 7),\ P(X \geq 5)$$

b) Poissonverteilung $n = 100$, $p = 1/20$

$$P(X = 11),\ P(X = 5),\ P(X \leq 11),\ P(X \geq 5)$$

c) Normalverteilung $n = 200$, $p = 1/20$

$$P(8 \leq X \leq 9),\ P(7 \leq X \leq 8),\ P(X \leq 9),\ P(X \geq 7)$$

Hinweis: Hier wird die Normalverteilung *nicht* als Näherung der Binomalverteilung benutzt, deshalb rechnen wir mit $P(X \geq x) = 1 - P(X \leq x)$

Lösung:

a) Binomialverteilungen, $n = 10$, $p = 1/3$;

$$P(X = 7) = 0.0163,\ P(X = 5) = 0.1366$$

$$P(X \leq 7) = 0.9966,\ P(X \geq 5) = 1 - P(X \geq 4) = 0.2131$$

b) Poissonverteilung $n = 100$, $p = 1/20$

$$P(X = 11) = 0.0082,\ P(X = 5) = 0.1755$$

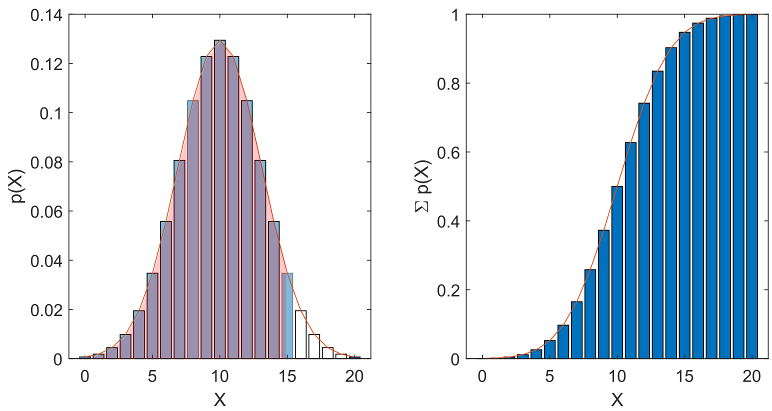
$$P(X \leq 11) = 0.9945,\ P(X \geq 5) = 1 - P(X \geq 4) = 0.5595$$

c) Normalverteilung $n = 200$, $p = 1/20$

$$P(8 \leq X \leq 9) = 0.1146,\ P(7 \leq X \leq 8) = 0.0930$$

$$P(X \leq 9) = 0.3728,\ P(X \geq 7) = 1 - P(X \leq 7) = 0.8348$$

Infobox 7.5 Normalverteilung: Näherung diskret vs. stetige Verteilung



- Bei diskreten Verteilungen ist $P(X \geq x) = 1 - P(X \leq x - 1)$ also z.B.

$$P(X \geq 16) = 1 - P(X \leq 15)$$

- Bei stetigen Verteilungen ist $P(X \geq x) = 1 - P(X \leq x)$ oder auch

$$P(X \geq x) = 1 - P(X < x)$$

also z.B.

$$P(X \geq 16) = 1 - P(X \leq 16)$$

- Wird die Normalverteilung als Näherung für eine diskrete Verteilung benutzt rechnen wir mit $P(X \geq x) = 1 - P(X \leq x - 1)$

Infobox 7.6 Stetigkeit der Normalverteilung

- Die Normalverteilung kann auch für $P(X \leq 9.1)$ ausgelesen werden, oder allgemein für $P(X \leq x)$, wobei $x \in \mathbb{R}$.
- Wir nennen die Normalverteilung deshalb eine *stetige* Verteilung.
- Weil sie so wichtig ist und früher kein Excel zur Verfügung stand, ist $X \sim \mathcal{N}(0, 1^2)$ in Tabellen angegeben.

7.3 Die Normalverteilung (mit der Tabelle)

In Abb. 7.3 **a)** sieht man: Viele Messgrössen (z.B. die Körpergrösse von Menschen) gehorchen der Normalverteilung. Gemäss Bundesamt für Statistik sind (2017) Frauen durchschnittlich $\mu = 164.7$ cm gross mit einer Standard-Abweichung $\sigma = 12$ cm, und Männer $\mu = 177.4$ cm mit einer Standard-Abweichung $\sigma = 15$ cm. Unabhängig davon, was gemessen wird: Bei diesen Verteilungen bleibt immer gleich, dass z.B. sehr viele Menschen eine Körpergrösse zwischen $\mu - \sigma$ und $\mu + \sigma$ haben (es sind 68.2%), aber nur wenige eine Körpergrösse über $\mu + 3\sigma$. Deshalb wird die Normalverteilung standardisiert: Abb. 7.3 **b)** auf der x-Achse wird die Grösse geteilt durch σ. Es entsteht die Variable z ohne physikalische Grösse. Ausserdem wird die Kurve (blau) so skaliert, dass die Gesamtfläche unter dem Graphen 1 ergibt. Wir erhalten eine Wahrscheinlichkeitsverteilung.

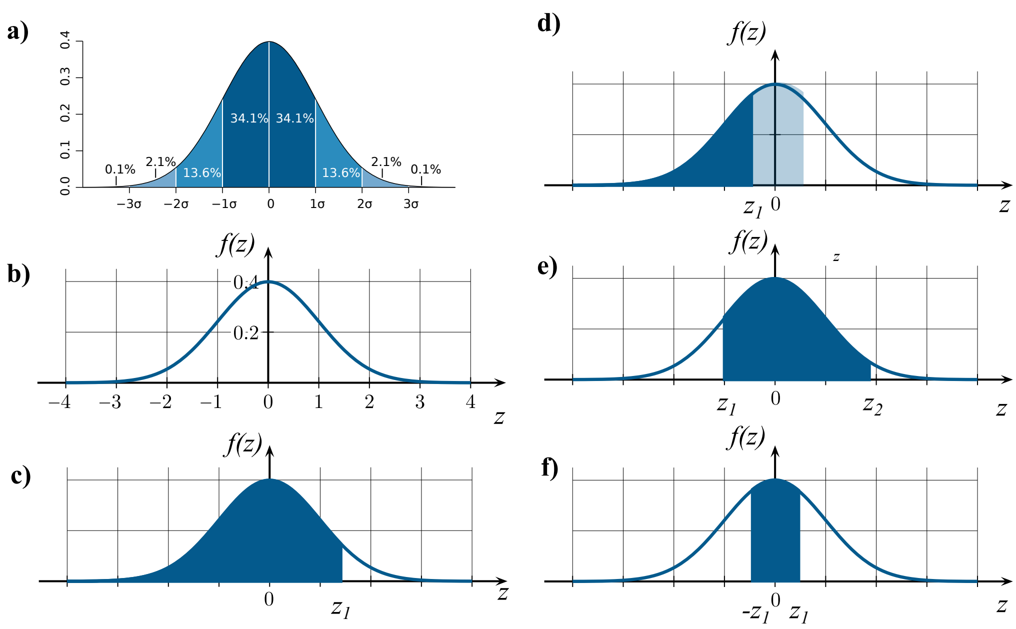


Abbildung 7.1: Darstellungen der Normalverteilung (b-f sind standardisiert)

Definition 7.1 Standardisierte Normalverteilung

Die standardisierte Normalverteilung besitzt eine Dichtefunktion der Form

$$f(z) = \phi(z, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

und ist für $-\infty < z < \infty$ definiert.

Sie wird in Excel mit `NORM.VERT(z;0;1;False)` erzeugt. Wir interessieren uns meistens für Flächen unter dem Graphen von $f(z)$. Die kann man leider *nicht analytisch* berechnen. Man verwendet zur Berechnung Excel mit `NORM.VERT(z;0;1;True)` oder die Tafel T.1 im Skript. Dort sind die Werte tabelliert als Funktion der oberen Grenze z:

$$P(Z \leq z) = \int_{-\infty}^{z} f(t)dt = \Phi(z, 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

Die Stammfunktion von $f(t)$ schreiben wir als $\Phi(z, 0, 1)$. Da es zu jeder Funktion viele Stammfunktionen gibt, legen wir zusätzlich fest $\Phi(0, 0, 1) = \frac{1}{2}$. Abb. 7.3 **c)** zeigt noch einmal, was wir aus der Tafel T.1 auslesen oder mit `NORM.VERT(z;0;1;True)` berechnen: Es ist die Fläche unter dem Graphen zwischen $-\infty$ und z, d.h. die Form des blauen Graphen $f(z)$ ist weniger wichtig.

In der Tabelle T1 ist die Fläche nur für $z_1 > 0$ angegeben. Abb. 7.3 **d)** zeigt wie man auch für $z_1 < 0$, die Fläche berechnen kann:

$$P(Z < z_1) = 1 - \Phi(Z = -z_1)$$

Abb. 7.3 **e)** zeigt, wie man Flächen berechnen kann, wenn eine obere und eine untere Grenze gegeben ist:

$$P(z_1 < Z < z_2) = \Phi(Z = z_2) - \Phi(Z = z_1)$$

Schliesslich zeigt Abb. 7.3 **f)**, wie man die Fläche für symmetrische Bereiche berechnen kann:

$$P(-z_1 < Z < z_1) = \Phi(Z = z_1) - \Phi(Z = -z_1) = \Phi(Z = z_1) - \big(1 - \Phi(Z = z_1)\big) = 2\Phi(Z = z_1) - 1$$

Beispiel 7.7 Normalverteilung: Oben begrenzte Wahrscheinlichkeiten **YZ-PJ14**

Bestimmen Sie die folgenden Wahrscheinlichkeiten einer standardisierten Normalverteilung, d.h. $Z \sim \mathcal{N}(0,1)$.

a) $P(Z \leq 2.34)$

b) $P(Z \leq 3.21)$

Lösung:
Wir lesen aus

a) $P(Z \leq 2.34) = 0.9904$

b) $P(Z \leq 3.21) = 0.9993$

Zweiseitig begrenzte Flächen werden berechnet über

$$P(z_1 \leq Z \leq z_2) = \int_{z_1}^{z_2} f(t)dt = \int_{-\infty}^{z_2} f(t)dt - \int_{-\infty}^{z_1} f(t)dt = \Phi(z_2, 0, 1) - \Phi(z_1, 0, 1)$$

Das Verfahren ist das selbe, wie bei Funktionen, die eine analtische Stammfunktion besitzen.

Beispiel 7.8 Normalverteilung: Oben begrenzte Wahrscheinlichkeiten, negative Grenzen **YZPJ14**

Bestimmen Sie die folgenden Wahrscheinlichkeiten einer standardisierten Normalverteilung, d.h. $Z \sim \mathcal{N}(0,1)$.

a) $P(Z \leq -2.34)$

b) $P(Z \leq -3.21)$

Lösung:
Wir lesen aus

a) $P(Z \leq 2.34) = 1 - 0.9904 = 0.0096$

b) $P(Z \leq 3.21) = 1 - 0.9993 = 0.0007$

Ist eine Grenze negativ, d.h. $-z < 0$, kann mit der Gegenwahrscheinlichkeit gerechnet werden:

$$P(Z \leq -z) = \int_{-\infty}^{-z} f(t)dt = 1 - \int_{-\infty}^{z} f(t)dt = 1 - \Phi(z, 0, 1)$$

Beispiel 7.9 Standardisierte Normalverteilung **046938**

Bestimmen Sie die folgenden Wahrscheinlichkeiten einer standardisierten Normalverteilung, d.h. $Z \sim \mathcal{N}(0,1)$.

a) Ein-σ-Bereich: $P(-1 \leq Z \leq 1)$

b) Zwei-σ-Bereich: $P(-2 \leq Z \leq 2)$

c) Drei-σ-Bereich: $P(-3 \leq Z \leq 3)$

d) $P(|Z| \leq 1)$

e) $P(|Z| \geq \frac{1}{2})$

f) $P(-3 \leq Z \leq 1)$

Lösung:

a)

$$P(-1 \leq Z \leq 1) = \Phi(1,0,1) - \Phi(-1,0,1) = 0.8413 - 0.1587 = 0.6827$$

b)

$$P(-2 \leq Z \leq 2) = \Phi(2,0,1) - \Phi(-2,0,1) = 0.9772 - 0.0228 = 0.9545$$

c)

$$P(-3 \leq Z \leq 3) = \Phi(3,0,1) - \Phi(-3,0,1) = 0.9987 - 0.0013 = 0.9973$$

d)

$$P(-1 \leq Z \leq 1) = \Phi(1,0,1) - \Phi(-1,0,1) = 0.8413 - 0.1587 = 0.6827$$

e)

$$\begin{aligned} P(|Z| \geq \frac{1}{2}) &= P(Z < -\frac{1}{2}) + P(Z > \frac{1}{2}) \\ &= \Phi(-0.5,0,1) - \Phi(-\infty,0,1) + \big[\Phi(\infty,0,1) - \Phi(0.5,0,1)\big] \\ &= 0.308538 - 0 + [1 - 0.6915] = 0.617075 \end{aligned}$$

f)

$$P(-3 \leq Z \leq 1) = \Phi(1,0,1) - \Phi(-3,0,1) = 0.8413 - 0.0013 = 0.8400$$

Infobox 7.7 Standardisierte Normalverteilung

- $\int_{-\infty}^{\infty} f(z) = 1$
- Wir nennen die standartisierte Normalverteilung $\Phi(z, 0, 1)$ auch $\mathcal{N}(0, 1)$ oder

$$\Phi(z, 0, 1) = \frac{1}{2}\text{erf}\left(\frac{z}{\sqrt{2}}\right) + \frac{1}{2}$$

- $\Phi(z, 0, 1)$ ist die Stammfunktion von $\varphi(z, 0, 1)$, deshalb gilt

$$P(z_1 \le Z \le z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{t^2}{2}} dt = \Phi(z_2, 0, 1) - \Phi(z_1, 0, 1)$$

- $\varphi(z, 0, 1)$ hat keine Elementare Stammfunktion, $\Phi(z, 0, 1)$ findet man aufgelistet in Tafel (z.B. T.1)
- für $z_1 < 0$ lese Tafel bei $z > 0$ aus:

$$P(Z \le z_1) = 1 - \Phi(-z_1, 0, 1)$$

7.3.1 Die Normalverteilung mit den Parametern μ und σ

Definition 7.2 Normalverteilung mit den Parametern μ und σ^2

Normalverteilung mit den Parametern μ und σ^2 besitzt die Wahrscheinlichkeitsdichte:

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

und ist für $-\infty < x < \infty$ definiert Die Verteilungsfunktion schreiben wir als $\Phi(x, \mu, \sigma^2)$:

$$P(X \le x) = \Phi(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Infobox 7.8 Normalverteilung

- Maximum von $f(x)$ liegt bei $x = \mu$ und beträgt $f(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}$
- $f(x)$ ist symmetrisch bezüglich dem Maximum
- Die Wendepunkte von $f(x)$ liegen bei $x = \mu \pm \sigma$
- Je grösser σ desto breiter $f(x)$

Beispiel 7.10 Transformation auf die standardisierte Normalverteilung BB-MXII

Führen Sie die Transformation $z = \frac{t-\mu}{\sigma}$ für das Integral aus und drücken Sie den Wert mit der standardisierten Normalverteilung aus:

$$P(x_1 \le X \le x_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x_1}^{x_2} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Lösung:

Wir erhalten $dt = dz \cdot \sigma$ und also

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\frac{x_1-\mu}{\sigma}}^{\frac{x_2-\mu}{\sigma}} e^{\frac{-z^2}{2}} \cdot \sigma \cdot dz \\ &= \Phi(\frac{x_2-\mu}{\sigma}, 0, 1) - \Phi(\frac{x_1-\mu}{\sigma}, 0, 1) \end{aligned}$$

Infobox 7.9 T1 benutzen für die Normalverteilung $X \sim \mathcal{N}(\mu, \sigma^2)$

$$P(x_1 \leq X \leq x_2) = \Phi(\frac{x_2-\mu}{\sigma}, 0, 1) - \Phi(\frac{x_1-\mu}{\sigma}, 0, 1)$$

wobei $\Phi(z, 0, 1)$ die Verteilfunktion der standardisierten Normalverteilung ist

Beispiel 7.11 Normalverteilung **013263**

Berechnen Sie

a) die Wahrscheinlichkeit $P(1 \leq X \leq 2.45)$ wenn $X \sim \mathcal{N}(2, 4)$;

b) die Wahrscheinlichkeit $P(-4 \leq X \leq 8)$ wenn $X \sim \mathcal{N}(2, 4)$;

c) die Wahrscheinlichkeit $P(2 \leq X)$ wenn $X \sim \mathcal{N}(1, 9)$;

d) die Wahrscheinlichkeit $P(|X| \leq 1)$ wenn $X \sim \mathcal{N}(-1, 16)$.

