I Wichtiges aus der Vorlesung

Merkmalstyp, Messniveau

Merkmalstyp	Messniveau	Beispiele
qualitatives Merkmal	Nominalskala	Familienstand, Geschlecht, Farbe, Bundesland
Rangmerkmale	Ordinalskala	Rangplätze bei Wettbewerben
quantitatives Merkmal (metrisches Merkmal)	Intervallskala	Nullpunkt NICHT absolut festgelegt, z. B. (∗)
diskret / stetig	Verhältnisskala	Nullpunkt eindeutig festgelegt, z. B. (∗∗)

(∗) Bei **Intervallskalen** sind Addition und Subtraktion der Messwerte zur Bestimmung eines Mittelwertes sinnvoll, NICHT jedoch die Multiplikation und Division der Werte untereinander.
(∗∗) Bei **Verhältnisskalen** sind auch Multiplikation und Division der Messwerte untereinander sinnvoll.

Merkmale mit nur zwei Ausprägungen (z. B. ja–nein, Erfolg–Misserfolg) heißen auch **dichotone Merkmale**.

1.1 Häufigkeiten

Sei $X\colon M \longrightarrow A$. Dann heißt

$$h(X=a) := \#\{X=a\} = \#\left\{m \in M \;\middle|\; X(m)=a\right\}$$

die **(absolute) Häufigkeit** der Ausprägung a und

$$r(X=a) := \frac{h(X=a)}{\#M}$$

die **relative Häufigkeit** von a.
Es gelten:

$$\sum_{a\in M} h(X=a) = \#M$$

$$\sum_{a\in M} r(X=a) = 1$$

1.2 arithmetisches Mittel

Sei $M=\{1,\ldots,n\}$, $X\colon M \longrightarrow A$ eine Abbildung und $A \subset \mathbb{R}$ (z. B. für die Ausprägungen einer Intervall- oder Verhältnisskala). Zu X gehört die Datenreihe $x_1,\ldots,x_n$. Dann heißt

$$\bar{x} := \frac{1}{n}(x_1 + \ldots + x_n)$$

das **arithmetische Mittel**[1] der Datenreihe. Es ist die am häufigsten verwendete Maßzahl, um den mittleren Wert einer Datenreihe zu charakterisieren.

[1] *englisch* arithmetic mean

Für $A = \{a_1, \ldots, a_n\}$ kann das arithmetische Mittel aus den absoluten oder relativen Häufigkeiten berechnet werden durch

(i) $\bar{x} = \frac{1}{n}\left(\sum_{j=1}^{k} a_j h(X = a_j)\right)$

(ii) $\bar{x} = \sum_{j=1}^{k} a_j r(X = a_j)$

Für $x_1, \ldots, x_n \in \mathbb{R}$ gilt: $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

1.3 Linearitätseigenschaft des arithmetischen Mittels

Vor: Es seien $x_1, \ldots, x_n \in \mathbb{R}$ und $y_1, \ldots, y_n \in \mathbb{R}$ zwei Datenreihen, $a, b \in \mathbb{R}$.

Beh:

(a) $z_i = x_i + y_i \implies \bar{z} = \bar{x} + \bar{y}$

(b) $z_i = a \cdot x_i + b \implies \bar{z} = a \cdot \bar{x} + b$

1.4 Minimalitätseigenschaft des arithmetischen Mittels bzgl. der quadratischen Abweichung

Vor: Es sei $x_1, \ldots, x_n \in \mathbb{R}$. Dann ist:

Beh:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \leq \sum_{i=1}^{n}(x_i - d)^2 \quad \forall\, d \in \mathbb{R}$$

1.5 Stichproben-Varianz, Standardabweichung, Streuungsmaßzahlen

Es sei $x_1, \ldots, x_n \in \mathbb{R}$ eine Datenreihe einer I- oder V-Skala.

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

heißt **Stichproben-Varianz** und

$$s := \sqrt{s^2}$$

heißt **Standardabweichung** der Datenreihe $x_1, \ldots, x_n$ bezüglich der Stichproben-Varianz. s^2 und s sind sogenannte **Streuungsmaßzahlen**.

Liegen die Daten (Messwerte) in Form einer Häufigkeitstabelle vor, wobei $A = \{a_1, \ldots, a_k\}$, so gilt:

$$s^2 = \frac{1}{n-1} \cdot \sum_{j=1}^{k}(a_j - \bar{x})^2 \cdot h(X = a_j)$$

Vor: Es sei $x_1, \ldots, x_n \in \mathbb{R}$ eine Datenreihe und $z_i = ax_i + b$ $(a, b \in \mathbb{R})$

Beh: $s_z^2 = a^2 s_x^2$ und $s_z = |a| \cdot s_x$

Anmerkung:
$a = 1$ bedeutet, dass die Datenreihe x_i um die Konstante b verschoben wird, d. h. das arithmetische Mittel ändert sich um den Wert b, Varianz und Standardabweichung ändern sich NICHT.

1.6 erwartungstreuer Schätzwert

(a) s ist ein sogenannter **erwartungstreuer Schätzwert** für die Standardabweichung einer Grundgesamtheit. Bei Gesamterhebungen und in manchen Lehrbüchern wird

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

als Varianz und

$$s' := \sqrt{s'^2}$$

als Standardabweichung verwendet.

(b) Umrechung:

$$s'^2 = \frac{n-1}{n} \cdot s^2 \quad , \quad s' = \sqrt{s'^2} = \sqrt{\frac{n-1}{n}} \cdot s$$

Bei sehr großen n ist der Unterschied zwischen s und s' vernachlässigbar, z. B. ist $n = 1000 \implies \frac{n-1}{n} = 0.999$, $\sqrt{\frac{n-1}{n}} = 0.9995$.

1.7 empirische Verteilungsfunktion, kumulierte relative Häufigkeit

X sei ein Merkmal mit Ausprägung $A = \{a_1, \ldots, a_k\}$ mit $a_1 < a_2 < \ldots < a_k$
($X \colon M \longrightarrow A$, M Individuen, Objekte)
Die Funktion $F \colon \mathbb{R} \longrightarrow [0, 1]$ definiert durch

$$F(x) = r(X \leq x) := \sum_{a_j \leq x} r(X = a_j)$$

heißt **(empirische) Verteilungsfunktion** des Merkmals X.

$r(X \leq x)$ heißt **kumulierte relative Häufigkeit**.

1.8 empirischer Median, Zentralwert, zentrales Wertepaar

$x_1, \ldots, x_n$ seien Werte einer O-, I- oder V-Skala.

(a) Der **empirische Median**[2] ist eine Lagemaßzahl

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

[2] *englisch* median

In manchen Lehrbüchern wird der Median auch folgendermaßen definiert:

$$\tilde{x}' := \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{falls } n \text{ ungerade} \\ x_{\left(\frac{n}{2}\right)} & \text{falls } n \text{ gerade} \end{cases}$$

(b) Für n ungerade heißt $x_{\left(\frac{n+1}{2}\right)}$ auch **Zentralwert** der Datenreihe und für n gerade heißt $\left(x_{\left(\frac{n}{s}\right)}, x_{\left(\frac{n+2}{2}\right)}\right)$ **zentrales Wertepaar** der Datenreihe.

1.9 Median

$X\colon \underbrace{\{1,\ldots,n\}}_{=M} \longrightarrow A.$

Eigenschaft des Median

$$\#\left\{j \in M \;\middle|\; x_j < \tilde{x}\right\} \leq \frac{n}{2}$$
$$\#\left\{j \in M \;\middle|\; x_j > \tilde{x}\right\} \leq \frac{n}{2}$$
$$\#\left\{j \in M \;\middle|\; x_j \leq \tilde{x}\right\} \geq \frac{n}{2}$$

Mindestens die Hälfte der Messwerte haben einen Wert kleiner oder gleich dem Median.

$$\#\left\{j \in M \;\middle|\; x_j < \tilde{x}\right\} \leq \frac{n}{2} \leq \#\left\{j \in M \;\middle|\; x_j \leq \tilde{x}\right\}$$

1. Fall

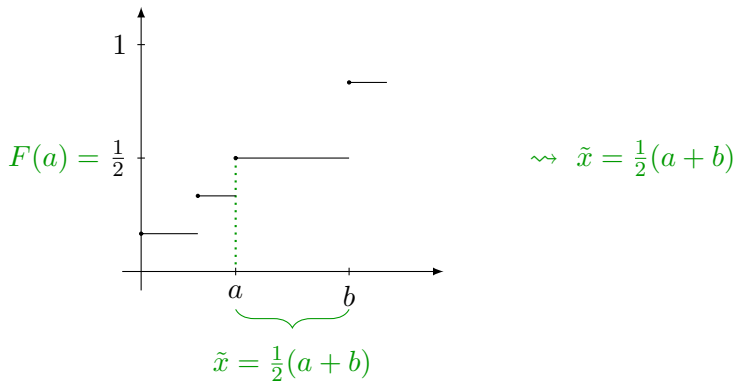


Abbildung 1.1: *z. B. n gerade, $n = 10$*

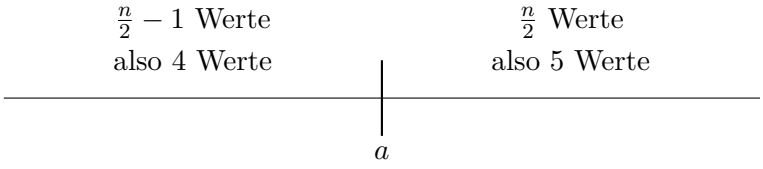


Abbildung 1.2: $F(a) = r(X \leq a) = \frac{5}{10} = \frac{1}{2}$, $\tilde{x} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}\right)$

2. Fall

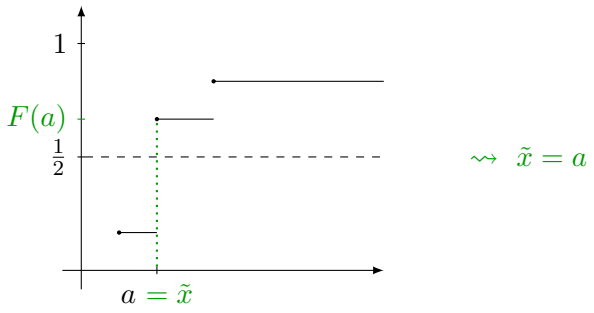


Abbildung 1.3: *z. B. n ungerade,* $n = 11$

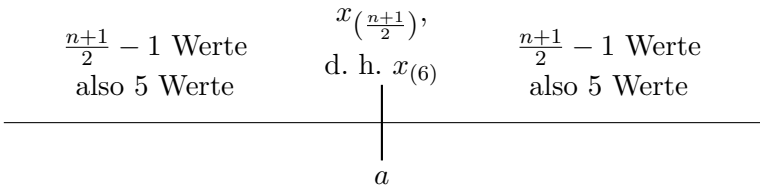


Abbildung 1.4: $F(a) = r(X \leq a) = \frac{6}{11} > \frac{1}{2} \quad , \quad \tilde{x} = x_{(\frac{n+1}{2})}$

1.10 Minimalitätseigenschaft des Median bzgl. der Abweichung zum Betrag

Es seine $x_1, \ldots, x_n \in \mathbb{R}$. Dann ist

$$\sum_{i=1}^{n} |x_i - \tilde{x}| \leq \sum_{i=1}^{n} |x_i - d| \quad \forall\, d \in \mathbb{R}$$

1.11 α-Quantil, r-tes Perzentil

$x_1, \ldots, x_n$ seien Werte einer O-, I- oder V-Skala.

(a) Für $0 < \alpha < 1$ ist das **α-Quantil** definiert durch

$$x_\alpha := \begin{cases} \frac{1}{2}\left(x_{(n\alpha)} + x_{(n\alpha+1)}\right) & \text{falls } n\alpha \text{ ganzzahlig ist} \\ x_{([n\alpha+1])} & \text{sonst} \end{cases}$$

wobei $[n\alpha + 1]$ größte ganze Zahl ist, die kleiner oder gleich $n\alpha + 1$ ist, z. B. $[7.2] = 7$, $[7.8] = 7$, $[7] = 7$.

