Herzlich willkommen zur Vorlesung

Statistik 1 (Teil 1)

Dr. Mirka Henninger

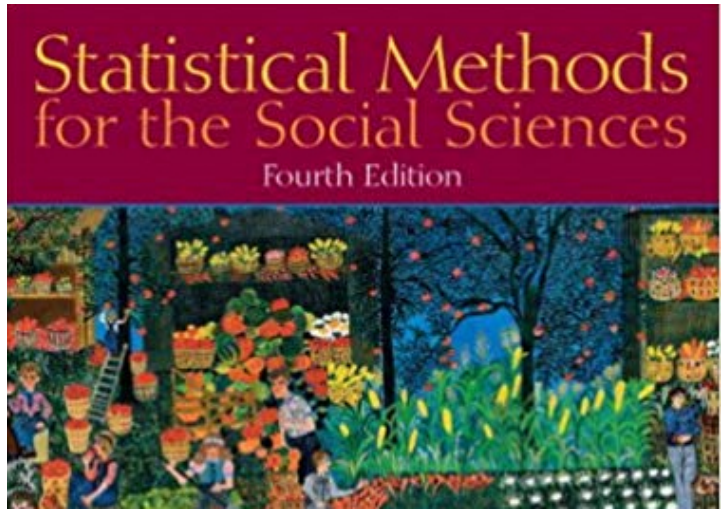
HS 2022

Organisatorisches

- Vorlesungsfolien, Formelsammlung und Übungsblätter im OLAT Kurs
- Vorlesung + Tutorate + Vor- und Nachbereitung zuhause = gute Prüfungsvorbereitung
- Beispiele für einführende Statistik-Lehrbücher

Bortz & Schuster

Agresti & Finlay



nur wenn Sie wollen! prüfungsrelevant sind ausschliesslich die Materialien im OLAT Kurs

Vorlesung

- Folien (ausgedruckt/elektronisch) + selber schreiben und zeichnen $\rightarrow$ Papier und Buntstifte mitbringen! (elektronisch nur wenn schnell genug)
- Podcast im OLAT $\rightarrow$ keine anderen Aufzeichnungen!
- nach der Pause pünktlich zurück
- Ruhe!
- Fragen: jederzeit zwischendurch melden oder nach der Vorlesung – aber aufgrund der hohen Teilnehmerzahl <u>nicht per email</u>
- ab einem gewissen Zeitpunkt: Nachbereitung zuhause nötig! Verpassen Sie den Anschluss nicht!

Tutorate

- bei jedem Tutorat in derselben Woche gleicher Stoff
- Aufgaben vorher zuhause versuchen zu lösen!
  (Übungsblätter im OLAT zum selber Ausdrucken)
- es werden keine offiziellen Musterlösungen herausgegeben
  (Diskussion zwecklos)
- es zwingt Sie niemand hinzugehen - aber wenn Sie hingehen
  verhalten Sie sich kollegial!
- mehrmals pro Semester Übungs-Quiz mit prüfungsähnlichen
  Multiple-Choice Fragen

Propädeutikums-Prüfung

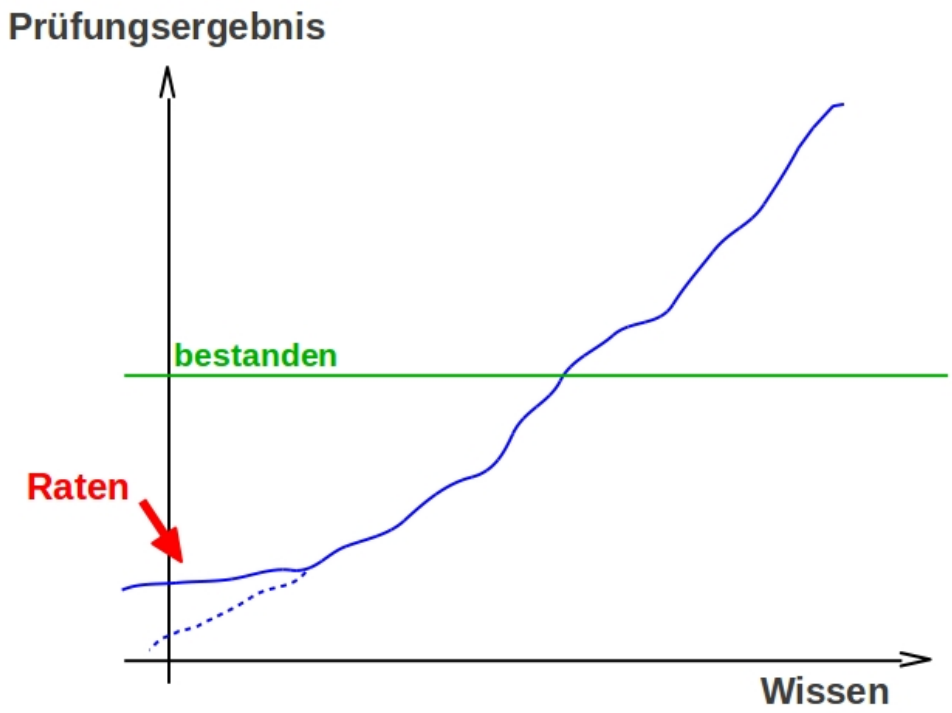
- zu Teil 1 und Teil 2 jeweils 10 Fragen
- d.h. pro Frage durchschnittlich 2 Minuten Zeit (für manche braucht man nur 30 Sekunden, für manche über 3 Minuten)
- Multiple-Choice mit Partial-Credit-Bewertung
  z.B. „Kreuzen Sie zwei richtige/alle richtigen Antworten an.“
- zugelassene Hilfsmittel
  - Formelsammlung (wie im OLAT, wird ausgegeben)
  - Taschenrechner (nur die zugelassenen Modelle!)
    `http://www.psychologie.uzh.ch/de/studium/bscmsc/studium/pruefungen/rundum.html`
- verbale Aufgaben und Rechenaufgaben (vgl. Quizze)
- bei Rechenaufgaben:
  - in jedem Rechenschritt auf mind. 3 Stellen runden
  - Zahlen sind so gewählt, dass Ergebnisse eindeutig
  - ggf. die Antwort ankreuzen, die am nächsten dran

Soll man in der Propädeutikums-Prüfung raten?

Multiple-Choice mit Partial-Credit-Bewertung, d.h.

- innerhalb einer Aufgabe Punktabzug wenn (zusätzlich zu den richtigen) falsche Antworten angekreuzt
- aber pro Aufgabe nie unter 0 Punkte

Aber Ihr Ziel sollte sein, nicht raten zu müssen, denn:



$\Rightarrow$ nicht die Ratestrategie entscheidet, sondern Ihre Vorbereitung!

Propädeutikums-Prüfung

Vorbereitung: Materialien aus dieser Veranstaltung

- Übungsaufgaben aus den Tutoraten
- Quizze aus den Tutoraten
- Aufgaben aus den Folien nachrechnen
- alte Propädeutikums (früher Assessment) Aufgaben ab 2012 (viele davon sind bereits in unsere Quizze integriert)

- nicht sinnvoll, z.T. sogar irreführend und fehlerhaft: Aufgaben von früheren Dozenten vor 2012, aus anderen Unis, Aufgaben aus Büchern, alte oder kommerziell angebotene Skripte, Webseiten, Karteikarten o.ä.

- sinnvoll: Lerngruppen bilden, Lernplan machen
- entscheidend sind: Stoff-Menge und Rechen-Routine

Sprachliches

- Sie dürfen Fragen gerne auf Schweizerdeutsch stellen, betreffende Folie/Blatt mitbringen hilft beim Erklären
- Beispiele sind aus didaktischen Gründen inhaltlich und sprachlich vereinfacht

Statistik – Warum tun wir Ihnen das an?!?

Inhalte – Statistik 1, Teil 1

1. Deskriptive Statistik
2. Wahrscheinlichkeitstheorie
3. Tests und Konfidenzintervalle
4. Kovarianz und Korrelation

Emprirische Forschung und Skalenniveaus

Phasen der empirischen Forschung

Erkundungsphase Theoretische Phase Planungsphase Untersuchungsphase	Vorlesung Forschungsmethoden
Auswertungsphase Entscheidungsphase	Vorlesung Statistik

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Wir haben eine interessante wissenschaftliche Studie durchgeführt und folgende Daten erhoben:

Person	Alter	Geschlecht	Lieblingsfarbe	Begeisterung für Sport
1	22	m	grün	mittel
2	53	w	gelb	hoch
3	37	x	grün	niedrig
4	31	w	rot	mittel
⋮	⋮	⋮	⋮	⋮

Welche Auswertungen machen Sinn?

- mittleres Alter der Befragten
- mittlere Lieblingsfarbe der Befragten
- am häufigsten genannte Lieblingsfarbe der Befragten

- Das Prinzip der Skalenniveaus wurde von dem Psychophysiker S. S. Stevens in einem Artikel von 1946 („On the Theory of Scales of Measurement“) explizit formuliert.
- Für Statistik von sehr großer Bedeutung: Je nach Skalenniveau sind nur bestimmte Auswertungsmethoden sinnvoll.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Skalenniveaus

Absolutskala
Identität

~~Differenzenskala~~
~~$a + x$~~

Verhältnisskala
$b \cdot x$

Intervallskala
$a + b \cdot x$

„metrische“ Skalen

Ordinalskala
streng monotone Transformationen

Nominalskala
eineindeutige Transformationen

Nominalskala

- Niedrigstes Skalenniveau.
- Messen auf Nominalskalenniveau ist Klassifikation.
- Grundlegende Operation ist die Feststellung der Gleichheit oder Ungleichheit.
- Beispiele: Farbe, Geschlecht, Nationalität, Religionszugehörigkeit, usw.
- Alle eineindeutigen Transformationen sind zulässig:
  rot $\rightarrow$ 1, blau $\rightarrow$ 2, grün $\rightarrow$ 3
  aber nicht: rot $\rightarrow$ 1 und blau $\rightarrow$ 1, grün $\rightarrow$ 2

Zuordnung

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Ordinalskala

- Objekte können in eine Rangreihe gebracht werden, aber die Abstände zwischen den Kategorien sind nicht gleich bzw. nicht bedeutsam.
- Grundlegende Operationen: $>$, $<$, $=$
- Beispiele: Rangordnung von Bewerbenden, Ränge beim Militär, Schulabschlüsse, Grad der Zustimmung.
- Alle streng monotonen (also ordnungserhaltenden) Transformationen sind zulässig:
  gut $\rightarrow$ 1, mittel $\rightarrow$ 2, schlecht $\rightarrow$ 3
  aber auch: gut $\rightarrow$ 42, mittel $\rightarrow$ 17, schlecht $\rightarrow$ 3
  Kodierung oft durch Zahlen, aber Abstände nicht interpretieren!
- häufiger Streitpunkt: ist eine psychologische Skala ordinal oder metrisch zu interpretieren? (vgl. spätere Vorlesungen zu Testtheorie und Skalierungsverfahren)

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Intervallskala

- Erste metrische Skala, d.h. die Skala besteht aus gleich großen Intervallen, den Einheiten.
- Beispiele: Temperatur, psychologische Messungen wie z.B. IQ-Werte.
- Zulässige Transformationen für Intervallskalen sind

$$a + b \cdot x \text{ (mit } b > 0)$$

  Verschiebung des Nullpunktes und Stauchung/Streckung der Einheiten (d.h. Nullpunkt und Einheit können willkürlich festgelegt werden).
- Umrechnungsformeln zwischen Grad Celsius $^\circ C$ und Grad Fahrenheit $^\circ F$:

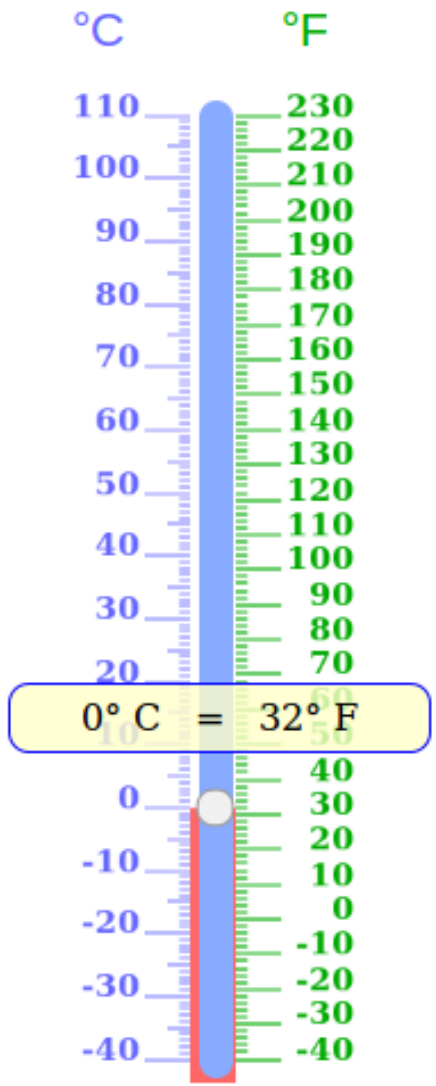
$$C = \frac{5}{9}(F - 32)$$

$$F = 32 + \frac{9}{5} \cdot C$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Intervallskala

°C	°F	Description
180	**356**	Moderate Oven
100	**212**	Water boils
40	**104**	Hot Bath
37	**98.6**	Body temperature
30	**86**	Beach weather
21	70	Room temperature
10	**50**	Cool Day
0	**32**	Freezing point of water
−18	0	Very Cold Day
−40	**−40**	Extremely Cold Day (and the same number!)
(**bold** are exact)		



Quelle: `http://www.mathsisfun.com`

Beispiel

- Die Temperatur gestern war $5^\circ C$ und heute haben wir $10^\circ C$.
- Ist die Aussage „Heute ist es doppelt so warm wie gestern" sinnvoll?
- Umrechnen von $5^\circ C$ und $10^\circ C$ zur Fahrenheit-Skala ergibt: $41^\circ F$ und $50^\circ F$.
- $50^\circ F$ ist nicht doppelt so hoch wie $41^\circ F$. Die Aussage ist auf Intervallskalenniveau <u>nicht</u> sinnvoll.
- Anstatt „nicht sinnvoll" sagt man auch „nicht bedeutsam".

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Zulässige Transformationen für Verhältnisskalen sind

$$b \cdot x \text{ (mit } b > 0)$$

  Verhältnisskalen haben einen festen Nullpunkt, aber die Größe der Einheiten kann durch Stauchung/Streckung verändert werden (d.h. die Einheit kann willkürlich festgelegt werden).
- Beispiele: Länge (in cm oder in m), Alter (in Jahren oder in Monaten), monatliches Nettoeinkommen (in CHF oder in EUR).
- Umrechnungsformeln zwischen Länge in cm und Länge in m:

$$cm = 100 \cdot m$$

- Aussagen über Verhältnisse bleiben unverändert.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Absolutskala

- Zulässige Transformation für die Absolutskala ist nur die Identität (also keine Veränderung).
- Beispiele: Anzahl gekaufter Produkte, Anzahl Kinder, Wahrscheinlichkeiten
- Höchstes Skalenniveau $\Leftrightarrow$ strengste Transformation.

metrische Skalen	Natürlicher Nullpunkt		natürliche Messeinheit	
Intervallskala	—	(c)	—	(b)
Verhältnisskala	✓	(-)	—	(b)
Absolutskala	✓	(-)	✓	(-)

Deskriptive Statistik

Wo sind wir?

1. Deskriptive Statistik
1.1 Maße der zentralen Tendenz (Lagemaße)
1.2 Maße der Variabilität (Streuungsmaße)
1.3 Lineare Transformationen
1.4 Diagramme

2. Wahrscheinlichkeitstheorie

3. Tests und Konfidenzintervalle

4. Kovarianz und Korrelation

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Maße der zentralen Tendenz (Lagemaße)

- Mittelwert (arithmetisches Mittel)
  ab metrischem Skalenniveau
- Median (50%-Quantil)
  ab Ordinalskala
- Modalwert (Modus)
  ab Nominalskala

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Mittelwert

- Für n Werte berechnet man:

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

- Numerisches Beispiel: Fünf Werte: 3, 5, 6, 8, 14
- Man errechnet

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(3+5+6+8+14)\\ &= 36/5 = 7{,}2.\end{aligned}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Eigenschaften des Mittelwerts

- Die Summe der Abweichungen vom Mittelwert ist null, d. h.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0.$$

- Der Mittelwert ist sensitiv gegenüber Ausreißern.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Der Mittelwert ist sensitiv gegenüber Ausreißern*

- Angenommen aufgrund eines Tippfehlers erhällt man: 3, 5, 6, 8, und 144.
- Die letzte Zahl sollte eigentlich 14 sein.
- Als Mittelwert ergibt sich 33,2 anstatt 7,2.
- Offensichtlich ist 33,2 keine gute Repräsentation der zentralen Lage der Daten.

* unter Ausreißern versteht man extreme Werte (das können Tippfehler sein, aber auch echte Werte von einzelnen Personen mit extremen Ausprägungen eines Merkmals)

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Mitdenk-Frage

Sie wollen auf die Geldknappheit unter Zürcher Studierenden aufmerksam machen. Dazu befragen Sie die Teilnehmenden Ihres Statistik I Tutoriums, wie viel Geld sie monatlich zur Verfügung haben, und berechnen daraus den Mittelwert.
Am Tag ihrer Umfrage ist auch ein Millionärs-Sohn anwesend.

- Was passiert?
- Macht es einen Unterschied, ob in Ihrer Tutoriums-Gruppe 20 oder 200 Leute sind?

Median

- Der Median „halbiert“ die Häufigkeitsverteilung.
- Mindestens 50% aller Werte sind $\leq$ der Median und mindestens 50% aller Werte sind $\geq$ der Median.
- Der Median ist das 50%-Quantil.
- Der Median setzt mindestens ordinales Skalenniveau voraus.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Berechnung des Medians

- Sortiere die Daten. (Nicht vergessen!)
- Sortierte Daten bezeichnet man oft durch: $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$.
- Falls n ungerade:

$$\text{Median} = x_{(\frac{n+1}{2})}$$

- Falls n gerade:

$$\text{Median} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$$

(bei ordinalen Daten ist diese Definition – die üblicherweise auch von Statistik-Programmen verwendet wird – nicht ideal; per Definition ist sowohl $x_{(\frac{n}{2})}$ als auch $x_{(\frac{n}{2}+1)}$ Median)

Beispiel

- Ungerader Stichprobenumfang, $n = 5$:
  - Unsortierte Werte: 5, 2, 3, 7, 8
  - Sortierte Werte: 2, 3, 5, 7, 8
  - Median $= x_{(\frac{n+1}{2})} = x_{(\frac{5+1}{2})} = x_{(3)} = 5$.
- Gerader Stichprobenumfang, $n = 6$:
  - Unsortierte Werte: 2, 8, 6, 4, 12, 10
  - Sortierte Werte: 2, 4, 6, 8, 10, 12
  - Median $= \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{(\frac{6}{2})} + x_{(\frac{6}{2}+1)}}{2} = \frac{x_{(3)} + x_{(4)}}{2} = \frac{6+8}{2} = 7$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Eigenschaften des Median

- Der Median ist nicht so sensitiv gegenüber Ausreißern wie der Mittelwert.
- Illustration mit vorherigem Beispiel: 3, 5, 6, 8, und 144 anstatt 14.
- Der Median (6) verändert sich hier nicht.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Modalwert (Modus)

- Der Modalwert ist der am häufigsten vorkommende Wert (d.h. das Maximum der Häufigkeitsverteilung).
- Der Modalwert ist per Definition nicht notwendigerweise eindeutig.
- Üblicherweise spricht man aber nur bei solchen Verteilungen von einem Modalwert, die ein eindeutiges Maximum besitzen.
- Beispiel: 2, 1, 3, 5, 2, 4, 3, 2
  Der Wert 2 kommt am häufigsten vor und ist daher Modalwert.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Maße der Variabilität (Streuungsmaße)

- Varianz
- Standardabweichung
- Variationsbreite (Range)
- Interquartilbereich

ab metrischem Skalenniveau

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Varianz und Standardabweichung

- Die Varianz ist definiert als

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

  (durchschnittliche quadrierte Abweichung vom Mittelwert).
- Die Wurzel der Varianz bezeichnet man als Standardabweichung, d. h. $s = \sqrt{s^2}$.
- Die Standardabweichung hat die gleiche Einheit wie die Rohwerte.
- Oft wird „Streuung" synonym für Standardabweichung verwendet. Andere Quellen verwenden „Streuung" allgemeiner als Bezeichnung für die Variabilität der Daten.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

Reaktionszeiten von Versuchspersonen in Sekunden:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
3,3	0,8	0,64
1,7	−0,8	0,64
2,0	−0,5	0,25
4,0	1,5	2,25
1,3	−1,2	1,44
2,0	−0,5	0,25
3,0	0,5	0,25
2,7	0,2	0,04
3,7	1,2	1,44
2,3	−0,2	0,04
1,7	−0,8	0,64
2,3	−0,2	0,04
		7,92

Mittelwert

$$\bar{x} = 2{,}5.$$

Varianz:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{7{,}92}{11} = 0{,}720$$

Standardabweichung:

$$s = \sqrt{0{,}72} = 0{,}849$$

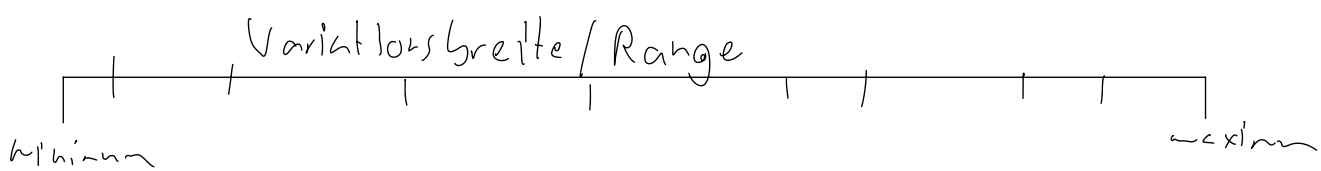
©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Variationsbreite

- $x_{(1)}$ = kleinster Wert, Minimum
  $x_{(n)}$ = größter Wert, Maximum
  Variationsbreite $= x_{(n)} - x_{(1)}$
- Die Variationsbreite wird auch Range genannt.
- Die Variationsbreite ist sensitiv gegenüber Ausreißern.
- Die Variationsbreite ist kein sehr nützliches Maß zur Charakterisierung der Variabilität.

Quartil $= \frac{1}{4}$

- Kontrolle von Minimum und Maximum kann aber hilfreich sein, um unplausible Werte zu entdecken, z.B.:
  Alter von Studierenden reicht von 11 bis 317 Jahren?!? $\rightarrow$ vermutlich Tippfehler im Datensatz.



Ausreisser empfindlich

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Interquartilbereich

- Weniger ausreißeranfällig: Bereich der mittleren 50% aller Werte.
- Dieser Bereich ist durch Werte (sog. Angelpunkte) begrenzt, die die unteren 25% bzw. die oberen 25% der Verteilung abschneiden.
- Der untere Angelpunkt, der die unteren 25% abschneidet, wird als erstes Quartil Q_1 (oder 25%-Quantil) bezeichnet.
- Der obere Angelpunkt, der die oberen 25% abschneidet, wird als drittes Quartil Q_3 (oder 75%-Quantil) bezeichnet.
- Interquartilbereich $IQR = Q_3 - Q_1$
- Der IQR drückt die Länge des Bereichs aus, über den die mittleren 50% einer Rohwerteverteilung streuen.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Quantile (Perzentile)

- Quantile sind Kennwerte, die die relative Position eines Messwertes innerhalb der Stichprobe zum Ausdruck bringen.
- Ein Quantil bezieht sich immer auf einen vorgegebenen Prozentsatz.
- Das Quantil x_p ist der Messwert, der p-Prozent der Werte „abschneidet".
- Vgl. Definition Median:
  Mindestens 50% aller Werte sind $\leq$ der Median und mindestens 50% aller Werte sind $\geq$ der Median.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Spezielle Quantile (Perzentile)

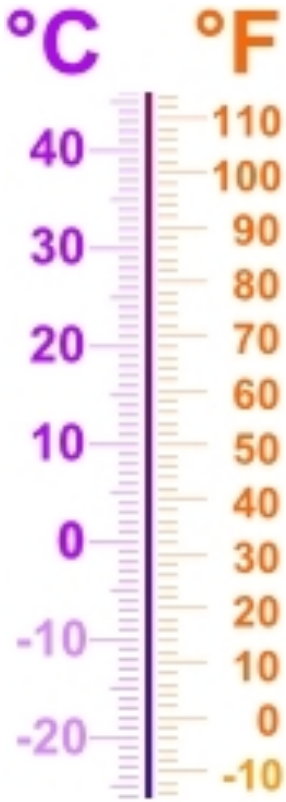
- x_{25}, x_{50}, x_{75}: Quartile
- x_{50}: Median
- x_{25}: 1. Quartil Q_1, x_{75}: 3. Quartil Q_3
- $x_{10}, x_{20}, \ldots, x_{90}$: Dezile

- Außer für den Median ist die Bestimmung von Quantilen per Hand nicht praktikabel.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Lineare Transformationen

Bei intervall- und verhältnisskalierten Daten sind sog. lineare Transformationen zulässig, z.B. Umrechnung von $^\circ C$ in $^\circ F$ $(a + b \cdot x)$ oder von CHF in EUR $(b \cdot x)$:



CHF	EUR
0	0
1	0,96
2	1,92
3	2,88
4	3,84
⋮	⋮
10	9,60

Quelle: `http://www.mathsisfun.com`; Wechselkurs: April 2015

Wie verändern sich Mittelwert und Varianz/Standardabweichung dadurch? Muß man Mittelwert und Varianz/Standardabweichung nochmal komplett neu berechnen?

- Erhobene Werte: x
- Transformierte Werte: y
- Transformation: $y = a + b \cdot x$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiele

- Fünf x-Werte:

$$2, 3, 4, 5, 6.$$

- Für diese Werte ermittelt man: $\bar{x} = 4$ und $s_x = 1{,}581$
- Für die lineare Transformation $y = 2 + x$ ergeben sich die Werte:

Verschiebung

$$4, 5, 6, 7, 8$$

- Mittelwert und Standardabweichung nochmals neu berechnen:
- $\bar{y} = 6$ und $s_y = 1{,}581$
- Für die lineare Transformation $y = 3 \cdot x$ ergeben sich die Werte:

Streckung

$$6, 9, 12, 15, 18$$

- Mittelwert und Standardabweichung nochmals neu berechnen:
- $\bar{y} = 12$ und $s_y = 4{,}743$
- Gibt es allgemeine Regeln, was sich als neuer Mittelwert und Varianz/Standardabweichung ergibt?

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Allgemeine Regeln

- Mittelwert: $\bar{y} = a + b \cdot \bar{x}$
- Varianz: $s_y^2 = b^2 \cdot s_x^2$
  Standardabweichung: $s_y = |b| \cdot s_x$
- Beispiel: Für die lineare Transformation $y = 2 + 3 \cdot x$ ergeben sich:

$$8, 11, 14, 17, 20$$

- Mittelwert und Standardabweichung nochmals neu berechnen:
- $\bar{y} = 14$ und $s_y = 4{,}743$
- ergibt sich auch durch: $\bar{y} = 2 + 3 \cdot 4$ und $s_y = |3| \cdot 1{,}581$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Spezielle lineare Transformation: z-Transformation

- z-Wert = Abweichung vom Mittelwert im Verhältnis zur Standardabweichung:

$$
\begin{aligned}
z &= \frac{x - \bar{x}}{s} \\
&= a + b \cdot x \\
&= -\frac{\bar{x}}{s} + \frac{1}{s} \cdot x
\end{aligned}
$$

- Der z-Wert gibt damit an, um wieviele Standardabweichungen ein Wert über dem Mittelwert liegt.
- Für z-transformierte Werte gilt: $\bar{z} = 0$ und $s_z^2 = 1$.
- z-Werte sind dimensionslose Zahlen. Vorteil: sie sind direkt vergleichbar, egal welche Einheit die ursprünglichen Werte haben.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Vergleich zweier Personen aus zwei Ländern mit unterschiedlichen Tests zur beruflichen Eignung.
- Person A hat 150 Punkte in einem Test mit Mittelwert 100 und Standardabweichung 50 (bestimmt an einer großen Stichprobe aus diesem Land).
- Person B hat 20 Punkte in einem Test mit Mittelwert 10 und Standardabweichung 5.
- Wen soll man z.B. für einen Job auswählen? Person A weil sie mehr Punkte hat?

$$z_A = \frac{150 - 100}{50} = 1$$

$$z_B = \frac{20 - 10}{5} = 2$$

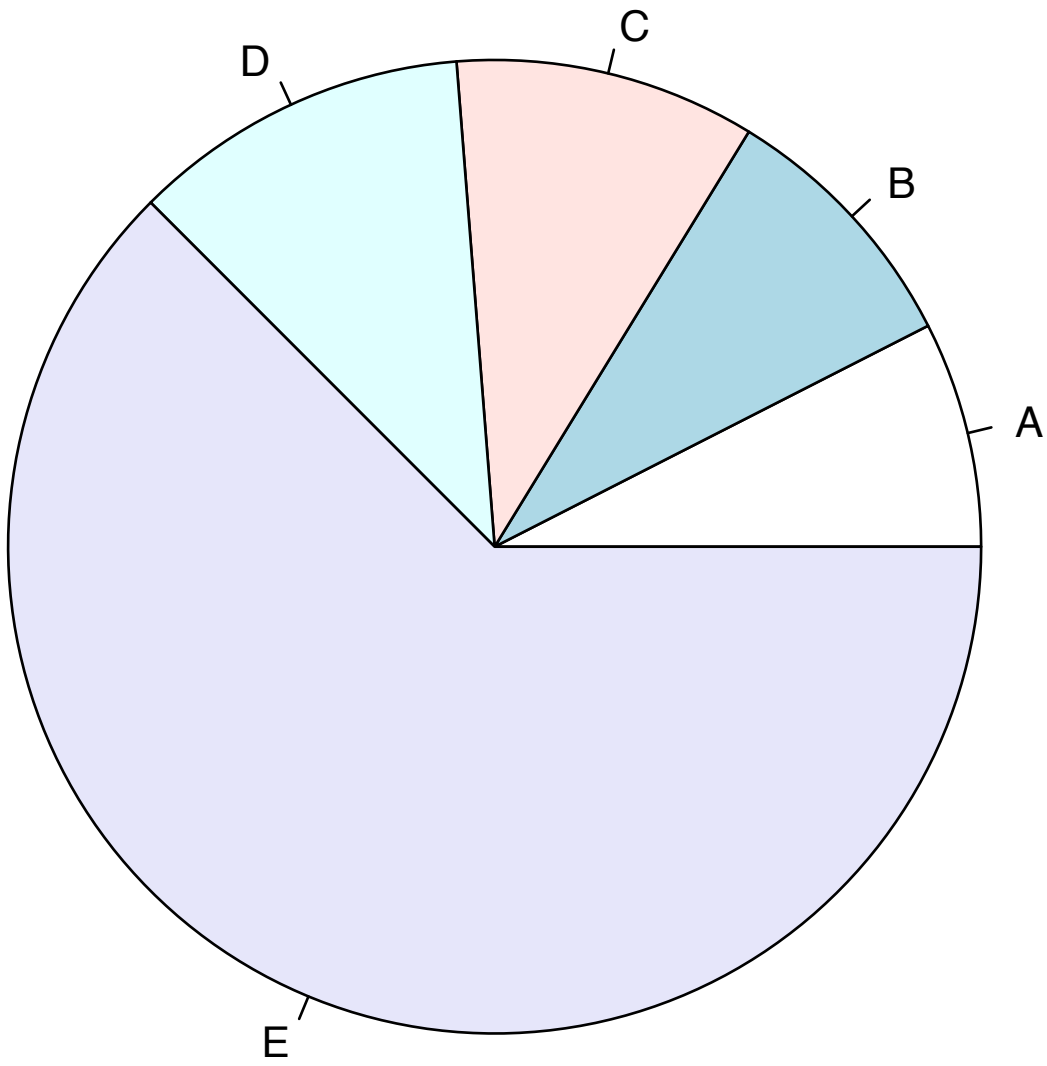
- Person A liegt nur eine Standardabweichung über dem Landesmittel, Person B hingegen 2 Standardabweichungen.
- Ich würde Person B auswählen.