Lösung:

a) $X \sim \mathcal{N}(2, 4)$, also $\mu = 2$ und $\sigma = \sqrt{4} = 2$

$$\begin{aligned} P(1 \leq X \leq 2.45) &= P(\frac{1-2}{2} \leq Z \leq \frac{2.45-2}{2}) \\ &= \Phi(0.225, 0, 1) - \Phi(-0.5, 0, 1) = \Phi(0.225, 0, 1) - [1 - \Phi(0.5, 0, 1)] \\ &= 0.5890 - 0.3085 = 0.2805 \end{aligned}$$

b) $X \sim \mathcal{N}(2, 4)$, also $\mu = 2$ und $\sigma = \sqrt{4} = 2$

$$\begin{aligned} P(-4 \leq X \leq 8) &= P(\frac{-4-2}{2} \leq Z \leq \frac{8-2}{2}) \\ &= \Phi(3, 0, 1) - \Phi(-3, 0, 1) = \Phi(3, 0, 1) - [1 - \Phi(3, 0, 1)] \\ &= 0.9987 - [1 - 0.9987] = 0.9974 \end{aligned}$$

c) $X \sim \mathcal{N}(1, 9)$, also $\mu = 1$ und $\sigma = \sqrt{9} = 3$

$$\begin{aligned} P(2 \leq X) &= P(\frac{2-1}{3} \leq Z \leq \frac{\infty-1}{3}) \\ &= \Phi(\infty, 0, 1) - \Phi(1/3, 0, 1) = 1 - 0.6293 = 0.3707 \end{aligned}$$

Vergleichen wir den Wert der Tabelle (oben) mit dem aus z.B. ($P = 0.369441$)), stellen wir eine Differenz von 0.0013 fest. Diese Differenz entsteht durch die Rundung der Integrationsgrenze. Wir können den Fehler beim Auslesen etwas verkleinern, indem wir die Werte der Tabelle linear interpolieren. Für den Bereich zwischen $z = 0.33$ und $z = 0.34$ gilt

$$\begin{aligned}\Phi(0.33 + dz) &= \Phi(0.33) + \frac{\Phi(0.34) - \Phi(0.33)}{0.01} \cdot dz \\ &= 0.6293 + \frac{0.6331 - 0.6293}{0.01} \cdot dz = 0.6293 + 0.38 \cdot dz\end{aligned}$$

Wir wollen also den Wert für $dz = 0.003\dot{3}$ annähern und erhalten:

$$\Phi(0.33 + 0.00\dot{3}) = 0.6293 + 0.38 \cdot 0.00\dot{3} = 0.630567$$

und also

$$1 - 0.630567 = 0.369433$$

Der Fehler beträgt nur noch $8 \cdot 10^{-6}$

d) $X \sim \mathcal{N}(-1, 16)$, also $\mu = -1$ und $\sigma = \sqrt{16} = 4$

$$\begin{aligned}P(|X| \leq 1) &= P(\frac{-1 - (-1)}{4} \leq Z \leq 8\frac{1 - (-1)}{4}) \\ &= \Phi(1/2, 0, 1) - \Phi(0, 0, 1) = 0.6915 - 0.5 = 0.1915\end{aligned}$$

Infobox 7.10 Für die Normalverteilung gilt

- $\mathrm{E}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu$
- $\mathrm{Var}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$

"Die Parameter der Normalverteilung lassen sich damit leicht deuten: μ ist der Erwartungswert der Zufallsgrösse X und σ^2 die Varianz."

Beispiel 7.12 Stammfunktion und Fläche **UNXL5Z**

Berechnen Sie mit der Stammfunktion $\Phi(x)$ das Integral

$$\int_0^6 \frac{1}{\sqrt{(x/2)^2 + 1}} dx$$

$$\begin{aligned}\varphi(x) &= \frac{1}{\sqrt{x^2 + 1}} \\ \Phi(x) &= \log\left(x + \sqrt{x^2 + 1}\right) = \int \varphi(x) dx\end{aligned}$$

Lösung:

Wir substituieren $z = \frac{x}{2}$, d.h. $dx = 2 \cdot dz$:

$$\begin{aligned}\int_0^6 \frac{1}{\sqrt{(x/2)^2 + 1}} dx &= \int_0^5 \frac{1}{\sqrt{(x/2)^2 + 1}} dx = 2 \int_0^3 \frac{1}{\sqrt{z^2 + 1}} dz \\ &= 2\big[\Phi(x)\big]_0^3 \\ &= 2\big(\Phi(3) - \Phi(0)\big) \approx 3.637\end{aligned}$$

Beispiel 7.13 Körpergrösse (2017) **LZPY8Y**

Frauen messen in der Schweiz durchschnittlich $\mu = 164.7$ cm mit einer Standard-Abweichung $\sigma = 12$ cm, und Männer $\mu = 177.4$ cm mit einer Standard-Abweichung $\sigma = 15$ cm. Anzahl Frauen $N_F = 3\,560\,275$ und Anzahl Männer $N_M = 3\,475\,924$.

a) Wieviel Prozent der Frauen sind 164.7 cm gross oder kleiner?

b) Wieviel Prozent der Männer sind zwischen $177.4 - 15$ cm und $177.4 + 15$ cm gross?

c) Wieviel Prozent der Frauen sind grösser als 172 cm?

d) Wieviel Prozent der Männer sind grösser als 194 cm?

e) Wieviele Frauen sind grösser als 200 cm?

f) Wieviele Männer sind kleiner als 150 cm?

Lösung:

a) $P(X < \mu) = 0.5$

b) $P(\mu - \sigma < X < \mu + \sigma) = 0.68$

c) $P(172 < X) = 1 - P(172 < X) = 0.27$

d) $P(194 < X) = 1 - P(194 < X) = 0.13$

e) $n = N_F * P(X > 200) = 5\,811$

f) $n = N_M * P(X < 150) = 117\,747$

7.3.2 Quantile der Normalverteilung

Infobox 7.11 Rücktransformation

Mit der Transformation

$$x = z \cdot \sigma + \mu$$

können Grenzen der standardisierten Normalverteilung für die Normalverteilung mit Parametern μ und σ "übersetzt" werden.
Die Quantile der standardisierten Normalverteilung finden sich in Tafeln, z.B. T.2.

Beispiel 7.14 Quantile

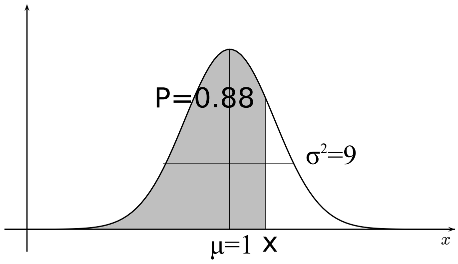
Berechnen Sie die Grenzen x

a) $P(x \geq X) = 0.88$ wenn $X \sim \mathcal{N}(1, 9)$;

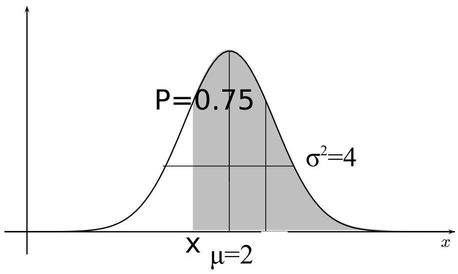
b) $P(X \geq x) = 0.75$ wenn $X \sim \mathcal{N}(2, 4)$;

c) $P = 0.36$, Grenzen symmetrisch um μ wenn $X \sim \mathcal{N}(-1, 16)$.

Lösung:

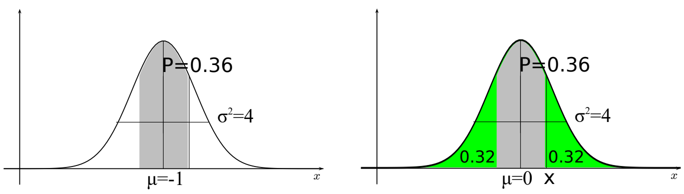
a) 

Wir nehmen zuerst eine Normalverteilung mit $\mu = 0$ und $\sigma^2 = 9$, d.h. die x-Werte sind symmetrisch um 0 verteilt.
In Excel berechnen wir $P(x \geq X) = 0.88 \quad \Rightarrow \quad x' = 3.525$ mit dem Befehl `NORMINV(0.88;0;3)`. Die gesuchte Grenze liegt bei $x = \mu + x' = 4.525$

b) 

$P(X - 2 \leq x) = 0.75$ wenn $X \sim \mathcal{N}(2, 4)$; Wir nehmen zuerst eine Normalverteilung mit $\mu = 0$ und $\sigma^2 = 4$, d.h. die x-Werte sind symmetrisch um 0 verteilt. Da wir eine untere Grenze suchen, arbeiten wir mit der Gegenwahrscheinlichkeit $P = 0.25$. In Excel berechnen wir $P(x' \leq X) = 0.25 \quad \Rightarrow \quad x' = -1.349$, `NORMINV(0.25;0;2)`. Die gesuchte Grenze liegt bei $x = \mu + x' = 2 - 1.349 = 0.651$

c) $P = 0.36$ wenn $X \sim \mathcal{N}(-1, 16)$. Wir betrachten eine Normalverteilung um 0 mit der selben Varianz.



Die Wahrscheinlichkeit, die ausserhalb der gesuchten Grenze liegt ist $P_1 = 1 - 0.36$. Unterhalb der gesuchten unteren Grenze wird der Anteil $P_2 = \frac{P_1}{2} = 0.32$ liegen. Dafür können wir die Grenze berechnen: $P(x \leq X) = 0.32 \quad \Rightarrow \quad x = -1.8708$ =`NORMINV(0.32;0;4)` Wir können die Grenzen auch angeben mit $-1 + x = -2.8708$ und $-1 - x = 0.870795$.

Beispiel 7.15 Normmöbel (2017) **6K3U08**

Frauen messen in der Schweiz durchschnittlich $\mu = 164.7$ cm mit einer Standard-Abweichung $\sigma = 12$ cm, und Männer $\mu = 177.4$ cm mit einer Standard-Abweichung $\sigma = 15$ cm. Anzahl Frauen $N_F = 3\,560\,275$ und Anzahl Männer $N_M = 3\,475\,924$.

a) Eine Türe ist so ausgelegt, dass sich 68% der Frauen nicht bücken müssen. Wie hoch ist sie?

b) Wie hoch ist die Türe, wenn sich 98% der Männer nicht bücken müssen?

c) Ein Sitz (z.B. der SBB) wird so ausgelegt, dass 85% der Frauen darin bequem sitzen können. Wie gross ist die kleinste Frau, die noch bequem sitzt? Wie gross die grösste?

d) Wenn der Sitz für 85% der Männer ausgelegt wäre, wie gross ist der kleinste Mann, der bequem sitzt? Wie gross der grösste?

Lösung:

a) 170.65 cm

b) 212.4 cm

c) Die Frauen sind zwischen 149.6 und 178.6 cm gross.

d) Die Männer sind zwischen 156.7 und 194.1 cm gross.

7.4 Normalverteilung mit Excel

Definition 7.3 Standardisierte Normalverteilung

$$Z \sim \mathcal{N}(0, 1^2)$$

Beispiel 7.16 Normalverteilung Wahrscheinlichkeiten

Berechnen Sie $P(Z \leq 1)$ für $Z \sim \mathcal{N}(0, 1)$ **Lösung:**
Mit `NORM.DIST(`$x = 1$; $\mu = 0$; $\sigma = 1$;`cumulative)` erhalten wir 0.841345

Beispiel 7.17 Normalverteilung Wahrscheinlichkeiten **P1XQ8T**

Berechnen Sie

a) $P(-1 \leq Z \leq 1)$ für $Z \sim \mathcal{N}(0, 1)$

b) $P(Z \leq 0.842)$ für $Z \sim \mathcal{N}(0, 1)$

c) $P(9.1 \leq X \leq 10.783)$ für $X \sim \mathcal{N}(9.1, 4)$

d) $P(X \leq 7.66)$ für $X \sim \mathcal{N}(3.45, 25)$

Lösung:

a) $P(-1 \leq Z \leq 1) = 0.6827$. Da $\sigma^2 = \sigma = 1$, haben wir die Wahrscheinlichkeit berechnet, für $P(\mu - \sigma \leq X \leq \mu + \sigma)$. Es ist praktisch, dass man sich für die Skizzen der Wahrscheinlichkeitsverteilung merkt; dass in diesem Bereich rund 2/3 der Wahrscheinlichkeit liegt bei der Normal-Verteilung.

b) $P(Z \leq 0.842) = 0.80$

c) $P(9.1 \leq X \leq 10.783) = 0.30$

d) $P(X \leq 7.66) = 0.80$

Infobox 7.12 Quantile der Normalverteilung

- Für die Aufgabe $P(X \leq x) = 0.8$ mit $X \sim \mathcal{N}(3.45, 25)$, wo x *obere Grenze gesucht* ist benutzen wir in Excel `NORMINV`(P,μ,σ) hier also `NORMINV`(0.8,3.45,5)
- Beachte, dass NORMINV mit der Standard-Abweichung σ und nicht mit der Varianz σ^2 rechnet.

Beispiel 7.18 Quantile der Normalverteilung **7NKWLE**

Berechnen Sie die Grenze x der Quantile

a) $P(X < x) = 0.9$ bei $X \sim \mathcal{N}(3.45, 25)$

b) $P(X < x) = \frac{2}{3}$ bei $X \sim \mathcal{N}(5, 100)$

c) $P(x < X) = 0.1$ bei $X \sim \mathcal{N}(3.45, 9)$ (untere Grenze gesucht).

d) $P(x < X) = 0.2$ bei $X \sim \mathcal{N}(8, 0.05^2)$ (untere Grenze gesucht).

Lösung:

a) $x = 9.8578$

b) $x = 9.3073$

c) Mit $P(X < x) = 0.9$ erhalten wir $x = 7.29$

d) Mit $P(X < x) = 0.8$ erhalten wir $x = 8.0421$

Oft wird eine zweiseitig begrenztes Intervall gesucht, das heisst zum Beispiel $P(x' \leq X \leq x) = 0.8$ mit $X \sim \mathcal{N}(3.45, 25)$ wobei x und x' symmetrisch um den Erwartungswert $\mu = 3.45$ liegen sollen. Für diesen Fall rufen wir uns in Erinnerung, dass

$$P(X \leq x') = 0.1$$

und wir also mit dem einseitig begrenzten Intervall

$$P(X \leq x) = 0.8 + 0.1$$

rechnen können.
Allgemein:

Infobox 7.13 Quantile symmetrisch

Für das symmetrische (zweiseitige Quantil) $P(x' \leq X \leq x) = 1 - \alpha$ und $X \sim \mathcal{N}(\mu, \sigma^2)$ rechnen wir mit dem einseitig begrenzten Quantil

$$P(X \leq x) = 1 - \frac{\alpha}{2}$$

Beispiel 7.19 Symmetrische Quantile **2Y8ZUF**

Berechnen Sie die Grenzen der symmetrischen, zweiseitig begrenzten Quantile. Geben Sie die unterte und die obere Grenze an.

a) $P = 0.9$ für $X \sim \mathcal{N}(3.45, 25)$

b) $P = \frac{2}{3}$ für $X \sim \mathcal{N}(5, 100)$

c) $P = 0.99$ für $X \sim \mathcal{N}(3.45, 9)$

d) $P = 0.8$ für $X \sim \mathcal{N}(8, 0.05^2)$

Lösung:

a) $x = 11.6743$, $x' = -4.7743$

b) $x = 14.6742$, $x' = -4.6742$

c) $x = 11.1775$, $x' = -4.2775$

d) $x = 8.0641$, $x' = 7.9359$

KAPITEL 9

Statistische Tests

Lernziele 9.1 Statistische Tests

- Die Studierenden können aus einer Aufgabenstellung Nullhypothese und Alternativhypothese formulieren.
- Sie können unter der Annahme der Nullhypothese die Wahrscheinlichkeitsverteilung berechnen.
- Die Studierenden können statistische Schlüsse ziehen unter Berücksichtigung des Signifikanzniveaus.