(b) Das $\frac{r}{100}$-Quantil (für $r = 1, \ldots, 99$) heißt auch ***r*-tes Perzentil**.

Das 50-te Perzentil $\left(\alpha = \frac{1}{2}\right)$ ist der Median.
Das 25-te Perzentil $\left(\alpha = \frac{1}{4}\right)$ heißt auch **unteres Quartil**.
Das 75-te Perzentil $\left(\alpha = \frac{3}{4}\right)$ heißt auch **oberes Quartil**.

Der **Quartilabstand** ist ein Streuungsmaß zum Median und ist definiert durch

$$\text{Quartilabstand} := \text{oberes Quartil} - \text{unteres Quartil} = x_{\frac{3}{4}} - x_{\frac{1}{4}}$$

(nur sinnvoll für I- oder V-Skala).

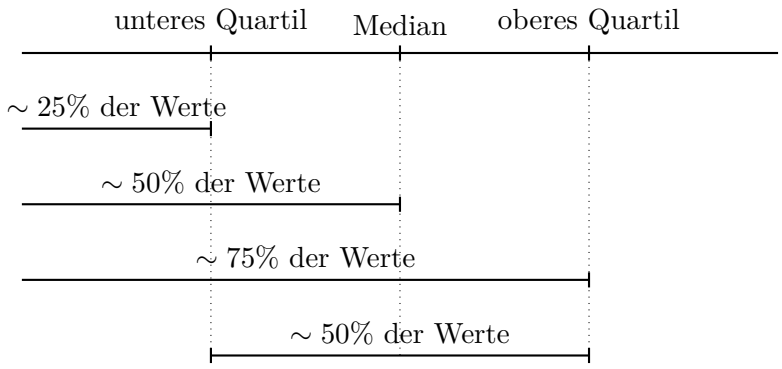


Abbildung 1.5: *Skizze zur groben Orientierung*

1.12 Grafiken zu den Quantilen

Für $0 < \alpha < 1$ gilt:

$$r(X < x_\alpha) \leq \alpha \leq r(X \leq x_\alpha) = F(x_\alpha)$$

Ablesen der Quantile aus einer Verteilungsfunktion zu einem Merkmal X.

1. Fall

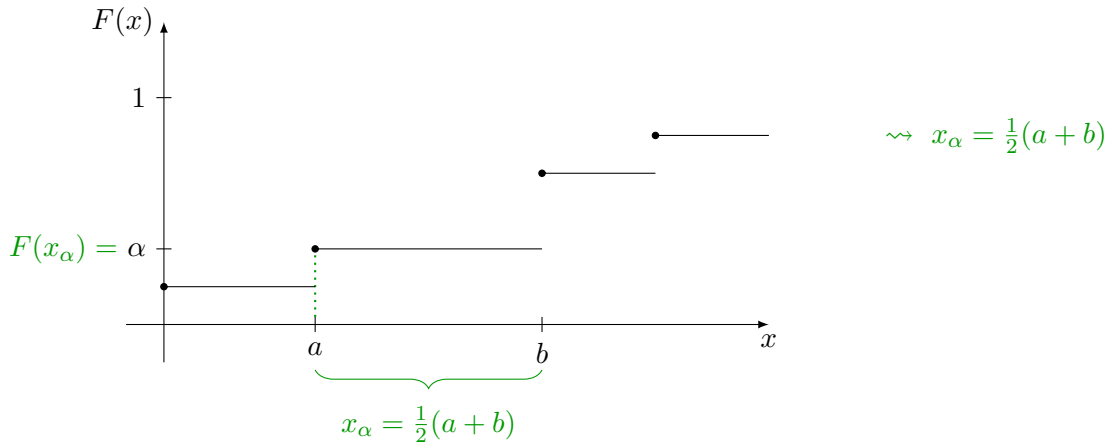


Abbildung 1.6

2. Fall

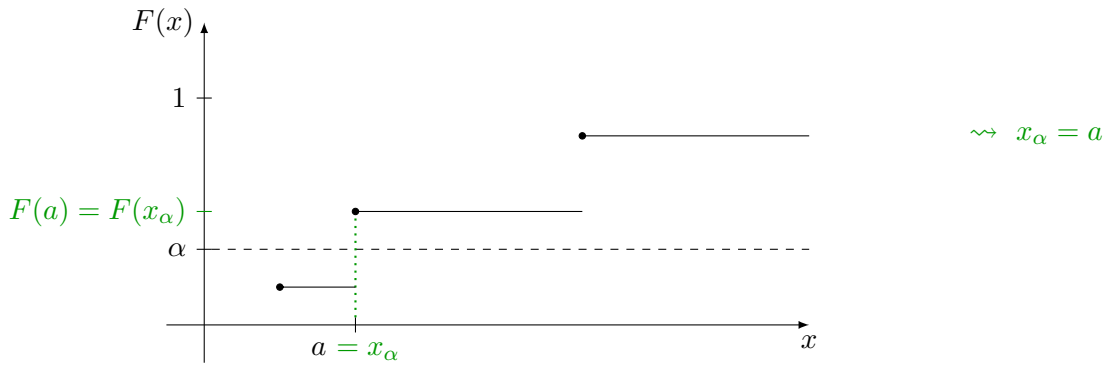


Abbildung 1.7

(a) Median ist unempfindlicher als das arithmetische Mittel gegenüber sogenannten Ausreißerwerten.
Beispiel:

(1) geordnete Messreihe: 2 5 7 7 8 10 12

$$n = 7 \quad ; \quad \tilde{x} = x_{\left(\frac{7+1}{2}\right)} = x_{(4)} = 7 \quad ; \quad \bar{x} = \frac{51}{7} \approx 7.28$$

(2) geordnete Messreihe: 2 5 7 7 8 10 12 30

$$n = 8 \quad ; \quad \tilde{x} = \frac{x_{(4)} + x_{(5)}}{2} = \frac{7+8}{2} = 7.5 \quad ; \quad \bar{x} = \frac{81}{8} = 10.125$$

(b) Das arithmetische Mittel ist ein Durchschnittswert für die Datenreihe (bzw. Schätzwert für den Erwartungswert einer Zufallsgröße). Der Median teilt die Datenreihe in 2 etwa gleich große Bereiche.
Beispiel:
Bei 50 Glühlampen wird gemessen, nach welcher Zeitspanne sie ausfallen.
arithmetisches Mittel $\Longrightarrow$ durchschnittliche Lebensdauer (Kosten!)
Median $\Longrightarrow$ Zeitspanne, zu der noch 50 % der Lampen in Ordnung sind (Ausfallsicherheit!)

(c) Verhalten von Median und arithmetischem Mittel bei verschiedenen Verteilungsfunktionen: „**symmetrische Verteilung**“, „**linkssteile Verteilung**“ und „**rechtssteile Verteilung**“

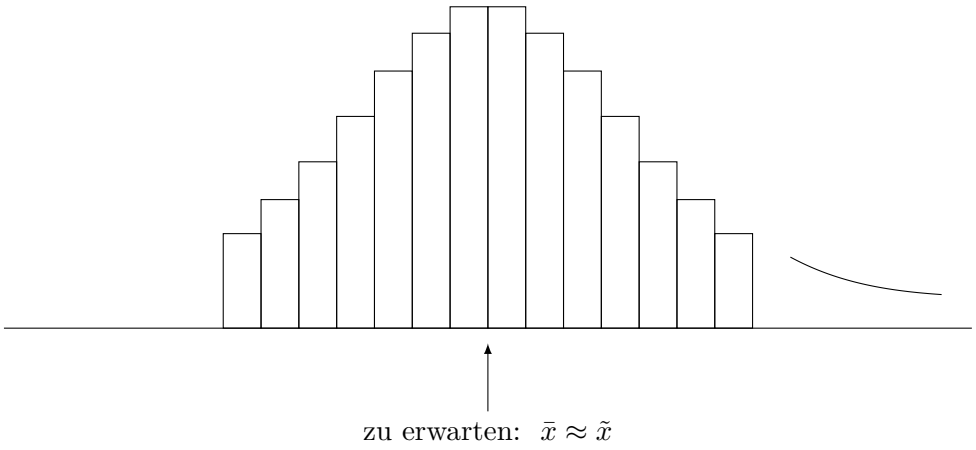


Abbildung 1.8: *symmetrische Verteilung*

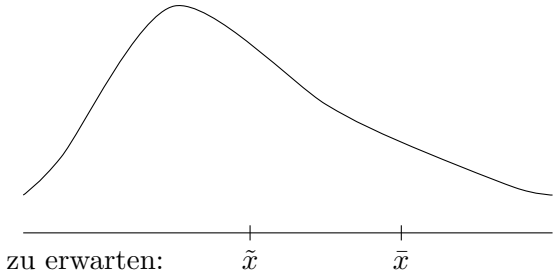


Abbildung 1.9: *linkssteile Verteilung*

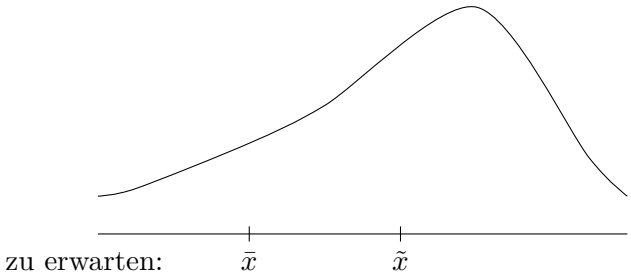


Abbildung 1.10: *rechtssteile Verteilung*

1.13 geometrisches und harmonisches Mittel

Es seien $x_1, \ldots, x_n \in \mathbb{R}$, $x_i \geq 0$ für alle $i \in \{1, \ldots, n\}$, Messwerte einer Verhältnisskala. Dann heißt

(a) $\bar{x}_G := \sqrt[n]{x_1 \cdot \ldots \cdot x_n}$ das **geometrische Mittel**

(b) $\bar{x}_H := n \cdot \frac{1}{\frac{1}{x_1} + \ldots + \frac{1}{x_n}}$ das **harmonische Mittel**.

Hgew steht für Gewichtung

(c) Variante des harmonischen Mittels:

$$\bar{x}_{Hgew} = \sum_{j=1}^{n} g_j \cdot \frac{1}{\frac{g_1}{x_1} + \ldots + \frac{g_n}{x_n}}$$

wobei g_j Zusatzinformationen zu x_j darstellen.

Es seien $x_1, \ldots, x_n \in \mathbb{R}$, $x_i > 0$ für alle $i \in \{1, \ldots, n\}$.

(a) $y_i = \ln(x_i) \implies \bar{x}_G = \exp(\bar{y})$ (logarithmische Koordinatentransformation)

(b) (i) $\bar{x}_G \leq \bar{x}$

(ii) $\bar{x}_G = \bar{x} \iff x_1 = \ldots = x_n$

Das geometrische Mittel wird vor allem bei Wachstumsprozessen (z. B. Zinseszins-Berechnungen) verwendet.
Das harmonische Mittel ist z. B. bei der Berechnung von Durchschnittsgeschwindigkeiten sinnvoll.

1.14 Modalwert

Vor allem bei Nominalskalen wird der **Modalwert** $\bar{x}_D$ als Maßzahl verwendet. $\bar{x}_D$ bezeichnet die Ausprägung, die am häufigsten vorkommt, d. h.

$$h(\bar{x}_D) = \max \left\{ h(a) \;\middle|\; a \in A \right\}$$

Gibt es mehrere Ausprägungen mit der größten Häufigkeit, so gibt es mehrere Modalwerte.

1.15 gewichtetes arithmetisches Mittel für I- oder V-Skala

$x_1, \ldots, x_n \in \mathbb{R}$ sei eine Datenreihe mit positiven Gewichten $g_i > 0$, $i = 1, \ldots, n$. Dann heißt

$$\bar{x}_{gew} = \frac{1}{\sum_{i=1}^{n} g_i} \cdot \sum_{i=1}^{n} g_i x_i$$

das **gewichtete arithmetische Mittel**.
Für $g_i = 1$ für alle $i \in \{1, \ldots, n\}$ ist $\bar{x}_{gew} = \bar{x}$.