Diagramme

Diagramme zur grafischen Darstellung von Daten

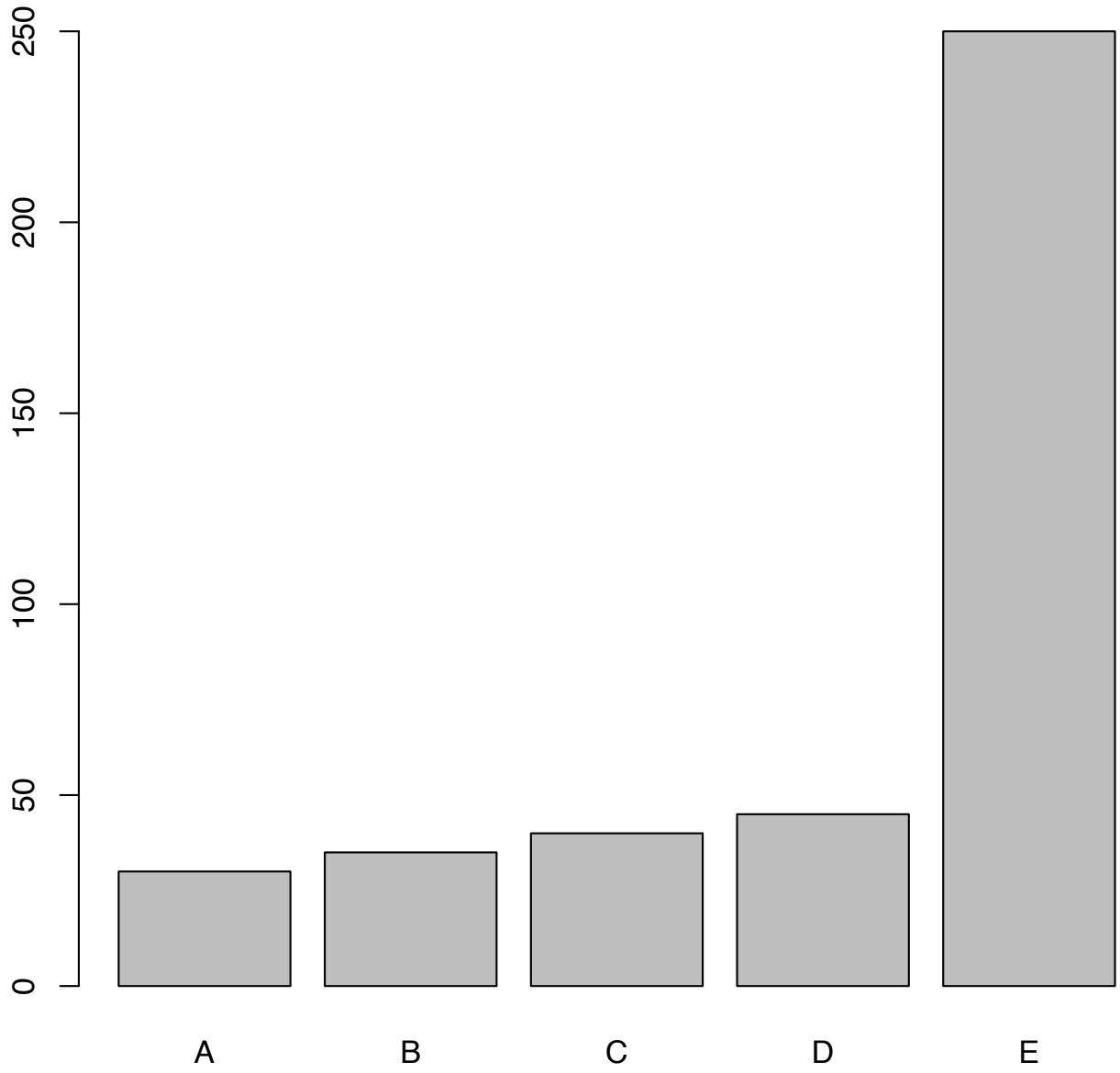
- für nominalskalierte Merkmale
  - Kreisdiagramm
  - Balkendiagramm
  - ...
- für ordinalskalierte Merkmale
  - Balkendiagramm
  - ...
- für metrische Merkmale
  - Boxplot
  - Histogramm
  - ...

Kreisdiagramm (alias „Tortendiagramm”)

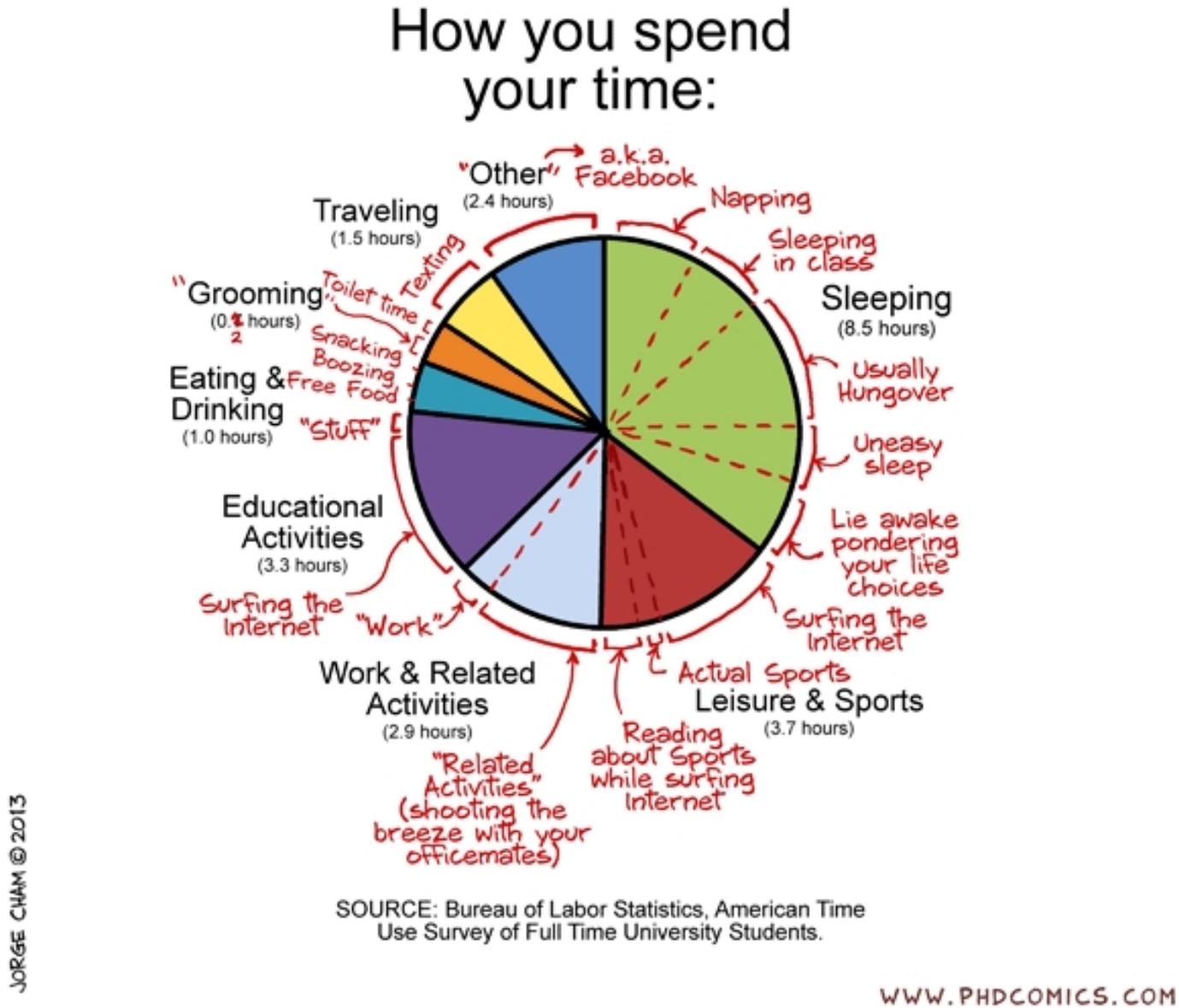


Winkel/Fläche entspricht Häufigkeit, aber schwer zu vergleichen.
Welches Stück ist größer: B oder C?

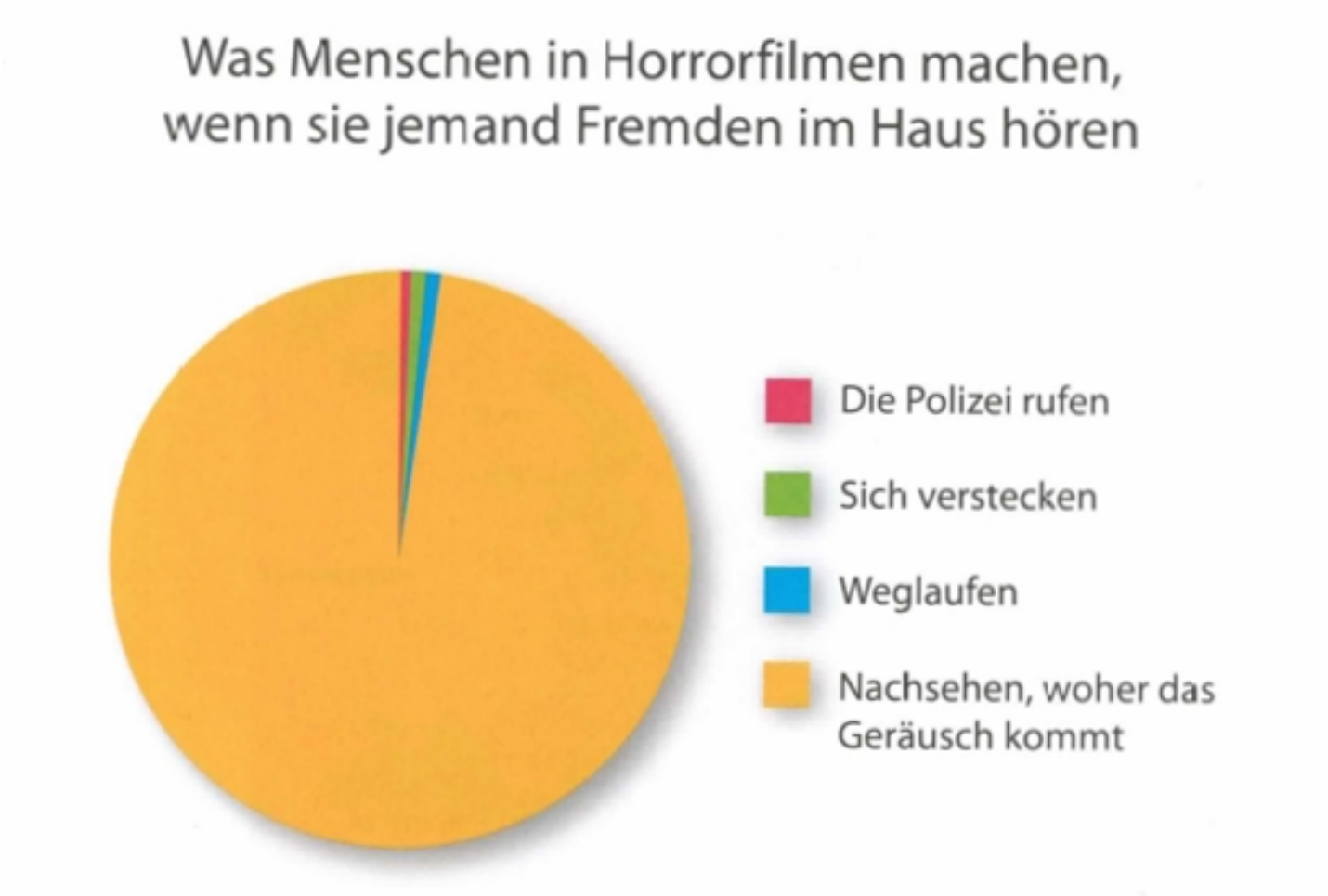
Balkendiagramm



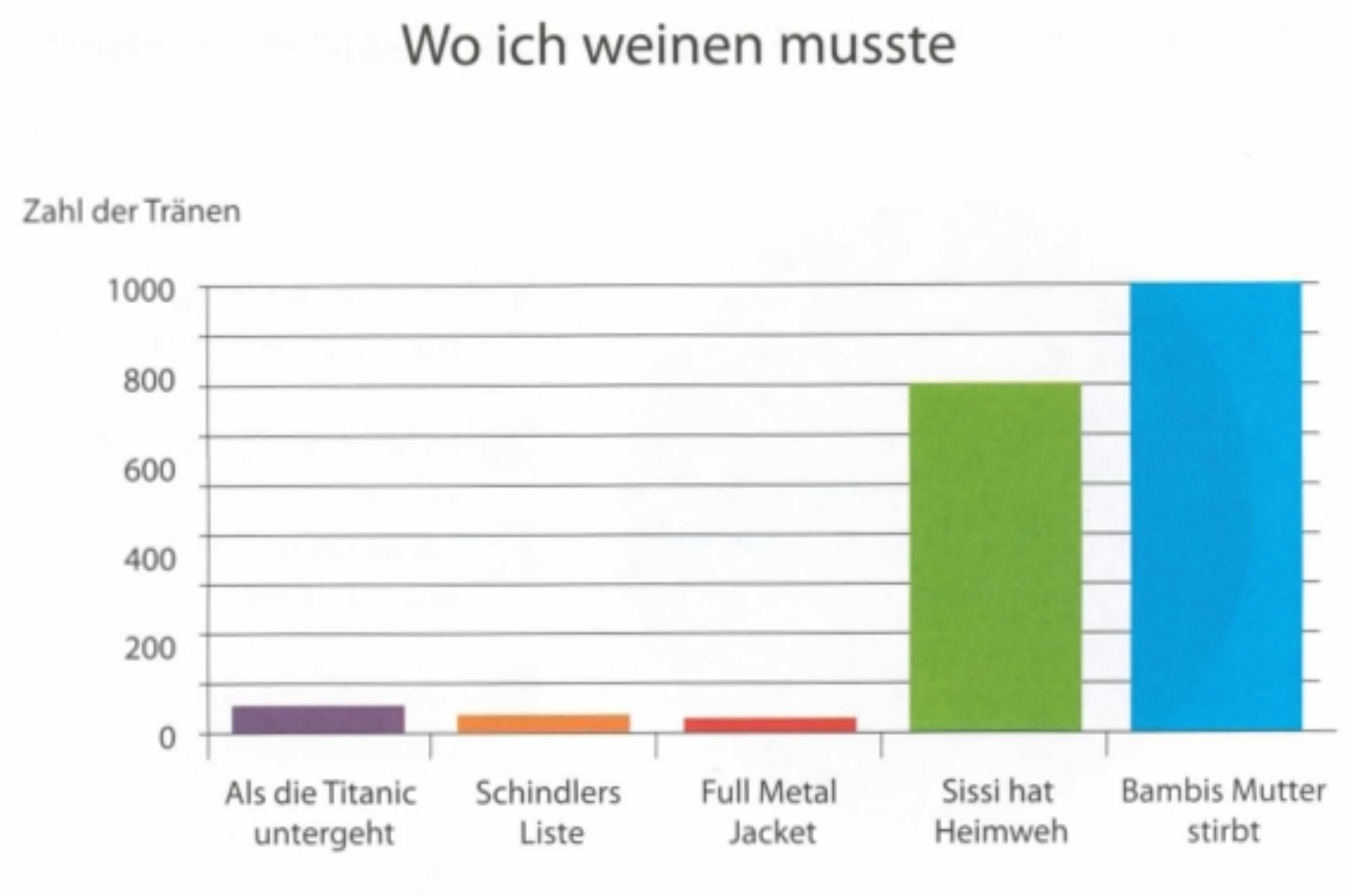
Höhe entspricht Häufigkeit, viel leichter zu vergleichen.



Quelle: `http://www.phdcomics.com`



Quelle: Katja Berlin & Peter Grünlich: *Was wir tun, wenn der Aufzug nicht kommt.*



Quelle: Katja Berlin & Peter Grünlich: *Was wir tun, wenn der Aufzug nicht kommt.*

Boxplot

- Eine Möglichkeit zur gleichzeitigen Veranschaulichung von Lage und Streuung einer Verteilung bietet der Boxplot.
- Die grundlegende Idee ist dabei, die mittleren 50% einer Verteilung durch eine „Box" zu repräsentieren, in die auch der Median eingezeichnet wird.
- Insofern kann man an der Box den Abstand der drei Quartile zueinander erkennen.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Gegeben sind folgende zehn Testergebnisse eines psychologischen Tests, wobei die Werte bereits nach der Größe sortiert wurden:

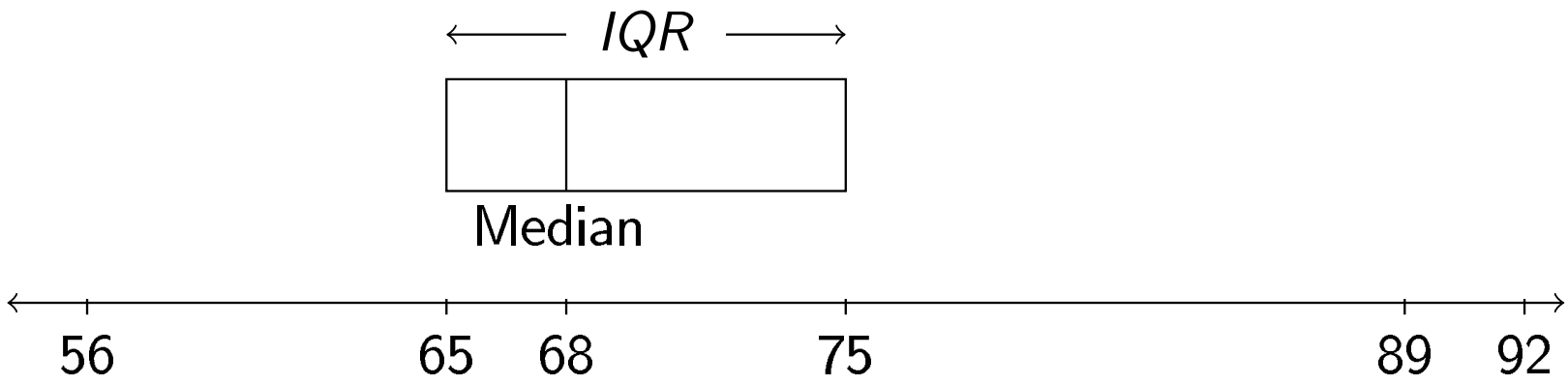
$$56, 57, 65, 66, 67, 69, 72, 75, 89 \text{ und } 92.$$

- Für die „Box“ des Boxplots werden folgende Größen benötigt:

$$\begin{aligned} Q_1 &= x_{(3)} = 65 \\ Median &= (x_{(5)} + x_{(6)})/2 = 68 \\ Q_3 &= x_{(8)} = 75 \\ IQR &= Q_3 - Q_1 = 10 \end{aligned}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Boxplot



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Ränder der Verteilung

- Die Werte an den Rändern der Verteilung werden ebenfalls in dem Boxplot repräsentiert.
- Dazu werden Striche eingezeichnet, die jeweils von einem Ende der Box ausgehen.
- Diese Striche werden „Whiskers" (Schnurrhaare) genannt.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Länge der Whiskers

- Wir betrachten z.B. den Whisker für den oberen Rand der Verteilung:
- Zunächst wird der IQR bestimmt und dann das 1,5-fache des IQR zu dem oberen Angelpunkt Q_3 addiert: $Q_3 + 1{,}5 \cdot IQR$
- Dieser Wert ist die Obergrenze für die Länge des Whiskers.
- Beobachtungen, deren Werte diese Grenze überschreiten, werden als Ausreißer betrachtet und als zusätzliche Punkte eingezeichnet.

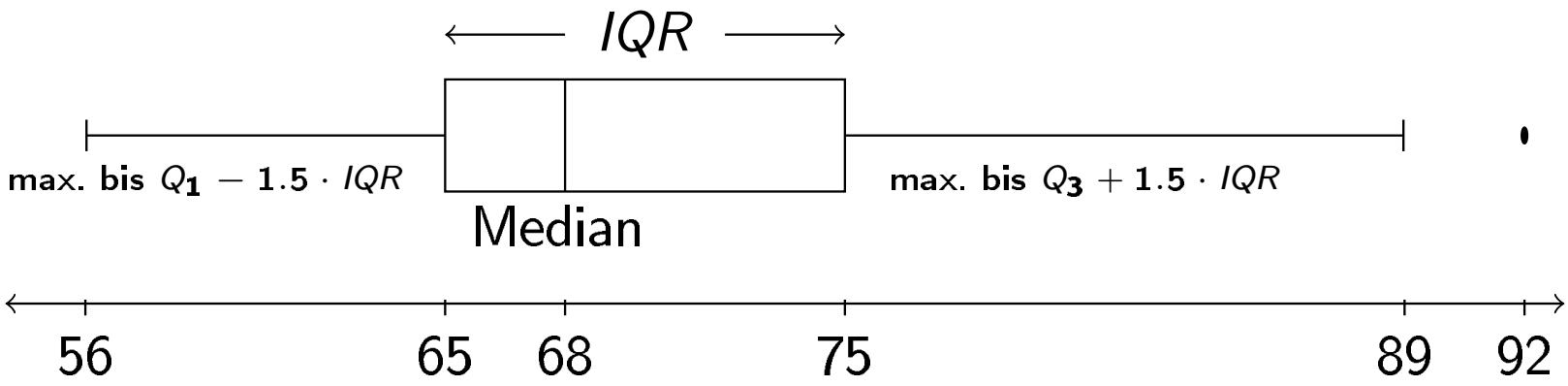
©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Um die Länge des unteren Whiskers zu bestimmen, berechnen wir die untere Grenze $Q_1 - 1{,}5 \cdot IQR = 50$.
- Die kleinste Beobachtung, die diese Grenze nicht unterschreitet, ist $x_{(1)} = 56$.
- Der Whisker endet deshalb bei diesem Wert.
- Die entsprechende Grenze, um den oberen Whisker zu bestimmen, ist $Q_3 + 1{,}5 \cdot IQR = 90$.
- Der größte beobachtete Wert, der diese Grenze nicht überschreitet, ist $x_{(9)} = 89$.
- Deshalb wird der obere Whisker bis zu dieser Stelle gezogen.
- Da die Beobachtung $x_{(10)} = 92$ die Grenze von 90 aber überschreitet, wird dieser Wert zusätzlich in den Plot eingezeichnet.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

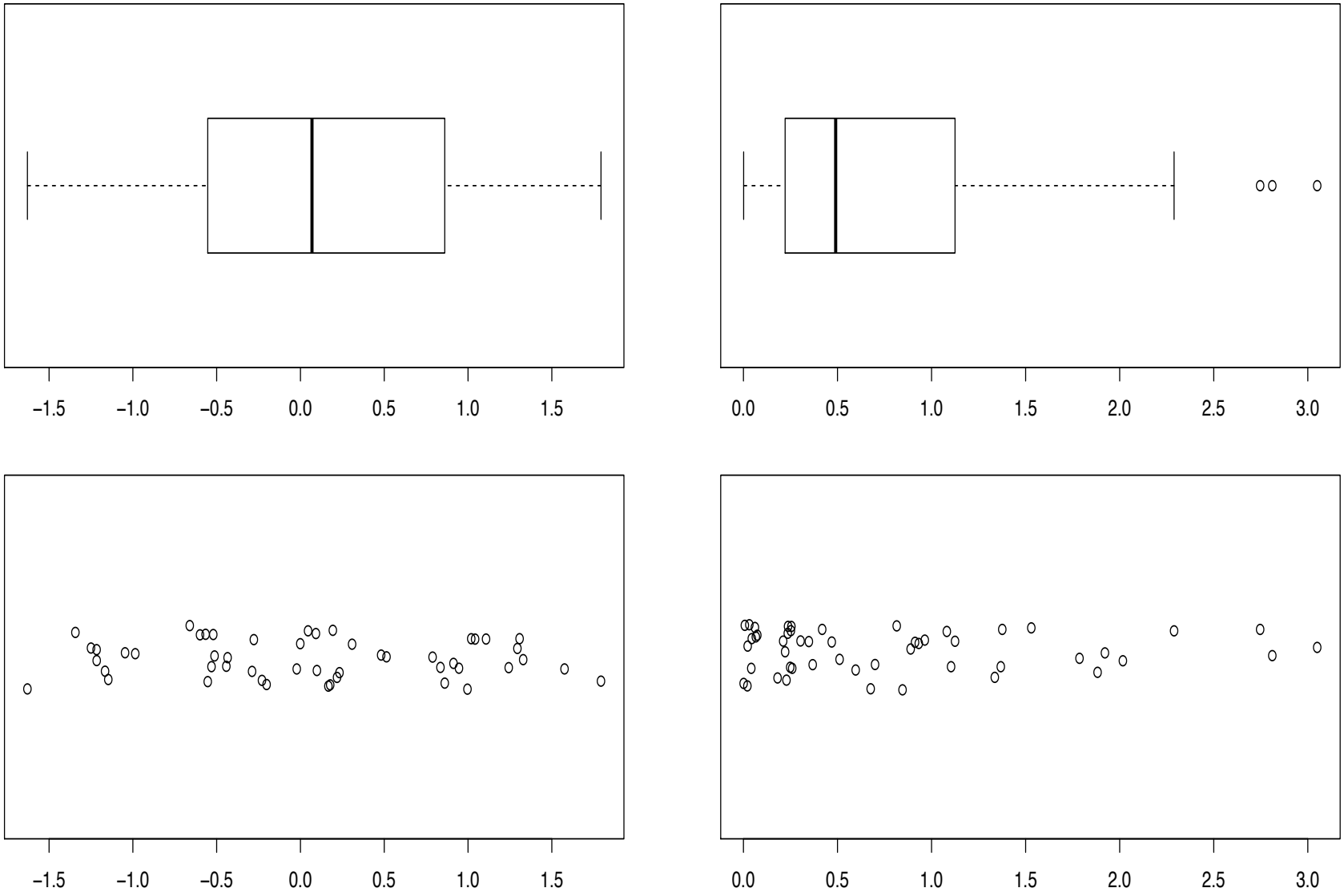
Boxplot



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

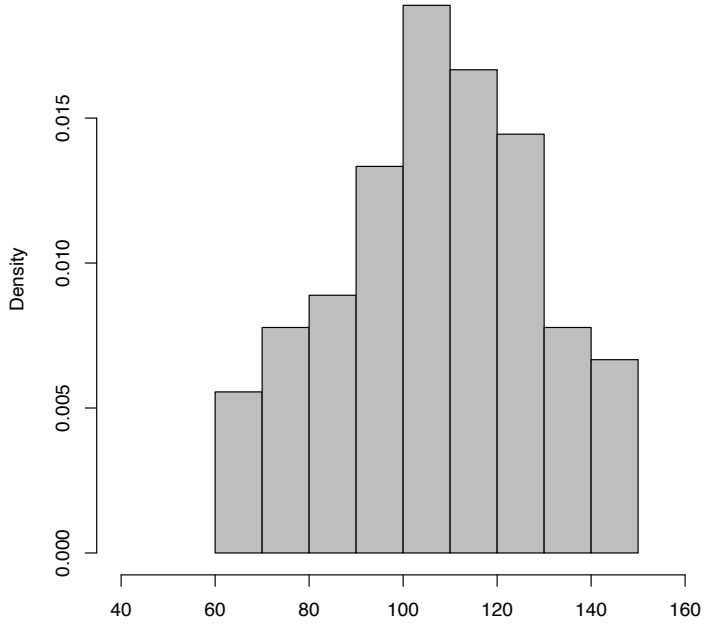
Boxplot

An der Form des Boxplots sieht man die Verteilung der Werte:



Histogramm

Zur grafischen Veranschaulichung der Verteilung eines metrischen Merkmals kann auch ein Histogramm angefertigt werden. Die Messwerte werden dazu in Kategorien eingeteilt.

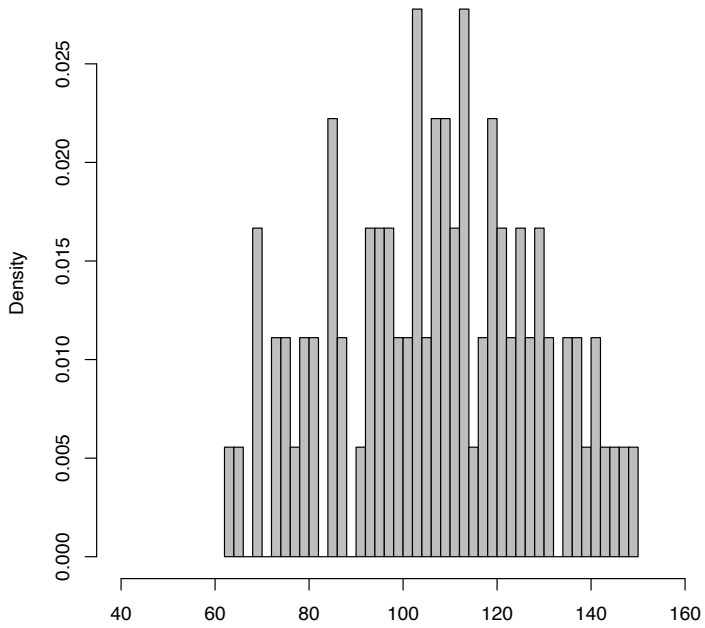


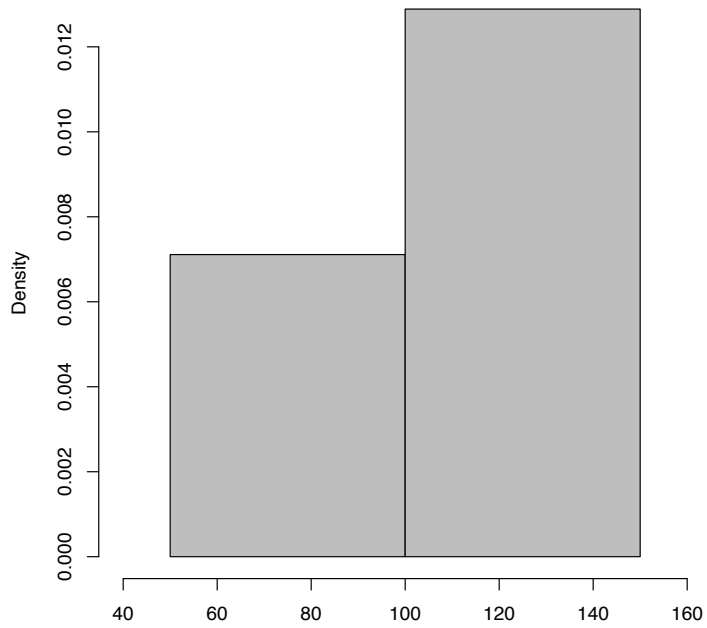
Beispiel

Kategorie	Anzahl	Anteil	Breite	Höhe = Anteil/Breite
$[60; 70[$	5	$0.05\overline{5}$	10	$0.005\overline{5}$
$[70; 80[$	7	$0.07\overline{7}$	10	$0.007\overline{7}$
$[80; 90[$	8	$0.08\overline{8}$	10	$0.008\overline{8}$
$[90; 100[$	12	$0.1\overline{3}$	10	$0.01\overline{3}$
$[100; 110[$	17	$0.1\overline{8}$	10	$0.01\overline{8}$
$[110; 120[$	15	$0.1\overline{6}$	10	$0.01\overline{6}$
$[120; 130[$	13	$0.1\overline{4}$	10	$0.01\overline{4}$
$[130; 140[$	7	$0.07\overline{7}$	10	$0.007\overline{7}$
$[140; 150[$	6	$0.06\overline{6}$	10	$0.006\overline{6}$
	90	1		

Histogramm – Anzahl/Breite der Kategorien

Anzahl und Breite der Kategorien sind nicht eindeutig festgelegt, es gibt nur einige Faustregeln. Je nach Anzahl der Kategorien sieht das Histogramm optisch anders aus:

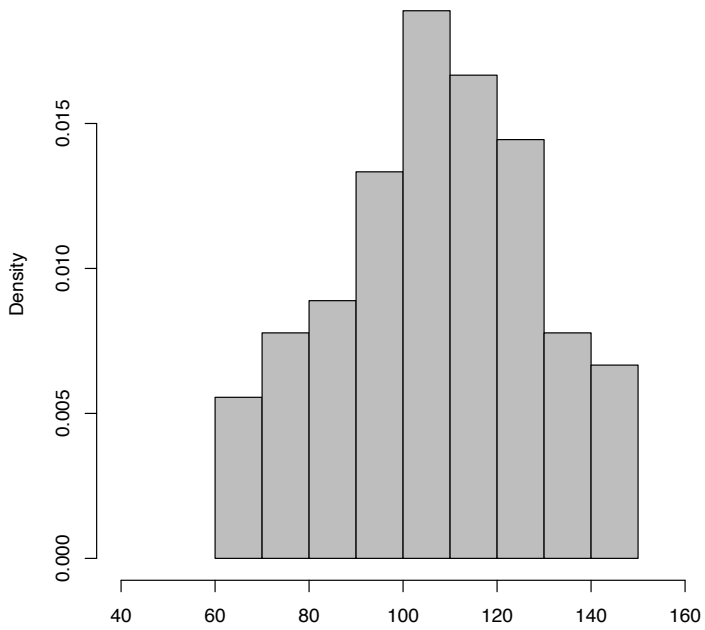


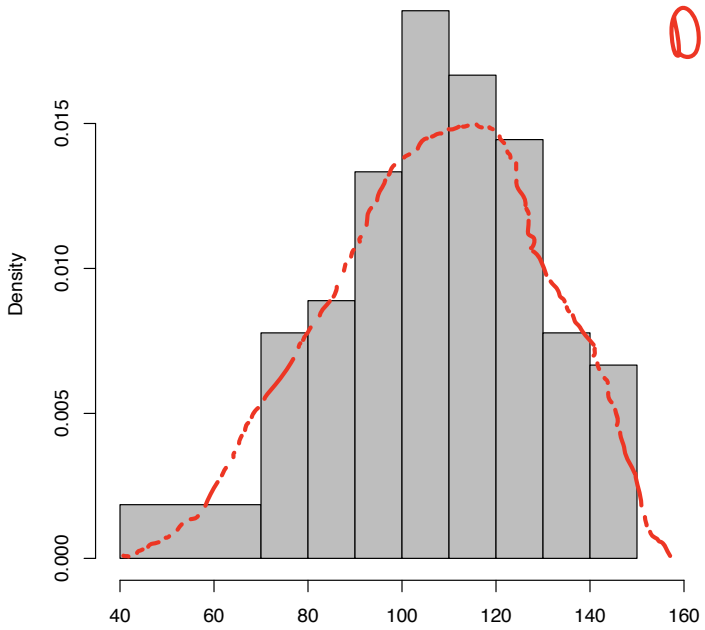


Zu viele Kategorien $\rightarrow$ sehr wackelig, zu wenige $\rightarrow$ sehr grob.
Empfehlung: Anzahl der Kategorien so wählen, daß die Form der Verteilung noch erkennbar ist (bei SPSS, R etc. automatisch).

Histogramm – Flächentreue

Beim Histogramm handelt es sich um einen Schätzer für die sog. Dichte (vgl. spätere Kapitel). Dabei ist die Fläche proportional zur Häufigkeit einer Kategorie (Flächentreue).
→ Wenn man z.B. eine Kategorie breiter macht, ohne dass dadurch mehr Beobachtungen hinzukommen, wird der zugehörige Balken flacher.