Beispiel 9.1 Übersinnliche Kräfte **447DHG**

Das Glücksrad mit 3 Zahlen (Trefferwahrscheinlichkeit $p = 1/3$) wird 10 Mal gedreht. Jemand errät 7 Zahlen richtig. Hat er übersinnliche Kräfte? Signifikanzniveau $\alpha = 0.05$.

a) Nullhypothese H_0 und Alternativhypothese H_1

b) Wir nehmen an H_0 stimmt: Berechne Wahrscheinlichkeiten

c) Statistischer Schluss

d) Irrtumswahrscheinlichkeit?

x	2	3	4	5	6	7	8	9
$P(X = x)$	0.1951	0.2601	0.2276	0.1366	0.0569	0.0163	0.0030	0.0003
$P(X \leq x)$	0.2991	0.5593	0.7869	0.9234	0.9803	0.9966	0.9996	1
$1 - P(x \leq x)$	0.7009	0.4407	0.2131	0.0766	0.0197	0.0034	0.0004	0

Lösung:

a) Nullhypothese H_0 und Alternativhypothese H_1:

- $H_0: \; p(X = 6) = 1/3$

- $H_1: \; p(X = 6) > 1/3$

Einseitiger Test (>)

b) Excel Binomial-Verteilung mit $n = 10$, $p = \frac{1}{3}$. Daraus folgen $\mu = 3.3$ und $\sigma^2 = 2.2$. Wir lesen aus $P(X \geq 7) = 1 - P(X \leq 6) = 0.0197$

c) Statistischer Schluss: Die Wahrscheinlichkeit für $P(X \geq 7)$ ist 2% und liegt damit unter dem Signifikanzniveau von $\alpha = 5\%$. Die Nullhypothese wird abgelehnt und die Alternativhypothese angenommen. D.h. die Person hat aussersinnliche Kräfte.

d) Irrtumswahrscheinlichkeit ist 2%.

Beispiel 9.2 Asymetrischer Würfel? **IMEX3S**

12000 Würfe davon $x = 2107$ Sechsen. Werden Sechsen bevorzugt gewürfelt? Signifikanzniveau $\alpha = 0.01$

a) Nullhypothese H_0 und Alternativhypothese H_1 formulieren

b) Wir nehmen an H_0 stimmt: Berechne Wahrscheinlichkeiten

c) Statistischer Schluss

d) Irrtumswahrscheinlichkeit

x	2102	2103	2104	2105	2106	2107	2108	2109
$P(X = x)$	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003	0.0003	0.0003
$P(X \leq x)$	0.9937	0.9941	0.9945	0.9949	0.9952	0.9956	0.9959	0.9962
$1 - P(x \leq x)$	0.0063	0.0059	0.0055	0.0051	0.0048	0.0044	0.0041	0.0038

Lösung:

a) Nullhypothese H_0 und Alternativhypothese H_1:

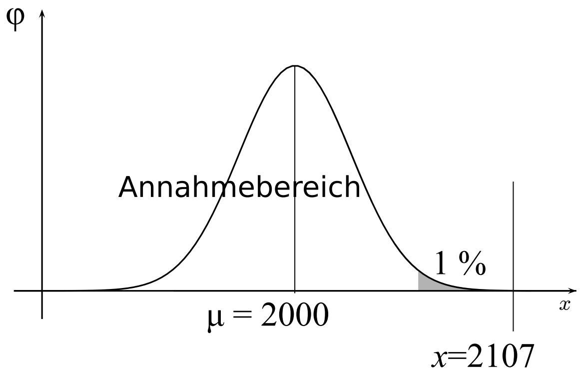
- $H_0: \; p(X = 6) = 1/6$
- $H_1: \; p(X = 6) > 1/6$

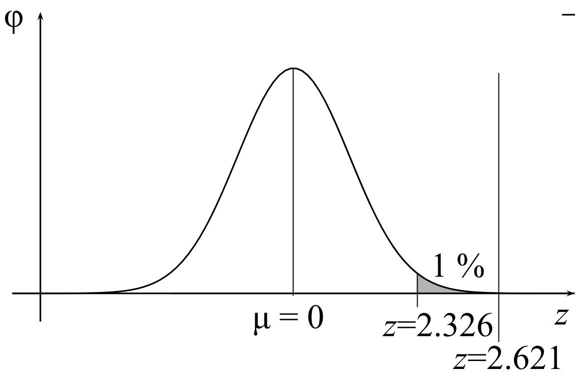
Einseitiger Test (>)

b) Excel Binomial-Verteilung mit $n = 12000$, $p = \frac{1}{6}$. Daraus folgen $\mu = 2000$ und $\sigma^2 = 1666.7$. Wir lesen aus $P(X \geq 2107) = 1 - P(X \leq 2106) = 0.0048$

c) Statistischer Schluss: Die Wahrscheinlichkeit für $P(X \geq 2107)$ ist 0.48% und liegt damit unter dem Signifikanzniveau von $\alpha = 1\%$. Die Nullhypothese wird abgelehnt und die Alternativhypothese angenommen.

d) Irrtumswahrscheinlichkeit ist 0.48%.





Im Skript wird mit $P(X \geq 2107) = 1 - P(X \leq 2107) = 0.0044$ gerechnet, was auch korrekt ist.

Beispiel 9.3 Weniger/mehr Halbtax-Abos? **807872**

2016: 75% haben eines
2017: 270 von 350 Befragte haben eines
Argumentation mit Tafel T.2 im z-Raum.

a) Nullhypothese H_0, und Alternativhypothese H_1

b) Näherung durch Normalverteilung?

c) Darstellung mit z (vorausgesetzt H_0)

d) Statistischer Schluss (Signifikanzniveau $\alpha = 0.1$).

Lösung:

a) Nullhypothese H_0, und Alternativhypothese H_1

- $H_0: \ p(X = \text{HA}) = 0.75$
- $H_1: \ p(X = \text{HA}) \neq 0.75$

Zweiseitiger Test ($\neq$)

b) Näherung durch Normalverteilung gerechtfertigt:

$$\mu = n \cdot p = 262.5, \ \sigma^2 = n \cdot p(1-p) = 218.75$$

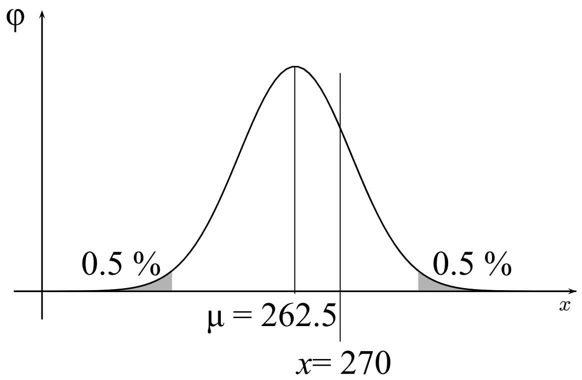
c) Darstellung mit z, (vorausgesetzt H_0): Annahmebereich für H_0 auslesen aus Tabelle T.2 (α, symmetrisch verteilen):

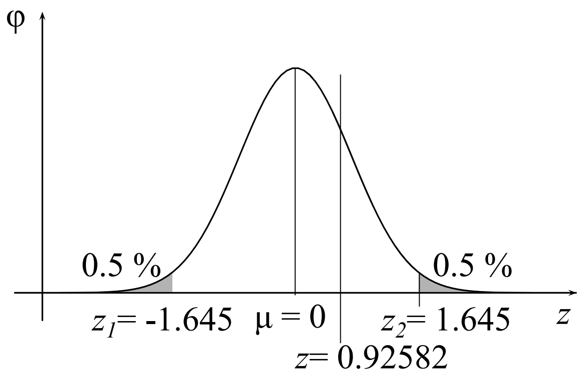
$$z(P = 0.95) = 1.645$$

$x = 270$ "übersetzen" in z:

$$z = \frac{270 - 262.5}{\sqrt{218.75}} = 0.92582$$

d) Statistischer Schluss: $z = 0.92582 < 1.645$ liegt im Annahmebereich $\Rightarrow$ H_0 annehmen.





Beispiel 9.4 Weniger/mehr Halbtax-Abos? **XL6W93**

Argumentiere für die vorherige Aufgabe im x-Raum. Vorgehen:

a) σ und μ berechnen

b) Grenzen Annamebereich berechnen mit Tabelle

c) Grenzen Annamebereich berechnen mit Excel

d) Statistischer Schluss

Lösung:

a) σ und μ berechnen:

$$\sigma^2 = 218.75, \; \mu = 262.5$$

b) Grenzen Annamebereich berechnen mit Tabelle:

$$z_1 = -1.645, \; z_2 = 1.645$$

Annahmebereich ausgedrück in x

$$x_1 = z_1 \cdot \sigma + \mu = 250, \; x_2 = z_2 \cdot \sigma + \mu = 276,$$

c) Grenzen Annamebereich berechnen mit Excel:

$$x_1 = 250, \; x_2 = 276,$$

d) Statistischer Schluss: Die Messung liegt im Annahmebereich: H_0 annehmen.

Achtung: In den Beispielen und wird die Normalverteilung nur als *Annährung* an die Binomial-Verteilung benutzt, so dass die Tabelle der Quantile der Normalverteilung benutzt werden kann.

KAPITEL 10

Prüfen von Erwartungswerten, Parametertests, t-Verteilung

Lernziele 10.1 Parametertests, t-Verteilung

- Die Studierenden kennen das statistische Verhalten des Mittelwertes als Funktion des Stichprobenumfangs.
- Sie können aus $\bar{x}$, s und dem Stichprobenumfang die Testgrösse t berechnen.
- Die Studierenden kennen die Student-t-Verteilung und wissen, dass sie die Verteilung der Mittelwerte von zufälligen Stichproben angibt.
- Sie können Erwartungswerte testen (ein Erwartungswert, zwei Erwartugnswerte mit gleichem oder verschiedenem σ, verbundene Stichproben).

Skrips: `wst_studentt_verteilung_aufg.m`

10.1 Die Student-t-Verteilung

Beispiel 10.1 Maltosegehalt **VC8BDB**

William Sealey Gosset bekommt zwei Lieferungen à 4 Säckenen mit Malz[a]. Für die Bierproduktion ist der Maltosegehalt ausschlagebend. Den misst er in g pro 100 g Malzmehl. Hier sind die Ergebnisse:

$$A = \{4, 9, 9, 13\} \text{ und } B = \{5, 9, 9, 7\}$$

a) Das langjährige Mittel der Lieferungen ergab einen Schnitt von $\mu = 8 \pm 2$ g Malzgehalt. Wir unterteilen die Lieferungen, in die die nahe am Durschnitt liegen (20%) und die weit vom Durchschnitt liegen (80%). Zeichnen Sie die Verteilung und die Grenzen gemäss dieser Kriterien.

b) Berechnen Sie die Mittelwerte der Lieferungen und zeichnen Sie diese in die Verteilung ein.

c) Berechnen Sie nun den Mittelwert der Menge $C = A \cup B$

d) Vergleichen Sie die Mittelwerte. Welche Mittelwerte sind representativ, welche nicht? Was könnten die Gründe sein, dass die Mittelwerte so ausfallen?

Lösung:

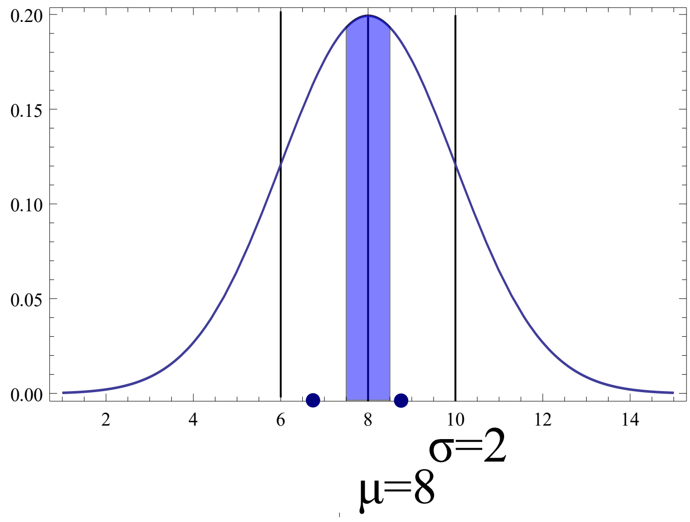
a) Wir erwarten 20% im Intervall $[7.5, 8.5]$, und 80% ausserhalb (wird berechnet mit den Quantilen der Normalverteilung).

b) Die Mittelwerte der Stichproben liegen bei

$$\begin{aligned} \overline{x}_A &= 8.75 \\ \overline{x}_B &= 6.75 \end{aligned}$$

c) Mittelwert der Menge $\overline{x}_C = 7.75$

d) Der Mittelwert von C fällt am besten aus. Das überrascht uns nicht, denn in einer kleinen Stichprobe kann es eher vorkommen, dass eine 'spezielle Auswahl' von Datanpunkten vorliegt. Beispielsweise bei A liegen die Datenpunkte zu hoch, bei B zu tief. Je grösser die Stichprobe, desto unwahrscheinlicher, dass man auf eine eine 'spezielle Auswahl' stösst und desto grösser die Wahrscheinlichkeit, dass der Mittelwert nahe am korrekten Mittelwert (Erwartungswert) liegt.



[a] Kurz gekeimtes und wieder getrocknetes Getreide, meist Gerste, Weizen, Roggen. Ein Teil der Stärke wird dabei in kleinere Moleküle (Disaccharide, wie Maltose) zerlegt.

Beispiel 10.2 Maltosegehalt II **77FF3E**

	0	1	2	3	4	5	6	7	8	9
0	12.13	8.4	7.31	6.59	10.53	6.79	6.95	7.97	7.95	8.13
10	6.8	8.33	8.49	9.68	3.47	10.62	7.84	5.73	9.86	6.91
20	8.54	7.34	8.49	9.25	5.98	4.34	9.23	8.41	11.65	6.74
30	5.68	7.25	6.49	11.59	4.93	4.07	6.33	7.7	6.59	6.46
40	10.81	10.17	10.79	8.56	8.58	10.27	8.55	9.48	9.19	6.44
50	6.14	3.36	8.21	7.11	9.95	6.	6.87	8.1	11.97	10.06
60	5.93	6.64	7.84	5.1	13.6	9.6	9.29	9.07	10.02	5.5
70	5.34	10.09	6.32	9.79	7.81	6.06	6.72	10.18	7.17	6.35
80	8.29	7.04	5.57	14.58	7.75	8.92	9.64	4.97	7.88	7.41
90	7.79	11.98	7.76	10.41	10.82	11.22	5.55	7.88	8.09	8.46

Maltosegehalt der letzen 100 gelieferten Säcke. Die 12. Lieferungen hatte z.B. einen Maltosegehalt von 8.33 g pro 100 g Mehl.

a) Wie sehen Lieferungen à 4 Säcken aus? Laden Sie sich die App "Zufallszahl UX" auf das Handy. Unter der Kategorie 'Nummer' können Sie Zahlen erzeugen zwischen 0 und 99 (Alternativ: Zufallszahl auf Google). Wählen Sie so zufällig 4 Säcke aus, und berechnen Sie den Mittelwert und die Standard-Abweichung.

b) Wie werden Mittelwert und Standard-Abweichung für grössere Lieferungen aussehen, z.B. 9, 16 oder 25 Säcke? Was erwarten Sie?

c) Benutzen Sie die App und erzeugen Sie nun zufällige Stichproben mit 4, 9, 16 und 25 Säcken. Berechnen Sie den Mittelwert und die Standard-Abweichung dieser Stichproben.

d) Welche Tendenzen lassen sich erkennen?

Lösung:

a) $\overline{x}_4 = 8.6075,\ s_4 = 2.46339$

b) Wie werden Mittelwert und Standard-Abweichung für grössere Lieferungen aussehen, z.B. 9, 16 oder 25 Säcke? Was erwarten Sie?

c)

$$\begin{aligned}
\overline{x}_9 &= 8.29111, & s_9 = &= 1.86713 \\
\overline{x}_{16} &= 8.13375, & s_{16} = &= 2.00631 \\
\overline{x}_{25} &= 8.0032, & s_{25} = &= 1.79237 \\
\overline{x}_{100} &= 8.0858, & s_{100} = &= 2.13055
\end{aligned}$$

d) Zwar wird der Mittelwert in den grösseren Stichproben besser. Dieser Effekt könnte aber auch zufällig sein, denn die Standard-Abweichung wird nicht systematisch besser.

Beispiel 10.3 Maltosegehalt III **JEV45E**

Wie verhält sieht nun die Verteilung von kleinen Stichproben aus?

a) Tauschen Sie mit ihren Mitstudierenden die Mittelwerte (nicht die Standard-Abweichungen). Sie erhalten also für jede Stichrprobengrösse ca. 20 Mittelwerte.

b) Berechnen Sie nun für jede Stichprobengrösse $N = \{4, 9, 16, 25\}$ Mittelwert und Standard-Abweichung *der Kategorie*. Damit können Sie folgende Tabelle verfollständigen (evtl. Arbeitsteilung in der Klasse):

N	4	9	16	25
μ'				
σ'				
$\sigma' \cdot \sqrt{N}$				

c) Beschreiben Sie nun in Worten: Was passiert mit der Mittelwert wenn wir die Stichprobe vergrössern oder verkleinern? Was passiert mit der Streung?

d) Bestimmen Sie einen Ausdruck für Mittelwert $\mu'(N)$ und Standard-Abweichung $\sigma'(N)$ einer Stichprobe vom Umfang N

Lösung:

a) Daten Austauschen

b) Tabelle ergänzen.

c) Der Mittelwert bleibt konstant, auch wenn die Stichprobe vergrössert wird. Die Streuung der Mittelwerte hingegen nimmt ab. Wir sagen: Grössere Stichproben sind representativer.

d) $\mu'(N) = \overline{x}$ und Standard-Abweichung $\sigma'(N) = s/\sqrt{N}$, dabei sind s berechnete Standardabweichung aus der Stichprobe (wir nennen Sie auch geschätze Standardabweichung) und $\overline{x}$ der Mittelwert der Stichprobe (auch geschätze Mittelwert).