1.16 Berechnungsvariante für die Stichprobenvarianz

Vor: $x_1, \ldots, x_n \in \mathbb{R}$ sei eine Datenreihe.

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

1.17 Boxplots

Geeignet für O-, I- und V-Skalen. Grafische Darstellung von Minimum, unteres Quartil, Median, oberes Quartil, Maximum.

Beispiel

Gegeben sei die geordnete Messreihe (Datenreihe) $x_{(1)}$ bis $x_{(11)}$:

2 , 3 , 3 , 4 , 5 , 5 , 6 , 7 , 8 , 11 , 14

Minimum: 2, Maximum: 14
unteres Quartil: $x_{\frac{1}{4}} = x_{(\lceil 11 \cdot \frac{1}{4} + 1 \rceil)} = x_{(3)} = 3$
Median: $\tilde{x} = x_{(\frac{11+1}{2})} = x_{(6)} = 5$
oberes Quartil: $x_{\frac{3}{4}} = x_{(\lceil 11 \cdot \frac{3}{4} + 1 \rceil)} = x_{(9)} = 8$

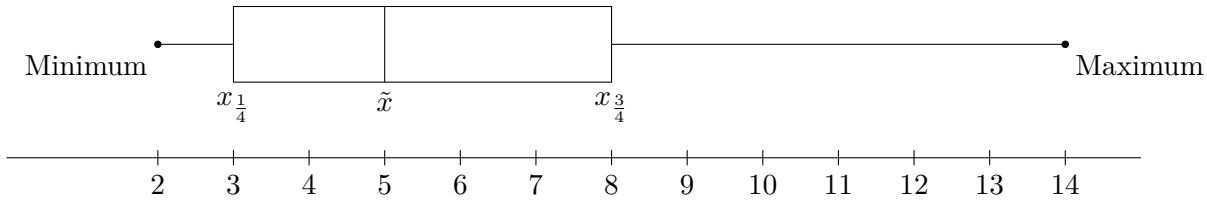


Abbildung 1.11: *Boxplot*

1.18 Punktwolke

Für die graphische Darstellung das Zusammenhangs zweier Merkmale werden **Punktwolken** (Streudiagramme[3]) verwendet. Dazu werden die Wertepaare (x_i, y_i) in ein Koordinatensystem eingetragen. Die Form der Punktwolke gibt einen Hinweis auf die Stärke des Zusammenhangs.

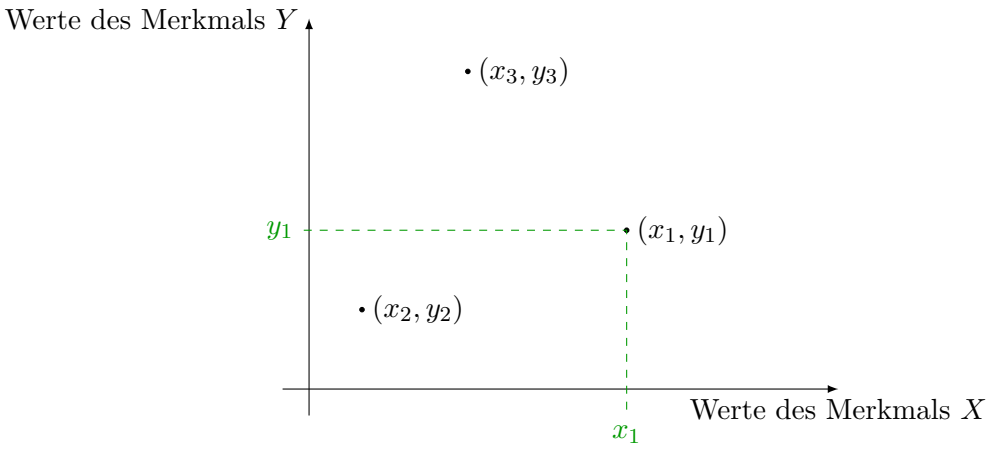


Abbildung 1.12: *Punktwolke (Streudiagramm)*

1.19 Pearsonscher Korrelationskoeffizient

Für metrische Merkmale wird als statistische Maßzahl für die Abhängigkeit (Korrelation) zweier Merkmale häufig der **Pearsonsche[4] Korrelationskoeffizient** verwendet. Er ist definiert durch

$$r = r_{x,y} := \frac{\overbrace{s_{x,y}}^{\substack{\text{enthält Informationen} \\ \text{über evtl. Zusammenhang}}}}{\underbrace{s_x \cdot s_y}_{\text{Normierungsfaktoren}}}$$

[3] *englisch* scatterplots

wobei

$$s_{x,y} := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \cdot \left(\left(\sum_{i=1}^{n} x_i y_i \right) - (n\bar{x} \cdot \bar{y}) \right)$$

empirische Kovarianz,

$$s_x := \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Standardabweichung der Datenreihe $(x_1, \ldots, x_n)$,

$$s_y := \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Standardabweichung der Datenreihe $(y_1, \ldots, y_n)$.

(a) r ist ein Maß für die lineare Abhängigkeit zwischen zwei Merkmalen.

(b) Es gilt:

(i) $-1 \leq r \leq 1$ (Beweis: Übungsaufgabe)

(ii) $r = 1 \iff y_i = ax_i + b$ mit $a, b \in \mathbb{R}$, $a > 0$

(iii) $r = -1 \iff y_i = ax_i + b$ mit $a, b \in \mathbb{R}$, $a < 0$

Sind $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n)$ Stichproben von zwei unabhängigen Merkmalen X und Y, so ist zu erwarten, dass r „nahe“ bei 0 liegt.

(c) Ist $r = 0$ oder „nahe“ bei 0, so bedeutet dies NICHT, dass die beiden Merkmale unabhängig sind (es besteht evtl. ein nichtlinearer Zusammenhang).

1.20 Verhalten von r bei linearen Transformationen der Messwerte

$a_1 x + b_1$ bedeute $x_i \xrightarrow{\text{transformiert in}} a_1 x_i + b_1$ für $i = 1, \ldots, n$

$a_2 y + b_2$ bedeute $y_i \xrightarrow{\text{transformiert in}} a_2 y_i + b_2$ für $i = 1, \ldots, n$

$$\implies r_{(a_1x+b_1),(a_2y+b_2)} = \left\{ \begin{array}{ll} r_{x,y} & \text{falls } a_1, a_2 > 0 \\ r_{x,y} & \text{falls } a_1, a_2 < 0 \\ -r_{x,y} & \text{falls } a_1 > 0,\ a_2 < 0 \\ -r_{x,y} & \text{falls } a_1 < 0,\ a_2 > 0 \end{array} \right\} = \frac{a_1 \cdot a_2}{|a_1 \cdot a_2|} \cdot r_{x,y}$$

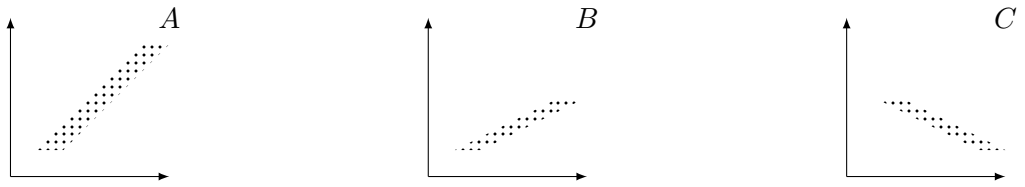


Abbildung 1.13: $r_A \approx r_B$; $r_C \approx -r_A$

Der Grad der Steigung der Punktwolke wirkt sich NICHT auf r aus.

1.21 Bemerkung zur Interpretation von r

Bei der statistischen Analyse des Zusammenhangs zweier Merkmale geht es NICHT um den Nachweis von Kausalitäten (Ursache $\longrightarrow$ Wirkung). Dieser Nachweis ist sachlogischen Argumenten vorbehalten.

r ist nur ein Maß dafür, wie stark die Werte der beiden Datenreihen gemeinsam variieren. Ein hoher Korrelationskoeffizient ist NICHT automatisch ein Hinweis auf einen direkten Zusammenhang zwischen den Merkmalen.

1.22 Lineare Regression

r ist ein Maßstab für die Stärke eines linearen Zusammenhangs. Ziel der Regressionsrechnung ist es, die Abhängigkeit des einen Merkmals von einem anderen Merkmal durch eine Funktionsgleichung zu beschreiben.

Man nennt dabei X **unabhängige Variable** und Y **abhängige Variable**.

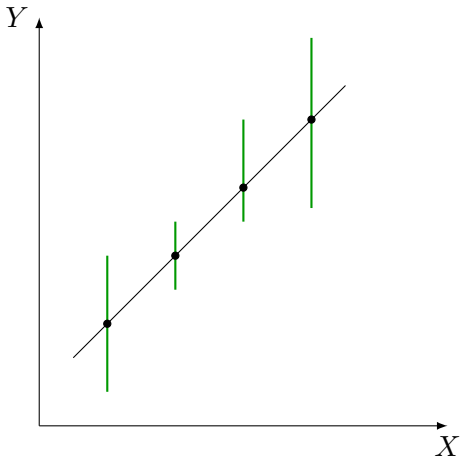
Welches Merkmal as unabhängig und welches als abhängig gelten soll, ergibt sich meistens aus der Problemstellung, z. B. Alter – Blutdruck: Alter unabhängig, Blutdruck abhängig.

Bei der **linearen Regression** soll die Form der Abhängigkeit durch eine **lineare Funktion** (eine sogenannte **Regressionsgerade** oder **Ausgleichsgerade**) dargestellt werden, d. h. in der Form

$$Y = a \cdot X + b \quad \text{mit } a, b \in \mathbb{R}\ ,\ a \neq 0$$

1.23 Modellvorstellung zur linearen Regression

(a) Jedem Wert x der Variablen X wird ein „wahrer Wert“ y der Variablen Y zugeordnet. Bei Messungen wird der zu einem x-Wert gewonnene y-Wert mehr oder minder stark vom wahren y-Wert abweichen (wegen Zufallseinflüssen).



„wahre Werte“ liegen auf der Gerade $Y = aX + b$

Streubereich der Werte für Y bei einem Experiment

Abbildung 1.14

(b) Bei einem Experiment erhält man eine zweidimensionale Datenreihe $(x_1, y_1), \ldots, (x_n, y_n)$

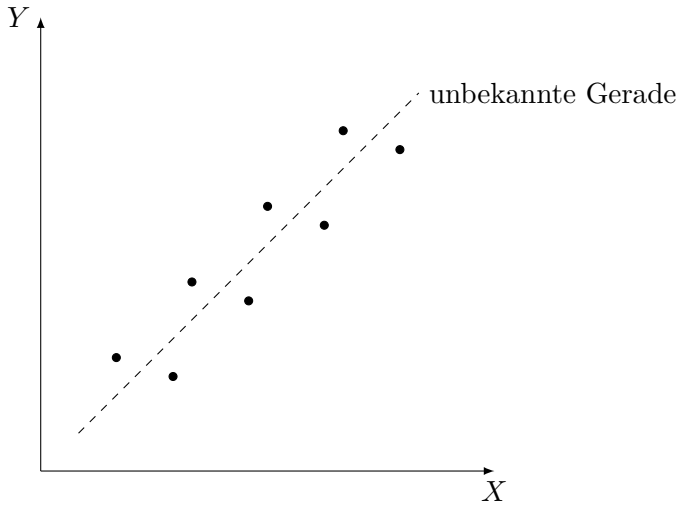


Abbildung 1.15

Wie erhält man aus der zweidimensionalen Datenreihe eine Gerade (Regressionsgerade), die den unbekannten linearen Zusammenhang $Y = aX + b$ „gut“ beschreibt?

Idee: Die Parameter a und b sollen so gewählt werden, dass die Gerade $Y = aX + b$ die Punktwolke möglichst „gut approximiert“.

1.24 Zur Bestimmung von a und b der Regressionsgerade

... verwendet man i. A. die **Methode der kleinsten Quadrate**[5]:

Die **Summe der Quadrate der Abstände** aller Punkte von der gesuchten Geraden soll minimal sein. Die Abstände werden dabei durch Parallelen zu Y-Achse bestimmt.