Dichte

Beispiel

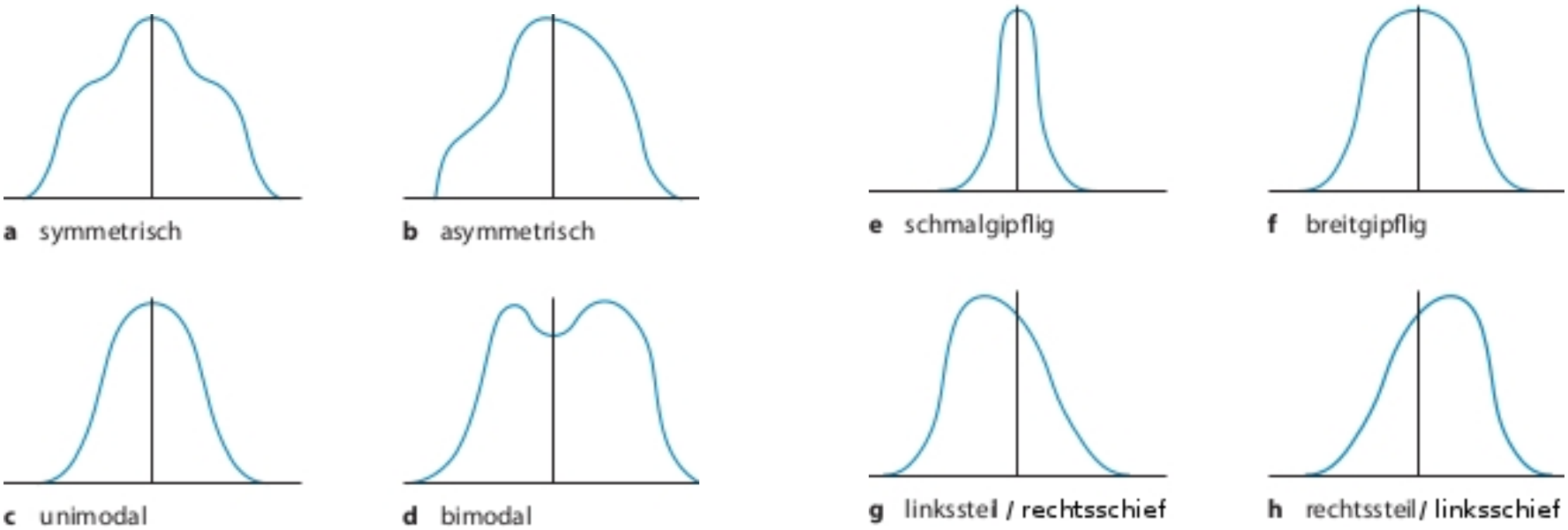
Kategorie	Anzahl	Anteil	Breite	Höhe = Anteil/Breite
$[60; 70[$	5	$0.05\overline{5}$	10	$0.005\overline{5}$
$[70; 80[$	7	$0.07\overline{7}$	10	$0.007\overline{7}$
$[80; 90[$	8	$0.08\overline{8}$	10	$0.008\overline{8}$
$[90; 100[$	12	$0.1\overline{3}$	10	$0.01\overline{3}$
$[100; 110[$	17	$0.1\overline{8}$	10	$0.01\overline{8}$
$[110; 120[$	15	$0.1\overline{6}$	10	$0.01\overline{6}$
$[120; 130[$	13	$0.1\overline{4}$	10	$0.01\overline{4}$
$[130; 140[$	7	$0.07\overline{7}$	10	$0.007\overline{7}$
$[140; 150[$	6	$0.06\overline{6}$	10	$0.006\overline{6}$
	90	1		

Beispiel

Kategorie	Anzahl	Anteil	Breite	Höhe = Anteil/Breite
[40; 70[	5	$0.05\overline{5}$	30	0.00185
[70; 80[	7	$0.07\overline{7}$	10	$0.007\overline{7}$
[80; 90[	8	$0.08\overline{8}$	10	$0.008\overline{8}$
[90; 100[	12	$0.1\overline{3}$	10	$0.01\overline{3}$
[100; 110[	17	$0.1\overline{8}$	10	$0.01\overline{8}$
[110; 120[	15	$0.1\overline{6}$	10	$0.01\overline{6}$
[120; 130[	13	$0.1\overline{4}$	10	$0.01\overline{4}$
[130; 140[	7	$0.07\overline{7}$	10	$0.007\overline{7}$
[140; 150[	6	$0.06\overline{6}$	10	$0.006\overline{6}$
	90	1		

- Wegen Flächentreue: Histogramme mit unterschiedlich großen Kategorien sind sehr schwer zu interpretieren $\rightarrow$ besser gleich große Kategorien, dann ist die Höhe wieder proportional zur Häufigkeit, wie beim Balkendiagramm.
- Anzahl der Kategorien wird so gewählt, daß die Form der Verteilung gut erkennbar ist.

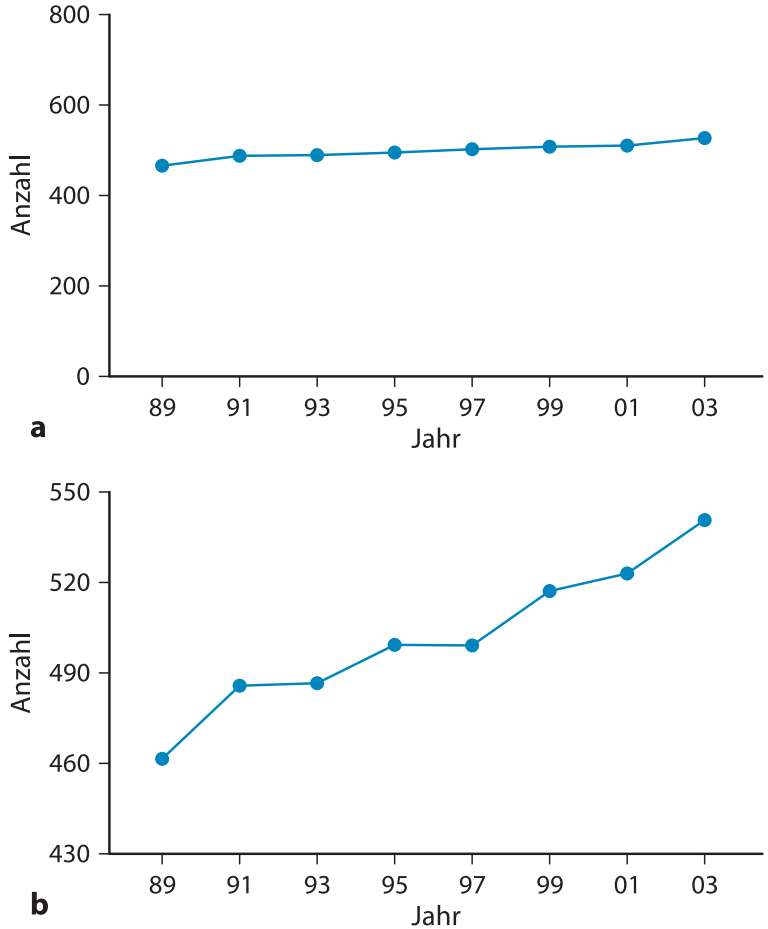
Verteilungsformen



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Suggestive Grafiken

z.B. durch Skalierung und Abschneiden der Achsen den Eindruck erwecken, es gäbe einen steilen Anstieg



Buchtipp: Walter Krämer: *So lügt man mit Statistik*

Webtipp: `www.stat4u.at` → Lehrmaterial → Fehlerquellen

REVITALIFT
LASER X3
1 MILDERT FALTEN
2 VERLEIHT DICHTE
3 REMODELLIERUNGS-EFFEKT
REVITALIFT
LASER X3
NACH 8 WOCHEN GEMILDERTE KRÄHEN-FÜSSE! VERBESSERUNG DES NIVEAUS IM BEREICH:
MICRO-RELIEF*
FEINE LINIEN
0%
-5%
-10%
-15%
-20%
-20%
-18%
-17%
-14%
NACH EINER CO2 LASER-SESSION
REVITALIFT LASER X3 CREME, WISSENSCHAFTLICHER TEST, 50 FRAUEN
DIE WISSENSCHAFT:
FRAGMENTIERTE HYALURONSÄURE

Suggestive Prozentangaben

Suggestive Prozentangaben

Sehr geehrte Frau Strobl,
vielen Dank für Ihre E-Mail und Ihre Anfrage bezüglich unseres OPTISCHEN 5 SEKUNDEN REDUZIERERS.
Der von Ihnen angefragte Wert - 38 % (verringerte Sichtbarkeit von Falten und Poren) ist der Prozentsatz der Selbstbewertung von 51 Frauen unseres Produktes.
Wir hoffen das wir Ihnen weiterhelfen konnten, und senden viele Grüße aus Düsseldorf,
L'Oréal Deutschland GmbH

$19/51 = 0.3725 \quad 20/51 = 0.3921 \quad 19.5/51 = 0.3824$

d.h. auch Männer sind teilbar...

Wahrscheinlichkeitstheorie

Wo sind wir?

1. Deskriptive Statistik

2. Wahrscheinlichkeitstheorie
2.1 Begriffe und Rechenregeln
2.2 Wahrscheinlichkeitsverteilungen
2.3 Stichprobenverteilung des Mittelwerts

3. Tests und Konfidenzintervalle

4. Kovarianz und Korrelation

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Frage: Wer oder was kann eine Wahrscheinlichkeit haben?
- Antwort: Ein „Ereignis". Ereignisse werden üblicherweise mit Großbuchstaben, $A, B, C, \ldots$, bezeichnet.
- Die Wahrscheinlichkeit, dass das Ereignis A eintritt, bezeichnet man als $\mathrm{P}(A)$.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Terminologie

- Elementarereignis: Ergebnis eines Zufallsexperimentes.
- Ereignismenge Ω: Menge der Elementarereignisse.
- Ereignis: Teilmenge von Ω.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiele für Elementarereignisse

- Münze: Kopf, Zahl.
- Würfel: Sechs Seiten des Würfels.
- Reißnagelspitze zeigt nach oben: ja, nein

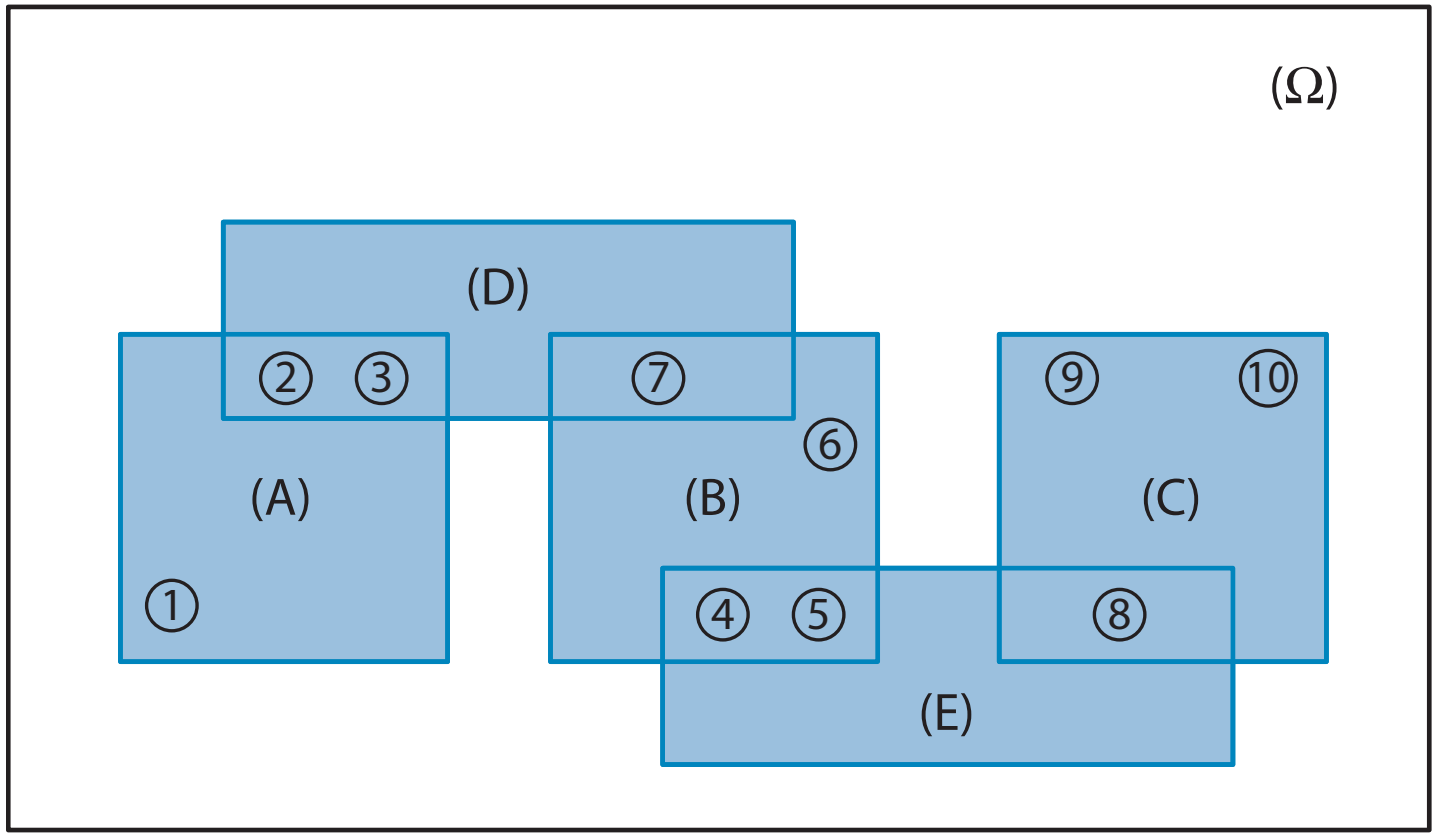
©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Ereignisse

- Sei $\Omega = \{1, 2, 3, \ldots, 10\}$
- Einige mögliche Ereignisse:
  - $A = \{1, 2, 3\}$
  - $B = \{4, 5, 6, 7\}$
  - $C = \{8, 9, 10\}$
  - $D = \{2, 3, 7\}$
  - $E = \{4, 5, 8\}$
- Bezeichnungen:
  - Vereinigung: $A \cup D = \{1, 2, 3, 7\}$
  - Schnitt: $A \cap D = \{2, 3\}$
  - Komplement: $\overline{A} = \{4, 5, 6, 7, 8, 9, 10\}$
  - Leere Menge: $A \cap B = \emptyset$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Veranschaulichung von Ereignissen durch Venn-Diagramme

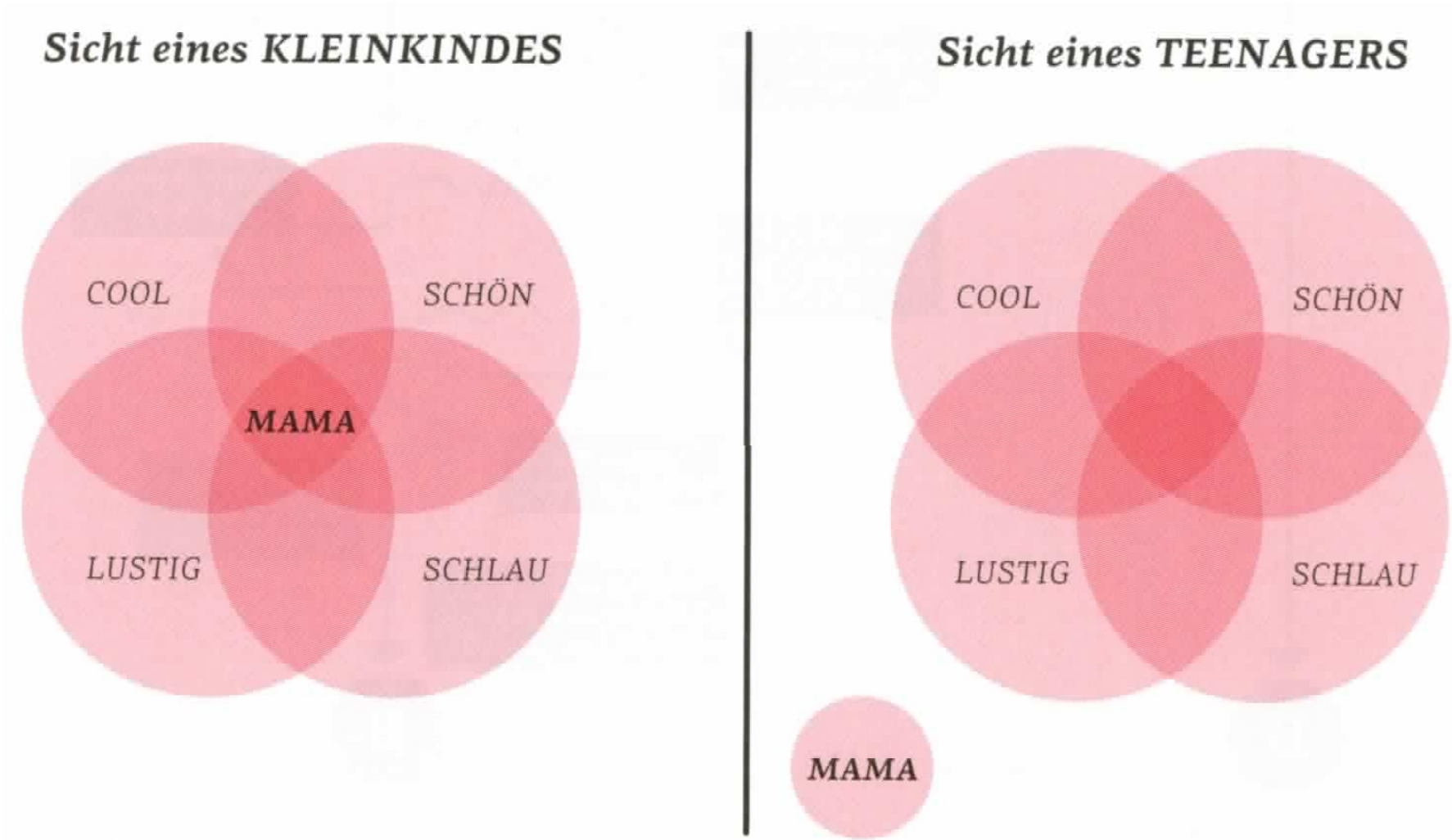


©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Zur Unterhaltung: Schnittmenge

Quelle: Katja Berlin & Peter Grünlich: *Was wir tun, wenn der Aufzug nicht kommt.*

Zur Unterhaltung: Schnittmenge



Quelle: Brigitte MOM, 1/2013

Additionstheorem allgemein

- Die Wahrscheinlichkeit des Eintretens von A oder B (oder beiden gemeinsam) ergibt sich als die Summe der Wahrscheinlichkeiten von A und B abzüglich der Wahrscheinlichkeit, dass beide Ereignisse gemeinsam eintreten.
- Warum muß man die gemeinsame Wahrscheinlichkeit einmal abziehen? $\rightarrow$ Wird im Venn-Diagramm klar!
- Als Formel ergibt sich:

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

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Additionstheorem für disjunkte Ereignisse

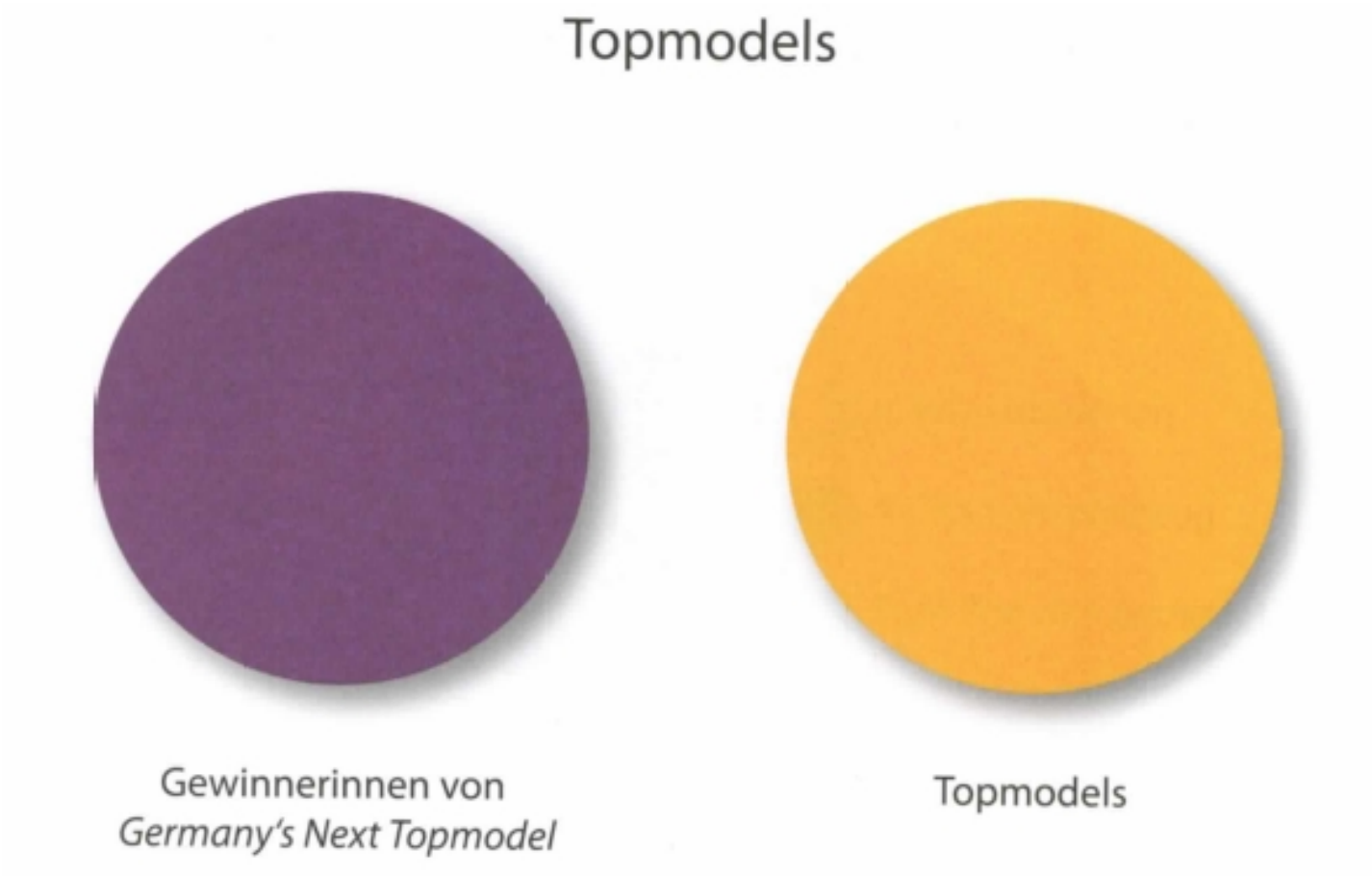
- Ein wichtiger Spezialfall des Additionstheorems ergibt sich, falls sich beide Ereignisse wechselseitig ausschließen.
- In diesem Fall:

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$$

  da in diesem Fall: $\mathrm{P}(A \cap B) = 0$
- Sich wechselseitig ausschließende Ereignisse werden auch „disjunkt" genannt.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Zur Unterhaltung: disjunkte Ereignisse



Quelle: Katja Berlin & Peter Grünlich: *Was wir tun, wenn der Aufzug nicht kommt.*

Mehr als zwei disjunkte Ereignisse

- Die Wahrscheinlichkeit, dass eines von k disjunkten Ereignissen eintritt, entspricht der Summe der Wahrscheinlichkeiten für die k Ereignisse.
- Als Formel:

$$\mathrm{P}(A_1 \cup A_2 \cup \cdots \cup A_k) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots + \mathrm{P}(A_k)$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Wahrscheinlichkeit des Komplements

- Das zu A komplementäre Ereignis ist „Nicht A" und wird mit $\overline{A}$ bezeichnet.
- Die Wahrscheinlichkeit von $\overline{A}$ kann mit Hilfe von $\mathrm{P}(A)$ bestimmt werden:

$$\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Multiplikationstheorem für unabhängige Ereignisse

- Ist das Eintreten zweier Ereignisse voneinander unabhängig ergibt sich die Wahrscheinlichkeit des gemeinsamen Auftretens von A und B als das Produkt der beiden Einzelwahrscheinlichkeiten:
- $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B)$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Multiplikationstheorem allgemein

- Die Wahrscheinlichkeit des gemeinsamen Auftretens von A und B ist gleich der Wahrscheinlichkeit für das Auftreten von A, multipliziert mit der bedingten Wahrscheinlichkeit des Auftretens von B, wenn A schon aufgetreten ist.
- Als Formel: $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B|A)$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Die Verallgemeinerung lautet:

$$\mathrm{P}(A \cap B \cap C \cap \ldots \cap Z) = \mathrm{P}(A) \cdot \mathrm{P}(B|A) \cdot \mathrm{P}(C|AB) \cdot \ldots \cdot \mathrm{P}(Z|AB\ldots)$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Bedingte Wahrscheinlichkeit

- Häufig tritt die Situation auf, dass wir uns für die Wahrscheinlichkeit des Ereignisses B interessieren, wobei wir aber bereits wissen, dass ein anderes Ereignis A eingetreten ist.
- Nutzen wir diese Kenntnis bei der Bestimmung der Wahrscheinlichkeit von B aus, so sprechen wir von der „bedingten Wahrscheinlichkeit", welche mit $\mathrm{P}(B|A)$ bezeichnet wird.
- $\mathrm{P}(B|A)$ bezeichnet also „die Wahrscheinlichkeit für B, falls A schon eingetreten ist".
- Wenn die Wahrscheinlichkeiten $\mathrm{P}(A \cap B)$ und $\mathrm{P}(A)$ bekannt sind, kann die bedingte Wahrscheinlichkeit $\mathrm{P}(B|A)$ mit folgender Formel berechnet werden:

$$\mathrm{P}(B|A) = \mathrm{P}(A \cap B)/\mathrm{P}(A),$$

wobei $\mathrm{P}(A) > 0$ gelten muss.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Unabhängige Ereignisse

- Zwei Ereignisse sind unabhängig, falls das Eintreten des einen Ereignisses keinerlei Effekt auf die Wahrscheinlichkeit für das Eintreten des anderen Ereignisses hat.
- Die Ereignisse A und B sind unabhängig, falls $\mathrm{P}(B|A) = \mathrm{P}(B|\overline{A}) = \mathrm{P}(B)$ und umgekehrt.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Frei erfundenes Beispiel: Studierende

Die besprochenen Rechenregeln für Wahrscheinlichkeiten kann man sich i.W. analog an Zufallsvariablen veranschaulichen:

Stellen Sie sich vor, aus einem Hörsaal mit 100 Studierenden ziehe ich zufällig eine Person (mit verbundenen Augen).

Zur Illustration der Rechenregeln für Wahrscheinlichkeiten interessieren uns folgende Ereignisse:

- Frau: ich erwische eine Frau
- lang: ich erwische eine Person mit langen Haaren
- Shirt: ich erwische eine Person mit schwarzem T-Shirt
- Gel: ich erwische eine Person mit Haargel

Frei erfundenes Beispiel: Studierende

Zum Rechnen brauchen wir einige Wahrscheinlichkeiten, die sich aus der Verteilung in der Grundgesamtheit ergeben:

		Frau: ja	Frau: nein	
lang	ja	40	5	45
	nein	10	45	55
		50	50	100

$$P(\text{Frau}) = 50/100 = 0.5$$
$$P(\text{lang}) = 45/100 = 0.45$$
$$P(\text{Frau} \cap \text{lang}) = 40/100 = 0.4$$
$$P(\text{lang}|\text{Frau}) = 40/50 = 0.8$$
$$P(\text{lang}|\overline{\text{Frau}}) = 5/50 = 0.1$$

		Haargel: ja	Haargel: nein	
lang	ja	0	45	45
	nein	15	40	55
		15	85	100

$$P(\text{Gel}) = 15/100 = 0.15$$

		Frau: ja	Frau: nein	
Shirt	ja	10	10	20
	nein	40	40	80
		50	50	100

$$P(\text{Shirt}|\text{Frau}) = 10/50 = 0.2$$
$$P(\text{Shirt}|\overline{\text{Frau}}) = 10/50 = 0.2$$
$$P(\text{Shirt}) = 20/100 = 0.2$$

- direkt aus Tabelle ablesen wenn angegeben:

  $\mathrm{P}(\text{lang}|\text{Frau}) = 40/50 = 0.8$

  von den 50 Frauen haben 40 lange Haare

- mit der Formel:

  $\mathrm{P}(\text{lang}|\text{Frau}) = \mathrm{P}(\text{Frau} \cap \text{lang})/\mathrm{P}(\text{Frau}) =$

  $$\frac{40/100}{50/100} = \frac{0.4}{0.5} = 0.8$$

  von allen 100 Personen sind 40 Frauen <u>und</u> haben lange Haare
  von allen 100 Personen sind 50 Frauen

  (aus Tabelle bzw. aus angegebenen Wahrscheinlichkeiten)

Wahrscheinlichkeit des Komplements

hier jeweils zwei sich ausschließende Möglichkeiten:
Frau vs. $\overline{\text{Frau}}$, lang vs. $\overline{\text{lang}}$

- bei „normalen“ Wahrscheinlichkeiten:

  $\mathrm{P}(\overline{\text{Frau}}) = 1 - \mathrm{P}(\text{Frau}) = 1 - 0.5 = 0.5$

- bei bedingten Wahrscheinlichkeiten (gleiche Bedingung!):

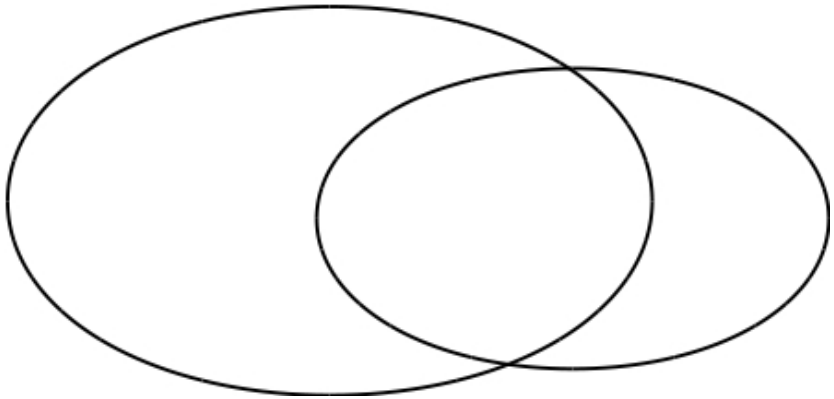
  $\mathrm{P}(\overline{\text{lang}}|\text{Frau}) = 1 - \mathrm{P}(\text{lang}|\text{Frau}) = 1 - 0.8 = 0.2$

(lassen sich auch direkt aus Tabelle ablesen wenn angegeben)

Additionstheorem allgemein

Wie hoch ist die Wahrscheinlichkeit, dass die Person, die ich ziehe, eine Frau ist oder lange Haare hat (oder beides)?

$$
\begin{aligned}
\mathrm{P}(\text{Frau} \cup \text{lang}) &= \mathrm{P}(\text{Frau}) + \mathrm{P}(\text{lang}) - \mathrm{P}(\text{Frau} \cap \text{lang}) \\
&= 0.5 + 0.45 - 0.4 \\
&= 0.55
\end{aligned}
$$



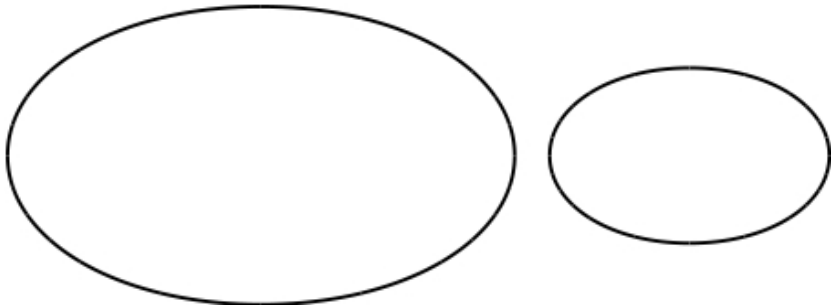
		Frau: ja	Frau: nein	
lang	ja	40	5	45
	nein	10	45	55
		50	50	100

Additionstheorem für disjunkte Ereignisse

Wie hoch ist die Wahrscheinlichkeit, dass die Person, die ich ziehe, lange Haare hat oder Haargel im Wet-Look trägt (oder beides)?

Es gibt keine Personen, die sowohl lange Haare als auch Haargel im Wet-Look tragen, d.h. $\text{lang} \cap \text{Gel} = \emptyset$ und $\mathrm{P}(\text{lang} \cap \text{Gel}) = 0$.

$$
\begin{aligned}
\mathrm{P}(\text{lang} \cup \text{Gel}) &= \mathrm{P}(\text{lang}) + \mathrm{P}(\text{Gel}) - \mathrm{P}(\text{lang} \cap \text{Gel}) \\
&= 0.45 + 0.15 - 0 \\
&= 0.60
\end{aligned}
$$



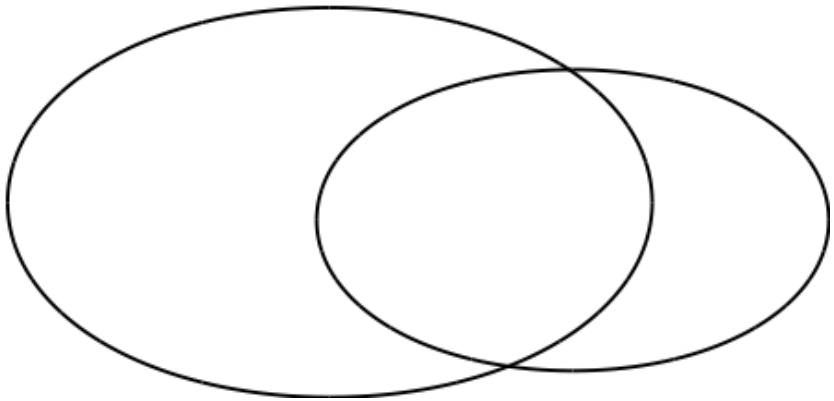
		Haargel		
		ja	nein	
lang	ja	0	45	45
	nein	15	40	55
		15	85	100

Multiplikationstheorem für unabhängige Ereignisse

Wie hoch ist die Wahrscheinlichkeit, dass die Person, die ich ziehe, eine Frau ist und ein schwarzes T-Shirt trägt?