Beispiel 10.4 Maltosegehalt IV **DCQC3U**

Hier sind die Resultate aus einer Klasse mit 10'000 Studierenden.

N	4	9	16	25
μ'	8.0201	7.9882	8.0089	8.0040
σ'	1.0002	0.6721	0.4942	0.3976
$\sigma' \cdot \sqrt{N}$	2.0004	2.0163	1.9767	1.9879

a) Passen die resltierenden Ausdrücke aus der vorigen Aufgabe auch auf die Tabelle, die wir in der Klasse generiert haben?

b) Wie könnten mögliche Abweichungen erklärt werden?

Lösung:

a) Die Ausdrücke passen gut.

b) Abweichungen spiegeln die statistischen Schwankungen wider.

In den vorangehenden Beispielen betrachten wir endliche Stichproben von normalverteilten Zufallsvariablen. Die Fragestellungen sind auf verschiedenen Hierarchiestufen, die wir mit der folgenden Notation unterscheiden:

- Grundgesamtheit mit Mittelwert μ und Standard-Abweichung σ. (Langjähriger Erfahrungswert)
- Stichprobe X; $\overline{x}$ Mittelwert der Stichprobe wird auch geschätzter Mittelwert genannt; s Standard-Abweichung der Stichprobe wird auch geschätzte Standard-Abweichung; Es sind Mittelwert und Standard-Abweichung von einer Lieferung.
- μ' Mittelwert der Mittelwerte $\overline{x}$ über mehrere Stichproben; σ' Standard-Abweichung der Mittelwerte $\overline{x}$ über mehrere Stichproben

Infobox 10.1 Mittelwerte von kleinen Stichproben

Mittelwerte von kleinen Stichproben sind nicht normalverteilt. Sie gehorchen der Student-t-Verteilung.

Infobox 10.2 Berechnung der Testgrösse

$$t = \frac{\overline{x} - \mu}{s} \cdot \sqrt{N}$$

Dabei sind

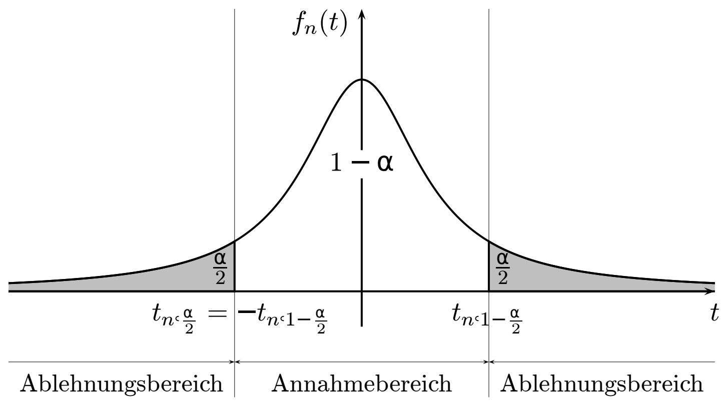
- $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$, der Mittelwert der Stichprobe (geschätzer Mittelwert)
- $s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$ geschätze Varianz (s, die geschätzte Standardabweichung)
- N, der Umfang der Stichprobe

Satz 10.1 Verteilung von kleinen Stichproben

Wir betrachten eine Zufallsgrösse X, die normalverteilt ist mit Mittelwert μ und Standard-Abweichung σ. Daraus entnehmen wir Stichproben mit Umfang N. Y ist die Menge der Mittelwerte von zufälligen Stichproben aus X. Die Testgrösse

$$T = \frac{Y - \mu}{\sigma} \cdot \sqrt{N}$$

ist dann eine Zufallsgrösse. Sie ist verteilt gemäss der Student-t-Verteilung mit $n = N - 1$ Freiheitsgraden.



Infobox 10.3 Mittelwert und Standardabweichung geschätzt

Mittelwert $\overline{x}$ und (geschätzte) Standardabweichung s einer Stichprobe hängen nicht vom Umfang der Stichprobe ab. Sie können benutzt werden um den Mittelwert μ und die Standardabweichung σ der Grundgesamtheit abzuschätzen.

Beispiel 10.5 Student-t-Verteilung **1UITC5**

$$X = \{5, -5, 7, 4, 15, -7, 5, 10, 18, 16\}$$

;

a) Wir betrachten die Stichprobe X und wollen entscheiden, ob sie tatsächlich den Mittelwert $\mu = 0$ hat. Signifikanzniveau α =5%. Formuliere Null-Hypothese und Alternativ-Hypothese.

b) Berechne Mittelwert und Standard-Abweichung der Stichprobe.

c) Berechne die Testgrösse $t = \frac{\overline{x}-\mu}{s} \cdot \sqrt{N}$

d) Bestimme die Annahmegrenzen des Konfidenzinvervalls zu $q = 1 - \alpha/2$ (zweiseitiger Test) oder $q = 1 - \alpha$ (einseitiger Test).
Benutze: In Excel `T.INV` oder `TINV`; Oder in Matlab `tinv`; Oder im Skript die Tafel T.3.

e) Führe dann den statischen Test durch.

f) Überprüfe auf die selbe Weise ob die Stichprobe

$$X = \{5, -5, 7, 4, 15, -7, 5, -10, -18, -16\}$$

den Mittelwert $\mu = 0$ hat.

Lösung:

a) Nullhypothese H_0: Die Stichprobe X stammt aus einer Grundgesamtheit mit $\mu = 0$ hat.
Alternativhypthese H_1: Die Stichprobe X stammt aus einer Grundgesamtheit mit $\mu \neq 0$ hat. (zweiseitiger Text)

b) $\overline{x} = 6.8$, $s = 8.4$

c) Testgrösse $t = \frac{\overline{x}-\mu}{s} \cdot \sqrt{N} = 2.567$

d) Annahmegrenzen des Konfidenzinvervalls zu $q = 1 - \alpha/2 = 0.9750$ (zweiseitiger Test $\Rightarrow$ α symmetrisch verteilen)

$$t_{\text{oben}} = 2.262$$

e) $|t| > |t_{\text{oben}}|$, die Nullhypothese wird verworfen, H_1 angenommen, d.h. $\mu \neq 0$

f) $\overline{x} = -2.0000$, $s = 10.8218$
Testgrösse $t = \frac{\overline{x}-\mu}{s} \cdot \sqrt{N} = -0.5844$
Das Konfidenzinvervall bleibt unverändert ($q = 1 - \alpha/2 = 0.9750$, zweiseitiger Test, $t_{\text{oben}} = 2.262$)
$|t| < |t_{\text{oben}}|$, also wird H_0 angenommen.

10.2 Test von Erwartungswerten

Übersicht

Allgemeine Testschritte	Test **eines** Erwartungswerts (1 Stichprobe)	Test **zwei** Erwartungswerte		
		gleiches σ^2	verschiedene σ^2	verbundene Stichproben
Aufstellen der Nullhypothese H_0	$H_0: \ \mu = \mu_0$	$H_0: \ \mu_1 = \mu_2$	$H_0: \ \mu_1 = \mu_2$	$H_0: \ \mu_1 = \mu_2$
Festlegen der Irrtumtswahrscheinlichkeit α	oft $\alpha = 1$ %, 5 % oder 10 %			
Berechnung der kritischen Grenzen	$\pm t_{N-1;1-\alpha/2}$	$n = N_1 + N_2 - 2$ $\pm t_{n;1-\alpha/2}$	$c = \frac{(s_1)^2/N_1}{(s_1)^2/N_1+(s_2)^2/N_2}$ $n = \left\lfloor \left(\frac{c^2}{N_1-1} + \frac{(1-c)^2}{N_2-1} \right)^{-1} \right\rfloor$ $\pm t_{n;1-\alpha/2}$	$\pm t_{N-1;1-\alpha/2}$
Berechnen der Prüfgrösse	$t = \frac{\overline{x}-\mu_0}{s} \cdot \sqrt{N}$	$s^2 = \frac{(N_1-1)\cdot(s_1)^2+(N_2-1)\cdot(s_2)^2}{N_1+N_2-2}$ $t = \frac{\overline{x}-\overline{y}}{s} \cdot \sqrt{\frac{N_1 \cdot N_2}{N_1+N_2}}$	$s^2 = \frac{(s_1)^2}{N_1} + \frac{(s_2)^2}{N_2}$ $t = \frac{\overline{x}-\overline{y}}{s}$	$d = x - y$ $\overline{d} = \mathrm{E}(d)$ $(s_d)^2 = \mathrm{Var}(d)$ $t = \frac{\overline{d}}{s_d} \cdot \sqrt{N}$

Skrips: `wst_parametertests_aufg.m`

Beispiel 10.6 Messreihen **621147**

Gegeben seien die Messreihen $X = \{1, 2, 3, 2, 1\}$ und $Y = \{2, 2, 4, 1\}$ aus normalverteilten Grundgesamtheiten. Wir testen unter der Voraussetzung $\sigma_x^2 = \sigma_y^2$ die Gleichheit der Erwartungswerte.

a) Zu welcher Kategorie gehört dieser Test?

b) Führen Sie den Test durch. Signifikanzniveau $\alpha = 0.05$.

Lösung:
Kategorie: unverbundene Stichproben, gleiches σ.
Wir testen die Hypothese $\mu_1 = \mu_2$ für x und y
x (bekommt Index 1) und y (bekommt Index 2): $n = N_1 + N_2 - 2 = 5 + 4 - 2 = 7$ und $t_{7;1-0.025} = t_{7;0.975} = 2.365$.
Varianz:

$$(s_1)^2 = 0.7 \text{ und } (s_2)^2 = 1.58333$$

also

$$s^2 = \frac{(5-1)\cdot 0.7 + (4-1)\cdot 1.5833}{7} = 1.07857$$

Testgrösse:

$$t = \frac{\overline{x} - \overline{y}}{s} \cdot \sqrt{\frac{5\cdot 4}{4+5}} = \frac{1.8 - 2.25}{1.03854} \cdot \sqrt{\frac{5\cdot 4}{9}} = -0.645925$$

Die Testgrösse liegt im Annahmebereich. Statistischer Schluss: $\overline{x} = \overline{y}$.

Beispiel 10.7 Bindemittel **602709**

Mit zwei verschiedenen Holzwerkstoffbindemitteln A und B werden Spanplatten hergestellt. Mit dem Bindemittel A erhalten wir 10 Prüfkörper, mit dem Mittel B deren 12. Alle Prüfkörper werden einem Querzugfestigkeitstest unterworfen. Folgende Werte wurden gemessen:

A	0.745	0.824	0.804	0.863	0.873	0.814	0.804	0.794	0.804	0.7453
B	0.745	0.686	1.049	1.059	0.873	0.834	0.735	0.971	0.932	0.932
	0.843	0.87								

a) Zu welcher Kategorie gehört dieser Test?

b) Führen Sie den Test durch. Signifikanzniveau $\alpha = 1\%$.

Lösung:
Kategorie: unverbundene Stichproben, verschiedene σ.
Wir testen die Hypothese $\mu_1 = \mu_2$ für A und B
A (bekommt Index 1) und B (bekommt Index 2): Varianz:

$$(s_1)^2 = 0.001746 \text{ und } (s_2)^2 = 0.0140632 \;\Rightarrow s^2 = \frac{0.001746}{10} + \frac{0.0140632}{12} = 0.00134653$$

$$c = \frac{0.001746/10}{0.001746/10 + 0.0140632/12} = 0.129667$$

und damit

$$n = \left\lfloor \left(\frac{c^2}{N_1 - 1} + \frac{(1-c)^2}{N_2 - 1} \right)^{-1} \right\rfloor = \left\lfloor \left(\frac{0.0168}{9} + \frac{0.757}{11} \right)^{-1} \right\rfloor = \left\lfloor (0.0707)^{-1} \right\rfloor = 14$$

Annahme-Bereich: $t_{14;0.995} = 2.977$
Testgrösse:

$$t = \frac{\overline{x} - \overline{y}}{s} = \frac{0.807 - 0.877667}{0.0366951} = -1.92578$$

Die Testgrösse liegt im Annahmebereich. Statistischer Schluss: $\overline{x} = \overline{y}$.

Beispiel 10.8 Widerstände **136503**

Zwei verschiedene Messmethoden für Widerstände sollen miteinander verglichen werden. Vergleichsmessungen an fünf Widerständen ergaben das folgende Messprotokoll:

i	1	2	3	4	5
1. Methode: x_i [in Ω]	100.5	102.4	104.3	101.5	98.4
2. Methode: y_i [in Ω]	98.2	99.1	102.4	101.1	96.2

Bestimmen Sie, ob beide Messmethoden als gleichwertig angesehen werden können.

a) Zu welcher Kategorie gehört dieser Test?

b) Führen Sie den Test durch. Signifikanzniveau α =1%.

Lösung:
Kategorie: verbundene Stichproben, gleiche σ.
Wir testen die Hypothese $\mu_1 = \mu_2$ für 1. Methode und und 2. Methode
Annahme-Bereich: $t_{5-1,0.995} = 4.603$
Differenz:

$$d = \begin{pmatrix} 100.5 \\ 102.4 \\ 104.3 \\ 101.5 \\ 98.4 \end{pmatrix} - \begin{pmatrix} 98.2 \\ 99.1 \\ 102.4 \\ 101.1 \\ 96.2 \end{pmatrix} = \{2.3, 3.3, 1.9, 0.4, 2.2\}$$

Mittelwert der Differenz:

$$\overline{d} = 2.02$$

Varianz der Differenz:

$$(s_d)^2 = 1.097$$

Testgrösse:

$$t = \frac{\overline{d}}{s} \cdot \sqrt{n} = \frac{2.02}{1.04738} \cdot \sqrt{5} = 4.31$$

Die Testgrösse liegt im Annahmebereich. Statistischer Schluss: $\overline{x} = \overline{y}$.

KAPITEL 15

Korrelation

Lernziele 15.1 Korrelation

- Die Studierenden kennen die Kovarianz und den Korrelationskoeffizienten.
- Die Studierenden können eine Korrelation auf Signifikanz überprüfen.
- Sie können einen Korrelationskoeffizienten deuten (positive/negative Korrelation) und wissen, dass eine Korrelation keinen kausalen Zusammenhang bedeutet.
- Sie wissen, dass die Varianz und die Kovarianz unabhängig vom Mittelwert sind. Sie wissen, dass die Kovarianz unabhängig von der Skalierung der Daten ist.
-

Skrips: `wst_korrelation_aufg.m`

Definition 15.1 Kovarianz

Die Kovarianz der Datenreihen

$$X = \{x_1, x_2, \ldots, x_n\} \text{ und } Y = \{y_1, y_2, \ldots, y_n\}$$

ist definiert als

$$\operatorname{cov}(X, Y) = s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

Infobox 15.1 Kovarianz und Varianz

Die (geschätze) Standard-Abweichung s_x einer Datenreihe X lässt sich mit der Kovarianz berechnen

$$s_x = \sqrt{s_{xx}} = \sqrt{\operatorname{cov}(X, X)}$$

Infobox 15.2 Kovarianz in Matlab

Matlab berechnet mit dem Befehl `cov(X,Y)` die Kovarianz-Matrix

$$C = \begin{pmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy} \end{pmatrix}$$

Definition 15.2 Korrelationskoeffizient

Der **Korrelationskoeffizient** ist definiert durch

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}}$$

Satz 15.1 Signifikante Korrelation

Die Korrelation zwischen den Datenreihen $\vec{x}$ und $\vec{y}$ ist signifikant, falls die Testgrösse

$$t_{\text{reg}} = \frac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1-(r_{xy})^2}} .$$

im Betrag grösser ist als die kritische Grösse

$$t_{n-2;1-\alpha/2}$$

der Student-t-Verteilung.

Beispiel 15.1 Glucosekonzentration **UH8WJW**

Proband	Alter	Glukosekonzentration
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

a) Visualisieren Sie die Daten (`sort, plot`)

b) Berechnen Sie den Korrelationskoeffizient zwischen Alter und Glucosekonzentration im Blut (`cov, sqrt`)

c) Formulieren Sie eine Nullhypothese und eine Alternativhypothese für den Korrelationskoeffizient.

d) Untersuchen Sie, ob die Korrelations signifikant ist (Signifikanzniveau $\alpha = 0.2$, Befehle: `tinv, sqrt`)

e) Eine erhöhte Glucosekonzentration im Blut kann ein Hinweis auf Diabetes mellitus sein. Erhöht sich also das Risiko an Diabetes zu erkranken mit dem Alter?