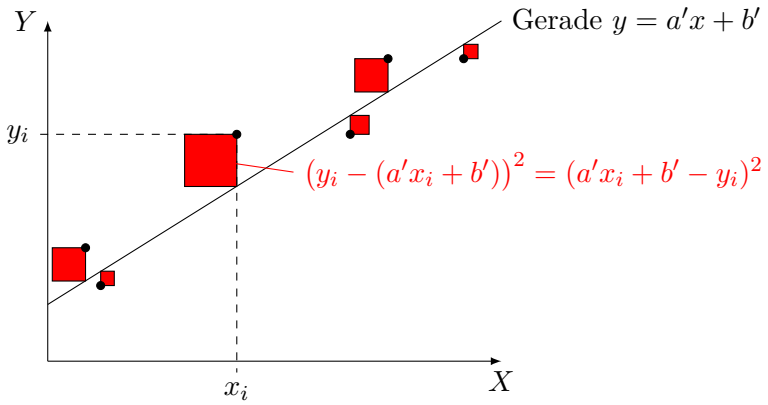


Abbildung 1.16

[5] *englisch* least square problem

Problem:
Gesucht sind $a, b \in \mathbb{R}$ mit der Eigenschaft

$$\sum_{i=1}^{n}(ax_i + b - y_i)^2 = \min_{(a',b')\in\mathbb{R}^2} \sum_{i=1}^{n}(a'x_i + b' - y_i)^2$$

Lösung:

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

wobei a **Regressionskoeffizient**[6] (Steigung der Geraden) und

$$b = \bar{y} - a\bar{x}$$

die **Regressionskonstante**[7] (Schnittpunkt mit y-Achse) darstellen. $\bar{x}$ und $\bar{y}$ sind die arithmetischen Mittel zu $(x_1, \ldots, x_n)$ bzw. $(y_1, \ldots, y_n)$.

1.25 Bemerkung (für $n \geq 2$)

(1) Für den Regressionskoeffizienten gilt:

$$a = \frac{(n-1)s_{x,y}}{(n-1)s_x^2} = \frac{s_{x,y}}{s_x^2}$$

(2) Es gilt: Der Punkt $(\bar{x}, \bar{y})$ liegt stets auf der Regressionsgeraden.
Genauer gilt: Für $\hat{y}_i = ax_i + b$ $(i = 1, \ldots, n;\ (a, b)$ aus der Regressionsgeraden) gilt

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

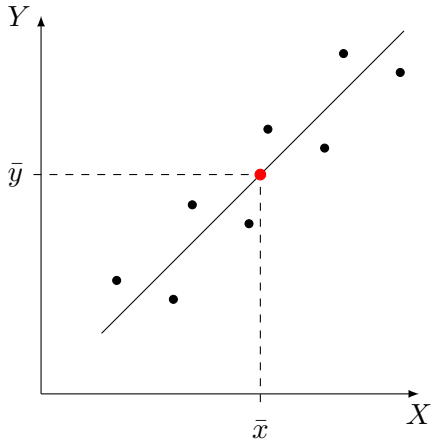


Abbildung 1.17

[6] *englisch* slope
[7] *englisch* intercept

1.26 Linearisierung durch Transformation

Durch Transformation der Messwerte lassen sich einige häufig vorkommende nichtlineare Zusammenhänge mit Hilfe der linearen Regression behandeln.

(a) **Exponentialfunktion** (einfach logarithmische Transformation):
Die Regressionsfunktion habe die Form

$$y = b \cdot a^x \quad \text{mit } b, a > 0 \ , \ x \in \mathbb{R}$$

Durch Logarithmieren auf beiden Seiten erhält man wegen

$$\underbrace{\ln(y)}_{=:y'} = \ln(b \cdot a^x) = \underbrace{\ln(b)}_{:=b'} + \underbrace{\ln(a)}_{=:a'} \cdot x$$
$$y' = b' + a' \cdot x$$

nach der der Transformation einen linearen Zusammenhang zwischen x und y'. Mit linearer Regressionsrechnung lassen sich a' und b' bestimmen und daraus erhält man

$$a = \exp(a') \quad , \quad b = \exp(b')$$

x	1	1.5	2	2.5	3	3.5	4
y	1	1.4	2	2.8	4	5.7	8
$\ln(y)$	0	0.3	0.7	1	1.4	1.7	2.1

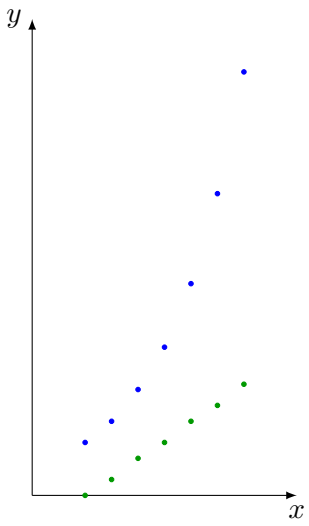


Abbildung 1.18

Durchführung:
Die zweidimensionale Datenreihe $(x_1, y_1), \ldots, (x_n, y_n)$ soll „möglichst gut“ durch eine Funktion der Form

$$y = b \cdot a^x$$

beschrieben werden.

1. Schritt: Erzeuge transformierte Datenreihe (Vor. $y_i > 0$):
$(x_1, \ln(y_1)), \ldots, (x_n, \ln(y_n))$ (einfache logarithmische Transformation)

2. Schritt: Berechne a' und b' zur transformierten Datenreihe mit linearer Regression

3. Schritt: Berechne $a = \exp(a')$ und $b = \exp(b')$

(b) **Potenzfunktion** (doppelt logarithmische Transformation)
Die Regressionsfunktion habe die Form

$$y = b \cdot x^a \quad \text{mit } b > 0\ ,\ x > 0\ ,\ a \in \mathbb{R}$$

Durch Logarithmieren auf beiden Seiten erhält man wegen

$$\underbrace{\ln(y)}_{=:y'} = \ln(b \cdot x^a) = \underbrace{\ln(b)}_{=:b'} + a \cdot \underbrace{\ln(x)}_{=:x'}$$
$$y' = b' + a \cdot x'$$

einen linearen Zusammenhang zwischen x' und y'.
Mit linearer Regression lassen sich a und b' bestimmen und $b = \exp(b')$ berechnen.

Durchführung:
Die zweidimensionale Datenreihe $(x_1, y_1), \ldots, (x_n, y_n)$ soll „möglichst gut“ durch eine Funktion der Form

$$y = b \cdot x^a$$

beschrieben werden.

1. Schritt: Erzeuge transformierte Datenreihe (Vor. $x_i, y_i > 0$):
$(\ln(x_1), \ln(y_1)), \ldots, (\ln(x_n), \ln(y_n))$ (doppelt logarithmische Transformation)

2. Schritt: Berechne a und b' zur transformierten Datenreihe mit linearer Regression

3. Schritt: Berechne $b = \exp(b')$

1.27 Anmerkung

Für Ordinalskalen wird oft der „Spearmansche Rangkorrelationskoeffizient“ verwendet.

1.28 Kontingenzkoeffizient

Bei der Analyse des Zusammenhangs zwischen nominalskalierten Merkmalen wird statt „Korrelation“ meisten der Begriff **Kontingenz** verwendet.

Zur Analyse des Zusammenhangs werden die Daten in einer sogenannten **Kontingenztafel** (Feldertafel) zusammengefasst.

Beispiel

Zusammenhang zwischen Augen- und Haarfarbe, $n = 100$ Personen:

y Haar / x Auge	blond	rot	braun	schwarz	Z_i Zeilensumme	Randwahrscheinlichkeit, relative Häufigkeit
grün	5	1	19	5	30	0.3
blau	24	1	3	0	28	0.28
braun	7	3	17	15	42	0.42
S_j Spaltensumme	36	5	39	20	100	
Randwahrscheinlichkeit, relative Häufigkeit	0.36	0.05	0.39	0.2		

Vorüberlegungen

Die erwartete Häufigkeit (unter der Voraussetzung, dass KEIN Zusammenhang vorliegt) sei mit E_{ij} bezeichnet. Für E_{ij} gilt:

$$E_{ij} = \underbrace{\frac{Z_i}{n} \cdot \frac{S_j}{n}}_{\substack{\text{Produkt rel.} \\ \text{Häufigkeiten}}} \cdot n = \frac{Z_i \cdot S_j}{n}$$

1.29 Kontingenzkoeffizient

(a) Als Hilfsgröße für den Zusammenhang bei nominalskalierten Merkmalen (aber auch bei Klassenbildung metrischer Merkmale) verwendet man

$$\chi^2 := \sum_{i=1}^{c_1} \sum_{j=1}^{c_2} \frac{(B_{ij} - E_{ij})^2}{E_{ij}}$$

(Chi-Quadrat), wobei c_1 die Anzahl der Zeilen und c_2 die Anzahl der Spalten der Kontingenztafel bezeichnen.

(b) Die Normierung von χ^2 führt zu einer Maßzahl für den Zusammenhang:

$$C := \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

heißt **Kontingenzkoeffizient**.

(1) C ist kleiner als 1 (natürlich ≥ 0).

(2) Der maximale Wert von C, bezeichnet mit $C_{\max}$, hängt von der Form der Kontingenztafel ab. Es gilt:

$$C_{\max} = \sqrt{\frac{m-1}{m}}$$

mit $m := \min(c_1, c_2)$.

(3) Wegen (2) verwendet man statt C auch den sogenannten korrigierten Kontingenzkoeffizienten C_k.

$$C_k := \frac{C}{C_{\max}} = \sqrt{\frac{\chi^2 \cdot m}{(\chi^2 + n) \cdot (m-1)}}$$

wobei n die Anzahl der Fälle und $m = \min(c_1, c_2)$ bezeichnen.

1.30 Wahrscheinlichkeitsrechnung: Bezeichnungen

Bei einem Zufallsexperiment seien $\omega_1, \ldots, \omega_n$ die möglichen **Ergebnisse**. Die Menge

$$\Omega := \{\omega_1, \ldots, \omega_n\}$$

heißt **Ergebnismenge**. Ziel ist es, für jedes $\omega \in \Omega$ eine reelle Zahl

$$P(\omega) \quad \text{mit } 0 \leq P(\omega) \leq 1$$

zu erklären, als Wahrscheinlichkeit dafür, dass das Ergebnis ω eintritt.

Eine Teilmenge $A \subset \Omega$ ($A = \Omega$ ist dabei auch möglich) wird als **Ereignis** bezeichnet. Gesucht ist wieder eine Zahl

$$P(A) \in \mathbb{R} \quad \text{mit } 0 \leq P(A) \leq 1$$

die Wahrscheinlichkeit dafür, dass $\omega \in A$, d. h. das Ereignis A, eintritt.
$A = \{\omega\}$ heißt **Elementarereignis**.

1.31 Relative Häufigkeiten

Ein Zufallsexperiment werde k-mal durchgeführt. Seien $\omega \in \Omega$ und $A \subset \Omega$.

(a) Häufigkeiten des Eintretens von ω bzw. A

$$H_k(\omega) = \text{Anzahl der Male, in denen } \omega \text{ eintritt}$$
$$H_k(A) = \text{Anzahl der Male, in denen } \omega \in A \text{ eintritt}$$

(b) Relative Häufigkeiten des Eintretens von ω bzw. A

$$R_k(\omega) := \frac{1}{k} \cdot H_k(\omega)$$
$$R_k(A) := \frac{1}{k} \cdot H_k(A) = \sum_{\omega \in A} R_k(\omega)$$

Mit obigen Bezeichnungen gelten:

(a) $0 \leq H_k(\omega) \leq k$, $0 \leq H_k(A) \leq k$

(b) $0 \leq R_k(\omega) \leq 1$, $\sum\limits_{\omega \in \Omega} R_k(\omega) = 1$

(c) $0 \leq R_k(A) \leq 1$ für $A \subset \Omega$

(d) $R_k(\Omega) = 1$

(e) $R_k(A \cup B) = R_k(A) + R_k(B)$ für $A, B \subset \Omega$, $\underline{A \cap B = \emptyset}$

1.32 Wahrscheinlichkeit

(Axiometische Definition der Wahrscheinlichkeit, Kolmogoroff 1933)

Ω sei eine (endliche) Menge und $\mathcal{A} \subset \mathcal{P}(\Omega)$, $\mathcal{P}(\Omega)$ ist die Potenzmenge von Ω, d. h. die Menge aller Teilmengen von Ω. Für $\mathcal{A}$ gelte: $\emptyset, \Omega \in \mathcal{A}$; $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$ und $\Omega \setminus A \in \mathcal{A}$.