Das Tragen von schwarzen T-Shirts ist unabhängig vom Geschlecht: $\mathrm{P}(\text{Shirt}|\text{Frau}) = \mathrm{P}(\text{Shirt}|\overline{\text{Frau}}) = \mathrm{P}(\text{Shirt}) = 0.2$

$$\begin{aligned} \mathrm{P}(\text{Frau} \cap \text{Shirt}) &= \mathrm{P}(\text{Frau}) \cdot \mathrm{P}(\text{Shirt}) \\ &= 0.5 \cdot 0.2 \\ &= 0.1 \end{aligned}$$



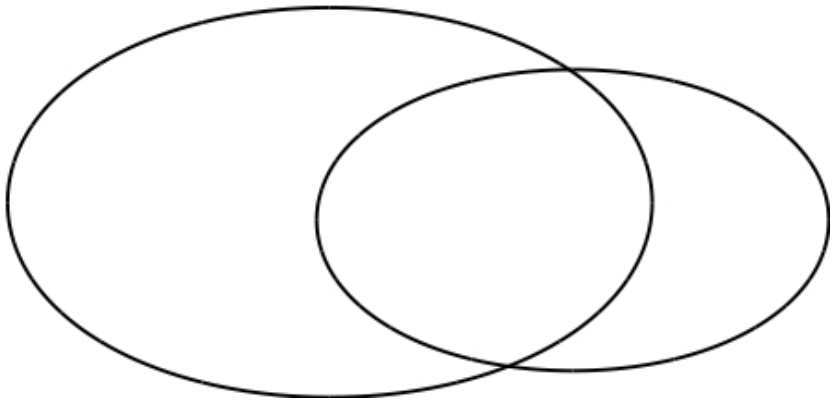
		Frau: ja	Frau: nein	
Shirt	ja	10	10	20
	nein	40	40	80
		50	50	100

Multiplikationstheorem allgemein

Wie hoch ist die Wahrscheinlichkeit, dass die Person, die ich ziehe, eine Frau ist und lange Haare hat?

Das Tragen von langen Haaren ist nicht unabhängig vom Geschlecht: $\mathrm{P}(\text{lang}|\text{Frau}) = 0.8 \neq \mathrm{P}(\text{lang}|\overline{\text{Frau}}) = 0.1$.

$$\begin{aligned} \mathrm{P}(\text{Frau} \cap \text{lang}) &= \mathrm{P}(\text{Frau}) \cdot \mathrm{P}(\text{lang}|\text{Frau}) \\ &= 0.5 \cdot 0.8 \\ &= 0.4 \end{aligned}$$

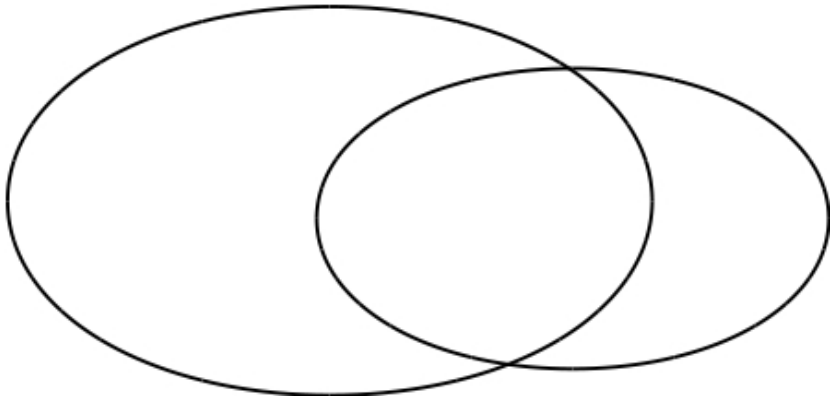


		Frau		
		ja	nein	
lang	ja	40	5	45
	nein	10	45	55
		50	50	100

Disjunktheit vs. Unabhängigkeit

Disjunktheit
Ist die Schnittmenge aus Frau und lang leer?
nein $\Rightarrow$ nicht disjunkt

Unabhängigkeit
Ist $\mathrm{P}(\text{lang}|\text{Frau})$ gleich $\mathrm{P}(\text{lang}|\overline{\text{Frau}})$?
nein $\Rightarrow$ nicht unabhängig



		Frau		
		ja	nein	
lang	ja	40	5	45
	nein	10	45	55
		50	50	100

Multiplikationstheorem aus psychologischer Sicht

Was ist größer: Die Wahrscheinlichkeit...

- ...dass die Person, die ich ziehe, eine Frau ist.
- ...dass die Person, die ich ziehe, eine Frau ist <u>und</u> lange Haare hat.

Wir haben gesehen: Frauen mit langen Haaren sind nur eine Teilmenge aller Frauen $\Rightarrow$ Wahrscheinlichkeit für verbundenes Ereignis ist kleiner (Multiplizieren zweier Wahrscheinlichkeiten ergibt geringere Wahrscheinlichkeit).

Multiplikationstheorem aus psychologischer Sicht

Berühmtes Experiment von Amos Tversky und Daniel Kahneman:

Versuchspersonen lasen einen Text über eine junge, intelligente Frau namens Linda, die Philosophie studiert hat, für soziale Gerechtigkeit eintritt und zu anti-Atomkraft-Demos geht. Danach wurden Sie gefragt:

Was ist größer: Die Wahrscheinlichkeit...

- ...dass Linda eine Bankangestellte ist.
- ...dass Linda eine Bankangestellte und in der Frauenrechtsbewegung aktiv ist.

Die meisten Versuchspersonen ließen sich aufgrund der sog. Repräsentativitäts-Heuristik verleiten, die zweite Antwort anzukreuzen. (Das Experiment wurde später aber oft kritisiert.)

Originalartikel: Tversky, A., & Kahneman, D. (1983). *Psychological Review*, *90*, 293-315.
vieldiskutiertes Buch: Daniel Kahneman: *Thinking, Fast and Slow.*

- Satz von Bayes:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A) \cdot \mathrm{P}(A)}{\mathrm{P}(B)}$$

- Der Nenner ist (Satz von der totalen Wahrscheinlichkeit):

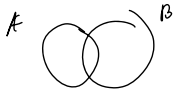
$$\mathrm{P}(B) = \mathrm{P}(B|A) \cdot \mathrm{P}(A) + \mathrm{P}(B|\overline{A}) \cdot \mathrm{P}(\overline{A})$$

- Daraus ergibt sich:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A) \cdot \mathrm{P}(A)}{\mathrm{P}(B|A) \cdot \mathrm{P}(A) + \mathrm{P}(B|\overline{A}) \cdot \mathrm{P}(\overline{A})}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Motivation



- Notation:
  - K Krankheit vorhanden
  - T Test positiv
- Terminologie:
  - $\mathrm{P}(K)$ Prävalenz
  - $\mathrm{P}(T|K)$ Sensitivität des Tests → Krank Positiv
  - $\mathrm{P}(\overline{T}|\overline{K})$ Spezifität des Tests → Gesund Negativ

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

HIV-Schnell-Tests bewilligt

QUICKIE ⇢ Die US-Gesundheitsbehörde FDA hat einen schnellen HIV-Test für zu Hause freigegeben.

Es ist in den USA der erste HIV-Test, der nach gut einer halben Stunde dem Patienten selbständig ein Ergebnis liefert: «Oraquick» soll in den USA ab Oktober für 60 Dollar frei in Apotheken verkauft werden. **Auch Schweizer könnten ihn theoretisch übers Internet bestellen.** Der Test wird in Amerika seit 2002 von Ärzten benutzt und liefert zu 99 Prozent das richtige Ergebnis.

Der Test funktioniert ganz einfach: Man nimmt das mitgelieferte Röhrchen in den Mund und befeuchtet es mit Speichel. **Die Wartezeit für das Testergebnis beträgt 20 bis 40 Minuten.** Ähnlich wie bei einem Schwangerschaftstest wir es auf dem Streifen sichtbar.

Mit dem Verfahren sollen Leute erreicht werden, die sich ansonsten vielleicht nicht testen lassen würden, schreibt die US-Gesundheitsbehörde FDA.

Sie macht in ihrer Erklärung jedoch deutlich, dass der Test nicht zu 100 Prozent sicher ist. Auch der Hersteller empfiehlt in jedem Fall noch einen regulären Test beim Arzt zu machen.

Die US-Regierung geht von 1,2 Millionen HIV-Infizierten in den USA aus. Etwa ein Fünftel der Betroffenen wissen nicht, dass sie das Virus in sich tragen. Jedes Jahr kämen etwa 50 000 HIV-Positive hinzu. **sax**

Instant-Test «Oraquick»-Anwenderin.

Quelle: *Blick am Abend*, 2012

- Angenommen bekannt sind:
  - $\mathrm{P}(K)$ Prävalenz
  - $\mathrm{P}(T|K)$ Sensitivität des Tests
  - $\mathrm{P}(\overline{T}|\overline{K})$ Spezifität des Tests
- Dann lässt sich mit dem Satz von Bayes $\mathrm{P}(K|T)$ und $\mathrm{P}(\overline{K}|\overline{T})$ ermitteln.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Satz von Bayes:

$$\mathrm{P}(K|T) = \frac{\mathrm{P}(T|K) \cdot \mathrm{P}(K)}{\mathrm{P}(T|K) \cdot \mathrm{P}(K) + \mathrm{P}(T|\overline{K}) \cdot \mathrm{P}(\overline{K})}$$

- Aus der Prävalenz folgt: $\mathrm{P}(\overline{K}) = 1 - \mathrm{P}(K)$
- Aus der Spezifität folgt: $\mathrm{P}(T|\overline{K}) = 1 - \mathrm{P}(\overline{T}|\overline{K})$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Kennwerte eines diagnostischen Tests:
  - Prävalenz: 0,001
  - Sensitivität: 0,98
  - Spezifität: 0,995
- Mit Hilfe des Satz von Bayes folgt:

$$\mathrm{P}(K|T) = \frac{0{,}001 \cdot 0{,}98}{0{,}001 \cdot 0{,}98 + 0{,}005 \cdot 0{,}999} = 0{,}164$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Wahrscheinlichkeitsverteilungen

Wo sind wir?

1. Deskriptive Statistik

2. Wahrscheinlichkeitstheorie
2.1 Begriffe und Rechenregeln
2.2 Wahrscheinlichkeitsverteilungen
2.3 Stichprobenverteilung des Mittelwerts

3. Tests und Konfidenzintervalle

4. Kovarianz und Korrelation

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Zufallsvariablen

- Eine Zufallsvariable ist eine Funktion, die Ereignissen reelle Zahlen zuordnet.
- Beispiele:
  - Würfel: Jeder Seite wird ihre Punkteanzahl zugeordnet.
  - Münze: Kopf wird als 1, Zahl als 0 kodiert.
  - Studierende: Ziehe eine Person und notiere deren Alter in Jahren.
- Zufallsvariablen können diskret oder stetig sein.
  - Diskret: Ergebnisse fallen in Kategorien.
  - Stetig: Werte können beliebig genau sein.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Zufallsvariablen besitzen eine Verteilung

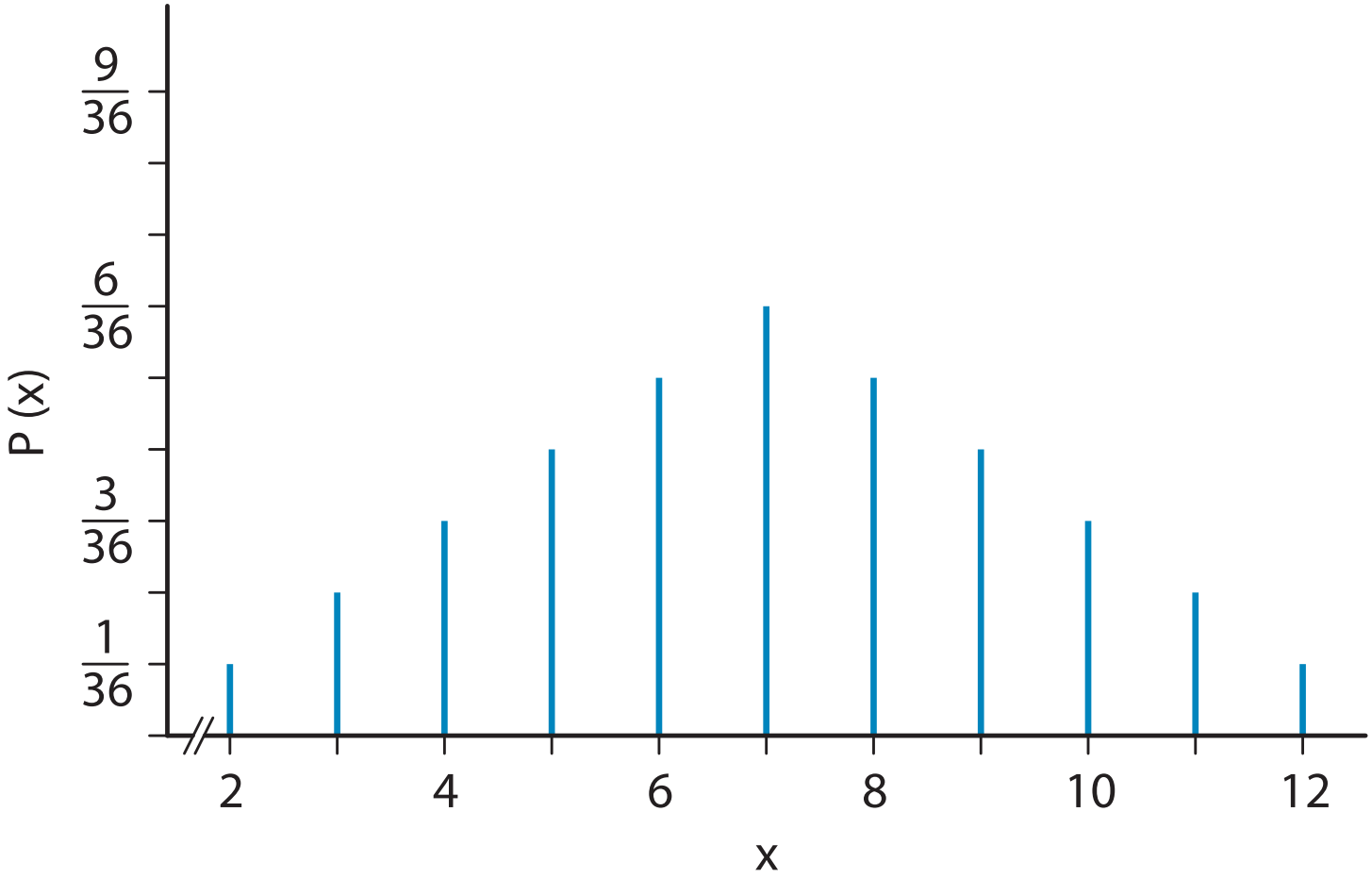
- Hier besprochene Beispiele für
  - diskrete Zufallsvariable: Binomialverteilung
  - stetige Zufallsvariable: Normalverteilung
- Es existieren noch zahlreiche andere Verteilungen.
- Wahrscheinlichkeitsverteilungen sind idealisierte statistische Modelle, die die echten Daten mehr oder weniger realistisch beschreiben.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Diskrete Zufallsvariable

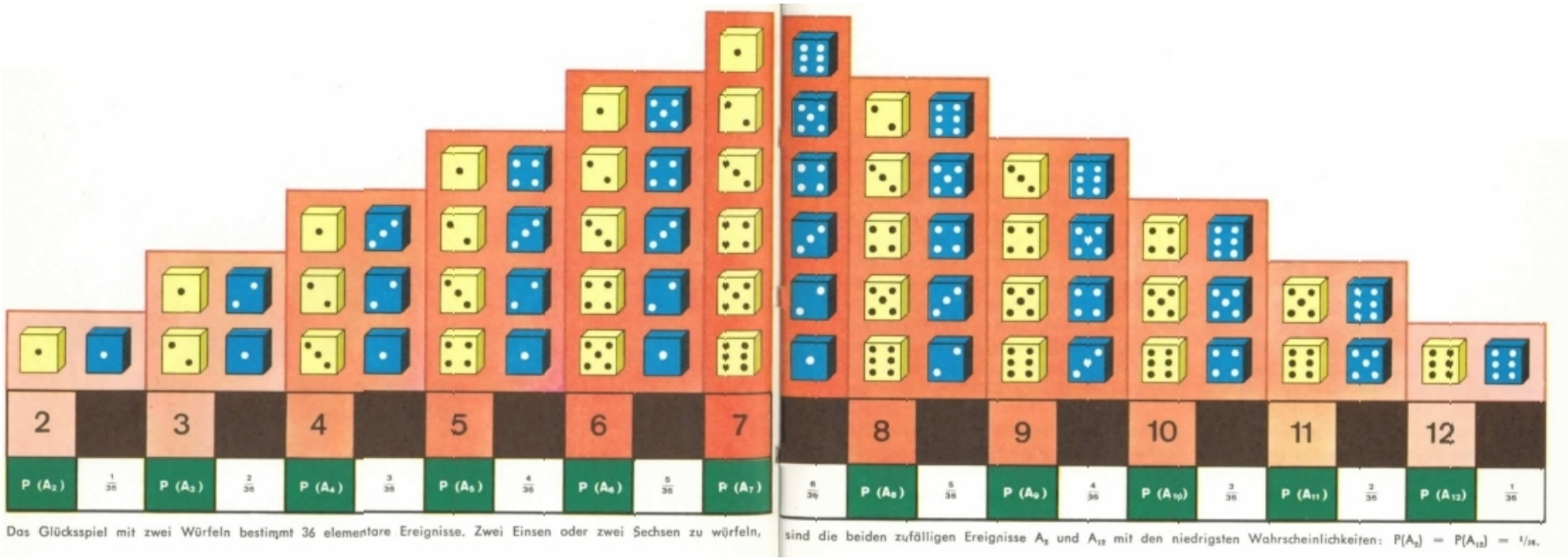
- Nur bestimmte Werte können auftreten.
- Jeder Wert hat eine gewisse Wahrscheinlichkeit.
- Die Wahrscheinlichkeiten addieren sich zu 1 auf.

Diskrete Zufallsvariable



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Zweimaliges Würfeln



Quelle: Walter Fuchs: *Knaurs Buch der modernen Mathematik.*

Erwartungswert, Varianz, Verteilungsfunktion

- Erwartungswert

$$\mu = \sum_{i=1}^{N} x_i \cdot \mathrm{P}(x_i)$$

- Varianz

$$\sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2 \cdot \mathrm{P}(x_i)$$

- Wahrscheinlichkeitsfunktion

$$\mathrm{P}(x_i)$$

- Verteilungsfunktion

$$F(x_i) = \sum_{j \leq i} \mathrm{P}(x_j)$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Wir betrachten eine Zufallsvariable, die nur zwei Werte annehmen kann.
- Solche Variablen bezeichnet man auch als binäre oder dichotome Variablen.
- Beispiele für dichotome Variablen sind:
  - Kopf versus Zahl beim Münzwurf
  - Hat eine Person ein Haustier? (ja/nein)

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Notation

- x Anzahl der Erfolge in n Versuchen.
- π Erfolgswahrscheinlichkeit pro Versuch.
- $\mathrm{P}(x)$ Wahrscheinlichkeit von x Erfolgen in n Versuchen.
- Kurzbezeichnung für Binomialverteilung: $B(n; \pi)$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Münzwurf

- Disjunkte Ereignisse: Kopf, Zahl
- $n = 1$
- $\pi = 0{,}5$
- Kopf zählt als Erfolg
- x = Anzahl von Würfen bei denen Kopf oben lag

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Münze einmal werfen

- Mögliche Werte für x: 0, 1
- Wahrscheinlichkeiten für x:
  - $\mathrm{P}(0) = 0{,}5$
  - $\mathrm{P}(1) = 0{,}5$
- $= B(1; 0{,}5)$ Verteilung

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Münze zweimal werfen

Fakts

- Mögliche Werte für x: 0, 1, 2
- Wahrscheinlichkeiten für x:
  - $\mathrm{P}(0) = 0{,}25$ (Zahl,Zahl)
  - $\mathrm{P}(1) = 0{,}5$ (Kopf,Zahl oder Zahl,Kopf)
  - $\mathrm{P}(2) = 0{,}25$ (Kopf,Kopf)
- $= B(2; 0{,}5)$ Verteilung

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Münze dreimal werfen

- Mögliche Werte für x: 0, 1, 2, 3
- Wahrscheinlichkeiten für x:
  - $\mathrm{P}(0) = 0{,}125$ (Zahl,Zahl,Zahl)
  - $\mathrm{P}(1) = 0{,}375$ (Kopf,Zahl,Zahl oder Zahl,Kopf,Zahl oder Zahl,Zahl,Kopf)
  - $\mathrm{P}(2) = 0{,}375$ (Kopf,Kopf,Zahl oder Kopf,Zahl,Kopf oder Zahl,Kopf,Kopf)
  - $\mathrm{P}(3) = 0{,}125$ (Kopf,Kopf,Kopf)
- $= B(3; 0{,}5)$ Verteilung

Im Prinzip ist es nicht schwierig, mit Hilfe des Multiplikations- und Additionstheorems beliebige Binomialverteilungen zu bestimmen, aber das Vorgehen ist nur praktikabel solange n klein ist.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Formel der Binomialverteilung

- Binomialverteilung: $\mathrm{P}(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$
- Binomial-Koeffizient (sprich: x aus n)

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

- $x!$ wird als „x Fakultät" bezeichnet.
- Die Fakultät einer positiven ganzen Zahl x ist definiert als

$$x! = 1 \cdot 2 \cdots (x-1) \cdot x$$

- Außerdem definiert man: $0! = 1$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Würfeln

- Wahrscheinlichkeit, genau einmal die Sechs zu werfen wenn man viermal würfelt?
- Man berechnet

$$\begin{aligned} \mathrm{P}(1) &= \binom{4}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^3 \\ &= 0{,}3858 \end{aligned}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Multiple-Choice Klausur

- Eine Multiple-Choice Klausur enthält 15 Fragen von denen jede fünf Antwortmöglichkeiten besitzt (d.h. die Ratewahrscheinlichkeit ist $\frac{1}{5} = 0{,}2$).
- Man kreuzt rein zufällig irgendwelche Antworten an.
- Die Klausur gilt als bestanden, falls neun oder mehr der Antworten richtig sind.
- Wie wahrscheinlich ist das Bestehen der Klausur?

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Multiple-Choice Klausur

- Parameter der Binomialverteilung: $n = 15$, $\pi = 0{,}2$
- Berechne

$$\mathrm{P}(9) + \mathrm{P}(10) + \mathrm{P}(11) + \mathrm{P}(12) + \mathrm{P}(13) + \mathrm{P}(14) + \mathrm{P}(15)$$

jeweils mithilfe der Binomialverteilungsformel

$$\mathrm{P}(9) = \binom{15}{9} \cdot (0{,}2)^9 \cdot (0{,}8)^{15-9} = 5005 \cdot (0{,}2)^9 \cdot (0{,}8)^6 \approx 0.000672$$

- Die Summe beträgt $\approx 0{,}000785$ (d.h. man sollte bei Multiple-Choice Klausuren also wirklich besser lernen, als zufällig irgendwelche Antworten anzukreuzen...)

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Erwartungswert und Varianz der Binominalverteilung

- Binomialverteilungen sind unimodale Verteilungen, die nur für $\pi = 1/2$ symmetrisch sind.
- Erwartungswert:

$$\mu = n \cdot \pi$$

- Varianz:

$$\sigma^2 = n \cdot \pi(1 - \pi)$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: 10x Münze werfen

- $n = 10$ und $\pi = 0{,}5$
- Erwartungswert: $\mu = 10 \cdot 0{,}5 = 5$ Erfolge
- Varianz: $\sigma^2 = 10 \cdot 0{,}5 \cdot 0{,}5 = 2{,}5$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

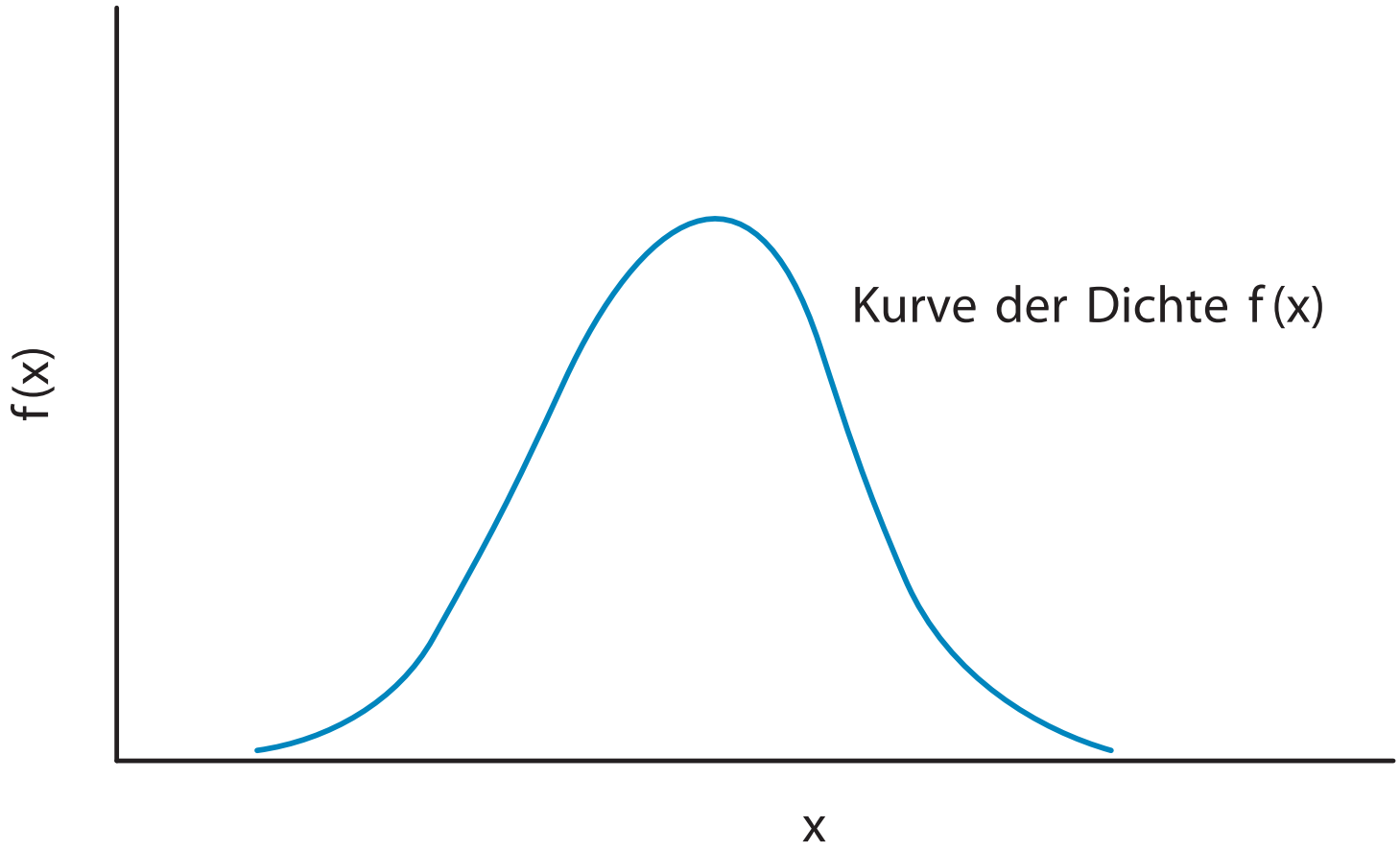
- Für stetige Zufallsvariablen wird die Wahrscheinlichkeitsverteilung durch die Dichtefunktion $f(x)$ beschrieben.
- Mit Hilfe der Dichte kann die Wahrscheinlichkeit von Intervallen bestimmt werden, also

$$\mathrm{P}(a < x \leq b) = \int_a^b f(x)\,\mathrm{d}x.$$

- Dabei gilt

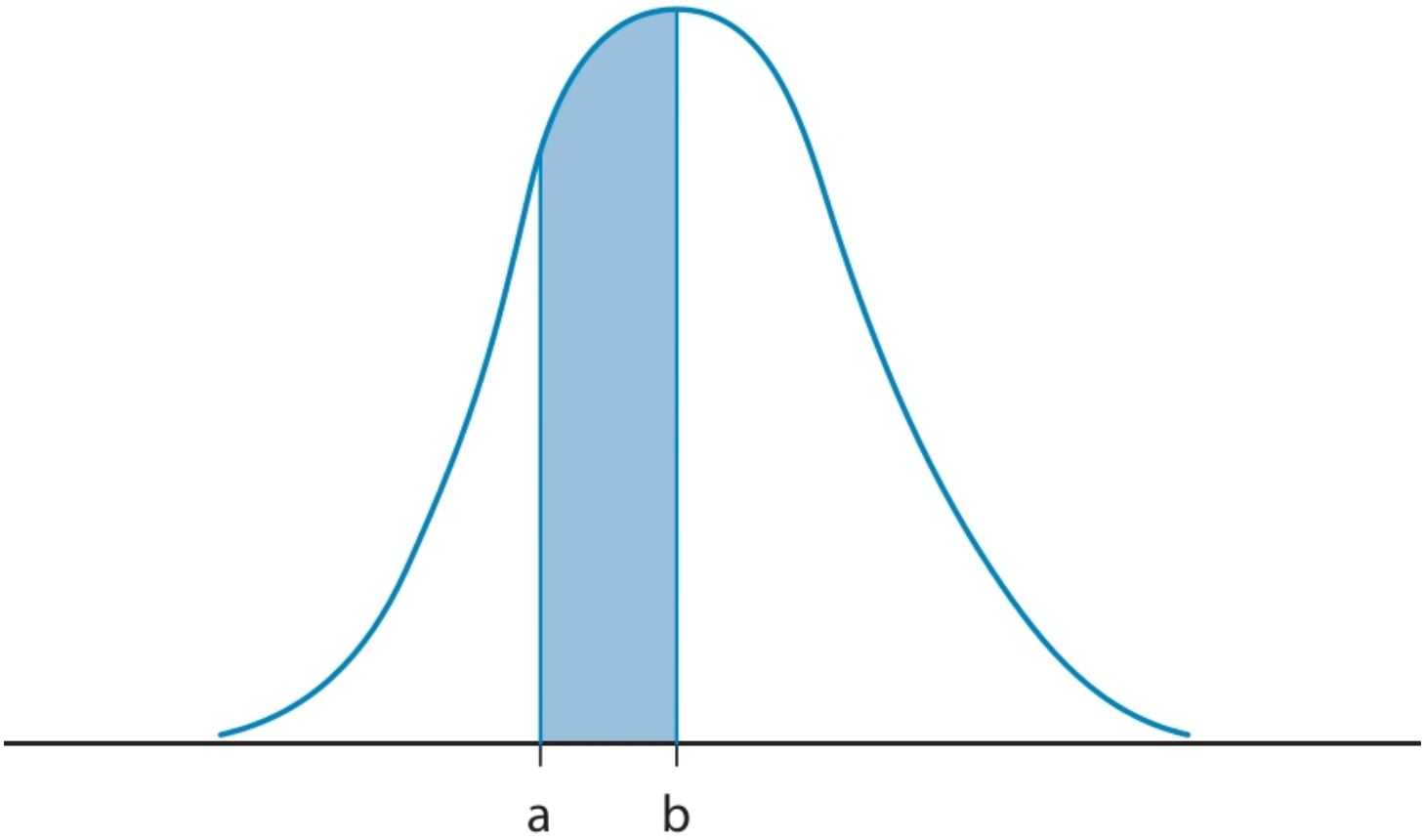
$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1.$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Da stetige Variablen überabzählbar viele Werte annehmen können, kann einzelnen Werten keine positive Wahrscheinlichkeit zugeordnet werden.
- Genauer gesagt geht die Wahrscheinlichkeit eines einzelnen reellen Wertes gegen Null.
- Für stetige Zufallsvariablen kann man deshalb nur Intervallen mit Hilfe der Dichtefunktion Wahrscheinlichkeiten zuzuordnen.