Lösung:

a) Visualisieren Sie die Daten (`sort, plot`)

b) Berechnen Sie den Korrelationskoeffizient zwischen Alter und Glucosekonzentration im Blut (`cov, sqrt`)

c) Nullhypothese H_0: $r = 0$, Alternativhypothese $r \neq 0$

d) Wir berechnen

$$t_{\text{reg}} = 1.2494 \,.$$

und

$$t_{\text{crit}} = t_{n-2;1-\alpha/2} = t_{6-2;1-0.2/2} = 1.5332$$

d.h. $|t_{\text{reg}}| < t_{\text{crit}}$ und die Nullhypothese wird verworfen. Die Daten sind unkorreliert.

e) Anhand dieser Daten gibt es keinen Hinweis darauf, dass sich im Alter das Risiko an Diabetes zu erkranken erhöht.

Der Korrelationskoeffizient kann auch mit dem Skalarprodukt berechnet werden

Infobox 15.3 Korrelationskoeffizient als Skalarprodukt

Wir wollen den Korrelationskoeffizient berechnen für die Daten-Vektoren

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ und } \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \,.$$

Dafür ziehen wir zuerst den Mittelwert ab

$$\vec{x}' = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} \text{ und } \vec{y}' = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \,,$$

und normieren die Vektoren

$$\vec{u} = \frac{\vec{x}'}{|\vec{x}'|} \text{ und } \vec{v} = \frac{\vec{y}'}{|\vec{y}'|} \,.$$

Dann ist der Korrelationskoeffizient das Skalarprodukt

$$r_{x,y} = \vec{u} \odot \vec{v}$$

Beispiel 15.2 Eigenschaften der Varianz DQNDPY

Wir betrachten die Datenreihe

$$X = \{-18, 8, 4, 0, 4, 7, 8, 12\}$$

a) Berechnen Sie den Mittelwert und die Varianz.

b) Erzeugen Sie eine neue Datenreihe Y, indem Sie zu jedem x_i eine feste Zahl z.B. 5 addieren

$$y_i = x_i + 5$$

Vergleichen Sie Mittelwert und Varianz der Datenreihen X und Y.

c) Wie beeinfluss der Mittelwert die Varianz einer Datenreihe? Formulieren Sie eine Hypothese.

d) Beweisen Sie Ihre Hypthese mathematisch. Benutzen Sie dafür die Definition der Varianz.

Lösung:

a) $s_{x,x} = 85.5536,\ \bar{x} = 3.1250$

b) $s_{y,y} = 85.5536,\ \bar{y} = 8.1250$ Durch die Addition einer festen Zahl ändert sich zwar der Mittelwert aber die Varianz ändert sich nicht.

c) Der Mittelwert einer Datenreihe beeinfluss die Varianz nicht.

d) Beweis:

$$\begin{aligned} s_{y,y} &= \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})\cdot(y_i-\bar{y}) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i+c-(\bar{x}+c))\cdot(x_i+c-(\bar{x}+c)) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(x_i-\bar{x}) \end{aligned}$$

Beispiel 15.3 Eigenschaften der Varianz **3PSMTG**

Wir betrachten die Datenreihe

$$X = \{-18, 8, 4, 0, 4, 7, 8, 12\}$$

a) Erzeugen Sie eine neue Datenreihe Z, indem Sie jedes x_i mit einer festen Zahl z.B. 3 multiplizieren

$$z_i = x_i \cdot 3$$

Vergleichen Sie Mittelwert und Varianz der Datenreihen X und Z.

b) Wie beeinflusst die Skalierung $Z = \lambda \cdot X$ die Varianz einer Datenreihe? Formulieren Sie eine Hypothese.

c) Beweisen Sie Ihre Hypthese mathematisch. Benutzen Sie dafür die Definition der Varianz.

Lösung:

a) $s_{z,z} = 769.9821,\ \bar{y} = 9.3750$ Durch die Multiplikation mit ändert sich wir der Mittelwert drei Mal grösser und die Varianz 9 Mal grösser als die entsprechenden Werte bei X.

b) $Z = c \cdot X \ \Rightarrow\ s_{z,z} = s_{x,x} \cdot c^2$

c) Beweisen Sie Ihre Hypthese mathematisch. Benutzen Sie dafür die Definition der Varianz.

$$\begin{aligned} s_{z,z} &= \frac{1}{n-1}\sum_{i=1}^{n}(z_i-\bar{z})\cdot(z_i-\bar{z}) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i\cdot c-(\bar{x}\cdot c))\cdot(x_i\cdot c-(\bar{x}\cdot c)) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}c\cdot(x_i-\bar{x})\cdot c\cdot(x_i\cdot -\bar{x}) \\ &= c^2\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(x_i-\bar{x}) = c^2\cdot s_{x,x} \end{aligned}$$

Beispiel 15.4 Eigenschaften der Kovarianz **FEWL9F**

Wir betrachten die Datenreihe

$$\begin{aligned} X &= \{-18, 8, 4, 0, 4, 7, 8, 12\} \\ Y &= \{9, 3, 5, 5, 4, 3, -7, -2\} \end{aligned}$$

a) Berechnen Sie die Kovarianz

$$\mathrm{cov}(X,Y) = s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})$$

b) Erzeugen Sie eine neue Datenreihe V, indem Sie jedes y_i mit einer festen Zahl z.B. 7 addieren

$$v_i = y_i + 7$$

Vergleichen $\mathrm{cov}(X,Y)$ und $\mathrm{cov}(X,V)$.

c) Erzeugen Sie eine neue Datenreihe W, indem Sie jedes y_i mit einer festen Zahl z.B. 2 multiplizieren

$$w_i = y_i \cdot 2$$

Vergleichen $\mathrm{cov}(X,Y)$ und $\mathrm{cov}(X,W)$.

d) Wie beeinflusst die Verschiebung des Mittelwerts $V = a+Y$ und die Skalierung $W = \lambda \cdot X$ die Kovarianz

$$\mathrm{cov}(X,V) \text{ und } \mathrm{cov}(X,W)$$

Formulieren Sie zwei Hypothesen, in denen $\mathrm{cov}(X,Y)$ vorkommt.

e) Beweisen Sie Ihre Hypthesen mathematisch.

f) Welche Konsequenzen haben Ihre Resultate für den Korrelationskoeffizient?

Lösung:

a) Kovarianz

$$\mathrm{cov}(X,Y) = s_{xy} = -31.9286$$

b) $\mathrm{cov}(X,V) = -31.9286$, durch die Addition einer Zahl Y (Änderung des Mittelwertes der Datenreihe) ändert sich die Kovarianz nicht.

c) $\mathrm{cov}(X,W) = -63.8571$. Bei der Transformation

$$Y \ \rightarrow \ W = Y \cdot \lambda$$

verändert sich die Kovarianz um den Faktor $c = 2$.

d) Wir vermuten

$$\mathrm{cov}(X,V) = \mathrm{cov}(X,Y) \text{ und } \mathrm{cov}(X,W) = \lambda \cdot \mathrm{cov}(X,Y)$$

C

e) Beweise:

$$\begin{aligned}
\mathrm{cov}(X,V) &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot((y_i+c)-(\bar{y}+c)) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i+c-\bar{y}-c) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y}) = \mathrm{cov}(X,Y)
\end{aligned}$$

$$\begin{aligned}
\mathrm{cov}(X,W) &= s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(\lambda\cdot y_i-\lambda\cdot\bar{y}) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot\lambda\cdot(y_i-\bar{y}) \\
&= \lambda\cdot\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y}) = \lambda\cdot\mathrm{cov}(X,Y)
\end{aligned}$$

f) Der Korrelationskoeffizient ändert sich nicht bei den Transformationen.

$$r_{xv} = \frac{s_{xv}}{\sqrt{s_{xx}\cdot s_{vv}}} = \frac{s_{xy}}{\sqrt{s_{xx}\cdot s_{yy}}}$$

und

$$r_{xw} = \frac{s_{xw}}{\sqrt{s_{xx}\cdot s_{ww}}} = \frac{\lambda\cdot s_{xy}}{\sqrt{s_{xx}\cdot\lambda^2\cdot s_{yy}}} = \frac{s_{xy}}{\sqrt{s_{xx}\cdot s_{yy}}}$$

C

Beispiel 15.5 Eigenschaften des Korrelationskoeffizienten **S3CFJX**

Benutzen Sie nun die Gesetze von oben um folgende Aussagen mathematisch zu beweisen. Benutzen Sie die Definition des Korrelationskoeffizienten. Schreiben Sie den Korrelationskoeffizienten als

$$r_{xy} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{cov}(X,X)\cdot\mathrm{cov}(Y,Y)}}$$

a) “Jede Datenreihe ist mit sich selber perfekt korreliert”

b) “Der Korrelationskoeffizient ist unabhängig von der Skalierung der Daten-Reihe”

c) “Der Korrelationskoeffizient ist unabhängig vom Mittelwert der Daten-Reihe”

Lösung:

a) “Jede Datenreihe ist mit sich selber perfekt korreliert”

$$r_{xx} = \frac{\operatorname{cov}(X,X)}{\sqrt{\operatorname{cov}(X,X)\cdot\operatorname{cov}(X,X)}} = 1$$

b) “Der Korrelationskoeffizient ist unabhängig von der Skalierung der Daten-Reihe” Wir betrachten die Datenreihen X, Y und W, mit $W = \lambda \cdot Y$ ($\lambda \in \mathbb{R}$). Dann gilt

$$r_{xw} = \frac{\operatorname{cov}(X,W)}{\sqrt{\operatorname{cov}(X,X)\cdot\operatorname{cov}(W,W)}} = \frac{\lambda\cdot\operatorname{cov}(X,Y)}{\sqrt{\operatorname{cov}(X,X)\cdot\lambda^2\cdot\operatorname{cov}(Y,Y)}} = r_{xy}$$

c) “Der Korrelationskoeffizient ist unabhängig vom Mittelwert der Daten-Reihe” Wir betrachten die Datenreihen X, Y und V, mit $W = Y + a$ ($a \in \mathbb{R}$). Dann gilt

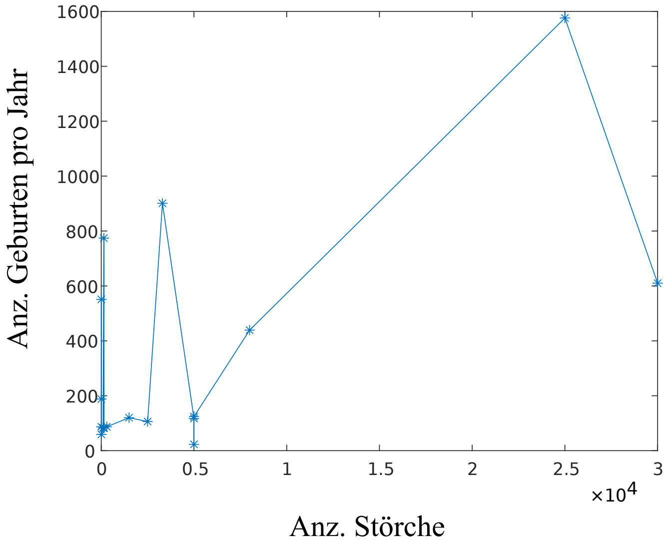
$$r_{xv} = \frac{\operatorname{cov}(X,V)}{\sqrt{\operatorname{cov}(X,V)\cdot\operatorname{cov}(V,V)}} = \frac{\lambda\cdot\operatorname{cov}(X,Y)}{\sqrt{\operatorname{cov}(X,X)\cdot\lambda^2\cdot\operatorname{cov}(Y,Y)}} = r_{xy}$$

15.1 Korrelation und Kausalität

Beispiel 15.6 Der Storch bringt die Babies **J5DYP8**

Blatex

Land	Fläche [km^2]	Störche [Paare]	Menschen [Mio]	Geburten [10^3/Jahr]
Albanien	28750	100	3.2	83
Belgien	30520	1	9.9	87
Bulgarien	111000	5000	9	117
Dänemark	43100	9	5.1	59
Deutschland	357000	3300	78	901
Frankreich	544000	140	56	774
Griechenland	132000	2500	10	106
Holland	41900	4	15	188
Italien	301280	5	57	551
Österreich	83860	300	7.6	87
Polen	312680	30000	38	610
Portugal	92390	1500	10	120
Rumänien	237500	5000	23	23
Spanien	504750	8000	39	439
Schweiz	41290	150	6.7	82
Türkei	779450	25000	56	1576
Ungarn	93000	5000	11	124



a) Berechnen Sie den Korrelationskoeffizient zwischen der Anzahl der Störche und der Geburten (`cov, sqrt`)

b) Visualisieren Sie die Daten (`sort, plot`)

c) Formulieren Sie eine Nullhypothese und eine Alternativhypothese für den Korrelationskoeffizient.

d) Untersuchen Sie, ob die Korrelations signifikant ist (Signifikanzniveau $\alpha = 0.1$, Befehle: `tinv, sqrt`)

e) Diskutieren Sie ihre Resultate. Beantworten Sie auch die Frage, ob die Störche die Babies bringen? Verwenden Sie in ihrer Antwort u.a. die Ausdrücke 'Korrelation', 'Kausalität', 'versteckte Variable'.

f) Entwerfen Sie einen zweiten Korrelations-Test, bei dem Sie den Einfluss der versteckten Variablen eliminieren? Bleibt die Korrelation bestehen?

Lösung:

a) $r_{x,y} = 0.6089$

b) Plot oben

c) Nullhypothese H_0: $r = 0$, Alternativhypothese $r_{x,y} \neq 0$

d) Beim Signifikanzniveau $\alpha = 0.1$ erhalten wir die kritische Grösse (Grenze)

$$t_{\text{crit}} = t_{17-2,1-\alpha/2} = 1.7531$$

und die Testgrösse

$$t_{\text{reg}} = \frac{r_{xy} \cdot \sqrt{n-2}}{\sqrt{1-(r_{xy})^2}} = 2.9727 \ .$$

$|t_{\text{reg}}| > t_{\text{crit}}$, die Nullhypothese wird verworfen, H_1 angenommen, d.h. $r_{x,y} \neq 0$

e) Wir beobachten also eine signifikante Korrelation. Hier ist klar, dass es sich nicht um eine 'Kausalität' handelt. Wir vermuten, dass die 'versteckte Variable' die Fläche des Landes ist: Sie führt dazu, dass die Bevölkerung gross ist, dass damit viele Geburten stattfinden aber auch dazu, dass es in einem Land viele Störche gibt.

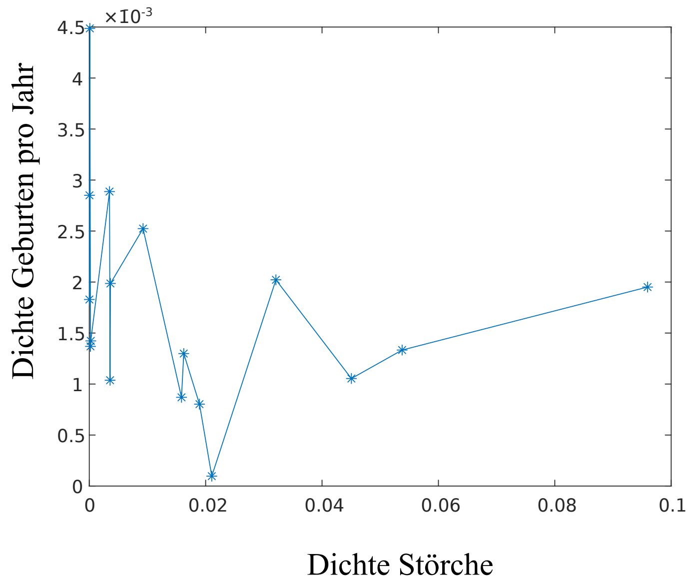
f) Wir erzeugen einen zweiten Datensatz,

$$X' = X/Z \text{ und } Y' = Y/Z$$

indem wir sowohl die Anzahl der Geburten X wie auch die Anzahl Störche Y durch die Fläche des Landes teilen Z. Damit berechnen wir $r'_{x,y} = -0.2171$ und

$$t'_{\text{reg}} = -0.8612\ .$$

Die Korrelation verschwindet statistisch, d.h. sie ist zwar nicht null aber sie fällt unter den kritischen Wert. D.h. Wir haben die versteckte Variable richtig bestimmt.