Eine Wahrscheinlichkeitsverteilung ist eine Abbildung

$$P \colon \mathcal{A} \longrightarrow \mathbb{R} \ , \ A \longmapsto P(A)$$

mit folgenden Eigenschaften:

(1) $0 \leq P(A) \leq 1$ für alle $A \in \mathcal{A}$

(2) $P(\Omega) = 1$

(3) $P(A \cup B) = P(A) + P(B)$ falls $A \cap B = \emptyset$, $A, B \in \mathcal{A}$

Für nicht endliche Ω sind die Eigenschaften an $\mathcal{A}$ und Bedingung (3) zu modifizieren.

1.33 Ω endlich

Sei Ω eine nicht leere Menge und $\mathcal{A} \subset \mathcal{P}(\Omega)$. $A_1, \ldots, A_r$ seien aus $\mathcal{A}$ und alle Operationen mit den A_i (Vereinigung, Durchschnitt usw.) seien ebenfalls aus $\mathcal{A}$. Erfüllt die Abbildung

$$P \colon \mathcal{A} \longrightarrow \mathbb{R} \ , \ A \longmapsto P(A)$$

die Eigenschaften (1) bis (3) aus 1.32, so gilt:

(a) $P(\emptyset) = 0$

(b) $P(A_1 \cup \ldots \cup A_r) = P(A_1) + \ldots + P(A_r)$, falls $A_i \cap A_j = \emptyset$ für $i \neq j$

(c) $P(\bar{A}) = 1 - P(A)$, wobei $\bar{A} = \Omega \setminus A$

(d) $A_1 \subset A_2 \implies P(A_1) \leq P(A_2)$

(e) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

(f) $P(A_1 \cup \ldots \cup A_r) \leq P(A_1) + \ldots + P(A_r)$

1.34 Gleichverteilung/Laplace-Verteilung

Es sei $\Omega = \{\omega_1, \ldots, \omega_n\}$. Durch

$$P(\omega) = \frac{1}{n} \quad \forall\ \omega \in \Omega$$

wird eine **Gleichverteilung** (oder **Laplace-Verteilung**) auf Ω definiert.

Ist $A \subset \Omega$, so gilt für die Wahrscheinlichkeit von A

$$P(A) = \frac{\#A}{\#\Omega}$$

Dabei bedeutet das Symbol # „Anzahl der Elemente von“.

1.35 Zufallsvariable/Zufallsgröße, Ausprägungen

(a) Ein wichtiger Grundbegriff ist die **Zufallsvariable** (oder **Zufallsgröße**). Das ist eine Abbildung

$$X \colon \Omega \longrightarrow \mathbb{R}$$

Man kann sich dabei vorstellen, dass jedes Ergebnis ω eine gewisse Zahl $X(\omega)$ (etwa Gewinn oder Verlust) auslöst. Wie bei den „Merkmalen“ nennt man die Werte $a = X(\omega) \in \mathbb{R}$ **Ausprägungen** von X

(b) Für $a \in \mathbb{R}$ sei

$$\{X = a\} := \left\{ \omega \in \Omega \;\middle|\; X(\omega) = a \right\} \subset \Omega$$

und

$$P(X = a) := P(\{X = a\}) \in [0, 1]$$

die Wahrscheinlichkeit dafür, dass X den Wert a annimmt.

1.36 Graphische Darstellung

(a) Darstellung von $P(X = a)$ durch ein Stabdiagramm.

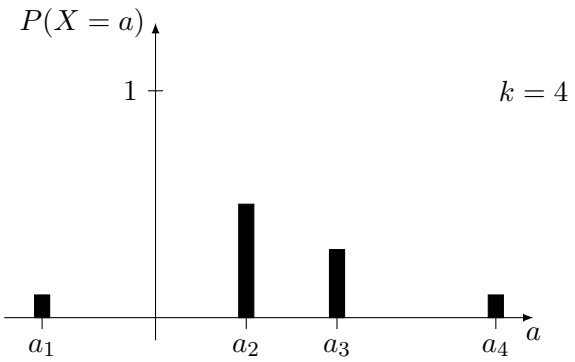


Abbildung 1.19

Dabei ist

$$\sum_{a \in \mathbb{R}} P(X = a) = 1$$

(Ω endlich, dann ist es eine endliche Summe)

(b) Darstellung durch eine Verteilungsfunktion

$$F_X \colon \mathbb{R} \longrightarrow [0, 1] \quad \text{mit } F_X(x) := P(X \leq x)$$

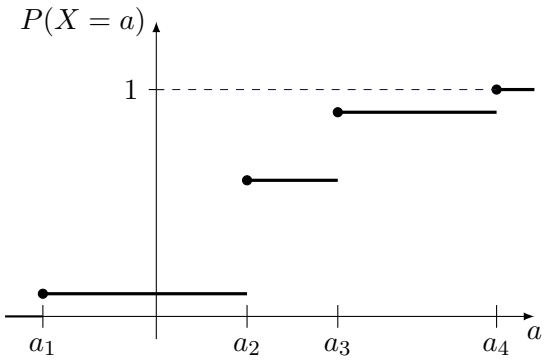


Abbildung 1.20

Ist eine endliche Ergebnismenge Ω mit einer Wahrscheinlichkeitsverteilung P und eine Zufallsvariable

$$X\colon \Omega \longrightarrow \mathbb{R}\ ,\ \omega \longmapsto X(\omega)$$

gegeben, so kann man den Wert $X(\omega)$ als ein neues von ω abhängiges Ergebnis betrachten; damit erhält man eine neuer Ergebnismenge

$$\Omega' := X(\Omega) \subset \mathbb{R}$$

Setzt man für X einen Wert $a \in \Omega'$

$$P'(a) := P(X = a)$$

so wird dadurch eine Wahrscheinlichkeitsverteilung P' auf Ω' erklärt.
Ein Vorteil von Ω' gegenüber Ω kann sein, dass $\Omega' \subset \mathbb{R}$. Die Verteilungsfunktion F_X von X ist durch P' bestimmt.

1.37 Urnenmodelle – Regeln mit Ergebnismengen

Sei $M = \{1, \ldots, n\}$. k Kugeln werden gezogen.

(1) Geordnet, mit Zurücklegen

$$\Omega_1 = \left\{(a_1, \ldots, a_k) \in M^k \;\middle|\; 1 \le a_i \le n\right\}$$

(2) Geordnet, ohne Zurücklegen

$$\Omega_2 = \left\{(a_1, \ldots, a_k) \in M^k \;\middle|\; a_i \neq a_j \text{ für } i \neq j\right\}$$

(3) Ungeordnet, mit Zurücklegen

$$\Omega_3 = \left\{(a_1, \ldots, a_k) \in M^k \;\middle|\; 1 \le a_1 \le \ldots \le a_k \le n\right\}$$

(4) Ungeordnet, ohne Zurücklegen ($k \leq n$)

$$\Omega_4 = \left\{ (a_1, \ldots, a_k) \in M^k \;\middle|\; 1 \leq a_1 < a_2 < \ldots < a_k \leq n \right\}$$

Geordnet bedeute dabei, dass im Ergebnis die Reihenfolge der Züge berücksichtigt wird.
Ungeordnet bedeute, dass die Reihenfolge der Ziehungen unwesentlich ist; das Ergebnis wird nach der Größe der Kugelnummern sortiert.

Es gelten:

(a)
$$\begin{array}{ccc} \Omega_1 & \supset & \Omega_2 \\ \cup & & \cup \\ \Omega_3 & \supset & \Omega_4 \end{array}$$

(b) $\Omega_4 = \Omega_2 \cap \Omega_3$

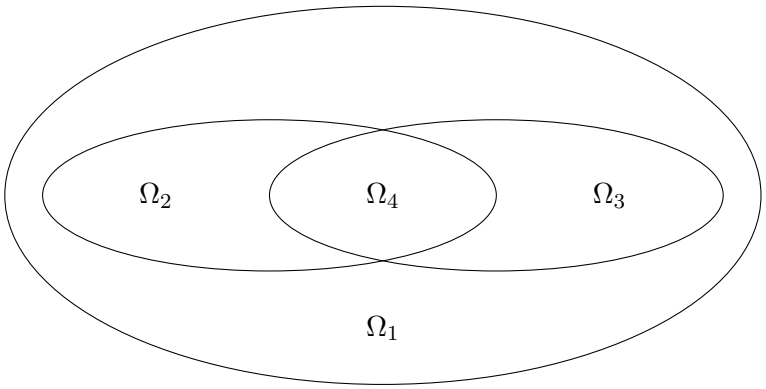


Abbildung 1.21: *Venndiagramm zu 1.37*

1.38 Fakultät, Binomialkoeffizient

$n! := n(n-1) \cdot \ldots \cdot 2 \cdot 1$ (n-Fakultät)
$\binom{n}{k} := \frac{n!}{(n-k)!k!}$ (n über k), heißt Binomialkoeffizient; dabei ist $0! = 1$ und für $k = 0$ oder $k > n$ wird $\binom{n}{k} := 0$ gesetzt.

Es gilt:

(a) $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ (Binomische Formel)

(b) $\binom{n}{k} = \binom{n}{n-k}$

(c) Die Binomialkoeffizienten sind ganzzahlig.

(d) $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

Ferner gilt ($\#M$ bezeichne die Anzahl der Elemente von M):

(a) $\#\Omega_1 = n^k$

(b) $\#\Omega_2 = \frac{n!}{(n-k)!}$

(c) $\#\Omega_3 = \binom{n+k-1}{k}$

(d) $\#\Omega_4 = \binom{n}{k}$

1.39 Baumdiagramme und Pfadregeln

Zweimaliger Wurf einer Münze.

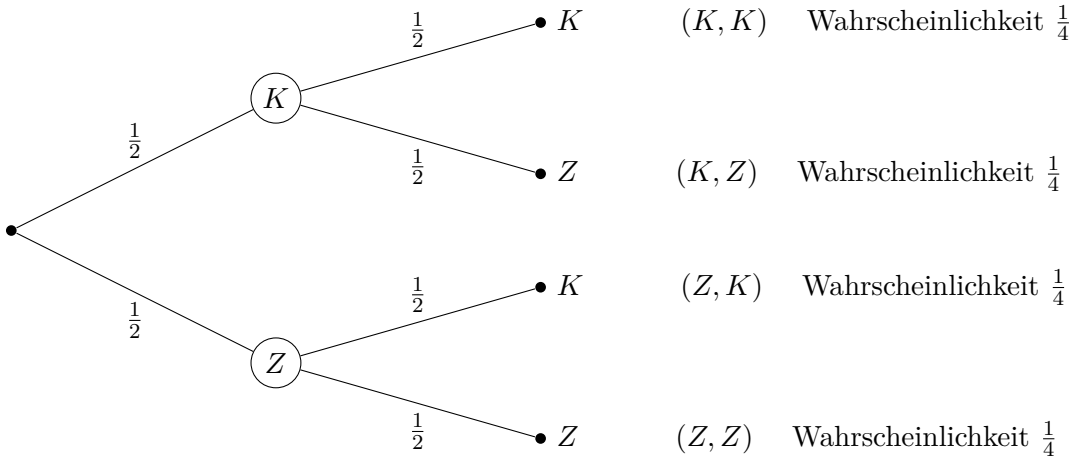


Abbildung 1.22: *Baumdiagramm*

Die Wahrscheinlichkeit, dass ein Ergebnis – z. B. (Z, K) – eintritt, berechnet man, indem man die Wahrscheinlichkeiten entlang des Pfades miteinander multipliziert.