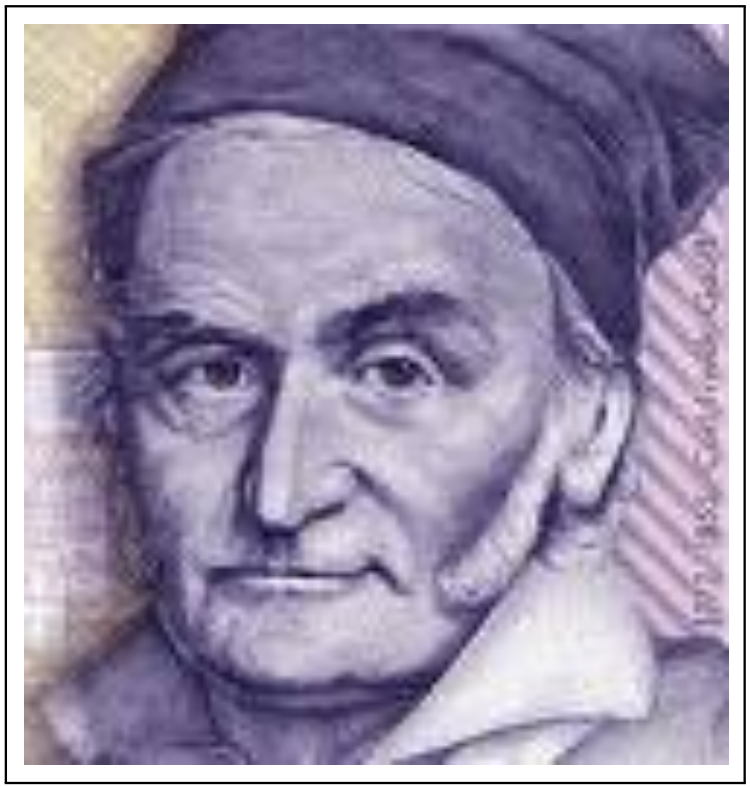


©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Erwartungswert
$$\mu = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x.$$
- Varianz
$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, \mathrm{d}x.$$
- Dichtefunktion $f(x)$.
- Verteilungsfunktion $F(x_p)$ gibt an, wie viel Prozent der Verteilung kleiner oder gleich einem bestimmten Wert x_p sind:
$$F(x_p) = \mathrm{P}(x \leq x_p).$$
- Vergleiche Quantile in der deskriptiven Statistik: z.B. 25% der Werte sind kleiner oder gleich Q_1.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Carl-Friedrich Gauß (1777-1855)

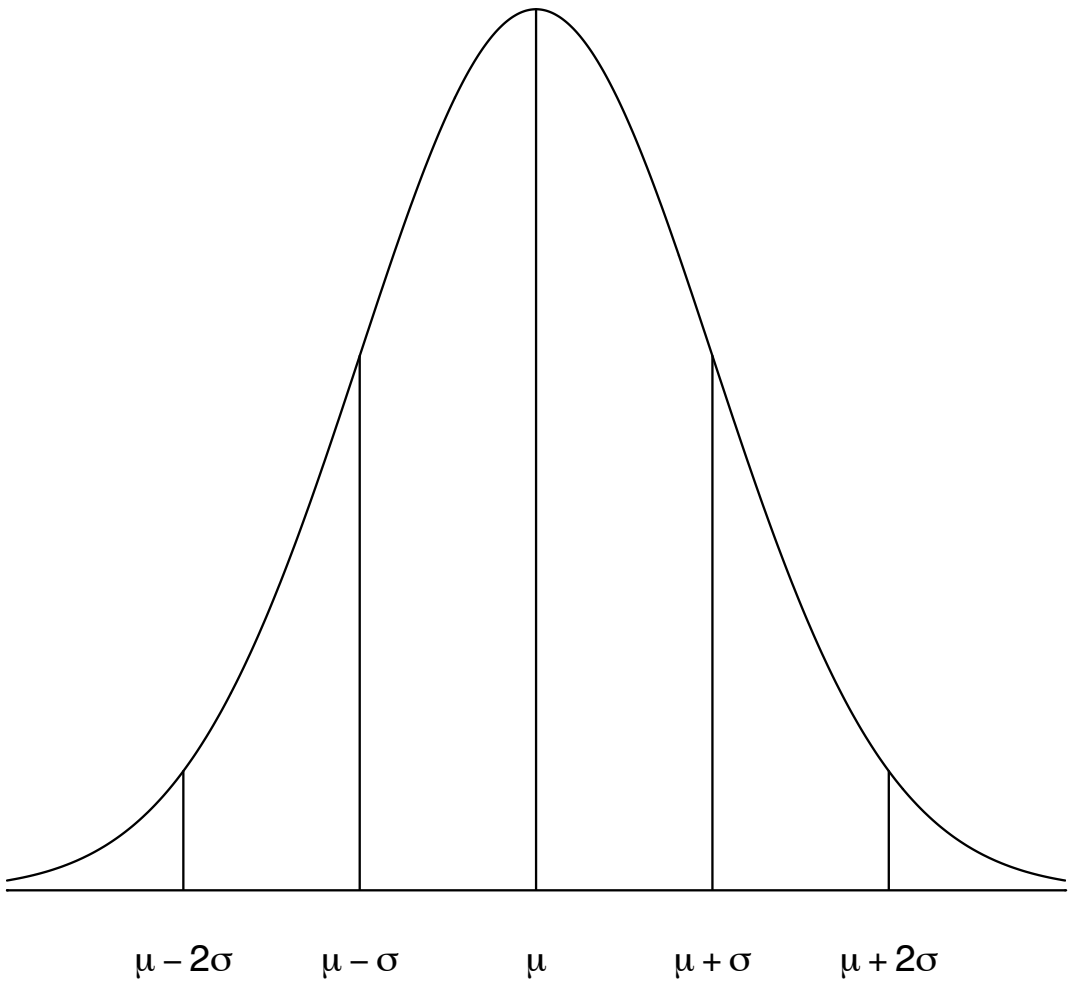


Dichte der Normalverteilung:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Normalverteilung



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Verteilungsfunktion der Normalverteilung

- Die Verteilungsfunktion der Normalverteilung kann nicht als einfache Formel ausgedrückt werden.
- Deshalb werden die Flächenanteile der Standardnormalverteilung tabelliert.
- Flächenanteile unter Normalverteilungen mit beliebigen Werten für μ und σ^2, können mit Hilfe der z-Transformation auf die tabellierten Flächenanteile der Standardnormalverteilung zurückgeführt werden $\Rightarrow$ nur eine Tabelle nötig.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

z-Transformation

- z-Transformation (Daten)

$$z = \frac{x - \bar{x}}{s}$$

- z-Transformation (Verteilung)

$$z = \frac{x - \mu}{\sigma}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Eigenschaften von z-Werten

- Die z-Transformation ändert den Verteilungstyp nicht.
- z-Werte haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- Wenn die Rohwerte normalverteilt sind, folgen die z-Werte einer Standardnormalverteilung.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Notation

- $N(\mu, \sigma^2)$ für beliebige Normalverteilung, z.B. $N(1,\ 4)$, $N(3,\ 4)$, $N(3, 25)$ etc.
- $N(0, 1)$ für Standardnormalverteilung

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Fragetyp 1:

- Bestimmung der Wahrscheinlichkeit = Fläche unter der Normalverteilungsdichte bis zu einem bestimmten Wert.
- IQ-Werte in der Gesamtbevölkerung sind normalverteilt mit $\mu = 100$, $\sigma = 15$.
- Frage: Welcher Prozentsatz der Gesamtbevölkerung hat einen IQ-Wert bis $x = 110$?

Lösung

- Bestimme z-Wert.
- $z = (110 - 100)/15 \approx 0{,}67$.
- Verwende die vereinfachte $N(0,1)$-Tabelle aus der Formelsammlung:

z_p	$F(z_p)$
0	0.50
0.68	0.75
1.28	0.90
1.65	0.95
1.96	0.975
2.33	0.99
2.58	0.995

- Nächster z-Wert ist 0,68, d.h. der Flächenanteil beträgt ca. 0,75.
- Ca. 75% der Gesamtbevölkerung haben einen IQ-Wert bis $x = 110$.

Fragetyp 2:

- Bestimmung des Quantils = Wert, der bestimmte Fläche „abschneidet".
- IQ-Werte in der Gesamtbevölkerung sind normalverteilt mit $\mu = 100$, $\sigma = 15$.
- Frage: Bis zu welchem Wert liegen die IQ-Werte von 99% der Gesamtbevölkerung?

Lösung

- Bestimme z-Wert des 99%-Quantils von $N(0,1)$ aus der vereinfachten $N(0,1)$-Tabelle in der Formelsammlung:

z_p	$F(z_p)$
0	0.50
0.68	0.75
1.28	0.90
1.65	0.95
1.96	0.975
2.33	0.99
2.58	0.995

- Man erhält: $z = 2{,}33$.
- z-Transformation „rückwärts" um den Wert auf der ursprünglichen Skala zu finden:

$$2{,}33 = \frac{x - 100}{15}$$

- Also $x = 100 + 2{,}33 \cdot 15 = 134{,}95$.
- 99% der Gesamtbevölkerung haben einen IQ-Wert bis ca. 135.

Umgekehrte Frage

- Bisher: „von links kommend" = Wahrscheinlichkeit <u>bis zu</u> einem bestimmten Wert.

Fragetyp 1b:

- Frage: Welcher Prozentsatz der Gesamtbevölkerung hat einen IQ-Wert <u>über</u> $x = 110$?

Lösung

- Bestimme z-Wert.
- $z = (110 - 100)/15 \approx 0{,}67$.
- Verwende die vereinfachte $N(0,1)$-Tabelle aus der Formelsammlung:

z_p	$F(z_p)$
0	0.50
0.68	0.75
1.28	0.90
1.65	0.95
1.96	0.975
2.33	0.99
2.58	0.995

- Nächster z-Wert ist 0,68, d.h. der Flächenanteil <u>von links kommend bis zum Wert 110</u> beträgt ca. 0,75.
- Auf der anderen Seite, d.h. <u>über 110</u> bleibt dann übrig: 1-0,75 = 0,25.
- Ca. 25% der Gesamtbevölkerung haben einen IQ-Wert über $x = 110$.

Frage zu Intervall

- „Von links kommend" = Wahrscheinlichkeit <u>bis zu</u> einem bestimmten Wert.
- Für Wahrscheinlichkeit eines Werte-Intervalls: Wahrscheinlichkeit bis zum rechten Rand − Wahrscheinlichkeit bis zum linken Rand.

Fragetyp 1c:

- Frage: Welcher Prozentsatz der Gesamtbevölkerung hat einen IQ-Wert zwischen $x = 100$ und $x = 110$?

Lösung

- Bestimme z-Werte und Wahrscheinlichkeiten:
- $z = (110 - 100)/15 \approx 0{,}67$.
- Nächster z-Wert ist 0,68, d.h. der Flächenanteil <u>von links kommend bis zum Wert 110</u> beträgt ca. 0,75.
- $z = (100 - 100)/15 = 0$.
- Der Flächenanteil <u>von links kommend bis zum Wert 100</u> beträgt 0,5.
- 0,75 - 0,5 = 0,25.
- Ca. 25% der Gesamtbevölkerung haben einen IQ-Wert zwischen $x = 100$ und $x = 110$.

Stichprobe und Grundgesamtheit

Grundgesamtheit

- Die Grundgesamtheit (Population) umfasst alle potenziell untersuchbaren Einheiten, die ein gemeinsames Merkmal aufweisen.
- Eine Stichprobe ist eine Teilmenge der Grundgesamtheit.
- Für statistische Überlegungen werden praktisch immer unendlich große Grundgesamtheiten vorausgesetzt (dadurch bleiben die Wahrscheinlichkeiten auch nach dem Ziehen gleich).

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Einfache Zufallsstichprobe

- Die Bezeichnung „einfache Zufallsstichprobe“ besagt, dass alle Mitglieder der Grundgesamtheit mit gleicher Wahrscheinlichkeit gezogen werden können.
- Sind alle Mitglieder der Grundgesamtheit bekannt, kann man leicht eine Zufallsstichprobe bestimmen, z. B. Urne mit Losen, Zufallszahlen am Computer generieren o.ä.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- *Definition*: Die Stichprobenverteilung ist eine theoretische Verteilung, die die möglichen Ausprägungen eines statistischen Kennwertes (z. B. $\bar{x}$) und deren Auftretenswahrscheinlichkeit beim Ziehen von ganz vielen Zufallsstichproben des Umfanges n beschreibt.
- Der englische Ausdruck für Stichprobenverteilung lautet „Sampling Distribution".

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Eine Stichprobenverteilung lässt sich für einen beliebigen Kennwert ermitteln.
- Beispiele für Kennwerte: Anzahl der Erfolge, Median, Mittelwert, Varianz, Standardabweichung, Interquartilbereich etc.
- Jeder dieser Kennwerte besitzt seine eigene Stichprobenverteilung.
- Im Folgenden wird die Stichprobenverteilung des Mittelwerts betrachtet.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Man bezeichnet Erwartungswert und Standardabweichung der Stichprobenverteilung des Mittels mit $\mu_{\bar{x}}$ und $\sigma_{\bar{x}}$.
- $\sigma_{\bar{x}}$ wird als *Standardfehler des Mittels* bezeichnet.
- $\mu_{\bar{x}} = \mu$.
- $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ bzw. $\sigma^2_{\bar{x}} = \sigma^2/n$.
- Der Standardfehler des Mittels verändert sich also proportional zur Standardabweichung des Merkmals in der Population und verringert sich mit zunehmendem Stichprobenumfang.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Falls es sich bei der Populationsverteilung um eine Normalverteilung handelt, so ist auch die Stichprobenverteilung des Mittelwerts eine Normalverteilung.
- Diese Aussage gilt für beliebig kleinen Stichprobenumfang n, auf dem die Berechnung des Mittelwertes basiert.
- Da Normalverteilungen durch ihren Erwartungswert sowie ihre Varianz eindeutig festgelegt sind, ist damit die Stichprobenverteilung des Mittelwerts vollständig bekannt.
- Wir können diese Aussage auch folgendermaßen ausdrücken:

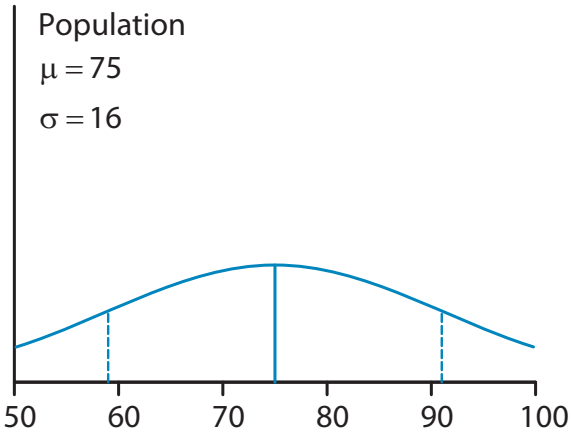
$$\bar{x} \sim N(\mu, \sigma^2_{\bar{x}})$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

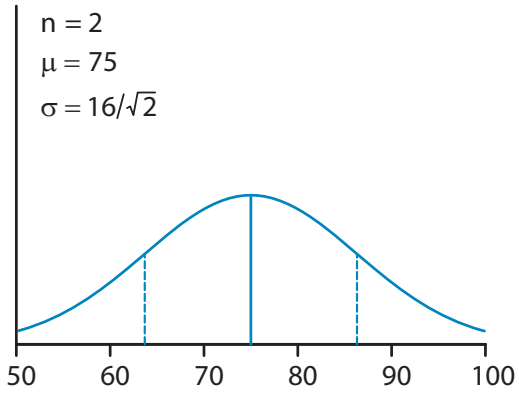
Stichprobenverteilungen bei normalverteilter Population

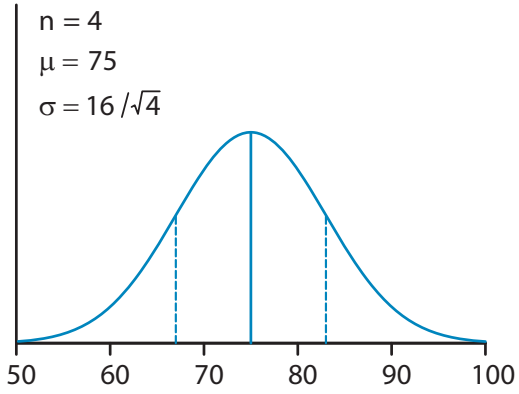
Verteilung der Werte

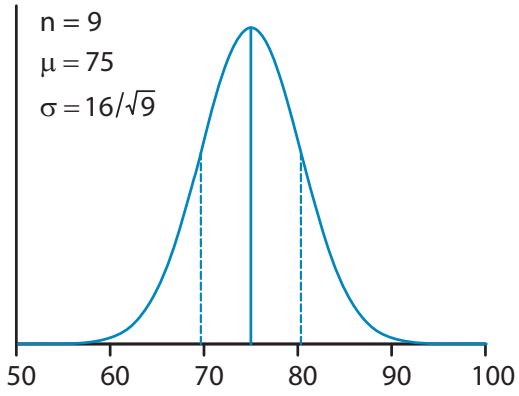
in der Grundgesamtheit



Stichprobenverteilung des Mittelwerts







©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Aufgrund des zentralen Grenzwertsatzes sind wir in der Lage, auch dann Aussagen über die Form der Stichprobenverteilung des Mittelwerts zu machen, wenn wir keine normalverteilte Population vorauszusetzen können – in diesem Fall aber nur, wenn der Stichprobenumfang groß ist.
- Zentraler Grenzwertsatz: Die Verteilung einer Summe (und damit auch eines Mittelwertes) von n unabhängig und identisch verteilten Zufallsvariablen geht mit wachsendem Stichprobenumfang n in eine Normalverteilung über.
- Mit anderen Worten: Für großes n ist die Stichprobenverteilung des Mittelwerts wieder eine Normalverteilung, auch wenn die einzelnen Werte nicht aus einer Normalverteilung stammen.

Wie groß muss n sein, damit $\bar{x}$ approx. normalverteilt ist?

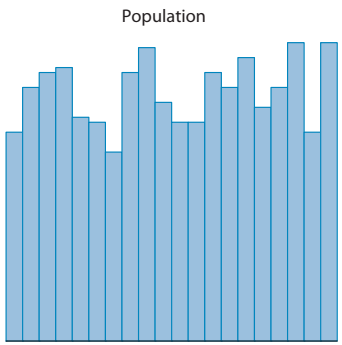
- Auf diese Frage gibt es keine einfache Antwort, denn der erforderliche Stichprobenumfang n hängt von der Rohwertverteilung ab.
- Je stärker die Rohwerteverteilung von einer Normalverteilung abweicht – insbesondere je schiefer die Rohwerteverteilung – umso größer muss n sein.
- Für praktische Zwecke werden wir davon ausgehen, dass ab $n > 30$ die Stichprobenverteilung des Mittelwerts approximativ normalverteilt ist.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

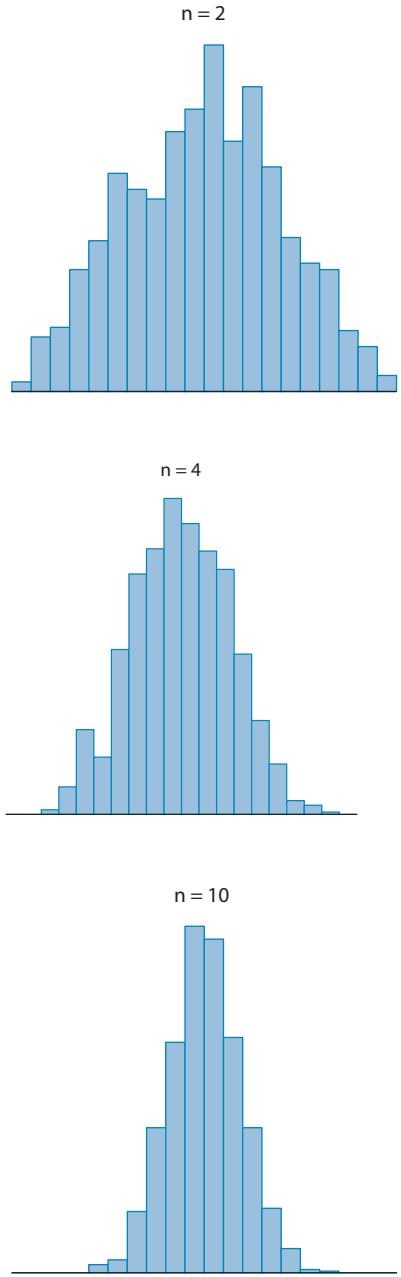
Rohwerte gleichverteilt

Verteilung der Werte

in der Grundgesamtheit



Stichprobenverteilung des Mittelwerts

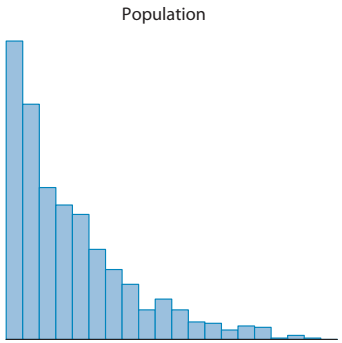


©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

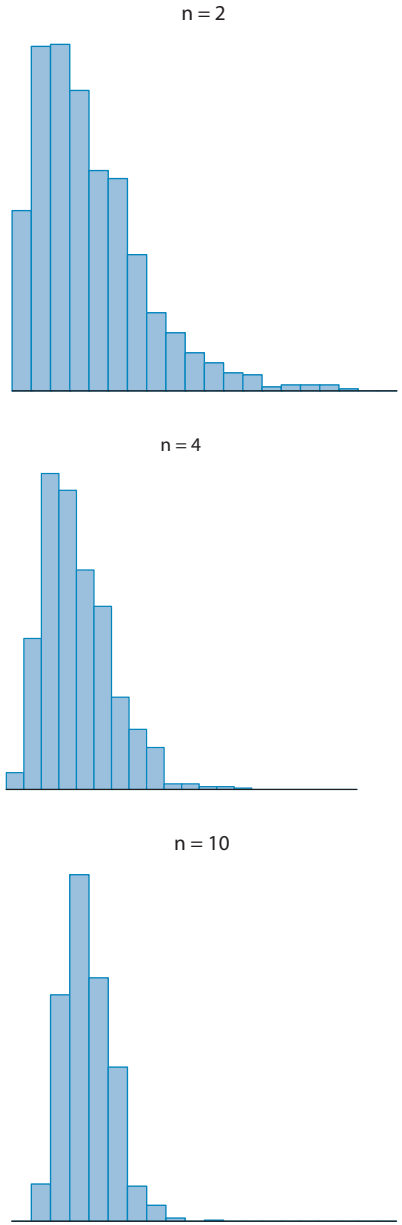
Schiefe Rohwerteverteilung

Verteilung der Werte

in der Grundgesamtheit



Stichprobenverteilung des Mittelwerts



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- $\bar{x}$ und s^2 sind statistische Kennwerte, die aus den für eine Stichprobe erhobenen Werten berechnet werden.
- μ und σ^2 sind unbekannte Parameter, die die Verteilung der Werte in der Grundgesamtheit (die aber zu groß ist, um alle Werte zu erheben) beschreiben.
- Statistische Kennwerte werden verwendet zur
  1. Beschreibung der Stichprobe,
  2. Schätzung von Populationsparametern.
- Wie wir anhand der Stichprobenverteilung des Mittelwerts gesehen haben ist $\bar{x}$ ein guter Schätzer für den wahren Mittelwert μ.
- Auch s^2 ist ein guter Schätzer für die wahre Varianz σ^2.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Schätzung des Standardfehlers

- Wenn die wahre Varianz bekannt ist (z.B. aus einer früheren Erhebung für die gesamte Population) kann man den Standardfehler des Mittelwerts direkt ausrechnen.
- Ist aber (realistischer) die wahre Varianz nicht bekannt sondern muß geschätzt werden, liegt es nahe, den Standardfehler des Mittelwerts

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

dadurch zu schätzen, dass man den Schätzer s^2 für σ^2 in die Formel einsetzt (sog. „Plug-in Schätzer“):

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Man unterscheidet zwischen Punkt- und Intervallschätzung.
- Beispiele für Punktschätzer sind $\bar{x}$ für μ und s^2 für σ^2.
- Zu jedem Punktschätzer kann man ein Konfidenzintervall kostruieren, um einen Eindruck von der Unsicherheit der Schätzung zu erhalten.

Konfidenzintervall

- Beispiel: 95%-Konfidenzintervall für den Mittelwert.
- Die Wahrscheinlichkeit 95% nennt man auch Überdeckungswahrscheinlichkeit oder Sicherheitsgrad.
- Ein 95%-Konfidenzintervall wird so konstruiert, dass es mit 95-prozentiger Wahrscheinlichkeit den wahren Wert des Parameters überdeckt.
- Vorstellung: Wenn man sehr viele Stichproben aus derselben Grundgesamtheit mit Mittelwert μ zieht, und aus jeder Stichprobe den Mittelwert $\bar{x}$ sowie das entsprechende Konfidenzintervall um $\bar{x}$ berechnet, dann überdecken 95% der Intervalle den wahren Mittelwert in der Grundgesamtheit μ, die restlichen 5% liegen daneben.

- Konfidenzintervall:

$$\begin{aligned} \text{untere Grenze} &= \bar{x} - z_{1-\alpha/2} \cdot \sigma_{\bar{x}} \\ \text{obere Grenze} &= \bar{x} + z_{1-\alpha/2} \cdot \sigma_{\bar{x}} \end{aligned}$$

- α bezeichnet die Wahrscheinlichkeit dafür, dass das Intervall „daneben liegt".
- Die Überdeckungswahrscheinlichkeit ist entsprechend $1 - \alpha$.
- Als Überdeckungswahrscheinlichkeit (Sicherheitsgrad) wählt man üblicherweise 95% oder 99%.
- Die Breite des Konfidenzintervalls hängt vom Sicherheitsgrad $(1 - \alpha)$, von der Varianz der einzelnen Werte (σ^2) und vom Stichprobemumfang (n) ab.

Beispiel

- Gesucht wird das 95% Konfidenzintervall für die durchschnittliche Klassengröße an schweizer Sekundarschulen.
- Rohwerte aus einer Stichprobe von 11 Schulklassen:

$$18, 17, 19, 21, 23, 25, 27, 29, 31, 33 \text{ und } 32$$

- Daten: $n = 11$, $\bar{x} = 25$, Annahme: $\sigma = 6$ bekannt
- Also $\sigma_{\bar{x}} = 6/\sqrt{11} = 1{,}809$
- 95%-KI:

  $z_{1-\alpha/2} = z_{1-0{,}05/2} = z_{1-0{,}025} = z_{0{,}975} = 1{,}96$

$$\text{untere Grenze} = 25 - 1{,}96 \cdot 1{,}809 = 21{,}454$$

$$\text{obere Grenze} = 25 + 1{,}96 \cdot 1{,}809 = 28{,}546$$

  KI $\approx$ [21,454; 28,546]

Was sagt das Konfidenzintervall aus?

- Das Konfidenzintervall gibt durch seine Lage und Breite einen Eindruck über den Wert und die Unsicherheit der Schätzung.
- Stellen Sie sich vor, Sie lesen in der Zeitung: „Schweizer Psychologie-Studierende sind überdurchschnittlich neurotisch“

  Hintergrund: Mittelwert aus einer Stichprobe von Studierenden $\bar{x} = 61{,}5$ auf einer Neurotizismus-Skala mit Mittelwert 50 in der Allgemeinbevölkerung.
- Szenario 1: KI $\approx$ [47,64; 75,36]

  Was denken Sie jetzt über die Schlagzeile?
- Szenario 2: KI $\approx$ [60,11; 62,89]

  Und jetzt?

Tests und Konfidenzintervalle

Wo sind wir?

1. Deskriptive Statistik

2. Wahrscheinlichkeitstheorie

3. Tests und Konfidenzintervalle
3.1 Einführung
3.2 Power und Effektstärke
3.3 Tests und Konfidenzintervalle für Mittelwerte
Ein-Stichproben t-Test
t-Test für zwei unabhängige Stichproben
t-Test für zwei verbundene Stichproben
3.4 Nicht-parametrische Tests
3.5 χ^2-Unabhängigkeitstest

4. Kovarianz und Korrelation

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Inhaltliche Hypothesen

Beispiele:

- Durch eine Therapie kann die Suizidneigung gesenkt werden.
- Eine neue Lehrmethode verbessert die Leistung.

Beim Hypothesentesten geht es darum, Unterschiede (vorher vs. nachher, Treatment- vs. Kontrollgruppe) nachzuweisen.

Arten von Hypothesen

- Alternativhypothese H_1 vs. Nullhypothese H_0
  - Alternativhypothese: Unterschied
    (z.B. neue Lehrmethode bewirkt Verbesserung der Leistung)
  - Nullhypothese: kein Unterschied
    (z.B. Leistung bleibt gleich)
- Gerichtete vs. ungerichtete Alternativhypothese
  - Gerichtete Alternativhypothese: Richtung festgelegt
    (z.B. neue Lehrmethode bewirkt <u>Verbesserung</u> der Leistung;
    Therapie bewirkt <u>Senkung</u> der Suizidneigung)
  - Ungerichtete Alternativhypothese: Richtung nicht festgelegt
    (z.B. neue Lehrmethode bewirkt Unterschied in der Leistung,
    d.h. Verbesserung oder Verschlechterung möglich)

Mittelwerte μ_0 und μ

- Statistische Hypothesen beziehen sich auf Populationsparameter, z. B. „wahre" Mittelwerte in der Grundgesamtheit aller Schüler.
- Frage: Verbessert die neue Lehrmethode die Leistung der Schüler?
- Dazu Mittelwerte vergleichen:
- Wir nehmen an, die durchschnittliche Leistung von Schülern, die nach der Standard-Methode unterrichtet werden, sei bekannt. Diesen Mittelwert bezeichnen wir mit μ_0.
- Die durchschnittliche Leistung von Schülern, die nach der neuen Lehrmethode unterrichtet werden, wird als μ bezeichnet.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Null- und Alternativhypothese

- Die gerichtete Alternativhypothese zur Frage: Verbessert die neue Lehrmethode die Leistung der Schüler? heißt damit $H_1 : \mu > \mu_0$.
- Die dazugehörige Nullhypothese lautet: $H_0 : \mu = \mu_0$.
- Bezogen auf den Vergleich zweier Mittelwerte sind die folgenden drei Hypothesenpaare denkbar:

$$H_0 : \mu = \mu_0 \quad \text{und} \quad H_1 : \mu > \mu_0$$
$$H_0 : \mu = \mu_0 \quad \text{und} \quad H_1 : \mu < \mu_0$$
$$H_0 : \mu = \mu_0 \quad \text{und} \quad H_1 : \mu \neq \mu_0$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Fehler der 1. und 2. Art

- Anhand der Daten einer Stichprobe von Schülern soll der statistische Test entscheiden, ob die Nullhypothese abgelehnt werden kann.
- Dabei können zwei Arten von Fehlern gemacht werden:

tatsächlich gilt in der Population	Testentscheidung anhand einer Stichprobe	
	H_0 beibehalten	H_0 ablehnen
H_0 gilt (kein Unterschied)	richtig	Fehler 1. Art
H_0 gilt nicht (Unterschied)	Fehler 2. Art	richtig

Fehler der 1. und 2. Art

- Fehler 1. Art (α-Fehler): Man lehnt H_0 ab obwohl sie eigentlich zutrifft.
  Es wird fälschlicherweise angenommen, die neue Lehrmethode sei besser als die alte Methode (z.B. weil man zur Evaluation der neuen Lehrmethode nur eine kleine Stichprobe von Schülern verwendet und diese zufällig mehrere besonders gute Schüler enthält).
- Fehler 2. Art (β-Fehler): Man behält H_0 bei obwohl eigentlich H_1 zutrifft.
  Es wird fälschlicherweise angenommen, dass sich die neue Lehrmethode von der herkömmlichen nicht unterscheidet (z.B. weil die Stichprobe mehrere besonders schlechte Schüler enthält).

Konsequenzen von Fehlentscheidungen am Beispiel

- Fehler 1. Art: Es wird fälschlicherweise angenommen, die neue Lehrmethode sei besser als die alte Methode.
- Konsequenzen: Neuanschaffung von Lehrmaterial, Umschulung der Lehrkräfte, Neugestaltung der Curricula usw.
- Fehler 2. Art: Es wird fälschlicherweise angenommen, dass sich die neue Lehrmethode von der herkömmlichen nicht unterscheidet.
- Konsequenzen: Eine Chance, den Unterricht zu verbessern, wird verpasst.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Signifikanzniveau

- Idealerweise würde man immer nur richtige Entscheidungen treffen – das geht aber nicht.
- Zudem hängen die beiden Fehlerarten zusammen:
  - Sehr vorsichtig beim Ablehnen von $H_0 \Rightarrow$ Fehler 1. Art klein aber Fehler 2. Art groß
  - Sehr mutig beim Ablehnen von $H_0 \Rightarrow$ Fehler 1. Art groß aber Fehler 2. Art klein
- Kann nur einen von beiden festhalten (aus mathematischen Gründen: Fehler 1. Art).

Signifikanzniveau

- Das Signifikanzniveau α bezeichnet die vom Forschenden festgelegte Obergrenze für die Wahrscheinlichkeit für einen Fehler 1. Art.
- Üblicherweise wird α auf 5% (oder manchmal auf 1%) festgelegt.
- Der Fehler 2. Art kann hingegen nicht direkt kontrolliert werden.

Prüfgröße

- Um aufgrund des festgelegten Signifikanzniveaus α eine Entscheidung über die Nullhypothese treffen zu können, wird eine *Prüfgröße* (auch genannt *Teststatistik*) aus den Daten der Stichproben berechnet.
- Die Prüfgröße spiegelt den interessierenden Unterschied wieder (z.B. Verbesserung der durchschnittlichen Leistung durch neue Lehrmethode).
- Die Verteilung der Prüfgröße unter der Nullhypothese muß zur Berechnung des Tests bekannt sein (aus theoretischen Annahmen oder Simulationen am Computer).
- Beispielsweise verwendet das folgende Beispiel einen z-Wert als Prüfgröße, von dem wir wissen dass er unter der Nullhypothese standardnormalverteilt ist.
- Da Hypothesentests in der Regel nach ihrer Prüfgröße benannt werden, wird dieser Test auch z-Test genannt.