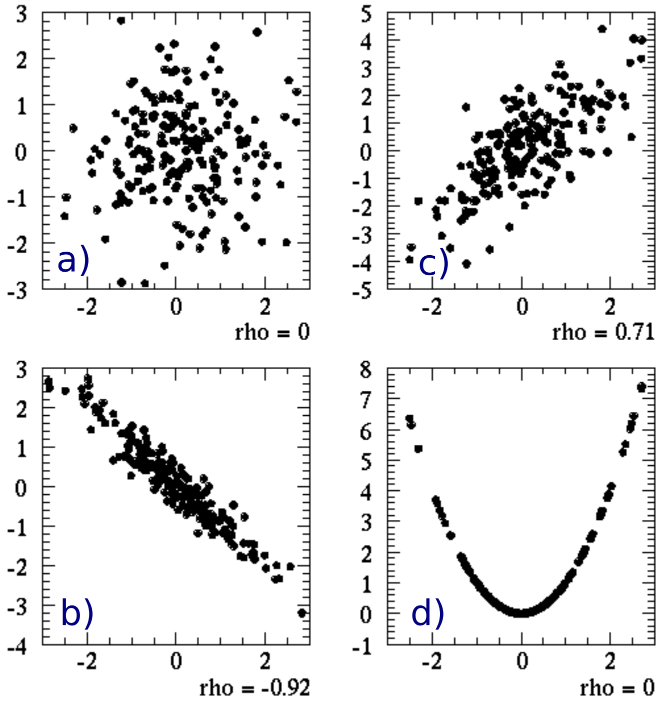
Beispiel 15.7 Korrelationskoeffizienten schätzen **85465E**

a) Schätzen Sie die Korrelationskoeffizienten (Betrag und Vorzeichen)

b) Beurteilen Sie ob der Korrelationskoeffizient einen möglichen Zusammenhang zwischen den Variablen korrekt anzeigt.

c) Entscheiden Sie für jede Graphik, ob die Daten korreliert sind oder funktional zusammenhängen.

Lösung:



Infobox 15.4 Positive und negative Korrelationskoeffizienten

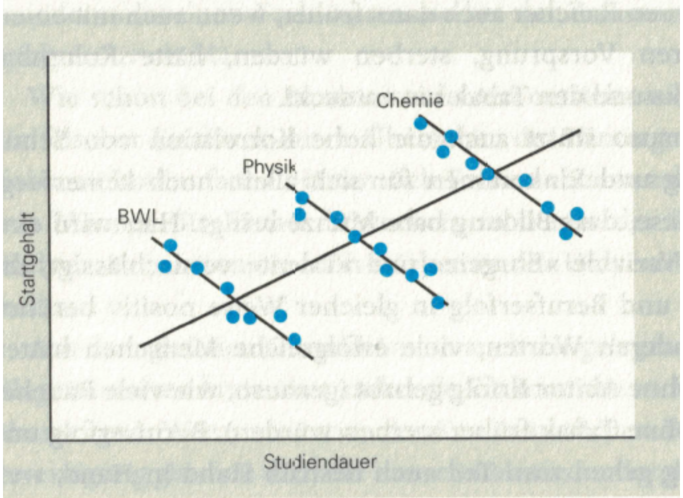
- Negative Korrelation: $r_{xy} < 0$, wenn x wächst, nimmt y ab
- Positive Korrelation: $r_{xy} > 0$, wenn x wächst, nimmt auch y zu

Infobox 15.5 Korrelation $\neq$ Kausalität

Korrelation entsteht nicht nur durch einen kausalen Zusammenhang. Oft besteht eine Korrelation durch einen Zusammenhang mit einer versteckten dritten Grösse.

Beispiel 15.8 Studiendauer-Lohn **N3UCB8**

Aus dem Handelsblatt: "Methusalems machen Kasse: Ein langes Studium zahlt sich in barer Münze aus. Zu diesem überraschenden Ergebnis kommt eine Studie über die Einstiegsgehälter von Berusanfägnern, für die die Deutsche Gesellschaft für Personenführung 44 Unternehmen befragt hat"



a) Welcher Faktor führt zu einem hohen Lohn? Welcher Faktor führt zu einer Lohnreduktion? Betrachten Sie die Grafik.

b) Welches ist die versteckte Variable?

c) Wie ist die Korrelation zustande gekommen?

d) Korrigieren Sie die die Aussage des Handelsblatts.

e) Handelt es sich um eine Korrelationen oder um eine Kausalitäten?

Nach Krämer [2015] **Lösung:**

a) Offenbar gibt es eine Korrelation zwischen Studiendauer und Einstiegslohn.

b) Versteckte Variable: Reputation des Studienfaches, d.h.
Studiendauer $\leftrightarrow$ Studienfach $\rightarrow$ Lohn

c) Studienfächer mit langer Studiendauer haben eine gute Reputation und (deshalb) auch einen hohen Einstiegslohn.

d) Für jedes Studienfach gilt aber: Eine lange Studienzeit mindert den Einstiegslohn.

e) Wir sehen hier eine Korrelationen, die durch die versteckte Variable (Reputation des Studienfaches) zustande kommt.

Beispiel 15.9 Kausalität oder Korrelation? **146PG2**

Beurteilen Sie folgende Aussagen. Identifizieren Sie mögliche Hintergrundvariablen.

a) “Ein 30-jähriger Raucher mit einem Konsum von ein bis zwei Päckchen Zigaretten am Tag stirbt 6 Jahre früher als ein Nichtraucher” (Bundeszentrale für gesundheitliche Aufklärung)

b) “Geliftete Frauen leben zehn Jahre länger als der Durchschnitt” (Mayo-Klinik in Rochester)

c) US-amerikansiche Ausgaben für das Weltraumprogramm sind mit Selbstmorden durch Erhängen und Ersticken korreliert ($r = 0.992$, `http://tylervigen.com`)

d) "Ehemänner leben länger: Junggesellen zwischen 45 und 54 Jahren sollten doch den Sprung ins Abenteur Ehe wagen. [...] 23% der ledigen Männer dieser Altergruppe sterben in den darauffolgenden 10 Jahren. Das Todesrisko verherateter Männer liegt dagegen nur bei 11%" (Epidemiologische Fakultät der Universität Kalifornien)

Nach Krämer [2015] **Lösung:**

Wir notieren mit $\oplus$ Argumente die für einen kausalen Zusammenhang sprechen und mit $\ominus$ Argumente, die dafür sprechen, dass die Korrelation zufällig oder durch versteckte Variablen entsteht.

a) $\oplus$ Es gilt als erwiesen, dass der Rauch Lunge und Atemwege schädigt.
$\ominus$ Versteckte Variable könnte die sogenannte 'Raucherpersönlichkeit' sein: Risikofreudige Personen rauchen eher. Sie würden auch früher sterben, wenn sie keine Zigaretten mehr bekommen würden.
$\ominus$ Versteckte Variable könnte das Geschlecht sein: Männer sterben (sowieso) früher und Männer rauchen auch eher. "Ein 30-jähriger Raucher mit einem Konsum von ein bis zwei Päckchen Zigaretten am Tag stirbt 6 Jahre früher als ein Nichtraucher" (Bundeszentrale für gesundheitliche Aufklärung)

b) $\ominus$ Versteckte Variable könnte der sozio-ökonomische Status sein: Es sind eher reiche Frauen, die sich liften lassen. Sie haben mehr Geld für die Gesundheitspflege zur Verfügung und würde auch ohne Lifting länger leben.

c) $\oplus$ Die Abhängigkeit könnte kausal sein: Je mehr Geld für die NASA ausgegeben wird, desto weniger bleibt für den sozialen Ausgleich übrig
$\ominus$ Versteckte Variable könnte die Zeit sein: Die Bevölkerung wächst mit der Zeit und in einer grösseren Bevölkerung gibt es auch mehr Suizide. Anderseits wachsen mit der Bevölkerung auch die Steuereinnahmen. Davon kann auch mehr für das Weltraumprogramm ausgegeben werden.

d) $\oplus$ Einsamkeit könnte dazu führen, dass Unfälle unentdeckt bleiben
$\ominus$ Versteckte Variable könnte der sozio-ökonomische Status sein: Reiche Männer finden eher eine Partnerin haben aber auch mehr Geld für die Gesundheitspflege zur Verfügung. Sie würden deshalb auch länger leben, wenn sie nicht heiraten könnten.

KAPITEL 16

Regression

Lernziele 16.1 Regression

- Die Studierenden können für einen Datensatz eine Regression berechnen.
- Die Studierenden kennen den Messvektor $\vec{y}$ und die Koeffizientenmatrix F.
- Sie können mit Hilfe von $\vec{y}$ und F das Normalensystem für einen Regression bestimmen.
- Die Studierenden können bei einer linearen Regression die Fehler der Regressions-Koeffizienten bestimmen.

16.1 Lineare Algebra und Matlab

Infobox 16.1 Voraussetzungen Lineare Algebra

- Matrix-Multiplikation, Transponierte
- Lineares Gleichungssystem schreiben mit der Koeffizienten-Matrix
- Lineares Gleichungssystem (LGS) lösen mit Hilfe der inversen Matrix
- LGS lösen mit Cramerschen-Regel
- Matlab: Funktionen für Linalg (`inv, *, transpose`)+ plotten

Beispiel 16.1 Matrix-Multiplikation **JEV45E**

$$\mathbf{M} = \begin{bmatrix} 5 & 0 \\ 2 & -6 \\ 0 & 3 \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} 3 & 0 & -4 \\ -5 & 0 & 6 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 5 \end{bmatrix}$$

a) Berechnen Sie die Matrix-Produkte falls möglich.

$$\mathbf{N} \odot \mathbf{M}, \ \mathbf{P} \odot \mathbf{M}^{\mathsf{T}}; \mathbf{M} \odot \mathbf{M}, \ \mathbf{P} \odot \mathbf{N}^{\mathsf{T}}$$

b) Unter welcher Voraussetzung existiert ein Matrix-Produkt?

Lösung:

a) Folgende Produkte sind definiert:

$$\mathbf{N} \odot \mathbf{M} = \begin{bmatrix} 15 & -12 \\ -25 & 18 \end{bmatrix} \text{ und } \mathbf{P} \odot \mathbf{N}^\mathsf{T} = \begin{bmatrix} 0 & 0 \\ -17 & 25 \end{bmatrix}$$

b) Ein Matrix-Produkt existiert, wenn die Matrix links gleich viele Spalten hat wie die Matrix rechts Zeilen hat.

Definition 16.1 Messvektor

Für eine Datensatz (x_i, y_i) fassen wir die abhängigen Variablen y_i im **Messvektor** zusammen.

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

Beispiel 16.2 Messvektor TQN9T2

$$\mathbf{M} = \begin{bmatrix} 4 & 0 & 2 & 5 & 3 \\ -3 & 0 & 1 & -5 & 1 \end{bmatrix}$$

i	x	y
1	-4	2
2	2	0
3	6	2
4	10	8
5	20	4

a) Bestimmen Sie den Messvektor $\vec{y}$.

b) Welche der folgenden Produkte existieren?

$$\vec{y} \odot \mathbf{M},\ \mathbf{M} \odot \vec{y},\ \vec{y} \odot \mathbf{M}^\mathsf{T},\ \mathbf{M}^\mathsf{T} \odot \vec{y},\ \vec{y}^\mathsf{T} \odot \mathbf{M},\ \mathbf{M} \odot \vec{y}^\mathsf{T},\ \vec{y}^\mathsf{T} \odot \mathbf{M}^\mathsf{T},\ \mathbf{M}^\mathsf{T} \odot \vec{y}^\mathsf{T}$$

Lösung:

a) Messvektor $\vec{y} = \begin{bmatrix} -5 \\ -9 \\ 0 \\ 2 \\ 2 \end{bmatrix}$.

b) Es existieren

$$\mathbf{M} \odot \vec{y} = \begin{bmatrix} -4 \\ 7 \end{bmatrix},\ \vec{y}^\mathsf{T} \odot \mathbf{M}^\mathsf{T} = \begin{bmatrix} -4 & 7 \end{bmatrix}$$

Beispiel 16.3 Inverse Matrix **1UITC5**

$$\mathbf{M} = \begin{bmatrix} 7 & -3 & 2 \\ -6 & 3 & -2 \\ 3 & -1 & 1 \end{bmatrix}, \mathbf{M}^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ -3 & -2 & 3 \end{bmatrix}, \mathbf{N} = \begin{bmatrix} -1 & -5 & 2 \\ 2 & 5 & -2 \\ -1 & -2 & 1 \end{bmatrix}, \mathbf{N}^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 1 & 3 & 5 \end{bmatrix},$$

a) Benennen Sie die Elemente des linearen Gleichungssystem (Koeffizientenmatrix, Inhomogenität, gesuchter Vektor)

$$\mathbf{M} \odot \vec{x} = \begin{bmatrix} 7 \\ -6 \\ 4 \end{bmatrix}$$

b) Lösen Sie das linearen Gleichungssystem (LGS) mit Hilfe der Inversen

c) Lösen Sie das das LGS

$$\mathbf{N} \odot \vec{x} = \begin{bmatrix} -23 \\ 29 \\ -12 \end{bmatrix}$$

Lösung:

a) Koeffizientenmatrix M, Inhomogenität $\begin{bmatrix} 7 \\ -6 \\ 4 \end{bmatrix}$, gesuchter Vektor $\vec{x}$

b) Lösung

$$\vec{x} = \mathbf{M}^{-1} \odot \begin{bmatrix} 7 \\ -6 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

c) Lösung

$$\vec{x} = \mathbf{N}^{-1} \odot \begin{bmatrix} -23 \\ 29 \\ -12 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix}$$

Beispiel 16.4 Regel von Cramers **1QY9NU**

a) Bestimmen Sie x_3 mit der Regel von Kramers

$$\mathbf{M} \odot \vec{x} = \begin{bmatrix} 7 \\ -6 \\ 4 \end{bmatrix} \text{ mit } \mathbf{M} = \begin{bmatrix} 7 & -3 & 2 \\ -6 & 3 & -2 \\ 3 & -1 & 1 \end{bmatrix}$$

b) Bestimmen Sie x_2 mit der Regel von Kramers

$$\mathbf{N} \odot \vec{x} = \begin{bmatrix} -23 \\ 29 \\ -12 \end{bmatrix} \text{ mit } \mathbf{N} = \begin{bmatrix} -1 & -5 & 2 \\ 2 & 5 & -2 \\ -1 & -2 & 1 \end{bmatrix}$$

Lösung:

a)

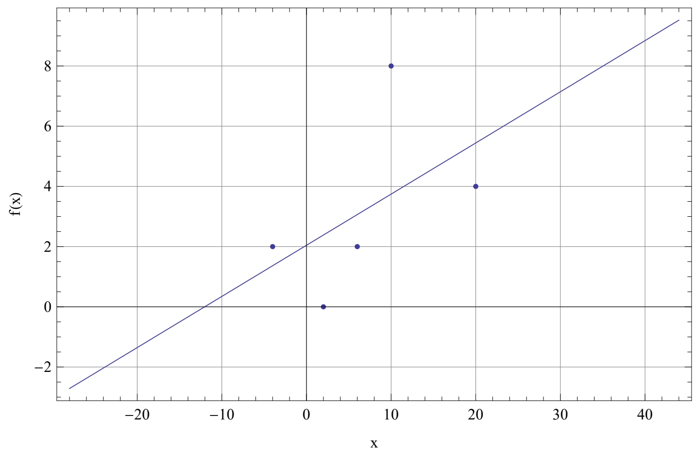
$$x_3 = \frac{\det(\begin{bmatrix} 7 & -3 & 7 \\ -6 & 3 & -6 \\ 3 & -1 & 4 \end{bmatrix})}{\det(\begin{bmatrix} 7 & -3 & 2 \\ -6 & 3 & -2 \\ 3 & -1 & 1 \end{bmatrix})} = 3$$

b)

$$x_2 = \frac{\det(\begin{bmatrix} -1 & -23 & 2 \\ 2 & 29 & -2 \\ -1 & -12 & 1 \end{bmatrix})}{\det(\begin{bmatrix} -1 & -5 & 2 \\ 2 & 5 & -2 \\ -1 & -2 & 1 \end{bmatrix})} = 5$$

16.2 Regressionsgerade

Beispiel 16.5 Regressionsgerade **MAXCIT**



i	x	y
1	-4	2
2	2	0
3	6	2
4	10	8
5	20	4

Erstellen Sie für die Punkte eine Regressionsgerade $y(x) = a \cdot x + b$. Gehen Sie dabei folgendermassen vor:

a) Stellen Sie das überbestimmte Gleichungssystem auf.

b) Bei x_1: quadrieren Sie den Abstand von $f(x_1)$ von y_1. Fahren Sie mit x_2 und x_3 fort und addieren Sie die Abstandsquadrate.