Die Wahrscheinlichkeit, dass ein Ereignis – z. B. $\{(Z, K); (K, Z)\}$ – eintritt, berechnet man, indem man die Pfad-Wahrscheinlichkeiten alles Punkte, die zu dem Ereignis gehören, addiert.

1.40 Bedingte Wahrscheinlichkeiten und unabhängige Ereignisse

Ω sei eine (endliche) Menge, $\mathcal{A} \subset \mathcal{P}(\Omega)$, $A, B \ldots \in \mathcal{A}$. $P\colon \mathcal{A} \longrightarrow \mathbb{R}$ eine Wahrscheinlichkeitsverteilung.

$A, B \in \mathcal{A}$, $P(B) > 0$. Dann heißt

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

die **bedingte Wahrscheinlichkeit** von A unter der Bedingung B.

Bemerkung: $P(B|B) = 1$

1.41 Multiplikationssatz

$A, B \in \mathcal{A}$, $P(B) > 0$. Dann gilt:

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

Sei $B \in \mathcal{A}$ und $P(B) > 0$. Durch $P_B(A) := P(A|B)$ wird eine Wahrscheinlichkeitsverteilung definiert.

$$P_B\colon \left\{ C \cap B \;\middle|\; C \in \mathcal{A} \right\} \longrightarrow \mathbb{R}$$

erfüllt die Axiome aus 1.32.

1.42 stochastisch unabhängig

Sei $B \in \mathcal{A}$ und $0 < P(B) < 1$. Dann heißt $A \in \mathcal{A}$ (stochastisch) **unabhängig** von B
$:\Longleftrightarrow \; P(A|B) = P(A|\bar{B})$

$A, B \in \mathcal{A}$, $0 < P(B) < 1$. Dann gilt:
A (stochastisch) unabhängig von B
$\Longleftrightarrow \; P(A|B) = P(A)$
$\Longleftrightarrow \; P(A \cap B) = P(A) \cdot P(B)$ (am häufigsten verwendet)

$A, B \in \mathcal{A}$ und $0 < P(A) < 1$, $0 < P(B) < 1$.
A unabhängig von $B \iff B$ unabhängig von A.

Bemerkung: $P(A|B) \overset{\text{i.A.}}{\neq} P(B|A)$

1.43 Disjunktion (Zerlegung) von Ω

$A_1, \ldots, A_n \in \mathcal{A}$ bilden eine Disjunktion (Zerlegung) von Ω
$:\Longleftrightarrow \; A_1 \cup \ldots \cup A_n = \Omega$ und $A_i \cap A_j = \emptyset$ für $i \neq j$.

Bemerkung: A, B heißen disjunkt $:\Longleftrightarrow \; A \cap B = \emptyset$

$A_1, \ldots, A_n$ Disjunktion von Ω, $P(A_i) > 0 \quad \forall\, i$ $\left(\text{damit ist } \sum_{i=1}^{n} P(A_i) = 1\right)$.
Dann gilt für $B \in \mathcal{A}$:

$$P(B) = \sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)$$

1.44 Bayessche Formel

$A_1, \ldots, A_n$ Disjunktionen von Ω, $P(A_i) > 0 \quad \forall\, i$ und $B \in \mathcal{A}$ mit $P(B) > 0$. Dann gilt:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)}$$

1.45 Verteilungen – Ziehen ohne Zurücklegen

Eine Zufallsvariable X mit Werten in $\{0, 1, \ldots, n\}$ heißt **binomialverteilt** mit den Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$, wenn gilt

$$(*) \quad b_{n;p}(k) = B(n; p; k) := P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

für $k \in \{0, 1, \ldots, n\}$

X sei binomialverteilt mit den Parametern n und p. Dann gilt:

(a) $P(l \leq X \leq m) = \sum_{k=l}^{m} P(X = k) = \sum_{k=l}^{m} b_{n;p}(k)$

(b) $P(0 \leq X \leq n) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \overset{\text{BLS}}{=} \big(p + (1-p)\big)^n = 1$

(c) $b_{n;0.5}(k) = \frac{1}{2^n}\binom{n}{k}$

(d) $P(X \geq k) = 1 - P(X < k) = 1 - \sum_{k=0}^{k-1} b_{n;p}(l)$

Bemerkung: $\sum_{k=0}^{k-1} b_{n;p}(l)$ findet man oft in Tabellen

Beispiel

Ein fairer Würfel wird n-mal geworfen. Zu berechnen sind die Wahrscheinlichkeiten folgender Ereignisse:

(a) Genau k-mal die Augenzahl 6.

(b) Mindestens k-mal die Augenzahl 1.

(c) Genau k-mal eine gerade Augenzahl.

(d) Mindestens k-mal eine gerade Augenzahl.

Lösung

(a) Die Zufallsvariable X zähle die Anzahl „Eine 6 wird gewürfelt".
X ist dann binomialverteilt mit den Parametern n und $p = \frac{1}{6}$.

$$P(X = k) = n_{n;\frac{1}{6}}(k) = \binom{n}{k}\left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k} = \binom{n}{k} \cdot \frac{5^{n-k}}{6^n}$$

z. B. $n = 10$, $k = 5$:

$$P(X = 5) = b_{10;\frac{1}{6}}(5) = \binom{10}{5} \cdot \frac{5^5}{6^{10}} \approx 0.013$$

(b) Die Zufallsvariable X zähle die Anzahl „Eine 1 wird gewürfelt".
X ist dann binomialverteilt mit den Parametern n und $p = \frac{1}{6}$.

$$P(X \geq k) = 1 - P(X < k) = 1 - \sum_{l=0}^{k-1} b_{n;p}(l)$$

z. B. $n = 10$, $k = 5$:

$$P(X \geq 5) = 1 - \underbrace{\sum_{l=0}^{4} b_{10;\frac{1}{6}}(l)}_{\approx 0.98454 \text{ (Tabelle)}} \approx 0.01546$$

(c) Die Zufallsvariable X zähle „Eine gerade Zahl wird gewürfelt“. X ist dann binomialverteilt mit den Parametern n und $p = \frac{1}{2}$.

$$P(X = k) = b_{n;\frac{1}{2}}(k) = \frac{1}{2^n} \cdot \binom{n}{k}$$

z. B. $n = 10$, $k = 3$:

$$P(X = 3) = \underbrace{b_{n;\frac{1}{2}}(3)}_{\approx 0.11719 \text{ (Tabelle)}} = \underbrace{\frac{1}{2^{10}} \cdot \binom{10}{3}}_{= \frac{1}{2^{10}} \cdot \frac{10!}{7! \cdot 3!} = \frac{1}{2^{10}} \cdot \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1}} \approx 0.1171825$$

(d) X, n, p wie bei (c).

$$P(X \geq k) = 1 - P(X < k) = 1 - \sum_{l=0}^{k-1} n_{n;\frac{1}{2}}(l)$$

z. B. $n = 10$, $k = 3$:

$$P(X \geq 3) = 1 - \underbrace{\sum_{l=0}^{2} b_{n;\frac{1}{2}}(l)}_{\text{(Tabelle)}} \approx 0.9453$$

1.46 Verteilungen von Zufallsvariablen

Seien $n, r, N \in \mathbb{N}$ mit $n, r \leq N$.

- hypergeometrisch verteilte Zufallsvariable X:

$$P(X = k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}$$

wobei $k \geq \max(0, n - N + r)$ und $k \leq \min(n, r)$.
Notation: $H(n; N; r)$-verteilt

- stetig verteilte Zufallsvariable X: X stetig verteilt mit Dichte f $:\Longleftrightarrow$ f integrierbar und für Verteilungsfunktion $F\colon \mathbb{R} \longrightarrow [0, 1]$ gilt:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \quad \forall\ x \in \mathbb{R}$$

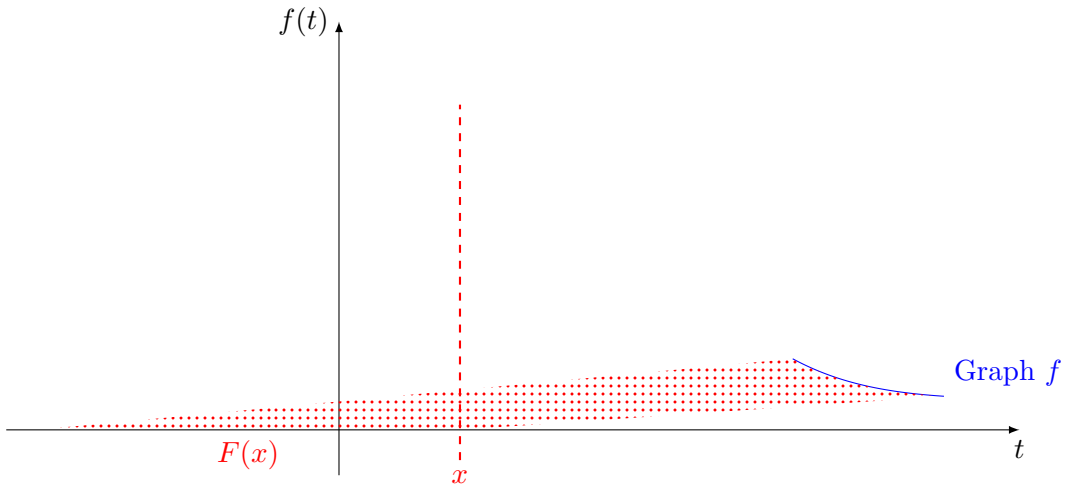


Abbildung 1.23

1.47 Veranschaulichung Wahrscheinlichkeiten $\longleftrightarrow$ Verteilungsfunktion

X stetig verteilte Zufallsgröße mit Dichte f. Dann gelten:

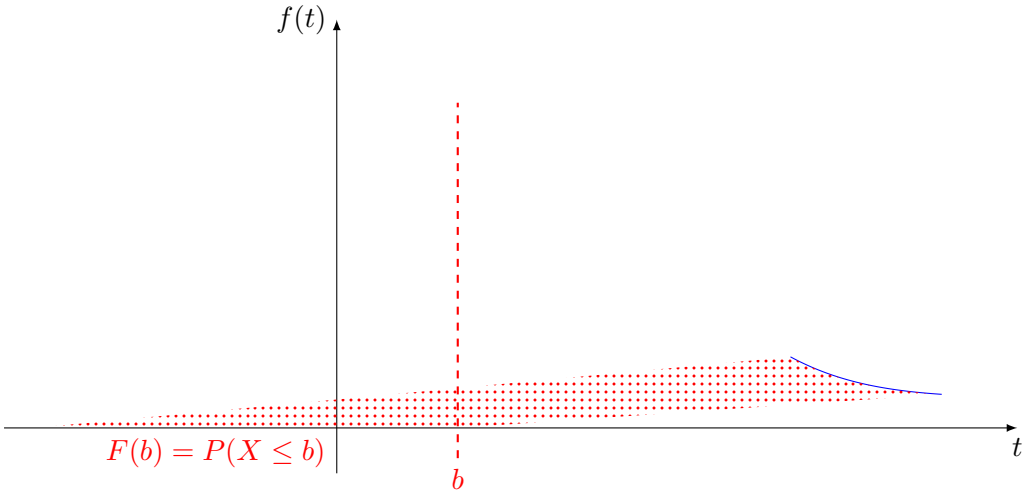


Abbildung 1.24: $F(b) = P(X \leq b)$

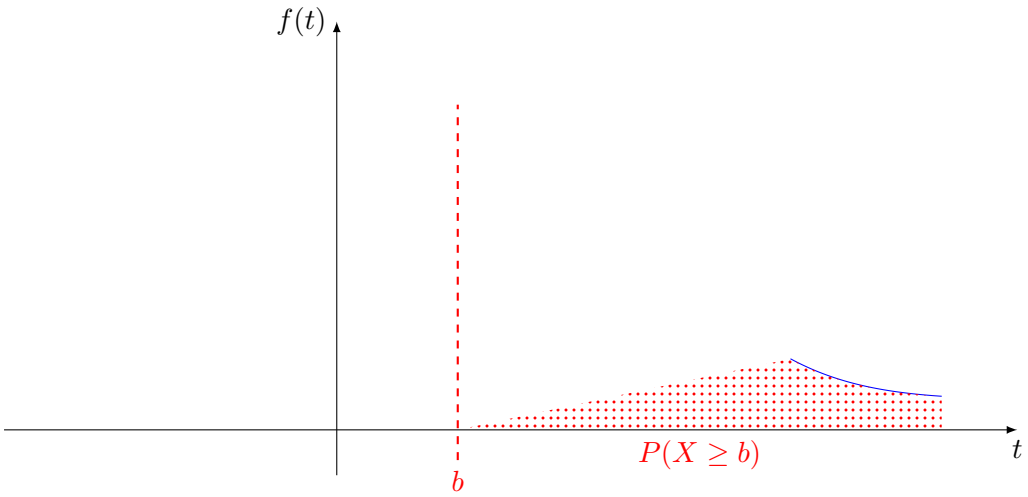


Abbildung 1.25: $P(X \geq b) = 1 - P(X < b) \overset{(*)}{=} 1 - P(X \leq b) = 1 - F(b)$
$(*)$ *X stetig verteilt*