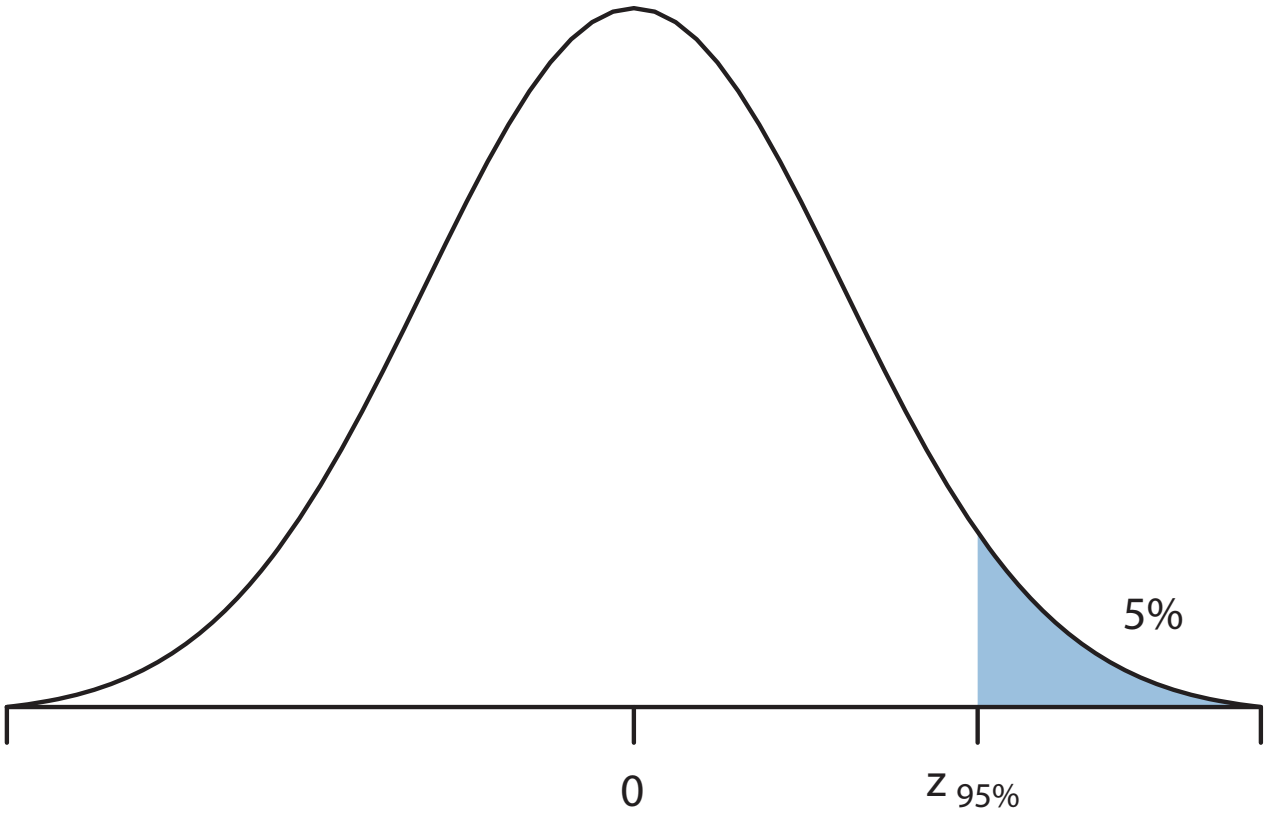
Ablehnbereich und kritische Werte

- Ist die Verteilung der Prüfgröße unter H_0 bekannt, dann lässt sich nach der Festlegung von α ein Bereich bestimmen, der diejenigen Werte der Prüfgröße umfasst, die zur Ablehnung der Nullhypothese führen sollen (z.B. starke Verbesserung der durchschnittlichen Leistung durch neue Lehrmethode spricht gegen H_0).
- Dieser Bereich wird *Ablehnbereich* der Nullhypothese genannt.
- Die Grenzen des Ablehnbereiches werden *kritische Werte* genannt.
- Die Form des Ablehnbereiches hängt davon ab, ob die Alternativhypothese gerichtet oder ungerichtet ist.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Ablehnbereich und kritischer Wert

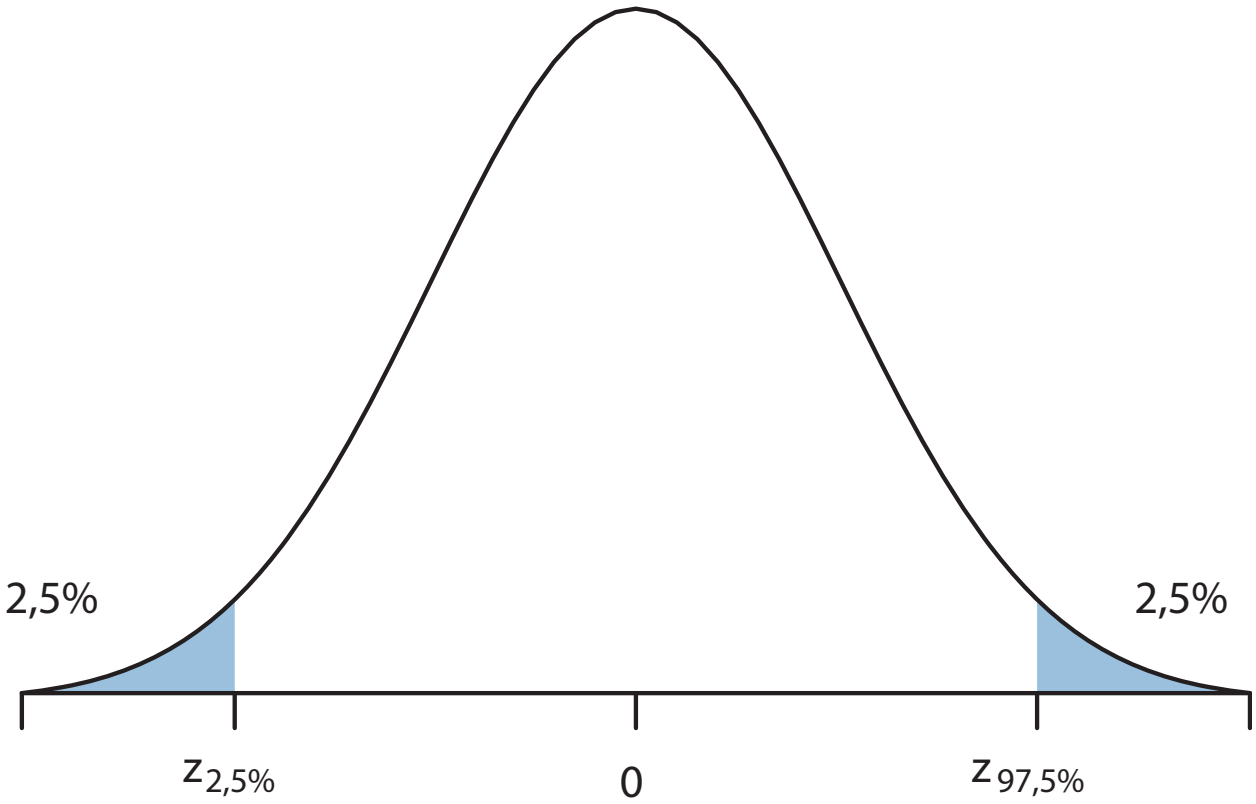
für den einseitigen z-Test auf 5%-Niveau



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Ablehnbereich und kritische Werte

für den zweiseitigen z-Test auf 5%-Niveau



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Es sei für die alte Lehrmethode bekannt, dass die Leistungen der Schüler (gemessen durch einen geeigneten Test) einen Mittelwert von $\mu_0 = 40$ und eine Standardabweichung von $\sigma = 4$ besitzen.
- Wir nehmen an, dass die Leistungen normalverteilt sind.
- Wir möchten die Hypothese überprüfen, dass die neue Lehrmethode die Leistungen der Schüler <u>verbessert</u>.
  Vgl. Formelsammlung:
  Fall a) $H_1 : \mu > \mu_0$
- (Die Varianz der Leistung soll sich aber durch die neue Methode nicht verändern.)

Beispiel

- Um die Hypothese zu prüfen wird eine Stichprobe von $n = 12$ Schülern nach der neuen Lehrmethode unterrichtet.
- Für diese Stichprobe ergibt sich eine durchschnittliche Leistung von $\bar{x} = 42$.
- Der Stichprobenmittelwert $\bar{x}$ ist ein Schätzer für den Mittelwert μ aller nach der neuen Lehrmethode unterrichteten Schüler.
- Soll man angesichts von $\bar{x} = 42$ die Nullhypothese $\mu = \mu_0 = 40$ beibehalten oder ablehnen?

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Erster Schritt: Prüfgrösse berechnen
  $z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\left(\frac{\bar{x} - \mu_0}{\sigma}\right) = \sqrt{12}\left(\frac{42-40}{4}\right)$
  $= 1{,}732$
- Zweiter Schritt: Kritischen Wert (= Grenze des Ablehnbereiches) aus der Tabelle ablesen
  Fall a) $z > z_{1-\alpha}$
  $z_{1-\alpha} = z_{1-0.05} = z_{0.95} = 1{,}65$
- Kontrolle: $1{,}732 > 1{,}65 \Rightarrow H_0$ ablehnen!
- Die neue Lehrmethode führt zu einer (auf dem 5%-Niveau) signifikanten Verbesserung der Leistung.

- Im Beispiel haben wir uns auf die einseitige Hypothese festgelegt, dass die neue (z.B. didaktisch besser konzipierte) Lehrmethode die durchschnittliche Leistung verbessert.
- Mit der einseitigen Hypothese legt man sich vorab auf die vermutete Richtung der Veränderung fest.
- Hat man hingegen weniger inhaltliches Vorwissen (z.B. über die didaktische Eignung der neuen Lehrmethode) kann man die zweiseitige Hypothese verwenden, um „sich beide Möglichkeiten offenzuhalten": eine Verbesserung oder eine Verschlechterung der Leistung.

- Der zweiseitige Test hat aber weniger *Power* zum Nachweis einer Verbesserung:
- Prüfgröße bleibt gleich
  $z = 1{,}732$
- Ablehnbereich jetzt
  Fall c) $z < z_{\alpha/2}$ oder $z > z_{1-\alpha/2}$ bzw. $|z| > z_{1-\alpha/2}$
  $z_{1-\alpha/2} = z_{1-0.025} = z_{0.975} = 1{,}96$
- Kontrolle: $1{,}732 \not> 1{,}96 \Rightarrow$ man kann H_0 nicht ablehnen!
- Fazit: Der zweiseitige Test „hält sich beide Möglichkeiten offen", hat aber dadurch weniger Power als der einseitige Test. Deshalb sollte man, wenn man aufgrund der inhaltlichen Hypothese eine bestimmte Richtung vorhersagen würde, den einseitigen Test verwenden.

Vor der Durchführung der Untersuchung:

- Signifkanzniveau festlegen
- Stichprobenumfang festlegen (s.u.)
- Hypothese festlegen

Eine Modifikation der Hypothese angesichts der gefundenen Daten und eine gleichzeitige Überprüfung der modifizierten Hypothese an denselben Daten ist unzulässig!

Sprachgebrauch

- Mit statistischen Mitteln kann man keine Hypothesen „beweisen”, weil es keine 100%-ige Sicherheit geben kann.
- Deswegen verwendet man die Formulierung: H_0 wird „beibehalten” oder zugunsten von H_1 „abgelehnt”.

Signifikanz und praktische Relevanz

- Statistische Signifikanz darf nicht mit praktischer Relevanz verwechselt werden!
- Ein signifikantes Testergebnis bedeutet: Es ist unwahrscheinlich, dass ein Unterschied dieser Größe nur durch Zufall zustande gekommen ist (obwohl in Wirklichkeit die Nullhypothese gilt).
- Jeder noch so kleine Unterschied kann aber z.B. mit extrem großen Stichproben als signifikant nachgewiesen werden.
- Ein signifikantes Testergebnis bedeutet deshalb noch lange nicht, dass der Unterschied so groß ist, dass er in der Praxis relevant ist (z.B. für den schulischen Erfolg oder für klinische Störungsbilder).
- Zur Größe des Unterschieds $\Rightarrow$ Effektstärke

p-Wert

- Berechnet man einen statistischen Test „per Hand", trifft man die Testentscheidung anhand des kritischen Wertes aus der Tabelle.
- Verwendet man ein Computer-Programm zur Berechnung wird stattdessen der *p-Wert* der Prüfgröße angegeben:

```
One Sample z-test
...
z = 1.7321
p-value = 0.04163
alternative hypothesis: true mean is greater than 40
...
```

- Der *p*-Wert entspricht der Wahrscheinlichkeit, diesen oder einen (in Richtung der Alternativhypothese) extremeren Wert für die Prüfgröße unter der Nullhypothese zu beobachten.

p-Wert

- Testentscheidung „per Hand": H_0 ablehnen wenn

  Prüfgröße im Ablehnbereich (vgl. Formelsammlung)

- äquivalente Testentscheidung über p-Wert: H_0 ablehnen wenn

  p-Wert $< \alpha$

Im obigen Beispiel:

- Testentscheidung „per Hand":
  $1{,}732 > 1{,}65 \Rightarrow H_0$ ablehnen
- äquivalente Testentscheidung über p-Wert:
  $0{,}042 < 0{,}05 \Rightarrow H_0$ ablehnen

(Auch bei zweiseitigen Tests enthält der üblicherweise ausgegebene zweiseitige p-Wert bereits die Wahrscheinlichkeit für beide Seiten, d.h. er kann auch direkt mit α verglichen werden.)

Power und Effektstärke

Power (Teststärke)

- α bezeichnet die Wahrscheinlichkeit eines Fehlers 1. Art (man lehnt H_0 ab, obwohl sie eigentlich zutrifft).
- β bezeichnet die Wahrscheinlichkeit eines Fehlers 2. Art (man behält H_0 bei, obwohl eigentlich H_1 zutrifft).
- $1 - \beta$ bezeichnet die *Power* (man leht H_0 zu Recht ab, wenn H_1 zutrifft).
- Die Power gibt also an, mit welcher Wahrscheinlichkeit ein tatsächlicher Unterschied auch entdeckt werden kann.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Power (Teststärke)

tatsächlich gilt in der Population	Testentscheidung anhand einer Stichprobe H_0 beibehalten	H_0 ablehnen
H_0 gilt	$1-\alpha$	Fehler 1. Art α
H_0 gilt nicht	Fehler 2. Art β	Power $1-\beta$

Größe des Unterschiedes

- Um die Power berechnen zu können, müßte die wahre Größe des Unterschiedes (*Effektgröße* oder *Effektstärke*) bekannt sein.
- Beispiel: Unterschied zwischen der durchschnittlichen Leistung von Schülern, die nach der neuen im Vergleich zur alten Lehrmethode unterrichtet wurden.
- Wahre Größe des Unterschiedes:

$$\mu - \mu_0$$

- Da die durchschnittliche Leistung mit der neuen Lehrmethode in der Population μ jedoch unbekannt ist, ist die wahre Größe des Unterschiedes $\mu - \mu_0$ ebenfalls unbekannt (und damit auch die Power).
- Man kann sich aber überlegen, wie groß der Unterschied sein müßte, um praktisch relevant zu sein, und dafür die Power berechnen.

Beispiel

- Wenn die neue Lehrmethode die durchschnittliche Leistung der Schüler um mindestens drei Punkte verbessern könnte, würde sich dies spürbar positiv auf ihre weitere schulische Laufbahn auswirken.
- Deshalb soll überprüft werden, ob die neue Lehrmethode eine Verbesserung der durchschnittliche Leistung um mindestens drei Punkte erzielen kann.
- Die durchschnittliche Leistung bei Verwendung der herkömmlichen Methode betrug $\mu_0 = 40$ Punkte. Mindestens drei Punkte Verbesserung entsprechen also einer durchschnittlichen Leistung von mindestens $\mu = 43$ Punkten.
- Der interessierende Unterschied entspricht also:

$$\mu - \mu_0 = 43 - 40 = 3$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Standardisierte Effektstärke

- Für weitere Berechnungen wird aus dem inhaltlich interessierenden Unterschied (z.B. Verbesserung um 3 Punkte) die *standardisierte Effektstärke* (= interessierender Unterschied in Standardabweichungen) berechnet:

$$\delta = \frac{\mu - \mu_0}{\sigma}$$

- Im Beispiel sind wir von einer bekannten Standardabweichung $\sigma = 4$ der Leistungswerte ausgegangen.
- Eine Verbessserung um 3 Punkte entspricht also der standardisierten Effektstärke:

$$\delta = \frac{43 - 40}{4} = 0{,}75$$

(d.h. einer Verbesserung um 0,75 Standardabweichungen).

- Die Festlegung eines interessierenden Unterschiedes ist notwendig zur Berechnung von β-Fehler und Power sowie des benötigten Stichprobenumfangs.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispielhaft für den z-Test:

- Bisher war immer nur die Verteilung der standardisierten Prüfgröße z unter der Nullhypothese bekannt:

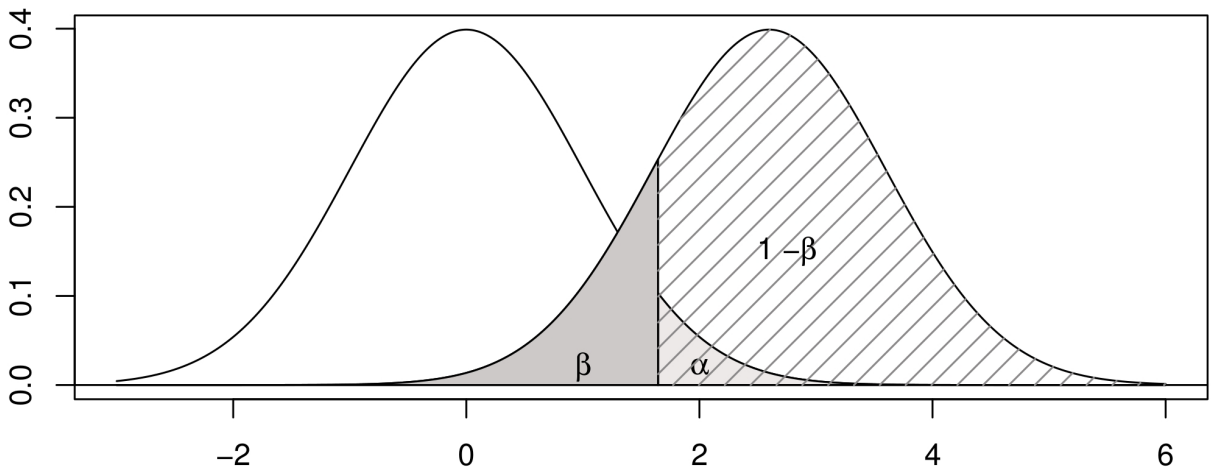
$$N(0,1)$$

- Interessiert man sich aber für einen bestimmten Unterschied und berechnet daraus die standardisierte Effektstärke $\delta = \frac{\mu-\mu_0}{\sigma}$, so ergibt sich als Verteilung der standardisierten Prüfgröße z unter der Alternativhypothese:

$$N(\sqrt{n}\cdot\delta, 1) = N(\sqrt{n}\cdot\frac{\mu-\mu_0}{\sigma}, 1)$$

Bestimmung von β-Fehler und Power

- Verteilungen der Prüfgröße unter H_0 und H_1:



α: Wahrscheinlichkeit unter H_0, dass Prüfgröße in Ablehnbereich fällt (nicht vorhandenen Unterschied „entdeckt“)

β: Wahrscheinlichkeit unter H_1, dass Prüfgröße nicht in Ablehnbereich fällt (vorhandenen Unterschied nicht entdeckt)

$1 - \beta$: Wahrscheinlichkeit unter H_1, dass Prüfgröße in Ablehnbereich fällt (vorhandenen Unterschied entdeckt)

- Die folgende Gleichung ist für die Berechnung der Power des z-Tests von zentraler Bedeutung, da sie die vier Größen, α, β, n und δ miteinander in Beziehung setzt:

$$z_\beta = z_{1-\alpha} - \sqrt{n} \cdot \delta$$

- Kennt man drei der vier Größen, kann man die vierte Größe bestimmen:
- β und $1 - \beta$ ergeben sich aus z_β.
- Für die Praxis ist außerdem die Bestimmung des Stichprobenumfangs n von großer Bedeutung.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Der Stichprobenumfang betrug $n = 12$.
- Ist man an der Entdeckung eines Leistungs-Unterschiedes von drei Punkten interessiert, ergibt sich der Wert $\delta = 0{,}75$ für die standardisierte Effektstärke.
- Der kritische Wert $z_{1-\alpha}$ ist: $z_{95\%} = 1{,}65$
- Somit errechnen wir:

$$z_\beta = 1{,}65 - \sqrt{12} \cdot 0{,}75 = -0{,}949$$

- Dieser z-Wert schneidet in der Standardnormalverteilung gerade 17% im unteren Rand der Verteilung ab (in unserer Tabelle aus Platzgründen nicht aufgelistet – wäre bei Aufgaben angegeben), d.h.: $\beta = 0{,}17$
- Somit hat die Power den Wert: $1 - \beta = 0{,}83$
- D.h. die Wahrscheinlichkeit einen Unterschied von 3 Punkten zu entdecken beträgt (bei der Stichprobengröße und Standardabweichung aus dem Beispiel) 83%.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Die Gleichung läßt sich auch so umformen, dass man den Stichprobenumfang bestimmen kann, der notwendig ist, um einen Effekt der Größe δ mit einer Power von $1 - \beta$ nachzuweisen:

$$n = \left(\frac{z_\beta - z_{1-\alpha}}{\delta} \right)^2$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Die an der Studie beteiligten Lehrkräfte könnten eine Power von mindestens $1 - \beta = 0{,}95$ fordern, um einen vorhandenen Effekt der Größe $\delta = 0{,}75$ mit großer Sicherheit entdecken zu können.
- Wie groß muss die Stichprobe von Schülern, die nach der neuen Lehrmethode unterrichtet werden, mindestens sein, um diese Power zu gewährleisten?

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Die Werte, die in die Gleichung eingesetzt werden müssen, sind: $\delta = 0{,}75$ und $z_\beta = -z_{1-\beta} = -z_{95\%} = -1{,}65$ (vgl. Formelsammlung)

$$n = \left(\frac{-1{,}65 - 1{,}65}{0{,}75}\right)^2 = 19{,}36$$

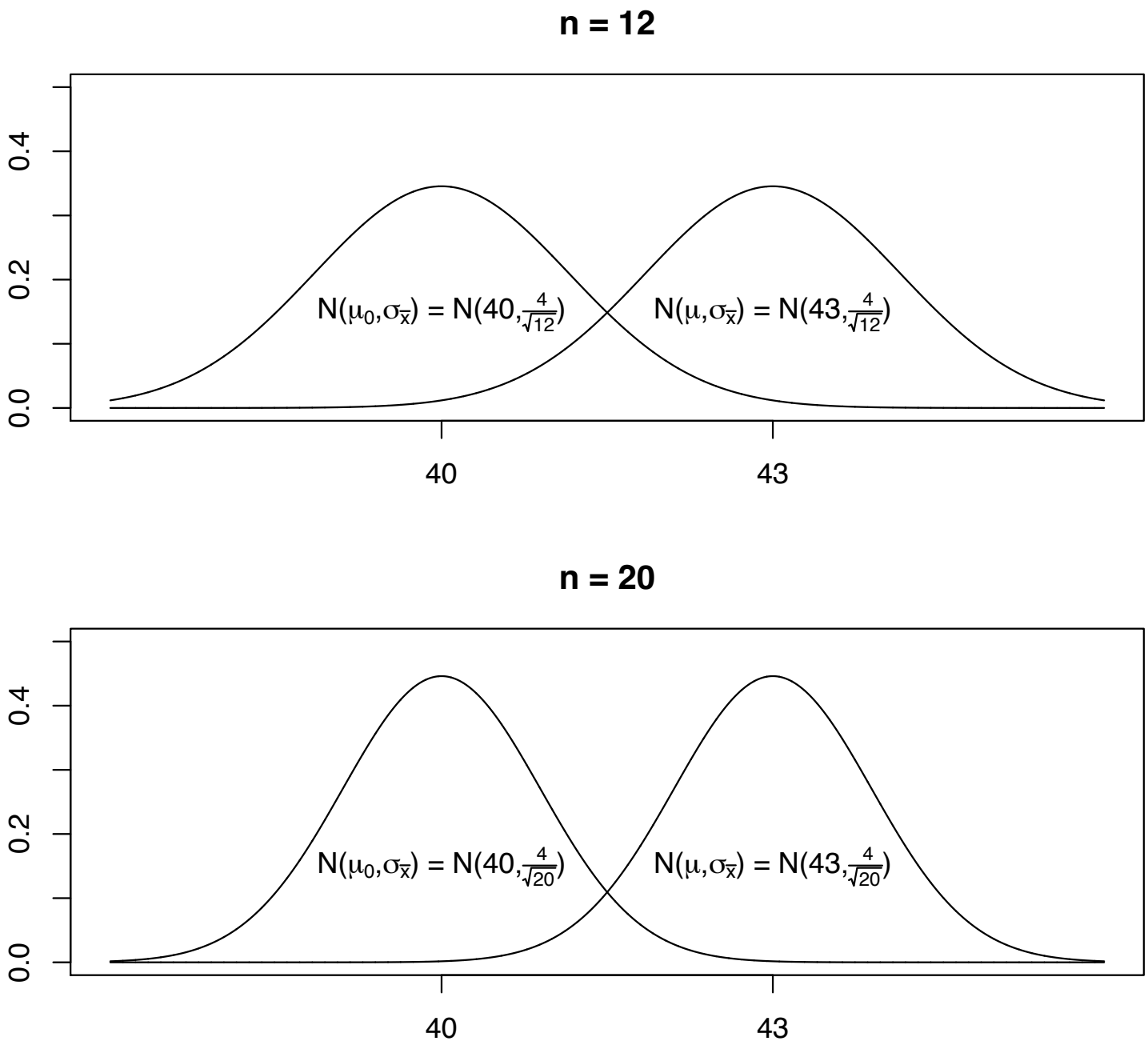
- Man benötigt also eine Stichprobe von mindestens 20 Schülern (nächst größere ganze Zahl) um den spezifizierten Effekt mit 95% Wahrscheinlichkeit entdecken zu können, wenn er existiert.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

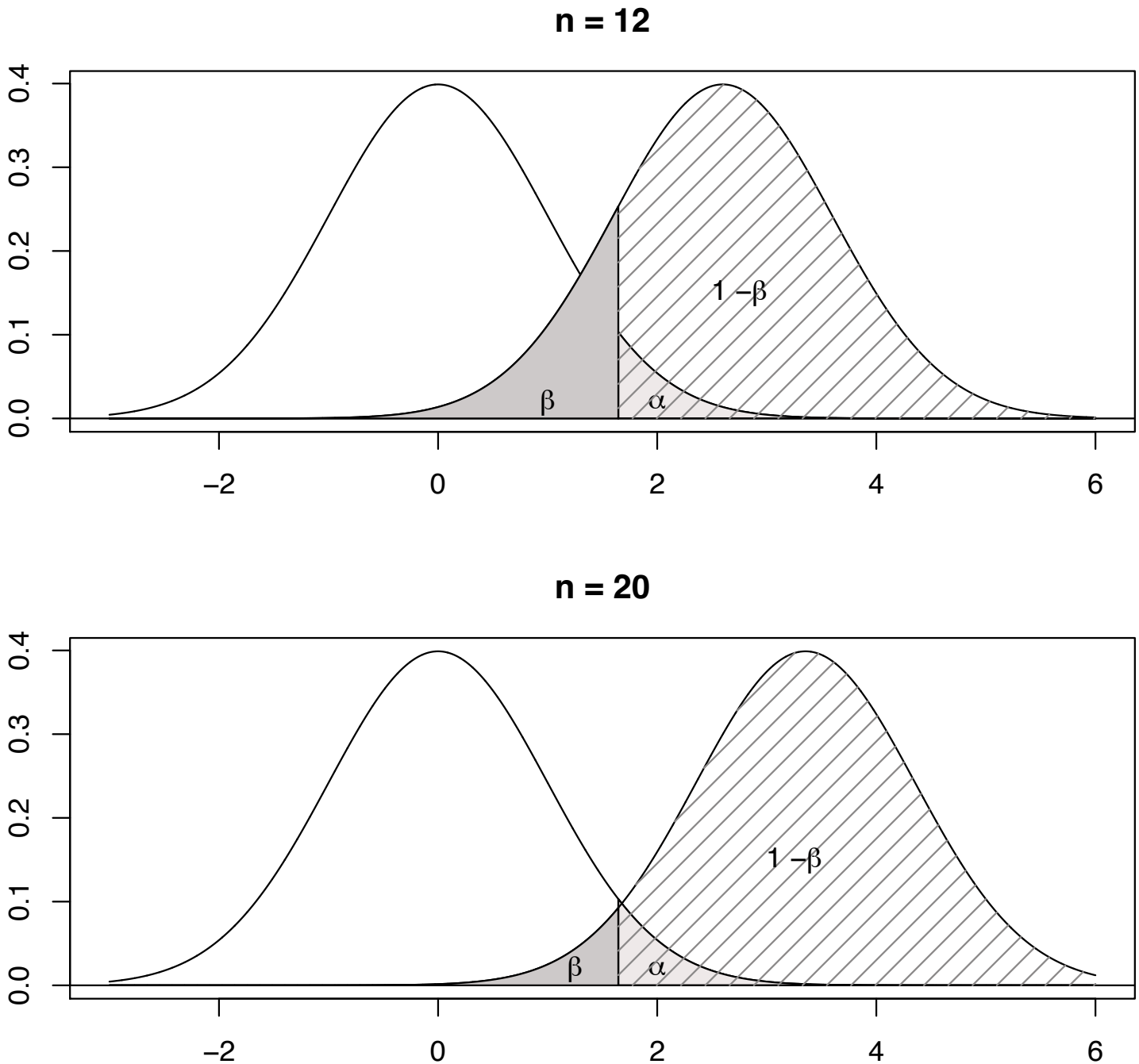
Determinanten der Power

- Die Power steigt bei größerem Stichprobenumfang n. Große Stichproben machen es leichter, einen Effekt nachzuweisen.
- Die Power steigt bei größerem Unterschied $|\mu - \mu_0|$. Große Effekte sind leichter nachzuweisen.
- Die Power sinkt bei niedrigerem Signifikanzniveau α. Vgl. Grafik: Strenger Test $\Leftrightarrow$ α kleiner $\Leftrightarrow$ kritischer Wert weiter rechts $\Leftrightarrow$ $1 - \beta$ auch kleiner.
- Der einseitige Test hat eine höhere Power als der zweiseitige Test (d.h. wenn man eine inhaltliche Hypothese über die Richtung des Unterschiedes hat, sollte man sie verwenden).