c) Wir benennen den entstandenen Ausdruck mit S. Leiten Sie S nach a ab

d) Leiten Sie S nach b ab

e) Da wir S minimieren wollen, setzen wir $\frac{\partial S}{\partial a} = 0$ und $\frac{\partial S}{\partial b} = 0$. Multiplizieren Sie die beiden vorherigen Ausdrücke aus, teilen Sie durch 2 und stellen Sie das lineare Gleichungssystem auf (Koeffizienten a und b links und Inhomogenität rechts)

f) Schreiben Sie das lineare Gleichungssystem mit einer Koeffizientenmatrix

g) Schreibe die Koeffizientenmatrix für n allgemeine datenpunkte (x_i, y_i)

h) Lösen Sie das Gleichungssystem

Lösung:

a) ÜberbestimmteS Gleichungssystem

$$\begin{array}{rcccccc} \text{bei } x_1: & a\cdot(-4) & + & b & = & 2 \\ \text{bei } x_2: & a\cdot 2 & + & b & = & 0 \\ \text{bei } x_3: & a\cdot 6 & + & b & = & 2 \\ \text{bei } x_4: & a\cdot 10 & + & b & = & 8 \\ \text{bei } x_5: & a\cdot 20 & + & b & = & 4 \end{array}$$

b) Abstandsquadrate:

$$\begin{array}{ll} [a\cdot(-4)+b-2]^2 & \\ & + \quad [2+b-0]^2 \\ & + \quad \dots \\ & + \quad [a\cdot 20+b-4]^2 \end{array}$$

c) Ableitung

$$\frac{dS}{da} = 2\cdot[a\cdot(-4)+b-2]\cdot(-4) + 2\cdot[a\cdot 2+b-0]\cdot 0 + \ldots + 2\cdot[a\cdot 20+b-4]\cdot 20$$

d) Ableitung

$$\frac{dS}{db} = 2\cdot[a\cdot(-4)+b-2] + 2\cdot[a\cdot 2+b-0] + \ldots + 2\cdot[a\cdot 20+b-4]$$

e) Gleich 0 setzen, durch 2 teilen, dann ausmultiplizieren

$$\frac{dS}{da} = a\cdot(-4)^2 + b\cdot(-4) - 2\cdot(-4) + 2^2 + b\cdot 2 - 0\cdot 2 + \ldots + a\cdot 20^2 + b\cdot 20 - 4\cdot 20 = 0$$

und

$$\frac{dS}{db} = a\cdot(-4) + b - 2 + 2 + b - 0 + \ldots + a\cdot 20 + b - 4 = 0$$

Lineares Gleichungssystem in Standard-Form

$$\left|\begin{array}{ccccc} [(-4)^2+2^2+\ldots+20^2]\cdot a & + & [(-4)+2+\ldots+20]\cdot b\cdot & = & -2\cdot(-4)+0\cdot 2+\ldots+4\cdot 20 \\ [(-4)+2+\ldots+20]\cdot a & + & \underbrace{[1+1+\ldots+1]}_{=n=5}\cdot b\cdot & = & 2+0+\ldots+4 \end{array}\right|$$

f) Koeffizientenmatrix

$$\begin{pmatrix} [(-4)^2 + 2^2 + \ldots + 20^2] & [(-4) + 2 + \ldots + 20] \\ [(-4) + 2 + \ldots + 20] & 5 \end{pmatrix} \odot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -2 \cdot (-4) + 0 \cdot 2 + \ldots + 4 \cdot 20 \\ 2 + 0 + \ldots + 4 \end{pmatrix}$$

also

$$\underbrace{\begin{pmatrix} 556 & 34 \\ 34 & 5 \end{pmatrix}}_{=\mathbf{M}} \odot \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} 164 \\ 16 \end{pmatrix}}_{\vec{y}'}$$

g) Koeffizientenmatrix allgemein

$$\begin{pmatrix} \sum_{i=1}^n (x_i)^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix} \odot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i \cdot y_i \\ \sum_{i=1}^n y_i \end{pmatrix}$$

h) Lösung des Gleichungssystems z.B. mit der Inversen

$$\begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{M}^{-1} \odot \vec{y}' = \begin{pmatrix} 0.1700 \\ 2.0443 \end{pmatrix}$$

Beispiel 16.6 Koeffizientenmatrix, Messvektor, Normalensystem QT4R27

$$\mathbf{F} = \begin{bmatrix} -4 & 1 \\ 2 & 1 \\ 6 & 1 \\ 10 & 1 \\ 20 & 1 \end{bmatrix}, \; \vec{y} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 8 \\ 4 \end{bmatrix}$$

a) Berechnen Sie

$$\mathbf{F}^\intercal \odot \mathbf{F} \text{ und } \mathbf{F}^\intercal \odot \vec{y}$$

b) Vergleichen Sie mit den Resultaten aus der vorherigen Aufgabe. Wie können Sie das Normalensystem mit Hilfe von $\mathbf{F}$ und $\vec{y}$ berechnen.

c) Betrachten Sie noch einmal die Lösung des Normalensystems $\vec{l} = \begin{pmatrix} 0.1700 \\ 2.0443 \end{pmatrix}$.

Welcher Koeffizient in $\vec{l}$ ist die Konstante? Welcher ist beschreibt die Steigung? Wie hängt die Reihenfolge dieser Koeffizienten in $\vec{l}$ mit der Struktur der Matrix $\mathbf{F}$ zusammen?

Lösung:

a)

$$\mathbf{F}^\intercal \odot \mathbf{F} = \begin{bmatrix} 556 & 34 \\ 34 & 5 \end{bmatrix} \text{ und } \mathbf{F}^\intercal \odot \vec{y} \begin{bmatrix} 164 \\ 16 \end{bmatrix}$$

b) Wir erhalten die Bestandteile des Normalensystems

$$\underbrace{\begin{pmatrix} 556 & 34 \\ 34 & 5 \end{pmatrix}}_{=\mathbf{M}=\mathbf{F}^\intercal \odot \mathbf{F}} \odot \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} 164 \\ 16 \end{pmatrix}}_{\vec{y}'=\mathbf{F}^\intercal \odot \vec{y}}$$

Das Normalensystem kann so mit mit Hilfe von $\mathbf{F}$ und $\vec{y}$ berechnet werden.

c) Die Reihenfolge der Koeffizienten in $\vec{l} = \begin{pmatrix} 0.1700 \\ 2.0443 \end{pmatrix}$ ergibt sich aus Struktur von **F**:

x	1
-4	1
2	1
6	1
10	1
20	1

Da die x-Werte in der ersten Spalte stehen, beschreibt der erste Koeffizient in $\vec{l}$ die Steigung, d.h. die Zahl vor x in der Regression und da die konstanten (1) in der zweiten Spalte stehen, beschreibt der zweite Koeffizient in $\vec{l}$ die Konstante.

Satz 16.1 Regressionsgerade

Die Gerade, die die Summe der Abstandsquadrate minimiert (**Regressionsgerade**) ist

$$y(x) = a \cdot x + b$$

ergibt sich aus dem Gleichungssystem

$$\mathbf{M} \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{F}^\intercal \odot \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

mit

$$\mathbf{F} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \\ x_n & 1 \end{pmatrix} \text{ und } \mathbf{M} = \mathbf{F}^\intercal \odot \mathbf{F}$$

Man kann zeigen, dass dies äquivalent ist zu

Satz 16.2 Regressionsgerade

Die Gerade, die die Summe der Abstandsquadrate minimiert (**Regressionsgerade**) ist

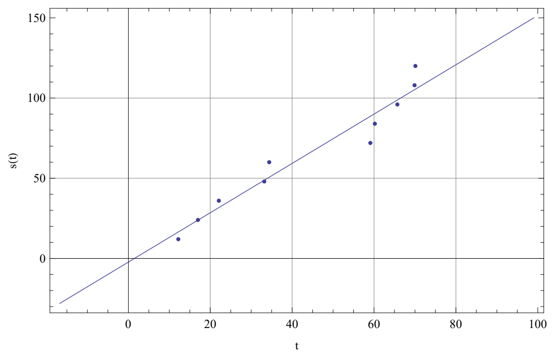
$$y(x) = a \cdot x + b$$

mit

$$a = \frac{s_{xy}}{(s_x)^2} \text{ und } b = \bar{y} - a \cdot \bar{x}$$

Beispiel 16.7 Karussel **DGGBKJ**

Zeit t [s]	Weg s [m]	Zeit t [s]	Weg s [m]
12.2	12	72	59.1
17	24	60.2	84
22.1	36	65.7	96
33.2	48	69.9	108
34.4	60	70.1	120



a) Wir betrachten ein Karussel, das sich gleichmässig dreht. Beschreiben Sie mathematisch: Welche Strecke legt ein Objekt auf dem Karussel in der Zeit t zurück? Welches ist dabei die unabhängige Variable? Welches ist die abhängige Variable?

b) Erstellen Sie eine Regression für die Daten $s(t) = a \cdot t + b$

c) Welche physikalischen Grössen sind in a und b enthalten?

Lösung:

a) Wir erwarten eine gleichmässige Bewegung $s(t) = v \cdot t + s_0$. Dabei ist der Ort s die abhängige Variable und t die unabhängige Variable.

b) Regression unten

c) a ist die Geschwindigkeit und b der Ort zum Zeitpunkt $t = 0$.

$$\mathbf{F} = \begin{pmatrix} 12.2 & 1 \\ 17.0 & 1 \\ 22.1 & 1 \\ 33.2 & 1 \\ 34.4 & 1 \\ 59.1 & 1 \\ 60.2 & 1 \\ 65.7 & 1 \\ 69.9 & 1 \\ 70.1 & 1 \end{pmatrix} \text{ und } \mathbf{M} = \mathbf{F}^{\mathsf{T}} \odot \mathbf{F} = \begin{pmatrix} 24445 & 444 \\ 444 & 10 \end{pmatrix}$$

ergeben das Normalensystem

$$\mathbf{M} \odot \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} 36588 \\ 660 \end{pmatrix}}_{= \mathbf{F}^{\mathsf{T}} \odot \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}}$$

mit der Lösung

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1.5379 \\ -2.2693 \end{pmatrix} \text{ für } s(t) = a \cdot t + b$$

16.3 Allgemeine lineare Regression

Beispiel 16.8 Allgemeine Regression Herleitung **UL1Q3E**

$f(x) = a_1 \cdot f_1(x) + a_2 \cdot f_2(x)$ soll möglichst nahe durch die gemessenen Punkte verlaufen.

$$\binom{x_1}{y_1}\,,\ \binom{x_2}{y_2}\,,\ \binom{x_3}{y_3}$$

a) Stellen Sie das überbestimmte Gleichungssystem auf.

b) Bei x_1: quadrieren Sie den Abstand von $f(x_1)$ von y_1. Fahren Sie mit x_2 und x_3 fort und addieren Sie die Abstandsquadrate.

c) Wir benennen den entstandenen Ausdruck mit S. Leiten Sie S nach a_1 ab

d) Leiten Sie S nach a_2 ab

e) Da wir S minimieren wollen, setzen wir $\frac{\partial S}{\partial a_1} = 0$ und $\frac{\partial S}{\partial a_2} = 0$. Multiplizieren Sie die beiden vorherigen Ausdrücke aus, teilen Sie durch 2 und stellen Sie das lineare Gleichungssystem auf (Koeffizienten a_1 und a_2 links und Inhomogenität rechts)

f) Schreiben Sie das lineare Gleichungssystem, das sogenannte Normalensystem, mit einer Koeffizientenmatrix.

g) Wir schreiben die Matix $\mathbf{F}$ und den Messvektor wie folgt

$$\mathbf{F} = \begin{pmatrix} f_1(x_1) & f_2(x_1) \\ f_1(x_2) & f_2(x_2) \\ f_1(x_3) & f_2(x_3) \end{pmatrix}\,,\ \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Wie können Sie das Normalensystem mit Hilfe dieser Definitionen bestimmen?

h) Verallgemeinern Sie das Resultat für $f(x) = \sum_{i=1}^{m} a_i \cdot f_i(x)$ und die Punkte $P_1(x_1, y_1), \ldots, P_n(x_n, y_n)$

Lösung:

a) Überbestimmtes Gleichungssystem

$$\begin{array}{lccccc} \text{bei } x_1: & a_1 \cdot f_1(x_1) & + & a_2 \cdot f_2(x_1) & = & y_1 \\ \text{bei } x_2: & a_1 \cdot f_1(x_2) & + & a_2 \cdot f_2(x_2) & = & y_2 \\ \text{bei } x_3: & a_1 \cdot f_1(x_3) & + & a_2 \cdot f_2(x_3) & = & y_3 \end{array}$$

b) Abstandsquadrate:

$$\begin{array}{rcl} [a_1 \cdot f_1(x_1) + a_2 \cdot f_2(x_1) - y_1]^2 & + & [a_1 \cdot f_1(x_2) + a_2 \cdot f_2(x_2) - y_2]^2 \\ & + & [a_1 \cdot f_1(x_3) + a_2 \cdot f_2(x_3) - y_3]^2 \end{array}$$

c) Ableitung von S nach a_1

$$\begin{array}{rcl} \frac{\partial S}{\partial a_1} = & & 2 \cdot [a_1 \cdot f_1(x_1) + a_2 \cdot f_2(x_1) - y_1] \cdot f_1(x_1) \\ & + & 2 \cdot [a_1 \cdot f_1(x_2) + a_2 \cdot f_2(x_2) - y_2] \cdot f_1(x_2) \\ & + & 2 \cdot [a_1 \cdot f_1(x_3) + a_2 \cdot f_2(x_3) - y_3] \cdot f_1(x_3) \end{array}$$

d) Ableitung von S nach a_2

$$\begin{aligned}\frac{\partial S}{\partial a_2} = \quad & 2 \cdot [a_1 \cdot f_1(x_1) + a_2 \cdot f_2(x_1) - y_1] \cdot f_2(x_1) \\ + \quad & 2 \cdot [a_1 \cdot f_1(x_2) + a_2 \cdot f_2(x_2) - y_2] \cdot f_2(x_2) \\ + \quad & 2 \cdot [a_1 \cdot f_1(x_3) + a_2 \cdot f_2(x_3) - y_3] \cdot f_2(x_3)\end{aligned}$$

e) Lineares Gleichungssystem :

$$\begin{aligned} & a_1 \cdot f_1(x_1) \cdot f_1(x_1) + a_2 \cdot f_2(x_1) \cdot f_1(x_1) - y_1 \cdot f_1(x_1) \\ + \quad & a_1 \cdot f_1(x_2) \cdot f_1(x_2) + a_2 \cdot f_2(x_2) \cdot f_1(x_2) - y_2 \cdot f_1(x_2) \\ + \quad & a_1 \cdot f_1(x_3) \cdot f_1(x_3) + a_2 \cdot f_2(x_3) \cdot f_1(x_3) - y_3 \cdot f_1(x_3) \quad = 0\end{aligned}$$

also

$$\begin{aligned} & a_1 \cdot [f_1(x_1) \cdot f_1(x_1) + f_1(x_2) \cdot f_1(x_2) + f_1(x_3) \cdot f_1(x_3)] \\ + \quad & a_2 \cdot [f_2(x_1) \cdot f_1(x_1) + f_2(x_2) \cdot f_1(x_2) + f_2(x_3) \cdot f_1(x_3)] \quad = y_1 \cdot f_1(x_1) + y_2 \cdot f_1(x_2) + y_3 \cdot f_1(x_3)\end{aligned}$$

und genau gleich

$$\begin{aligned} & a_1 \cdot [f_1(x_1) \cdot f_2(x_1) + f_1(x_2) \cdot f_2(x_2) + f_1(x_3) \cdot f_2(x_3)] \\ + \quad & a_2 \cdot [f_2(x_1) \cdot f_2(x_1) + f_2(x_2) \cdot f_2(x_2) + f_2(x_3) \cdot f_2(x_3)] \quad = y_1 \cdot f_2(x_1) + y_2 \cdot f_2(x_2) + y_3 \cdot f_2(x_3)\end{aligned}$$

f) lineare Gleichungssystem mit Koeffizientenmatrix:

$$\begin{bmatrix} f_1(x_1)\, f_1(x_1) + f_1(x_2)\, f_1(x_2) + f_1(x_3)\, f_1(x_3) & f_2(x_1)\, f_1(x_1) + f_2(x_2)\, f_1(x_2) + f_2(x_3)\, f_1(x_3) \\ f_1(x_1)\, f_2(x_1) + f_1(x_2)\, f_2(x_2) + f_1(x_3)\, f_2(x_3) & f_2(x_1) f_2(x_1) + f_2(x_2)\, f_2(x_2) + f_2(x_3)\, f_2(x_3) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$= \begin{bmatrix} y_1\, f_1(x_1) + y_2\, f_1(x_2) + y_3\, f_1(x_3) \\ y_1\, f_2(x_1) + y_2\, f_2(x_2) + y_3\, f_2(x_3) \end{bmatrix}$$

g) Normalensystem

$$\underbrace{\mathbf{M}}_{\mathbf{F}^\intercal \odot \mathbf{F}} \odot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \underbrace{\vec{y}'}_{\mathbf{F}^\intercal \odot \vec{y}}$$

h) Werden die Funktionen $f_i(x)$ nacheinander auf die unabhängige Variable x_i angewendet und die Resultate in die Spalten von $\mathbf{F}$ geschrieben, lässt sich das Resultat für beliebig viele Funktionen verallgemeinern.
Hier noch die Verallgemeinerung für $f(x) = \sum_{i=1}^m a_i \cdot f_i(x)$ und die Punkte $P_1(x_1, y_1), \ldots, P_n(x_n, y_n)$

$$\begin{bmatrix} \sum_{i=1}^n f_1(x_i)\, f_1(x_i) & \ldots & \sum_{i=1}^n f_m(x_i)\, f_1(x_i) \\ \vdots & & \vdots \\ \sum_{i=1}^n f_1(x_i)\, f_m(x_i) & \ldots & \sum_{i=1}^n f_m(x_i)\, f_m(x_i) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i\, f_1(x_i) \\ \vdots \\ \sum_{i=1}^n y_i\, f_m(x_i) \end{bmatrix}$$