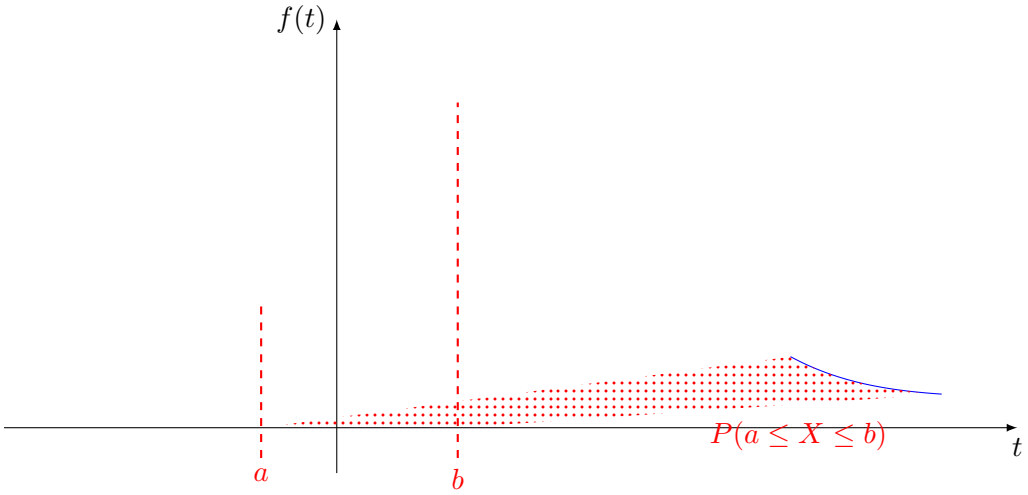


Abbildung 1.26: $P(a \leq X \leq b) = F(b) - F(a) = P(X \leq b) - P(X \leq a) = P(X \leq b) - P(X < a)$

1.48 Definition ((standard-)normalverteilte Zufallsvariable)

(a) Es seien $\mu, t \in \mathbb{R}$ und $\sigma > 0$. Eine Zufallsvariable X heißt **normalverteilt** mit Parametern μ und σ (kurz: $N(\mu, \sigma)$-verteilt), falls X stetig verteilt ist mit Dichte f definiert durch

$$f(t) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) \quad \forall\, t \in \mathbb{R}$$

μ Erwartungswert von X; σ Standardabweichung von X.

(b) Eine $N(0,1)$-verteilte Zufallsvariable heißt **standard-normalverteilt**.

(a) Ist X eine $N(0,1)$-verteilte Zufallsvariable, so gilt für die Verteilungsfunktion Φ von X

$(\Phi\colon \mathbb{R} \longrightarrow [0,1])$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} \exp\left(-\frac{1}{2}t^2\right) \mathrm{d}t$$

(b) Es gilt:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}x^2\right) \quad \forall\, x \in \mathbb{R}$$

ist die **Dichtefunktion** zur $N(0,1)$-verteilten Zufallsvariable.

sehr wichtige Formeln

(c) Es gilt:

$$\varphi(-x) = \varphi(x) \quad \forall\, x \in \mathbb{R}$$

und

$$\Phi(-x) = 1 - \Phi(x) \quad \forall\, x \in \mathbb{R}$$

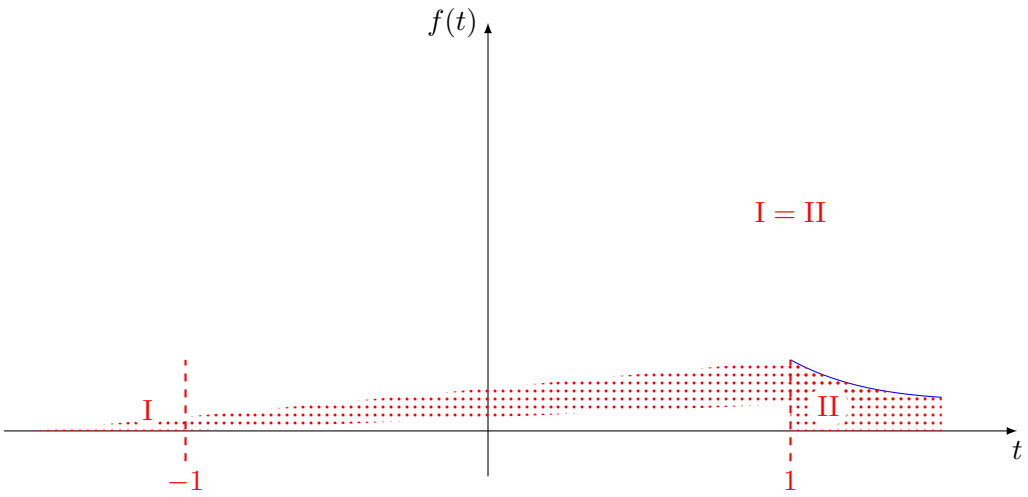


Abbildung 1.27: *Skizze zur Dichtefunktion*

(a) Ist X eine $N(\mu,\sigma)$-verteilte Zufallsvariable, so ist

$$U := \frac{X-\mu}{\sigma}$$

eine $N(0,1)$-verteilte Zufallsvariable.

(b) Ist X eine $N(\mu,\sigma)$-verteilte Zufallsvariable, so gilt:

$$P(X < b) = P(X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Für die Bestimmung von Wahrscheinlichkeiten zu normalverteilten Zufallsvariablen genügt die Kenntnis der Verteilungsfunktion Φ der $N(0,1)$-Verteilung für Werte $x \geq 0$.

Tabelle zu Werten von Φ

x	0.00	0.25	0.5	0.75	1	1.25	1.5	2	3
$\approx \Phi(x)$	0.5	0.5987	0.6915	0.7734	0.8413	0.8944	0.9332	0.9772	0.999

1.49 Approximation der Binomialverteilung durch die Normalverteilung

Für $p \in]0,1[$ und „große n“ (mindestens $n \cdot p \cdot (1-p) > 9$) ist

$$B(n;p;k) \approx \frac{1}{\sigma} \cdot \varphi\left(\frac{k-\mu}{\sigma}\right)$$

wobei $\mu = n \cdot p$, $\sigma = \sqrt{np(1-p)}$.

1.50 Definition (Erwartungswert)

Unter dem **Erwartungswert** einer Zufallsvariablen

$$X : \Omega \longrightarrow \mathbb{R}$$

versteht man die Zahl

$$E(X) := \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) = \sum_{k=1}^{r} a_k \cdot P(X = a_k)$$

wenn $a_1, \ldots, a_r \in \mathbb{R}$ die Ausprägungen, d. h. die paarweise verschiedenen Werte von X bezeichnen.

Für stetige Zufallsvariablen mit Dichtefunktion f gilt:

$$E(X) = \int_{-\infty}^{\infty} t \cdot f(t)\,\mathrm{d}t$$

1.51 Definition

Sind zwei Zufallsvariablen $X, Y : \Omega \longrightarrow \mathbb{R}$ und $\lambda \in \mathbb{R}$ gegeben, so kann man daraus neue Zufallsvariablen erklären durch

$$\begin{aligned}(X+Y)(\omega) &:= X(\omega) + Y(\Omega)\\ (\lambda X)(\omega) &:= \lambda \cdot X(\omega)\\ (X \cdot Y)(\omega) &:= X(\omega) \cdot Y(\omega)\end{aligned}$$

1.52 Rechenregeln für Erwartungswerte

Seien $X, Y : \Omega \longrightarrow \mathbb{R}$ Zufallsvariablen und $\lambda \in \mathbb{R}$. Dann gilt:

$$\begin{aligned}E(X+Y) &= E(X) + E(Y)\\ E(\lambda X) &= \lambda E(X)\end{aligned}$$

Im Allgemeinen ist $E(X \cdot Y) \overset{\text{i.A.}}{\neq} E(X) \cdot E(Y)$ $(*)$.
Gleichheit gilt in $(*)$, wenn die Zufallsvariablen „(stochastisch) unabhängig“ sind (siehe später).

1.53 Definition (Varianz)

Sei $X\colon \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable und $\mu := E(X)$.
Dann ist die **Varianz** von X definiert durch

$$\operatorname{Var}(X) = \begin{cases} \sum\limits_{\omega\in\Omega} \big(X(\omega)-\mu\big)^2 \cdot P(\omega)\ ,\ P(\omega)\neq 0 & \text{für endlich viele } \omega\in\Omega \\ \int\limits_{-\infty}^{\infty} (t-\mu)^2 f(t)\,\mathrm{d}t & \text{für stetige Zufallsvariablen mit Dichte } f \end{cases}$$

Für Ω endlich oder $X(\Omega)$ endlich kann man gleiche Werte von X zusammenfassen und erhält

$$\operatorname{Var}(X) = \sum_{k=1}^{r} (a_k-\mu)^2 \cdot P(X=a_k)$$

wobei $a_1,\ldots,a_r$ die verschiedenen Werte von X sind.

1.54 Definition (Standardabweichung)

Die Zahl

$$r_x := \sqrt{\operatorname{Var}(X)}$$

heißt **Standardabweichung** von X.

1.55 Rechenregeln für die Varianz

Es sei X eine Zufallsvariable, $\mu = E(X)$ und $\lambda\in\mathbb{R}$. Dann gilt:

(a) $\operatorname{Var}(X) = E\big((X-\mu)^2\big) = E(X^2)-\mu^2$

(b) $\operatorname{Var}(\lambda X) = \lambda^2\cdot\operatorname{Var}(X)$

(c) $\operatorname{Var}(X+\lambda) = \operatorname{Var}(X)$

Im Allgemeinen ist $\operatorname{Var}(X+Y) \overset{\text{i.A.}}{\neq} \operatorname{Var}(X)+\operatorname{Var}(Y)$.

1.56 Definition und Notiz

X sei eine Zufallsvariable, $\mu = E(X)$ und σ_X die Standardabweichung von X. Die Zufallsvariable

$$X^{\varphi} := \frac{X-\mu}{\sigma_X}$$

heißt die **Standardisierung** von X. Allgemein heißt X standardisiert, wenn $E(X)=0$ und $\operatorname{Var}(X)=1$.

1.57 Definition (Covarianz, Korrelationskoeffizient)

Sei Ω endlich, $X,Y\colon \Omega\in\mathbb{R}$ seien Zufallsvariablen, $\mu := E(X)$, $\nu := E(Y)$.
Dann heißt

$$\sigma_{x,y} = \operatorname{Cov}(X,Y) = \sum_{\omega\in\Omega} \big(X(\omega)-\mu\big)\big(Y(\omega)-\nu\big) = E\big((X-\mu)(Y-\nu)\big)$$

die **Covarianz** von X und Y und

$$\varrho = \varrho_{x,y} := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_x\cdot\sigma_y}$$

heißt **Korrelationskoeffizient** von X und Y.