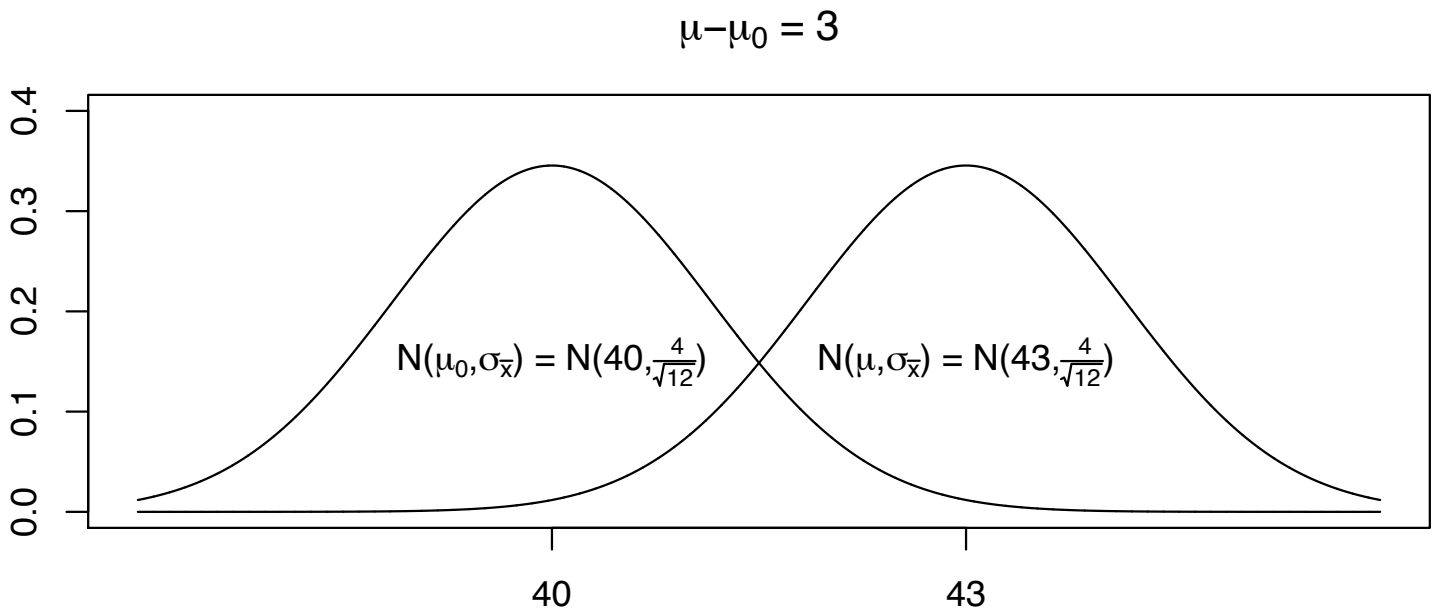
Stichprobenverteilung des Mittelwerts unter H_0 und H_1
in Abhängigkeit von der Größe des Stichprobenumfangs n

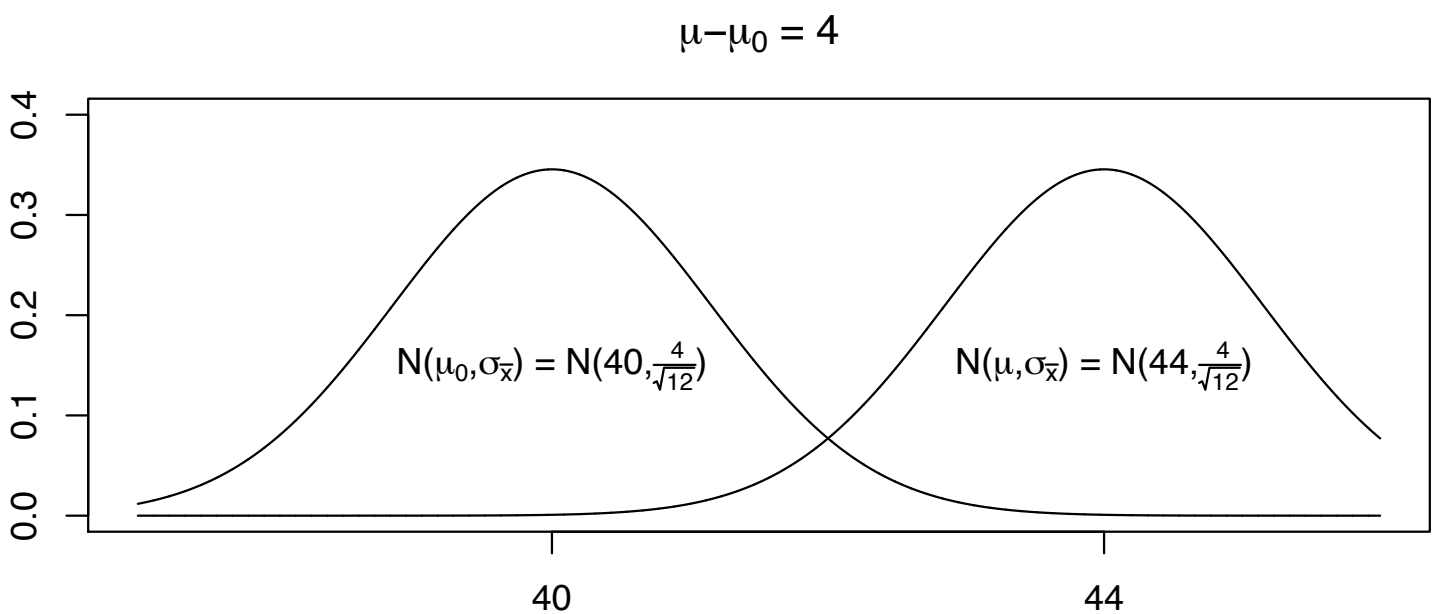


Power in Abhängigkeit von der Stichprobengröße n

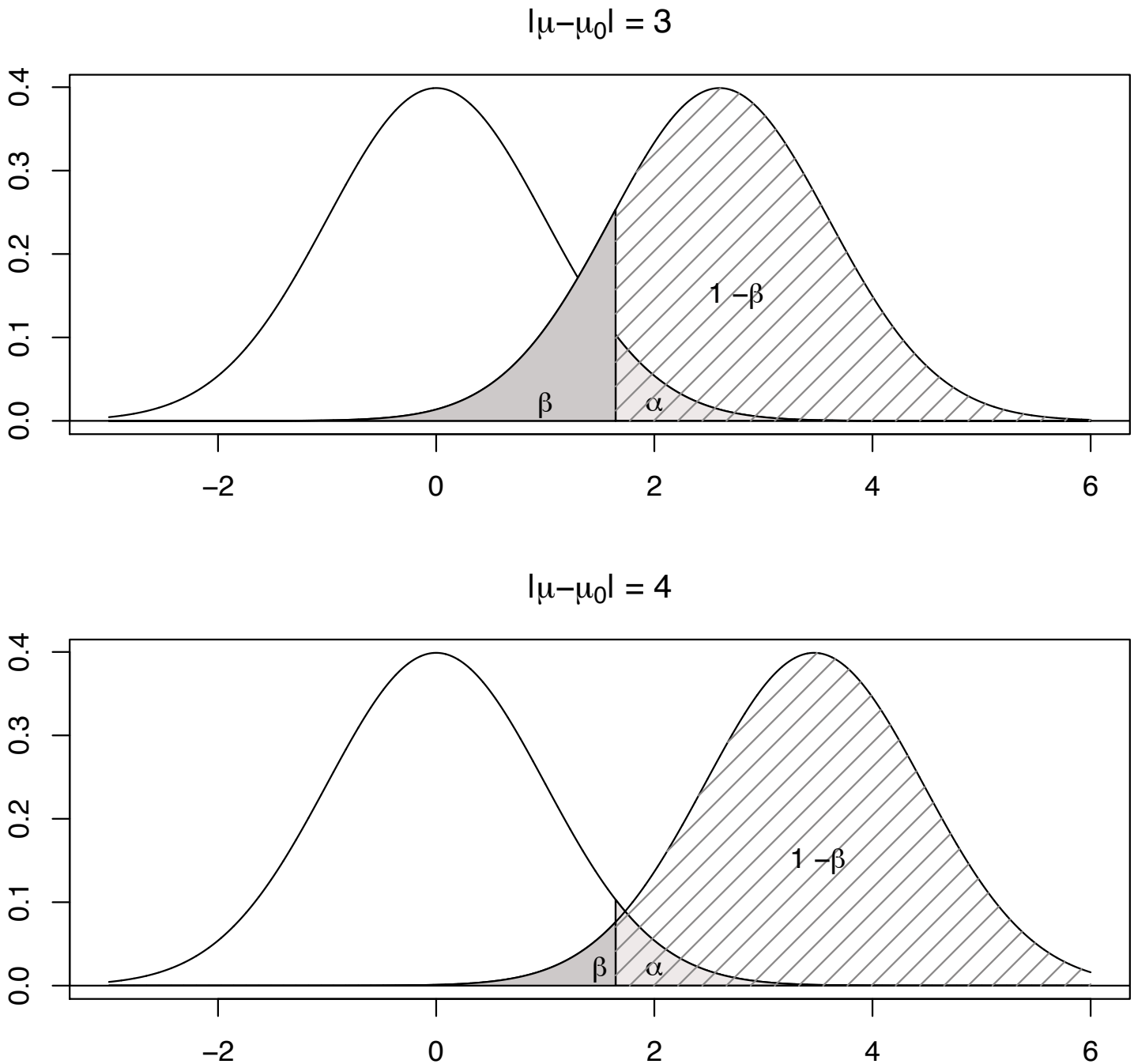


Stichprobenverteilung des Mittelwerts unter H_0 und H_1
in Abhängigkeit von der Größe des Unterschiedes $|\mu - \mu_0|$

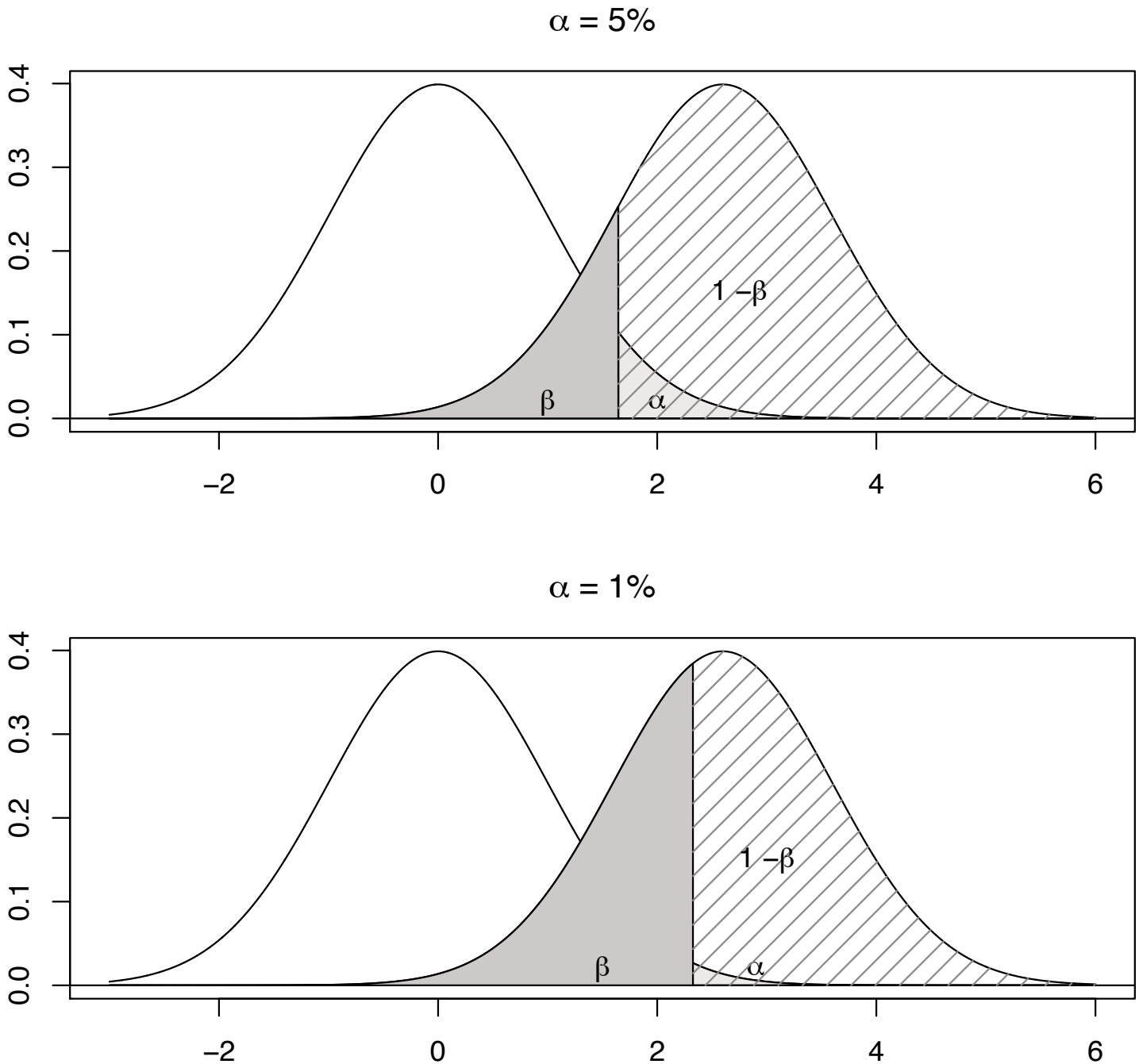




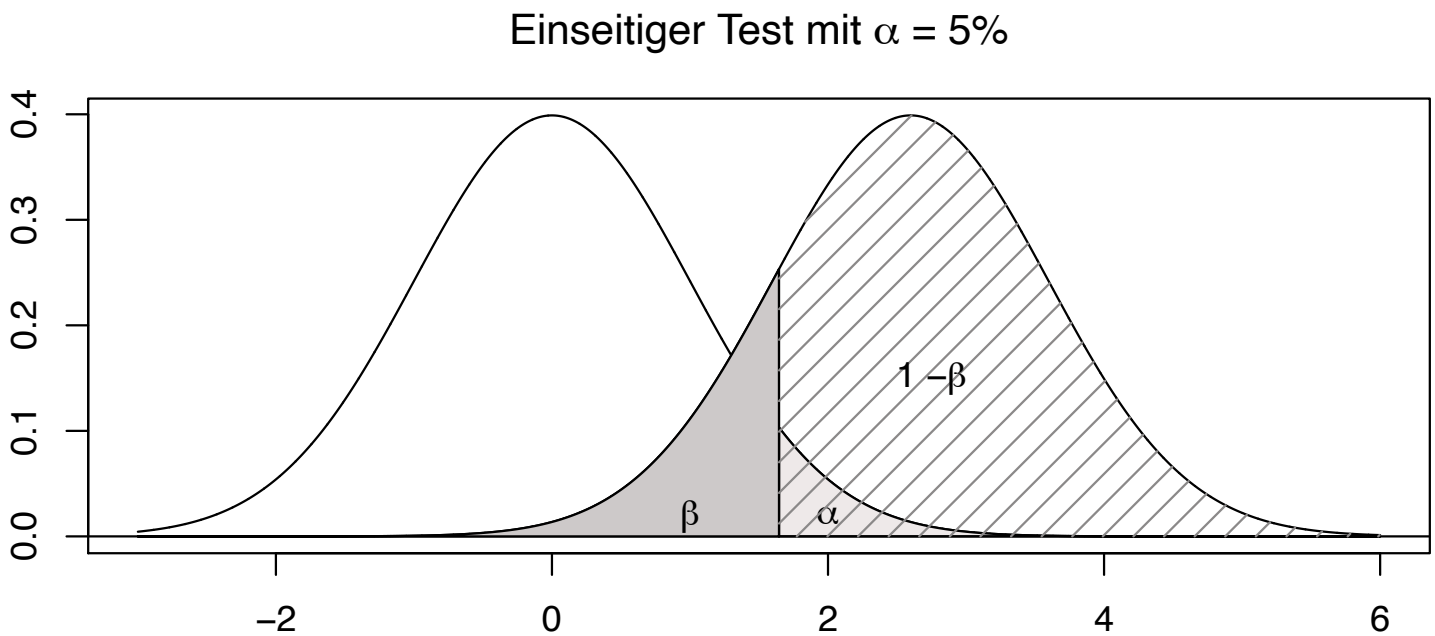
Power in Abhängigkeit von der Größe des Unterschiedes $|\mu-\mu_0|$

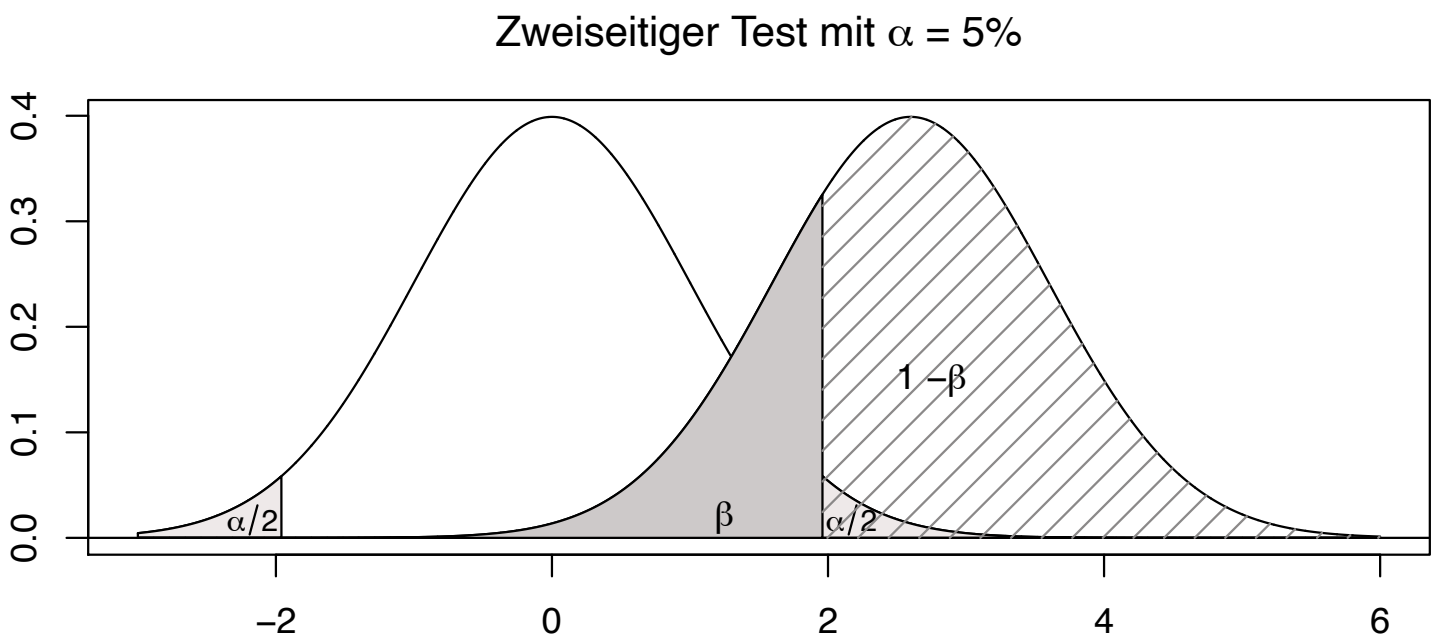


Power in Abhängigkeit vom Signifikanzniveau α



Power für einseitige und zweiseitige Tests





Tests und Konfidenzintervalle für Mittelwerte

Drei Arten von Fragestellungen

Vergleich ...

- eines Stichprobenmittelwertes mit einem Populationsmittelwert (Ein-Stichproben Tests)
- zweier Stichprobenmittelwerte für unabhängige Beobachtungen (Tests für unabhängige Stichproben)
- zweier Stichprobenmittelwerte für Beobachtungspaare (Tests für abhängige/verbundene Stichproben)

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Der im letzten Teil besprochene Test wird häufig z-Test genannt, weil die Nullhypothese anhand der Standardnormalverteilung überprüft wird.
- Für den z-Test musste die Populationsstreuung σ bekannt sein – was in der Realität üblicherweise nicht der Fall ist.
- Der t-Test verwendet die t-Verteilung und lässt sich auch für unbekannte Populationsstreuung durchführen.
- Beide Tests prüfen die Hypothese

$$H_0 : \mu = \mu_0$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Prüfverteilungen: Falls $H_0 : \mu = \mu_0$ gilt, dann ...

- Rohwerte normalverteilt und Populationsstreuung bekannt:

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\left(\frac{\bar{x} - \mu_0}{\sigma}\right) \sim N(0,1)$$

- Rohwerte normalverteilt aber Populationsstreuung unbekannt:

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \sqrt{n}\left(\frac{\bar{x} - \mu_0}{s}\right) \sim t(n-1)$$

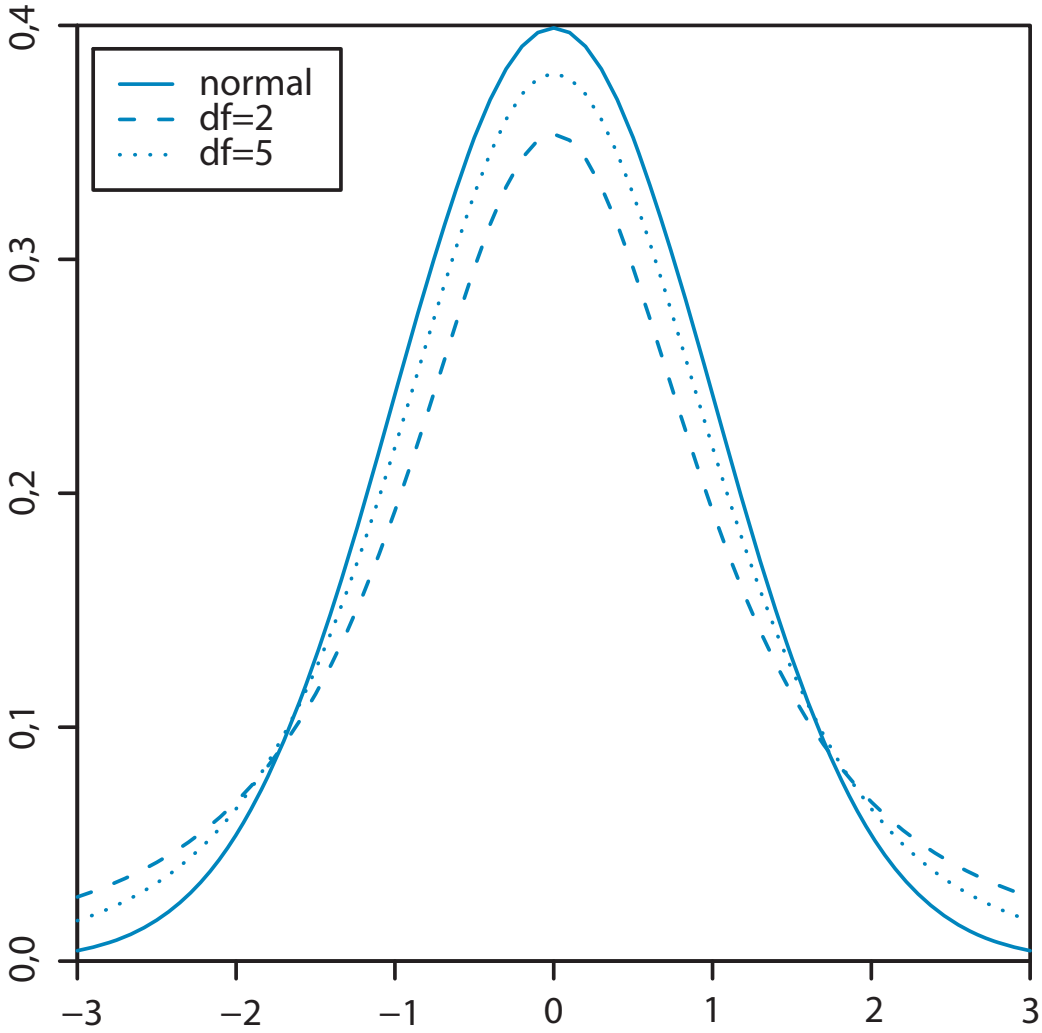
mit s als Schätzer für σ

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Eine t-Verteilung wird durch die Anzahl von sog. Freiheitsgraden genauer festgelegt.
- Die Prüfgröße des Ein-Stichproben t-Tests folgt z.B. einer t-Verteilung mit $n-1$ Freiheitsgraden (die Anzahl der Freiheitsgrade hängt von der Stichprobengröße und der Anzahl der zu schätzenden Parameter ab).
- Das englische Wort für Freiheitsgrade ist „degrees of freedom".
- Deswegen schreiben wir auch: $df = n - 1$
- Die korrekte Anzahl der Freiheitsgrade ist in der Formelsammlung mit angegeben.
- Für großen Stichprobenumfang ($n > 30$) ist die t-Verteilung der Standardnormalverteilung sehr ähnlich.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

t-Verteilungen und $N(0,1)$



©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

t-Tabelle

df	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.325	1.376	3.078	6.314	12.706	31.820	63.657	318.309	636.619
2	0.289	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
27	0.256	0.855	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.855	1.312	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.854	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.854	1.310	1.697	2.042	2.457	2.750	3.385	3.646
∞	0.25	0.84	1.28	1.65	1.96	2.33	2.58	3.09	3.29

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Zirkadianer Rhythmus

- Es soll überprüft werden, ob der 24-stündige Tagesrhythmus des Menschen auch ohne Tageslicht aufrecht erhalten wird.
- Versuch: Sieben Freiwillige werden für vier Tage im Schlaflabor ohne Tageslicht einquartiert.
- Jede Versuchsperson ist während der vier Tage alleine und erhält keine Informationen zur Tageszeit.
- Die Person muss unmittelbar vor dem Zu-Bett-Gehen einen Knopf betätigen, wodurch die Uhrzeit festgehalten wird.
- Als Indikator für den selbstgewählten Tagesrhythmus wird die Dauer (in Minuten) zwischen dem Zu-Bett-Gehen am dritten Versuchstag und dem Zu-Bett-Gehen am vierten Versuchstag verwendet.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Zirkadianer Rhythmus

- Wenn durch die experimentellen Bedingungen die innere Uhr der Personen nicht beeinträchtigt wird, sollte der wahre Mittelwert $24 \cdot 60\,\text{min} = 1440\,\text{min}$ betragen.
- D.h. die Nullhypothese lautet:

$$H_0 : \mu = 1440$$

- Es soll überprüft werden, ob sich der Tagesrhythmus verändert (länger oder kürzer wird).
- D.h. die Alternativhypothese formulieren wir zweiseitig:

$$H_1 : \mu \neq 1440$$

- = Fall c) in der Formelsammlung

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Zirkadianer Rhythmus

- Folgende Werte wurden für die sieben Versuchspersonen ermittelt:

  1452, 1438, 1487, 1439, 1454, 1461 und 1476

- Berechnet man den Mittelwert und die Standardabweichung der Werte, ergeben sich: $\bar{x} = 1458{,}143$ und $s = 18{,}197$.
- Das Signifikanzniveau wird auf $\alpha = 0{,}05$ festgelegt.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Zirkadianer Rhythmus

- Erster Schritt: Berechnung der Prüfgröße:

$$t = \sqrt{7}\ \left(\frac{1458{,}143 - 1440}{18{,}197}\right) = 2{,}638$$

- Zweiter Schritt: Kritischer t-Wert aus der Tabelle ablesen: Fall c) $t < t_{\alpha/2}(n-1)$ oder $t > t_{1-\alpha/2}(n-1)$ bzw. $|\,t\,| > t_{1-\alpha/2}(n-1)$

$$t_{1-\alpha/2}(n-1) = t_{0.975}(6) = 2{,}447$$

- Kontrolle: $2{,}638 > 2{,}447 \Rightarrow H_0$ ablehnen!
- Der Tagesrhythmus verändert sich durch das fehlende Tageslicht signifikant.
- (Am Mittelwert kann man deskriptiv die Richtung der Veränderung ablesen: Der Abstand zwischen dem Zu-Bett-Gehen ist länger.)

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Die früheren Ausführungen gingen von einer bekannten Populationsvarianz aus und führten zu folgender Berechnung der Intervallgrenzen:

$$\bar{x} \pm z_{1-\alpha/2} \cdot \sigma/\sqrt{n}$$

- Ist die Populationsvarianz nicht bekannt, so ist die Berechnung der Konfidenzgrenzen in zwei Punkten zu modifizieren:
  - Der Standardfehler des Mittels $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ wird durch $s_{\bar{x}} = s/\sqrt{n}$ ersetzt.
  - Das Perzentil der Standardnormalverteilung wird durch ein Perzentil der t-Verteilung ersetzt. D.h. anstatt $z_{1-\alpha/2}$ verwenden wir $t_{1-\alpha/2}(n-1)$.
- Die Konfidenzintervallgrenzen lauten somit:

$$\text{untere Grenze} = \bar{x} - t_{1-\alpha/2}(n-1) \cdot s/\sqrt{n}$$
$$\text{obere Grenze} = \bar{x} + t_{1-\alpha/2}(n-1) \cdot s/\sqrt{n}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Zirkadianer Rhythmus

- Die relevanten Kennwerte liegen für dieses Beispiel bereits vor: Mittelwert und Standardabweichung der $n = 7$ Rohwerte lauten: $\bar{x} = 1458{,}143$ und $s = 18{,}197$
- Wir berechnen nun ein 95% Konfidenzintervall für μ, wobei μ die „wahre" mittlere Dauer (in Minuten) zwischen dem Zu-Bett-Gehen am dritten Versuchstag und dem Zu-Bett-Gehen am vierten Versuchstag in der Grundgesamtheit bezeichnet (die sich nur ermitteln ließe, wenn man die gesamte Grundgesamtheit 4 Tage ins Schlaflabor sperren könnte...)
- Das benötigte Perzentil lautet: $t_{0.975}(6) = 2{,}447$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel: Zirkadianer Rhythmus

- Somit ergeben sich die Grenzen:

$$\text{untere Grenze} = 1458{,}143 - 2{,}447 \cdot 18{,}197/\sqrt{7} = 1441{,}313$$
$$\text{obere Grenze} = 1458{,}143 + 2{,}447 \cdot 18{,}197/\sqrt{7} = 1474{,}973$$

- Ein so konstruiertes Intervall überdeckt mit 95% Sicherheit den wahren Mittelwert.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Dualität von Test und Konfidenzintervall

- Man kann das Konfidenzintervall auch verwenden, um den zweiseitigen Test für die ungerichtete Alternativhypothese

$$H_1 : \mu \neq \mu_0$$

durchzuführen.
- Liegt der Wert μ_0 innerhalb des $(1-\alpha)$-Konfidenzintervalls für μ, so führt der zweiseitige Test zum Signifikanzniveau α zur Beibehaltung der Nullhypothese.
- Liegt der Wert μ_0 außerhalb des $(1-\alpha)$-Konfidenzintervalls für μ, so führt der zweiseitige Test zum Signifikanzniveau α zur Ablehnung der Nullhypothese.
- Hier: Der Wert $\mu_0 = 1440$ liegt außerhalb des Konfidenzintervalls $\Rightarrow H_0$ ablehnen!

t-Test für zwei unabhängige Stichproben

Wo sind wir?

1. Deskriptive Statistik

2. Wahrscheinlichkeitstheorie

3. Tests und Konfidenzintervalle
3.1 Einführung
3.2 Power und Effektstärke
3.3 Tests und Konfidenzintervalle für Mittelwerte
Ein-Stichproben t-Test
t-Test für zwei unabhängige Stichproben
t-Test für zwei verbundene Stichproben
3.4 Nicht-parametrische Tests
3.5 χ^2-Unabhängigkeitstest

4. Kovarianz und Korrelation

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

t-Tests für zwei Stichproben

- Forschungsfrage: Gibt es einen Unterschied zwischen zwei Gruppen?
- Vergleiche der Mittelwerte von zwei (später auch mehr) Gruppen kommen häufiger vor als Vergleiche eines Mittelwerts mit einem vorgegebenen Wert μ_0.
- Es gibt zwei Varianten:
  - unabhängige Stichproben
  - verbundene Stichproben

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Man vergleicht zwei Gruppen von Personen bzgl. einer Zielgröße.
- Beispiel: Vergleich von Treatment- und Kontroll-Gruppe bzgl. Gedächtnisleistung.
- Die wahren Mittelwerte und Standardabweichungen in den beiden Gruppen werden mit μ_1, μ_2, σ_1 und σ_2 bezeichnet.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- In der Grundgesamtheit liegen zwei Gruppen vor, aus denen die Stichproben gezogen werden (Beobachtungsstudie), z.B.:
  - Singles und in Partnerschaft lebende Personen
  - Psychologie- und BWL-Studierende
  - Personen, die eine Behandlung erhalten, und Personen, die keine Behandlung erhalten

  bzw. idealerweise: zufällige Zuordnung der gezogenen Personen zu den Treatments (Experiment).

Anmerkung: Ein statistisch signifikanter Unterschied zwischen den Gruppen allein ist kein Nachweis eines kausalen Einflusses der Gruppierungsvariable! In einer Beobachtungsstudie kann es noch weitere Unterschiede zwischen den Gruppen geben, die für den Effekt verantwortlich sind. Nur ein randomisiertes Experiment stellt sicher, dass sich die Gruppen in keiner anderen Variable systematisch unterscheiden.

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 > \mu_2 \qquad \text{(einseitig)}$$
$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 < \mu_2 \qquad \text{(einseitig)}$$
$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2 \qquad \text{(zweiseitig)}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Prüfgröße

- Prüfgröße des t-Test für zwei unabhängigen Stichproben:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

- Dabei bezeichnet $s_{\bar{x}_1 - \bar{x}_2}$ die Standardabweichung der Mittelwertdifferenzen.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Standardabweichung der Mittelwertsdifferenzen

- Der Nenner der t-Statistik ist

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

- Dabei wird die Varianz der Rohwerte aus beiden Stichproben zusammen (daher sog. „gepoolte" Varianz) folgendermaßen geschätzt:

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

- Dabei bezeichnet n_1 und n_2 die Gruppengrößen.
- Falls $n_1 = n_2$ ergibt sich die vereinfachte Formel:

$$s_p^2 = \frac{s_1^2+s_2^2}{2}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Es soll überprüft werden, ob Frühaufsteher oder Spätaufsteher belastbarer sind (zweiseitiger Test, $\alpha = 0{,}05$).
- Neun Früh- und neun Spätaufsteher wurden untersucht.
- Rohdaten:

Früh	Spät
86	97
91	87
96	113
103	93
121	115
86	108
121	123
105	124
112	93

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

Erster Schritt: Prüfgröße berechnen

- Man erhält: $\bar{x}_1 = 102{,}333$ und $\bar{x}_2 = 105{,}889$ sowie $s_1^2 = 187{,}500$ und $s_2^2 = 190{,}861$
- Da $n_1 = n_2$ ergibt sich mit der vereinfachten Formel: $s_p^2 = (187{,}500 + 190{,}861)/2 = 189{,}181$
- Durch Einsetzen dieses Wertes ergibt sich:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{189{,}181 \cdot \left(\frac{1}{9} + \frac{1}{9}\right)} = 6{,}484$$

- Also ergibt sich als Wert der Prüfgröße:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{102{,}333 - 105{,}889}{6{,}484} = -0{,}548$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

Zweiter Schritt: Kritischen Wert aus der Tabelle ablesen
Fall c) $t < t_{\alpha/2}(n_1 + n_2 - 2)$ oder $t > t_{1-\alpha/2}(n_1 + n_2 - 2)$ bzw. $|\,t\,| > t_{1-\alpha/2}(n_1 + n_2 - 2)$

- Anzahl der Freiheitsgrade: $n_1 + n_2 - 2 = 9 + 9 - 2 = 16$.
- Der kritische t-Wert lautet also:
  $t_{1-\alpha/2}(n_1 + n_2 - 2) = t_{0.975}(16) = 2{,}120$.

Kontrolle:

- Da $|-0{,}548| = 0{,}548 \not> 2{,}12$ wird die Nullhypothese nicht abgeleht.
- Es lässt sich also kein signifikanter Unterschied in der Belastbarkeit zwischen Früh- und Spätaufstehern nachweisen.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Zum Vergleich derselbe Test mit dem Statistik-Programm R berechnet:

```
Two Sample t-test
data:  frueh and spaet
t = -0.5484, df = 16, p-value = 0.591
alternative hypothesis: true difference in means
is not equal to 0
...
sample estimates:
mean of x mean of y
 102.3333  105.8889
```

Der p-Wert ist nicht kleiner als 5%, d.h. die Nullhypothese muss beibehalten werden.

Annahmen des t-Test für unabhängigen Stichproben

- Unabhängige Stichproben.
- Die beiden Populationsvarianzen sind gleich, d. h. $\sigma_1^2 = \sigma_2^2$.
- Die Rohwerte in beiden Populationen, aus denen die Stichproben gezogen werden, sind normalverteilt.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

t-Test für zwei verbundene Stichproben

Wo sind wir?