Beispiel 16.9 Allgemeine Regression (forts.) **CFQDP5**

Multiplizieren Sie die Transponierte der Koeffizientenmatrix $\mathbf{F}^\intercal$ mit dem Messvektor $\vec{y}$

	$f_1(x)$	$f_2(x)$
x_1 :	$f_1(x_1)$	$f_2(x_1)$
x_2 :	$f_1(x_2)$	$f_2(x_2)$
x_3 :	$f_1(x_3)$	$f_2(x_3)$

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Lösung:

$$\mathbf{F}^\intercal \odot \vec{y} = \begin{pmatrix} y_1 \cdot f_1(x_1) + y_2 \cdot f_1(x_2) + y_3 \cdot f_1(x_3) \\ y_1 \cdot f_2(x_1) + y_2 \cdot f_2(x_2) + y_3 \cdot f_2(x_3) \end{pmatrix}$$

Beispiel 16.10 Allgemeine Regression (forts.) **B3C86W**

Multiplizieren Sie die Transponierte der Koeffizientenmatrix $\mathbf{F}^\intercal$ mit der Koeffizientenmatrix $\mathbf{F}$

$$\mathbf{F} = \begin{bmatrix} f_1(x_1) & f_2(x_1) \\ f_1(x_2) & f_2(x_2) \\ f_1(x_3) & f_2(x_3) \end{bmatrix}$$

Lösung:

$$\mathbf{F}^\intercal \odot \mathbf{F} = \begin{pmatrix} f_1(x_1)\cdot f_1(x_1) + f_1(x_2)\cdot f_1(x_2) + f_1(x_3)\cdot f_1(x_3) & f_2(x_1)\cdot f_1(x_1) + f_2(x_2)\cdot f_1(x_2) + f_2(x_3)\cdot f_1(x_3) \\ f_1(x_1)\cdot f_2(x_1) + f_1(x_2)\cdot f_2(x_2) + f_1(x_3)\cdot f_2(x_3) & f_2(x_1) f_2(x_1) + f_2(x_2)\cdot f_2(x_2) + f_2(x_3)\cdot f_2(x_3) \end{pmatrix}$$

16.4 Fehler der Regressions-Koeffizienten

Definition 16.2 Mittlere Streuung der Messwerte um die Regressionsgerade

Aus der Abweichung von der Regressionsgerade

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{pmatrix} = \begin{pmatrix} a\cdot x_1 + b - y_1 \\ a\cdot x_2 + b - y_2 \\ \dots \\ a\cdot x_n + b - y_n \end{pmatrix}$$

kann die Mittlere Streuung der Messwerte um die Regressionsgerade berechnet werden:

$$\sigma_y = \sqrt{\frac{1}{n-2}\cdot\sum_{i=1}^{n}(\epsilon_i)^2}$$

Satz 16.3 Fehler der Regressions-Koeffizienten

Für den Datensatz (x_i, y_i) mit n Punkten und den linearen Regressions-Koeffizienten a und b ($y = a\cdot x + b$), sind die Standard-Fehler der Regressions-Koeffizienten

$$s_a = \sigma_y\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = \sigma_y\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i)^2 - n\overline{x}^2}}$$

und C

$$s_b = s_a\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i)^2}$$

Beispiel 16.11 Standard-Fehler der Regressions-Koeffizienten 1 **EY6K99**

i	x	y	x^2	ϵ^2
1	-4	2	16.	0.40382
2	2	0	4.	5.68458
3	6	2	36.	1.13218
4	10	8	100.	18.1149
5	20	4	400.	2.08326

Für diesen Datensatz (x_i, y_i) mit $n = 5$ Punkten finden wir die Regressions-Gerade $y = a\cdot x + b = 0.1700\cdot x + 2.0443$.

a) Berechnen Sie den Standard-Fehler auf a

b) Berechnen Sie den Standard-Fehler auf b

Lösung:

a) $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 324.8$, $\sigma_y = 3.2000$, $\overline{x} = 6.8$ und damit

$$s_a = \sigma_y \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 0.167747$$

b)

$$s_b = s_a \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i)^2} = 1.3520$$

C

Beispiel 16.12 Standard-Fehler der Regressions-Koeffizienten 2 A21295

i	t [s]	s [m]	t^2 [s^2]	$(\epsilon)^2$ [m^2]
1	12.2	12	148.84	20.1927
2	17	24	289.	0.0154376
3	22.1	36	488.41	18.3247
4	33.2	48	1102.24	0.624767
5	34.4	60	1183.36	87.6854
6	59.1	72	3492.81	276.329
7	60.2	84	3624.04	39.8776
8	65.7	96	4316.49	7.69262
9	69.9	108	4886.01	7.65675
10	70.1	120	4914.01	209.077

Für diesen Datensatz (x_i, y_i) mit $n = 10$ Punkten finden wir die Regressions-Gerade $y = 1.53794 \cdot t - 2.26927$

a) Berechnen Sie den Standard-Fehler auf a

b) Berechnen Sie den Standard-Fehler auf b

Lösung:

a) $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 4740.5$, $\sigma_y = 14.9162$, $\overline{x} = 44.39$ und damit

$$s_a = \sigma_y \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 0.2166$$

b)

$$s_b = s_a \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i)^2} = 6.7707$$

Infobox 16.2 Konfidenz-Intervalle für die Regressionskoeffizienten

Konfidenz-Intervalle zur Irrtums-Wahrscheinlichkeit α und dem Quantil $t_{n-2;\,1-\alpha/2}$ der Student-t-Verteilung sind

$$\begin{aligned} a &\in [a - s_a \cdot t_{n-2;\,1-\alpha/2},\; a + s_a \cdot t_{n-2;\,1-\alpha/2}] \\ b &\in [b - s_b \cdot t_{n-2;\,1-\alpha/2},\; b + s_b \cdot t_{n-2;\,1-\alpha/2}] \end{aligned}$$

Beispiel 16.13 Konfidenzintervalle der Regressionsparameter **RIMLG8**

Wir betrachten die Regressionsgerade $y = a \cdot x + b$ mit

$$\begin{aligned} a &= 0.17 \pm 0.05 \\ b &= 2.0 \pm 0.7 \end{aligned}$$

Sie wurde aus $n = 5$ Datenpunkten bestimmt.

a) Berechnen Sie die Konfidenzintervalle der Regressionsparameter zur Wahrscheinlichkeit 90%.

b) Berechnen Sie die Konfidenzintervalle der Regressionsparameter zur Wahrscheinlichkeit 50%.

Lösung:

a) Aus der Tabelle T.3 oder mit tinv bestimmen wir

$$t_{n-2;\,1-\alpha/2} = t_{5-2;\,1-0.1/2} = 2.35336$$

also ergibt sich

$$\begin{aligned} a &\in [0.052,\; 0.288] \\ b &\in [0.352646,\; 3.64735] \end{aligned}$$

b) Aus der Tabelle T.3 oder mit tinv bestimmen wir

$$t_{n-2;\,1-\alpha/2} = t_{5-2;\,1-0.5/2} = 0.7649$$

also ergibt sich

$$\begin{aligned} a &\in [0.1318,\; 0.2082] \\ b &\in [1.4646,\; 2.5354] \end{aligned}$$

C

KAPITEL 17

Fehlerfortpflanzung

Lernziele 17.1 Fehlerfortpflanzung

- Die Studierenden können den zu erwartenden Fehler einer Funktion in mehreren Veränderlichen abschätzen.

Satz 17.1 Fehlerfortpflanzung

Wir betrachten die Grössen $F(a,b)$, $S(a,b)$ und $P(a,b)$, die von den Variablen

$$\begin{aligned} a &= \overline{a} \pm \Delta a \\ b &= \overline{b} \pm \Delta b \end{aligned}$$

abhängen, deren Mittelwert $\overline{a}$ und Standardabweichung (Δa) bekannt sind. Daraus können wir den zu erwartenden Fehler berechnen:

- Fehler bei funktionaler Abhängigkeit

$$F(a) \;\Rightarrow\; \Delta F = \left.\frac{\partial F(a)}{\partial a}\right|_{a=\overline{a}} \cdot \Delta a$$

- Mehrere Veränderliche $F(a,b)$

$$\Delta F = \left.\sqrt{\left(\frac{\partial F(a)}{\partial a} \cdot \Delta a\right)^2 + \left(\frac{\partial F(b)}{\partial b} \cdot \Delta b\right)^2}\,\right|_{a=\overline{a};\; b=\overline{b}}$$

- Fehler bei Summe/Differenz $S(a,b) = a + b$ oder $S(a,b) = a - b$

$$|\Delta S| = \sqrt{(\Delta a)^2 + (\Delta b)^2}$$

- Relativer Fehler bei Produkt/Quotient $P(a,b) = a^{\alpha} \cdot b^{\beta}$ und $\alpha, \beta \in \mathbb{Z}$

$$\left|\frac{\Delta P}{P}\right| = \sqrt{\left(\alpha \frac{\Delta a}{a}\right)^2 + \left(\beta \frac{\Delta b}{b}\right)^2}$$

Beispiel 17.1 Oberfläche eines Zylinders **2S72IS**

$$\overline{r} = 10.5 \pm 0.2\,\mathrm{cm};\ \overline{h} = 15.5 \pm 0.3\,\mathrm{cm}$$

Berechne Standardabweichung der Oberfläche $A = 2\pi r^2 + 2\pi r \cdot h$

a) Ableitungen berechnen

b) Mittelwerte einsetzen

Lösung:

$$\frac{\partial A}{\partial r} = 4\pi r + 2\pi h;\ \frac{\partial A}{\partial h} = 2\pi r$$

Mittelwerte einsetzen:

$$\frac{\partial A}{\partial r} = 4\pi\overline{r} + 2\pi\overline{h} = 226.19\,\mathrm{cm};\ \frac{\partial A}{\partial h} = 2\pi\overline{r} = 65.97\,\mathrm{cm}$$

$$\begin{aligned} \Delta A &= \sqrt{(226.19\,\mathrm{cm} \cdot 0.2\,\mathrm{cm})^2 + (65.97\,\mathrm{cm} \cdot 0.3\,\mathrm{cm})^2} \\ &= 49\,\mathrm{cm}^2 \end{aligned}$$

Beispiel 17.2 Federkonstante **L8YMS5**

$$c = 4\pi^2 \frac{m}{T^2}$$

Berechne Standardabweichung der Federkonstante

$$\overline{m} = 200 \pm 0.68\,\mathrm{g};\ \overline{T} = 2.00 \pm 0.0105\,\mathrm{s}$$

Lösung:

$$\begin{aligned} \frac{\partial c}{\partial m} &= 4\pi^2 \cdot \frac{m}{T^2} = 9.8696\,\mathrm{s}^{-2} \\ \frac{\partial c}{\partial T} &= -8\pi^2 \cdot \frac{m}{T^3} = -1973.9\,\mathrm{g/s}^3 \\ \Delta c &= \sqrt{(9.8696\,\mathrm{s}^{-2} \cdot 0.68\,\mathrm{g})^2 + (-1973.9\,\mathrm{g/s}^3 \cdot 0.0105\,\mathrm{s})^2} \\ &= 21.79\,\mathrm{g/s}^2 = 0.022\,\mathrm{N/m} \end{aligned}$$

Federkonstante:

$$c = 1.974 \pm 0.022\,\mathrm{N/m}$$

Beispiel 17.3 Federkonstante (kurze Rechnung) **68ZWST**

$$c = 4\pi^2 \frac{m}{T^2}$$

$$\overline{m} = 200 \pm 0.68\,\mathrm{g};\ \overline{T} = 2,00 \pm 0.0105\,\mathrm{s}$$

Berechne Standardabweichung der Federkonstante

Lösung:

$$\begin{aligned}\Delta c &= c \cdot \sqrt{(1 \cdot 0.68/200)^2 + (-2 \cdot 0.0105/2)^2} \\ &= 0.022\end{aligned}$$

Federkonstante:

$$c = 1.974 \pm 0.022\,\mathrm{N/m}$$

Beispiel 17.4 Reihenschaltung Widerstände **77FF3E**

$$R_1 = 100 \pm 2\Omega;\ R_2 = 150 \pm 2\Omega;\ R_3 = 50 \pm 1\Omega$$

Berechnen Sie nacheinander

a) Gesamtwiderstand

b) Fehler des Gesamtwiderstands

Lösung:

a) Gesamtwiderstand

$$R_{\text{tot}} = 100 + 150 + 50 = 300$$

b) Fehler

$$\Delta R_{\text{tot}} = \sqrt{(2\Omega)^2 + (2\Omega)^2 + (1\Omega)^2} = 3\ \Omega$$

$$R_{\text{tot}} = 300 \pm 3\ \Omega$$

Beispiel 17.5 Karussel **TKHAGL**

Wir berechnen die Geschwindigkeit eines Autos auf einem Kinderkarussell. Das Auto verfügt über kein Tachometer. Unten finden sie Umlaufszeiten und zurückgelegte Strecken (ergibt sich aus Radius des Karussels).

Weg s [m]	Zeit t [s]	Weg s [m]	Zeit t [s]
12	12.2	72	59.1
24	17	84	60.2
36	22.1	96	65.7
48	33.2	108	69.9
60	34.4	120	70.1

a) Erstellen Sie eine Regression für die Daten $t(s) = a \cdot s + b$

b) Berechnen Sie die Fehler der Regressionskoeffizienten

Lösung:

a) Regression

$$\mathbf{F} = \begin{pmatrix} 12 & 1 \\ 24 & 1 \\ 36 & 1 \\ 48 & 1 \\ 60 & 1 \\ 72 & 1 \\ 84 & 1 \\ 96 & 1 \\ 108 & 1 \\ 120 & 1 \end{pmatrix} \text{ und } \mathbf{M} = \mathbf{F}^\intercal \odot \mathbf{F} = \begin{pmatrix} 55440 & 660 \\ 660 & 10 \end{pmatrix}$$

ergeben das Normalensystem

$$\mathbf{M} \odot \begin{pmatrix} a \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} 36588 \\ 444 \end{pmatrix}}_{= \mathbf{F}^\intercal \odot \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}}$$

mit der Lösung

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0.6137 \\ 3.8867 \end{pmatrix} \text{ für } t(s) = a \cdot s + b$$

b) Wir erhalten

$$s_a = \sigma_y \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \overline{x})^2}} = 5.77 \cdot 0.0092 = 0.0529$$

mit den Einheiten s/m und

$$s_b = s_a \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i)^2} = 0.0529 \cdot 74.4580 = 3.9417$$

in den Einheiten s.

Beispiel 17.6 Karussel Geschwindigkeit Fehler **MUY23W**

Berechnen Sie aus den Regressions-Koeffizienten und den dazugehörigen Fehlern den Fehler der Geschwindigkeit des Autos. **Lösung:**

Die Geschwindigkeit ist

$$v = 1/a = 1.6295 \, \frac{\mathrm{m}}{\mathrm{s}}$$

Der Fehler pflanzt sich wie folgt fort

$$\Delta v = \sqrt{(\frac{dv}{da} \cdot s_a)^2} = \left| -\frac{1}{a^2} \cdot s_a \right| = \frac{1}{a^2} \cdot s_a = 0.1406$$

in m/s. Also ist die Angabe für die Geschwindigkeit

$$v = 1.62 \pm 0.14 \, \mathrm{m/s}$$

d.h. wir runden den Fehler auf maxiamal zwei Stellen (oft sogar auf eine Stelle).

Literaturverzeichnis

Walter Krämer. *So lügt man mit Statistik*. Campus Verlag, 2015.

Marcel Steiner. *Wahrscheinlichkeitstheorie und Statistik (wst)*. Hochschulskript, 2015.

Ibrahim Wazir, Tim Garry, Peter Ashbourne, Paul Barclay, Peter Flynn, Kevin Frederick, and Mike Wakeford. *Mathematics 2012 Edition-Higher Level: Developed Specifically for the IB Diploma*. Pearson, 2012.