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X)$$

1.58 Rechenregeln im Zusammenhang mit der Covarianz

X und Y seien Zufallsvariablen. Dann gilt:

(a) $E(X\cdot Y) = E(X)\cdot E(Y) + \mathrm{Cov}(X,Y)$

(b) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\big(\mathrm{Cov}(X,Y)\big) + \mathrm{Var}(Y)$

(c) Sind X und Y (stochastisch) unabhängig, d. h.

$$P(X=a, Y=b) = P(X=a)\cdot P(Y=b)$$

so ist

$$E(X\cdot Y) = E(X)\cdot E(Y) \quad \text{und} \quad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

(d) $X_1,\ldots,X_n$ seien unabhängige Zufallsvariablen, d. h.

$$P(X_1=a_1,\ldots,X_n=a_n) = P(X_1=a_1)\cdot\ldots\cdot P(X_n=a_n)$$

so gilt

$$\mathrm{Var}(X_1+\ldots+X_n) = \mathrm{Var}(X_1)+\ldots+\mathrm{Var}(X_n)$$

1.59 Erwartungstreue

Vor: $X_1,\ldots,X_n$ seien paarweise unabhängige Zufallsvariablen mit $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$ für $i = 1,\ldots,n$. Weiter sei $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$.

Beh: $E(S^2) = \sigma^2$, d. h. S^2 ist erwartungstreu.

1.60 Signifikanztests

Ein Test, bei dem entschieden werden soll, ob eine Hypothese H_0 abgelehnt wird oder nicht, heißt **Signifikanztest**. Die zu testende Hypothese H_0 heißt **Nullhypothese**.

Bei einem Signifikanztest wählt man eine Zahl $\alpha \in \mathbb{R}$ mit $0 < \alpha < 1$. Der Test ist dann so zu konstruieren, dass die Wahrscheinlichkeit(H_0 wird abgelehnt, obwohl H_0 richtig ist)$\leq \alpha$ ist.
α heißt **Signifkanzniveau**.
Übliche Werte für α sind 0.05 und 0.01.

1.61 Anwendungsbeispiel zum Signifikanztest

(i) Jemand bezweifelt, dass beim Werfen einer bestimmten Münze die beiden Ereignisse W:„Wappen liegt oben“ und Z:„Zahl liegt oben“ gleich wahrscheinlich sind.

(ii) $P(W)$ und $P(Z)$ sind unbekannt.
Nullhypothese H_0: $P(W) = P(Z) = \frac{1}{2}$.
Als Signifikanzniveau α wählen wir $\alpha = 0.05$.

(iii) Für eine Entscheidung soll die Münze 200 mal („unabhängig“) geworfen werden (unabhängige Wiederholung eines Zufallsexperiments). Gezählt wird, wie oft das Wappen oben liegt. X_{200} sei die dadurch erhaltene Zahl.

(iv) Für welche Werte von X_{200} soll die Nullhypothese H_0 abgelehnt werden?
Vorgabe: $\alpha = 0.05$
Ist H_0 richtig, so ist X_{200} eine binomialverteilte Zufallsvariable, d. h.

$$P_{H_0}(X_{200} = k) = \binom{200}{k} \left(\frac{1}{2}\right)^{200}$$

und der Wert X_{200} wird in der Nähe von 100 liegen.

(v) Gesucht ist deshalb ein $c \in \mathbb{N}$, $c \in [0, 200]$, mit der Eigenschaft

$$P_{H_0}(|X_{200} - 100| > c) = P_{H_0}(X < 100 - c) + P_{H_0}(X > 100 + c) \leq \alpha$$

Bei diesem Beispiel handelt es sich um einen zweiseitigen Signifikanztest.

(vi) Berechnung der Wahrscheinlichkeiten:
Es gilt: $n \cdot p(1-p) = 200 \cdot \frac{1}{2} \cdot \frac{1}{2} = 50 > 9$
d. h. zur Berechnung für die binomialverteilte Zufallsvariable X_{200} kann die Normalverteilung verwendet werden.
Für X_{200} gilt:

$$\mu = n \cdot p = 100$$
$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{50}$$

Es gilt:

$$P_{H_0}(|X_{200} - 100| > c) \leq \alpha \iff 1 - P_{H_0}(|X_{200} - 100| > c) = P_{H_0}(|X_{200} - 100| \leq c) \geq 1 - \alpha$$

$$P_{H_0}(|X_{200}-100|\le c) = P_{H_0}(100-c\le X_{200}\le 100+c)$$
$$\overset{(\sun)}{\approx} \Phi\left(\frac{100+c-\overbrace{100}^{=\mu}}{\underbrace{\sqrt{50}}_{=\sigma}}\right) - \Phi\left(\frac{100-c-100}{\sqrt{50}}\right)$$
$$= \Phi\left(\frac{c}{\sqrt{50}}\right) \underbrace{-\Phi\left(\frac{-c}{\sqrt{50}}\right)}_{=\Phi\left(\frac{c}{\sqrt{50}}\right)-1} = 2\Phi\left(\frac{c}{\sqrt{50}}\right) - 1$$

($\sun$) Standardisierung; Approximation durch Normalverteilung
Also gilt:

$$P_{H_0}(|X_{200}-100|>c)\le\alpha$$
$$\iff 2\Phi\left(\frac{c}{\sqrt{50}}\right)-1\ge 1-\alpha$$
$$\iff \Phi\left(\frac{c}{\sqrt{50}}\right)\ge 1-\frac{\alpha}{2} \quad (*)$$

Gesucht ist das kleineste $c\in[0,200]$, $c\in\mathbb{N}$, mit der Eigenschaft

$$\Phi\left(\frac{c}{\sqrt{50}}\right)\ge 1-\frac{\alpha}{2}$$

(vii) Bestimmung von c
Gesucht ist ein c mit

$$(*)\quad \Phi\left(\frac{c}{\sqrt{50}}\right)\approx 1-\frac{\alpha}{2}=0.975$$

($\geq$, kleinstes c mit dieser Eigenschaft)

Dazu betrachten wir die Tabelle der $N(0,1)$-Verteilung

$$\Phi(X)=0.975 \text{ für } X\approx 1.96$$

Mit $(*)$ folgt

$$\frac{c}{\sqrt{50}}\approx 1.96 \quad \text{d. h. } c\approx 13.8$$

Wähle $c=14$.

(viii) Entscheidungsregel
H_0 wird abgelehnt für

$$X_{200}<100-14=86$$

oder

$$X_{200}>100+14=114$$

Für $X_{200}\in[86,114]$ wird H_0 NICHT abgelehnt. Dies heißt NICHT, dass für $X_{200}\in[86,114]$ H_0 beseitigt ist.

II Wichtige Excel-Befehle

Ein `$` Zeichen vor der Spalte bzw. Zeile stellt einen absoluten Bezug dar.

Summe der Zellen B2 bis B81	`=SUMME(B2:B81)`
Anzahl der beschriebenen Zellen zwischen B2 und B81	`=ANZAHL(B2:B81)`
Mittelwert der Zellen B2 bis B81	`=MITTELWERT(B2:B81)`
Median der Zellen B2 bis B81	`=MEDIAN(B2:B81)`
Maximum der Zellen B2 bis B81	`=MAX(B2:B81)`
Minimum der Zellen B2 bis B81	`=MIN(B2:B81)`
Absolute Häufigkeit des Wertes 1 in den Zellen B2 bis B81	`=ZÄHLENWENN($B$2:$B$81;1)`
Relative Häufigkeit des Wertes 1 in den Zellen B2 bis B81	`=ZÄHLENWENN($B$2:$B$81;1)/ANZAHL($B$2:$B$81)`
Varianz der Zellen B2 bis B81	`=VARIANZ(B2:B81)`
Standardabweichung der Zellen B2 bis B81	`=STABW(B2:B81)`
Wurzel des Wertes in Zelle B5	`=WURZEL(B5)`
unteres Quartil der Zellen B2 bis B81	`=QUARTILE(B2:B81;1)`
oberes Quartil der Zellen B2 bis B81	`=QUARTILE(B2:B81;3)`
a-tes Perzentil der Zellen B2 bis B81	`=QUANTIL(B2:B81;a/100)`
b-Quantil der Zellen B2 bis B81	`=QUANTIL(B2:B81;b)`
Kovarianz (x-Werte in A2 bis A8, y-Werte in B2 bis B8)	`=KOVAR(A2:A8;B2:B8)`
Korrelationskoeffizient	`=KORREL(A2:A8;B2:B8)`
Steigung der Regressionsgeraden	`=STEIGUNG(B2:B8;A2:A8)`
y-Achsenabschnitt der Regressionsgeraden	`=ACHSENABSCHNITT(B2:B8;A2:A8)`

Erstellen eines Punktdiagramms mit Regressionsgerade

(1) Tabelle mit Werten markieren $\longrightarrow$ Einfügen $\longrightarrow$ Diagramm

(2) Diagrammtyp: Punkt(XY) $\longrightarrow$ weiter $\longrightarrow$ weiter $\longrightarrow$ Fertig stellen

(3) Rechtsklick auf Punkt im Diagramm $\longrightarrow$ Trendlinie hinzufügen $\longrightarrow$ Regressionstyp linear

(4) ggf. Haken setzen bei Gleichung anzeigen

(5) Diagramm nach Belieben beschriften, skalieren und formatieren

III Wichtige R-Befehle

Wurzel von a	`sqrt(a)`
7. Spalte von dat wird GRB zugeordnet	`GRB<-dat[,7]`
Maximum von GRB	`max(GRB)`
Minimum von GRB	`min(GRB)`
Summe von GRB	`sum(GRB)`
arithmetisches Mittel von GRB	`mean(GRB)`
Median von GRB	`median(GRB)`
Standardabweichung von GRB	`sd(GRB)`
Varianz von GRB	`var(GRB)`
Sortierung von GRB	`sort(GRB)`
absolute Häufigkeitstabelle von GRB	`table(GRB)`
relative Häufigkeitstabelle von GRB	`table(GRB/length(GRB))`
Länge von GRB	`length(GRB)`
zusammenfassende Statistiken von GRB	`summary(GRB)`
Quantile von GRB	`quantile(GRB)`
a-tes Perzentil von GRB	`quantile(GRB, a/100, type=1)`
b-Quantil von GRB	`quantile(GRB, b, type=1)`
Datenreihe dat manuell mit Daten füllen	`dat<-c(149, 147, 158)`
empirische Verteilungsfunktion aus dat erstellen und Fn zuordnen	`Fn<-ecdf(dat)`
Fn zeichnen	`plot(Fn, main="Überschrift", xlab="Beschriftung x-Achse", ylab="Beschriftung y-Achse")`
Kovarianz nach Pearson der Datenreihen x und y	`cov(x,y)`
Korrelationskoeffizient nach Pearson	`cor(x,y)`
lineare Regression	`reg<-lm(y~x)`
Punktediagramm plotten (`lwd` ist Linienstärke)	`plot(y~x, main="Punktediagramm", col="blue", lwd=4)`
Regressionsgerade plotten	`abline(ref, lwd=3)`
kumulative Summe der Datenreihe x	`cumsum(x)`
Boxplot der Datenreihe x	`boxplot(x, main="Überschrift")`

- VorlDreiWuerfel.txt (im selben Ordner wie R-Datei gespeichert) in R einlesen und `dat` zuordnen funktioniert mit `dat<-read.table(file="VorlDreiWuerfel.txt",header=TRUE)`

- Histogramm wird geplottet mit `hist(GRB, xlab="Beschriftung x-Achse", ylab="Beschriftung y-Achse", main="Überschrift", cex.main=1.9, col="gray75", border="black")`
  `cex.main` skaliert dabei die Größe der Überschrift, `col` definiert die Füllfarbe des Plots und `border` legt die Farbe des Plots fest

- Datenmatrix `dat` manuell mit Daten füllen; findet Spaltenweise statt, `ncol` definiert die Anzahl der Spalten, `nrow` die Anzahl der Zeilen: `dat<-matrix(c(1,2,3,4,5,6,7,8,9,7,4,1,2,5,8), ncol=5, nrow=3))`
  Matrix sieht dann wie folgt aus:

1	4	7	7	2
2	5	8	4	5
3	6	9	1	8