1. Deskriptive Statistik

2. Wahrscheinlichkeitstheorie

3. Tests und Konfidenzintervalle
3.1 Einführung
3.2 Power und Effektstärke
3.3 Tests und Konfidenzintervalle für Mittelwerte
Ein-Stichproben t-Test
t-Test für zwei unabhängige Stichproben
t-Test für zwei verbundene Stichproben
3.4 Nicht-parametrische Tests
3.5 χ^2-Unabhängigkeitstest

4. Kovarianz und Korrelation

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

t-Test für verbundene Stichproben (Beobachtungspaare)

- Es werden Beobachtungspaare untersucht, wobei eine Beobachtung in die erste und die andere in die zweite Gruppe gehört.
- Beispiele für Beobachtungspaare:
  - Messwiederholung: Messung an derselben Person vor und nach der Behandlung (im Experiment: within-subjects Design statt between-subjects Design)
  - Testergebnisse für linkes und rechtes Auge derselben Person
  - Ehepaare, Geschwister
  - Matched Pairs (in Fall-Kontroll-Studien)
- Es ist davon auszugehen, dass die Beobachtungen eines Paares sich ähnlicher sind als Beobachtungen, die kein Paar bilden.
- Berücksichtigung des „Ausgangsniveaus".

- Die Hypothesen lauten wie beim t-Test für unabhängige Stichproben:

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 > \mu_2 \qquad \text{(einseitig)}$$
$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 < \mu_2 \qquad \text{(einseitig)}$$
$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2 \qquad \text{(zweiseitig)}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Durch die paarweise Zuordnung der Beobachtungen ist es jetzt aber möglich, die Differenz d_i für jedes Beobachtungspaar (z.B. Messung vor und nach Treatment) zu bilden:

$$d_i = x_{i1} - x_{i2}$$

- Die Nullhypothese gleicher Mittelwerte entspricht dann Differenzen mit Mittelwert 0:

$$\begin{aligned} H_0 : \mu_d = 0, \quad & H_1 : \mu_d > 0 && \text{(einseitig)} \\ H_0 : \mu_d = 0, \quad & H_1 : \mu_d < 0 && \text{(einseitig)} \\ H_0 : \mu_d = 0, \quad & H_1 : \mu_d \neq 0 && \text{(zweiseitig)} \end{aligned}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Prüfgröße

- Die aus den Differenzen bestimmte Prüfgröße lautet entsprechend:

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \sqrt{n}\left(\frac{\bar{d}}{s_d}\right)$$

- Die Standardabweichung der Differenzen wird dabei in der üblichen Form geschätzt:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Es soll überprüft werden, ob Examenskandidaten in der Lage sind, ihre eigene Leistungsfähigkeit richtig einzuschätzen.
- Vor Durchführung einer Klausur mit 70 Aufgaben sollen 15 Kandidaten angeben, wie viele Aufgaben sie glauben richtig lösen zu können.
- Die Anzahl der tatsächlich richtig gelösten Aufgaben wird mit der eingeschätzten Anzahl durch einen t-Test für Beobachtungspaare verglichen.
- Wir bezeichnen die eingeschätzte Anzahl mit x_{i1} und die Anzahl tatsächlich richtig gelöster Aufgaben mit x_{i2}.
- Wir gehen davon aus, dass die Differenzen zwischen den eingeschätzten und den tatsächlichen Leistungen normalverteilt sind.
- Die H_1 wird ungerichtet formuliert ($\alpha = 0{,}05$).

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

Vpn	x_{i1}	x_{i2}	d_i
1	40	48	-8
2	60	55	5
3	30	44	-14
4	55	59	-4
5	55	70	-15
6	35	36	-1
7	30	44	-14
8	35	28	7
9	40	39	1
10	35	50	-15
11	50	64	-14
12	25	22	3
13	10	19	-9
14	40	53	-13
15	55	60	-5

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

Erster Schritt: Prüfgröße berechnen

- Das Mittel der 15 Differenzen lautet
  $\bar{d} = \sum_i d_i / n = -96/15 = -6{,}4$
- Für die Standardabweichung der Differenzwerte berechnen wir
  $s_d = 7{,}854$
- Somit ergibt sich als Prüfgröße
  $t = \sqrt{15} \cdot \bar{d} / s_d = \sqrt{15} \cdot (-6{,}4/7{,}854) = -3{,}156$

Zweiter Schritt: Kritischen Wert aus der Tabelle ablesen
Fall c) $t < t_{\alpha/2}(n-1)$ oder $t > t_{1-\alpha/2}(n-1)$ bzw.
$|\,t\,| > t_{1-\alpha/2}(n-1)$

- Anzahl der Freiheitsgrade: $n - 1 = 14$

  n ist die Anzahl der <u>Differenzen</u> =
  Anzahl der Beobachtungs<u>paare</u> =
  z.B. Anzahl der an der Studie beteiligten Ehepaare bzw.
  Anzahl der Versuchspersonen bei Messwiederholungen

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Kritischer Wert: $t_{1-\alpha/2}(n-1) = t_{0.975}(14) = 2{,}145$

Kontrolle:

- $|-3{,}156| = 3{,}156 > 2{,}145 \Rightarrow H_0$ ablehnen
- Es besteht ein signifikanter Unterschied zwischen eingeschätzten und den tatsächlichen Leistungen.
- (An den Mittelwerten kann man deskriptiv ablesen, dass die Personen ihre Leistung im Mittel unterschätzen.)

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Annahmen des t-Test für verbundene Stichproben

- Für die Differenzen gelten die gleichen Annahmen wie beim Ein-Stichproben t-Test:
- Um den Test durchführen zu können, muss eine Stichprobe von Beobachtungspaaren vorliegen, deren Differenzen d_i normalverteilt sind.
- Die wahre Varianz der Differenzwerte, σ_d^2, braucht für die Testdurchführung nicht bekannt zu sein, da sie durch s_d^2 geschätzt werden kann.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Große Stichproben

- Alle drei besprochenen t-Tests lassen sich auch bei kleinem Stichprobenumfang anwenden, wenn ihre Annahmen – insbesondere die Annahme der Normalverteilung der Werte in den Gruppen – erfüllt sind.
- Wenn der Stichprobenumfang ausreichend groß ist, halten die t-Tests das festgelegte Signifikanzniveau aber auch dann ein, wenn die Werte in den Gruppen nicht normalverteilt sind.
- Natürlich sind große Stichprobenumfänge ohnehin wünschenswert, da dann die Power höher ist.
- Als grober Orientierungspunkt wird oft angegeben, dass in der Praxis mehr als 30 Beobachtungen pro Gruppe vorliegen sollten, um t-Tests zuverlässig einsetzen zu können.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Nicht-parametrische Tests

Parametrische und nicht-parametrische Tests

- Parametrische Tests machen bestimmte Verteilungsannahmen. Beispielsweise gehen t-Tests von normalverteilten Werten in den Gruppen aus.
- Nicht-parametrische Tests werden auch als „verteilungsfrei" bezeichnet, weil sie in der Regel keine Annahmen über die Verteilung der Werte machen.
- Nicht-parametrische Tests können eingesetzt werden, wenn die Verteilungsannahmen der parametrischen Tests verletzt sind.
- Wie beim t-Test gibt es die Varianten:
  - für zwei unabhängige Stichproben: Wilcoxon-Rangsummentest (äquivalent: Mann-Whitney U-Test)
  - für zwei verbundene Stichproben: Wilcoxon-Test für verbundene Stichproben

Beispiel

- Verlangsamt ein Medikament die Reaktionszeiten der Versuchspersonen?
- Diese Alternativhypothese soll auf dem 5% Signifikanzniveau getestet werden.
- An einem Reaktionszeit-Messgerät werden dazu 10 Personen, die das Medikament eingenommen haben (Gruppe 1), und 10 Personen, die einen Placebo erhalten haben (Gruppe 2), getestet.
- Da Reaktionszeiten häufig nicht normalverteilt sind, soll ein nicht-parametrisches Verfahren verwendet werden.
- zwei unabhängige Stichproben $\Rightarrow$ Wilcoxon-Rangsummentest

Wilcoxon-Rangsummentest

- Idee: Anstatt die Werte der Personen direkt zu vergleichen wird nur die Reihenfolge der Werte berücksichtigt.
- Beispiel: Haben die Personen, die das Medikament bekommen haben, längere Reaktionszeiten? Dann haben ihre Reaktionszeiten die höheren Ränge $\Rightarrow$ höhere Rangsumme.
- Der Rang ist die Platzierung eines Wertes, wenn man alle Werte (aus beiden Gruppen) in aufsteigender Reihenfolge sortiert.

Beispiel

Mit Medikament		Mit Placebo	
Reaktionszeit (ms)	Rangplatz	Reaktionszeit (ms)	Rangplatz
85	3	96	8
106	13	108	15
118	18	86	4
138	20	84	2
90	6	99	10
112	17	101	11
119	19	78	1
107	14	97	9
95	7	87	5
103	12	109	16
	Rangsumme $T_1 = 129$		Rangsumme $T_2 = 81$

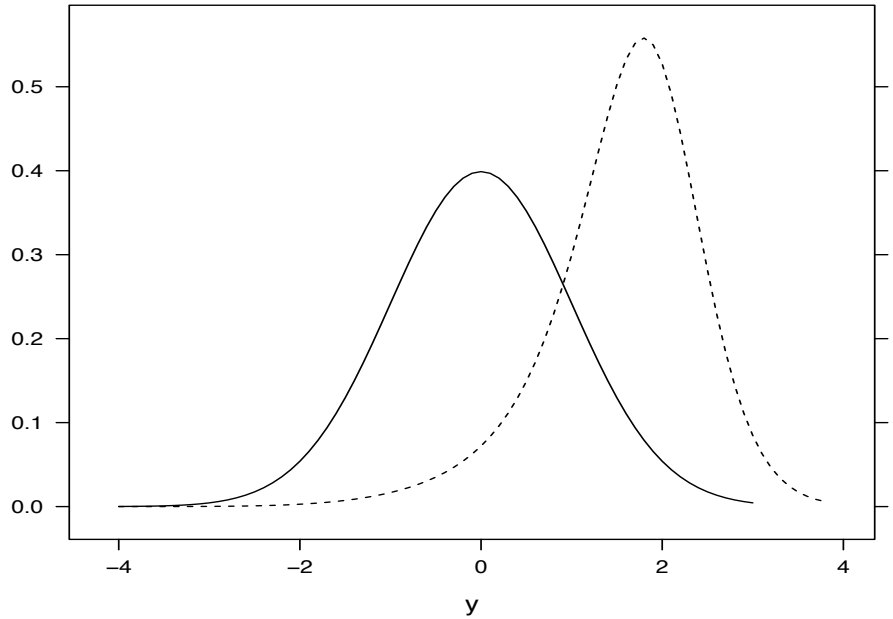
Beispiel

- Aus den Rang-Summen kann die Prüfgröße W (entspricht der Mann-Whitney U-Statistik) berechnet werden.
- Aus deren exakter Verteilung (oder mithilfe einer Normalverteilungs-Approximation) lassen sich p-Werte bestimmen.
- Für das Beispiel ergibt sich:

```
Wilcoxon rank sum test
...
W = 74, p-value = 0.03763
alternative hypothesis: true location shift
is greater than 0
```

- Der p-Wert ist kleiner als 5%, d.h. die Nullhypothese kann abgelehnt werden.
- Das Medikament verlangsamt die Reaktionszeiten signifikant.

- Sind die Annahmen des t-Tests erfüllt, hat der Wilcoxon-Rangsummentest eine etwas geringere Power als der t-Test, ansonsten eine höhere.
- Der t-Test ist sensitiv gegen eine reine Mittelwertsverschiebung (bei gleicher Varianz und Normalverteilung in beiden Gruppen).
- Der Wilcoxon-Rangsummentest ist sensitiv gegen eine Verschiebung der gesamten Verteilung in Lage und Form.



χ^2-Unabhängigkeitstest

Wo sind wir?

1. Deskriptive Statistik

2. Wahrscheinlichkeitstheorie

3. Tests und Konfidenzintervalle
3.1 Einführung
3.2 Power und Effektstärke
3.3 Tests und Konfidenzintervalle für Mittelwerte
3.4 Nicht-parametrische Tests
3.5 χ^2-Unabhängigkeitstest

4. Kovarianz und Korrelation

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

χ^2 Unabhängigkeitstest

- Werden zwei nominalskalierte Variablen betrachtet, interessiert oft ob ein Zusammenhang zwischen den Variablen besteht.
- Wir bezeichnen die Variablen mit A und B und bezeichnen die Kategorien jeder Variable mit A_i bzw. B_j.
- Zur Erinnerung: Bei Unabhängigkeit gilt

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

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- H_0: Die beiden Variablen sind unabhängig.
- H_1: Die beiden Variablen sind nicht unabhängig, d.h. es besteht ein Zusammenhang zwischen A und B.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

Sind die Variablen Sport (Person treibt regelmässig Sport) und Vegan (Person ernährt sich vegan) unabhängig voneinander?

beobachtete Häufigkeiten n_{ij}:

		Vegan		
		ja	nein	
Sport	ja	35	15	50
	nein	25	25	50
		60	40	100

Punktnotation

- Häufigkeiten am Beispiel einer 2×2-Tabelle:

n_{11}	n_{12}	$n_{1.}$
n_{21}	n_{22}	$n_{2.}$
$n_{.1}$	$n_{.2}$	n

- Randhäufigkeiten der Zeilen:
  - $n_{1.} = n_{11} + n_{12} = \sum_j n_{1j}$
  - $n_{2.} = n_{21} + n_{22} = \sum_j n_{2j}$
- Randhäufigkeiten der Spalten:
  - $n_{.1} = n_{11} + n_{21} = \sum_i n_{i1}$
  - $n_{.2} = n_{12} + n_{22} = \sum_i n_{i2}$
- Stichprobenumfang: $n = n_{11} + n_{12} + n_{21} + n_{22} = \sum_i \sum_j n_{ij}$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Erwartete Häufigkeiten

- Bei Unabhängigkeit würden sich die Wahrscheinlichkeiten für jede Zelle als Produkt der Randwahrscheinlichkeiten ergeben.
- Es gilt also für die unter Unabhängigkeit erwarteten Häufigkeiten:

$$m_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

Sind die Variablen Sport und Vegan unabhängig voneinander?

unter Unabhängigkeit erwartete Häufigkeiten m_{ij}:

		Vegan		
		ja	nein	
Sport	ja	$50 \cdot 60/100$	$50 \cdot 40/100$	50
	nein	$50 \cdot 60/100$	$50 \cdot 40/100$	50
		60	40	100

ergibt:

		Vegan		
		ja	nein	
Sport	ja	30	20	50
	nein	30	20	50
		60	40	100

Prüfgröße

- Zur Berechnung der Prüfgröße werden die in der Stichprobe beobachteten Häufigkeiten n_{ij} mit den unter Unabhängigkeit erwarteten Häufigkeiten m_{ij} verglichen:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

- Ist die Summe der quadrierten Abweichungen groß, weicht die gemeinsame Verteilung der beiden Variablen stark von der unter Unabhängigkeit erwarteten Verteilung ab.
- Entsprechend ist der Ablehnbereich der Nullhypothese:

$$\chi^2 > \chi^2_{1-\alpha}((k-1) \cdot (l-1))$$

wobei: k = Anzahl Zeilen
l = Anzahl Spalten

Beispiel

- Erster Schritt: Prüfgröße berechnen:

$$\chi^2 = \frac{(35-30)^2}{30} + \frac{(15-20)^2}{20} + \frac{(25-30)^2}{30} + \frac{(25-20)^2}{20} = 4{,}167$$

- Zweiter Schritt: Kritischer Wert aus der χ^2-Tabelle:

$$\chi^2_{0.95}\left((2-1)\cdot(2-1)\right) = \chi^2_{0.95}(1) = 3{,}84$$

- Kontrolle: $4{,}167 > 3{,}84 \Rightarrow$ Nullhypothese ablehnen!
- Es liegt ein signifikanter Zusammenhang zwischen regelmässig Sport treiben und veganer Ernährung vor.

Annahmen des χ^2-Test

- Die Beobachtungen stammen aus einer einfachen Zufallsstichprobe.
- Alle *erwarteten* Häufigkeiten m_{ij} sind größer als 5.
- Wenn letzteres nicht erfüllt ist, ist die asymptotische χ^2-Verteilung der Prüfgröße nicht mehr gewährleistet. In diesem Fall sollte Fishers exakter Test angewendet werden, der auf einer Permutation der Zellenhäufigkeiten bei gegebenen Randhäufigkeiten basiert.

Vereinfachte Prüfgröße für 2×2-Tafeln

- In 4-Felder-Tafeln lässt sich die Prüfgröße auch mit einer einfacheren Formel berechnen:

$$\chi^2 = \frac{n \cdot (ad - bc)^2}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}$$

mit den Bezeichnungen:

a	b	a+b
c	d	c+d
a+c	b+d	n

Vereinfachte Prüfgröße für 2×2-Tafeln

- Für unser Beispiel:

$$\chi^2 = \frac{100 \cdot (35 \cdot 25 - 15 \cdot 25)^2}{50 \cdot 50 \cdot 60 \cdot 40} = 4{,}167$$

Es ergibt sich also auch mit der einfacheren Formel derselbe Wert für die Prüfgröße und dieselbe Testentscheidung.

- Neben χ^2 (Wertebereich: 0 bis ∞) gibt es noch diverse weitere Zusammenhangs-Maße für nominalskalierte Merkmale.
- χ^2 drückt nur die Stärke des Zusammenhangs aus, einige andere Maße auch die Richtung des Zusammenhangs,
- z.B. der Phi-Koeffizient (Wertebereich: -1 bis 1):

$$\phi = \frac{ad - bc}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}}$$

- Ein weiteres bekanntes Zusammenhangs-Maß für 2×2-Tafeln ist der Odds Ratio.

Beispiel

		Vegan		
		ja	nein	
Sport	ja	35	15	50
	nein	25	25	50
		60	40	100

$$\chi^2 = 4{,}167$$

zeigt Zusammenhang zwischen regelmässig Sport treiben und veganer Ernährung.

$$\phi = \frac{35 \cdot 25 - 15 \cdot 25}{\sqrt{50 \cdot 50 \cdot 60 \cdot 40}} = 0.204$$

zeigt zusätzlich die Richtung des Zusammenhangs, nämlich dass Personen, die regelmässig Sport treiben, sich im Verhältnis <u>häufiger</u> vegan ernähren.

Kovarianz und Korrelation

Wo sind wir?

1. Deskriptive Statistik

2. Wahrscheinlichkeitstheorie

3. Tests und Konfidenzintervalle

4. Kovarianz und Korrelation

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

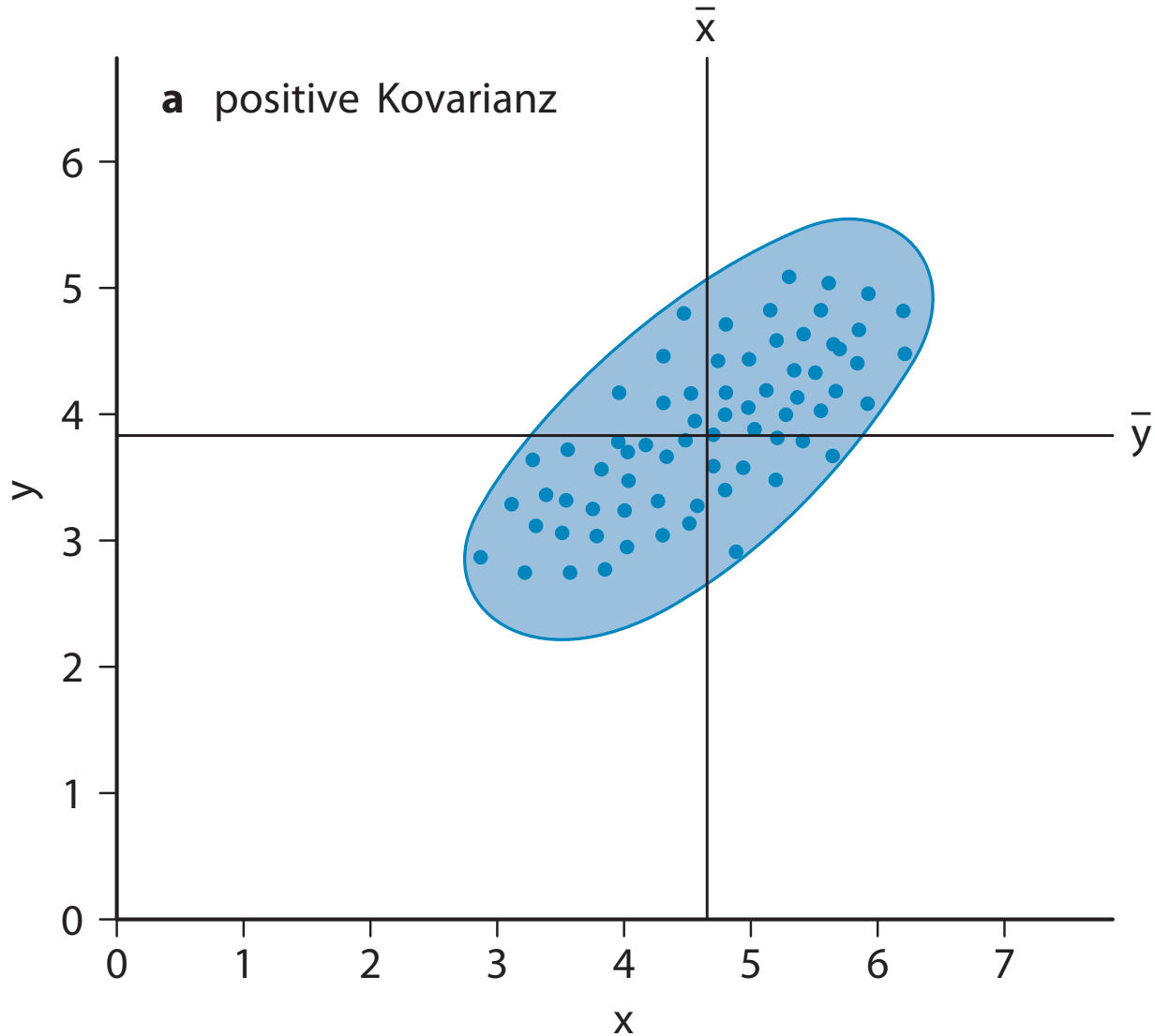
Kovarianz

- Die Stichproben-Varianz s_x^2 beschreibt die Streuung der Werte in Richtung einer Variablen X.
- Die Stichproben-Kovarianz s_{xy} beschreibt die Streuung der Werte in Richtung zweier Variablen X und Y.
- Sie charakterisiert den linearen Zusammenhang zwischen zwei metrischen Variablen.

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

- Die Kovarianz ist symmetrisch, d. h.: $s_{xy} = s_{yx}$

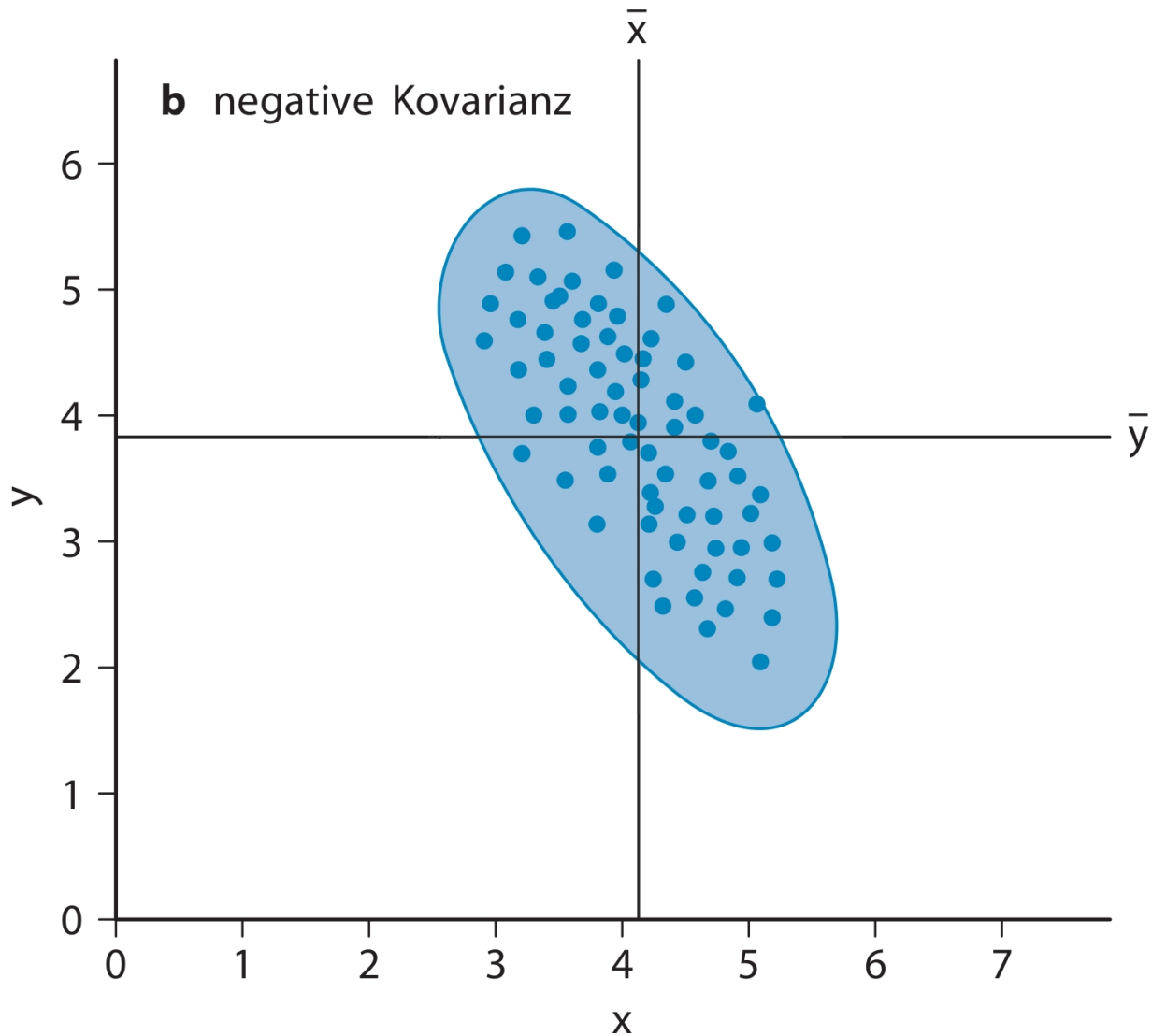
Positive Kovarianz



inhaltliches Beispiel: X = Lernzeit, Y = Punktezahl in Klausur

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

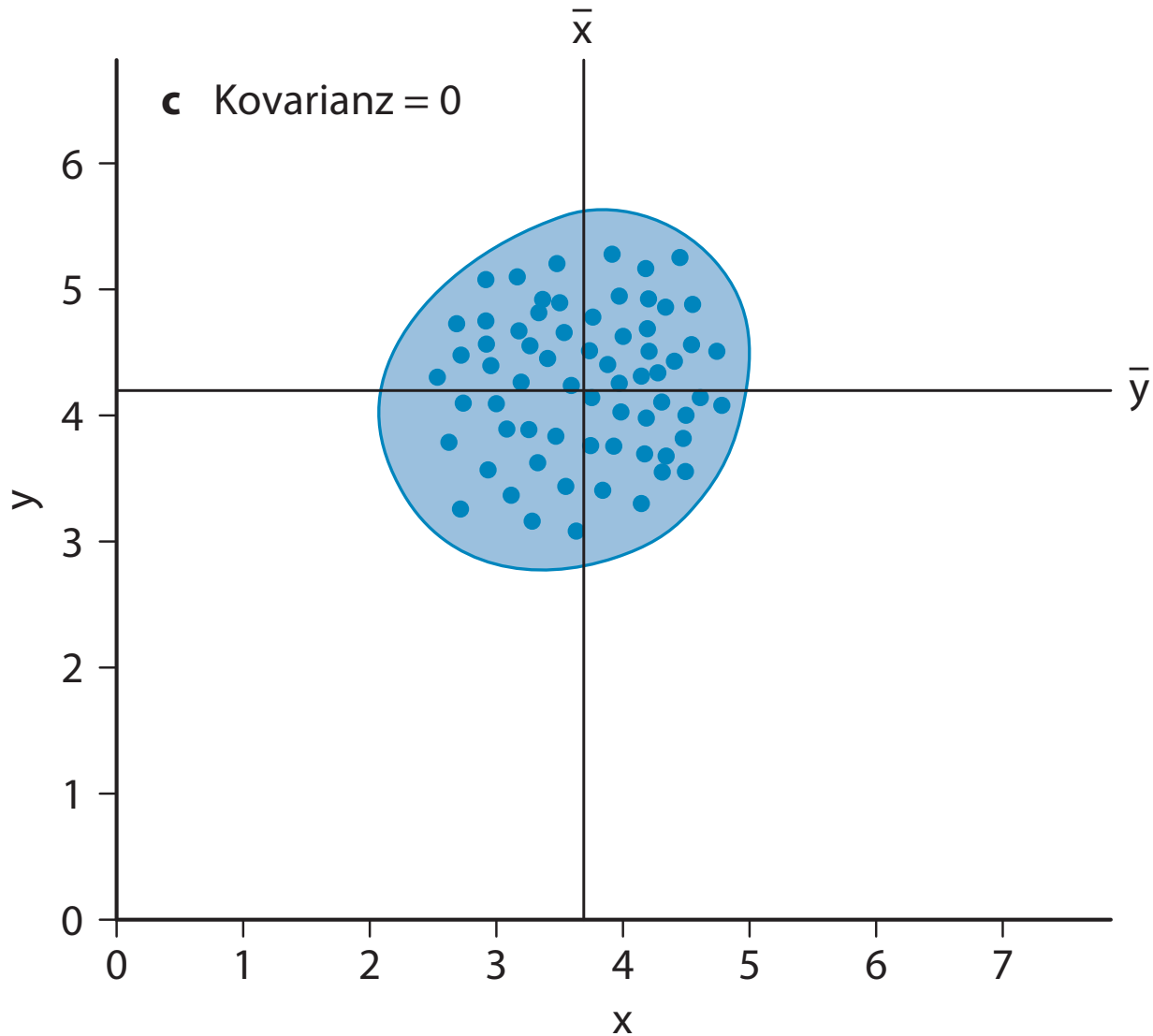
Negative Kovarianz



inhaltliches Beispiel: X = Dosis Medikament, Y = Blutdruck

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Unabhängige Variablen



inhaltliches Beispiel: X = Schuhgröße, Y = Punktezahl in Klausur

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Eigenschaften der Kovarianz

- Die Kovarianz ist von den Messeinheiten beider Variablen abhängig.
- Sie ist deshalb zur Beschreibung des Zusammenhangs unterschiedlicher Merkmale wenig geeignet.
- Für lineare Transformationen der Form $u = a + b \cdot x$ gilt

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

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Es wurde deshalb ein weiteres Maß zur Kennzeichnung von Zusammenhängen entwickelt, das gegenüber Maßstabsveränderungen der untersuchten Merkmale invariant ist, der *Korrelationskoeffizient* r:

$$r = \frac{s_{xy}}{s_x \cdot s_y}$$

- Der Korrelationskoeffizient r wird manchmal auch *Korrelationskoeffizient nach Bravais-Pearson* oder *Produkt-Moment-Korrelation* genannt.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Die Korrelation kann auch mit Hilfe von z-Werten dargestellt werden:
- Die Korrelation zweier Variablen entspricht der Kovarianz der z-standardisierten Variablen (denn deren Standardabweichung ist 1).

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Wertebereich

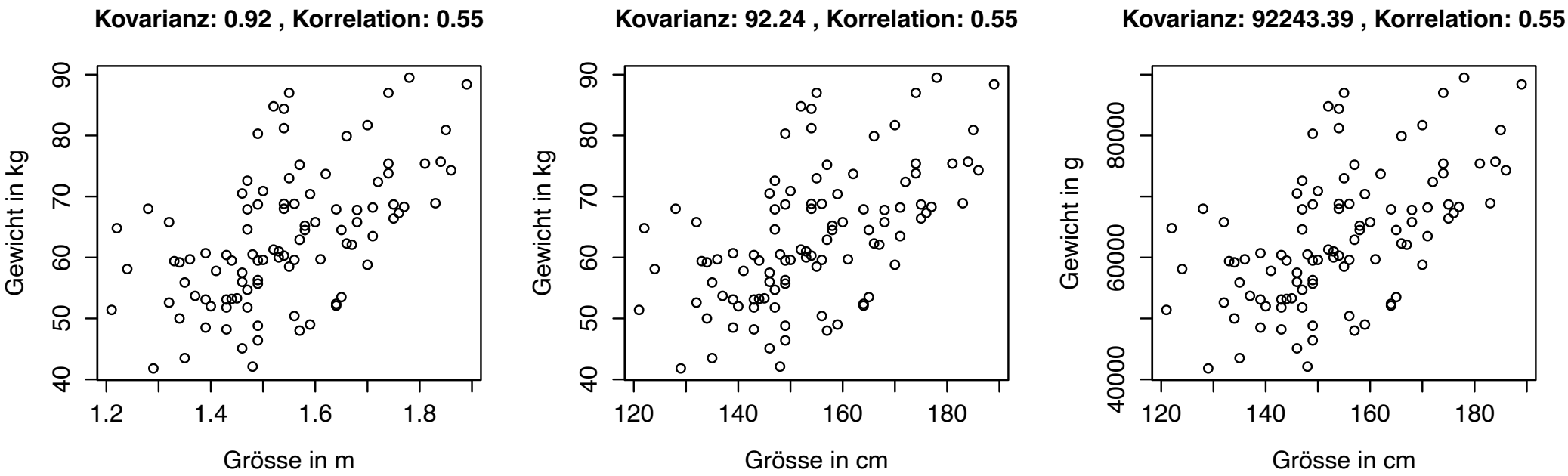
- Der Wertebereich des Korrelationskoeffizienten ist das Intervall von $-1{,}0$ bis $+1{,}0$.
- Eine Korrelation von $+1{,}0$ entspricht einem perfekten positiven Zusammenhang.
- Eine Korrelation von $-1{,}0$ entspricht einem perfekten negativen Zusammenhang.
- Ist die Korrelation null, besteht kein linearer Zusammenhang.
- Die Korrelation ist invariant gegenüber linearen Transformationen der Merkmale, d. h. falls $u = a + b \cdot x$, dann gilt:

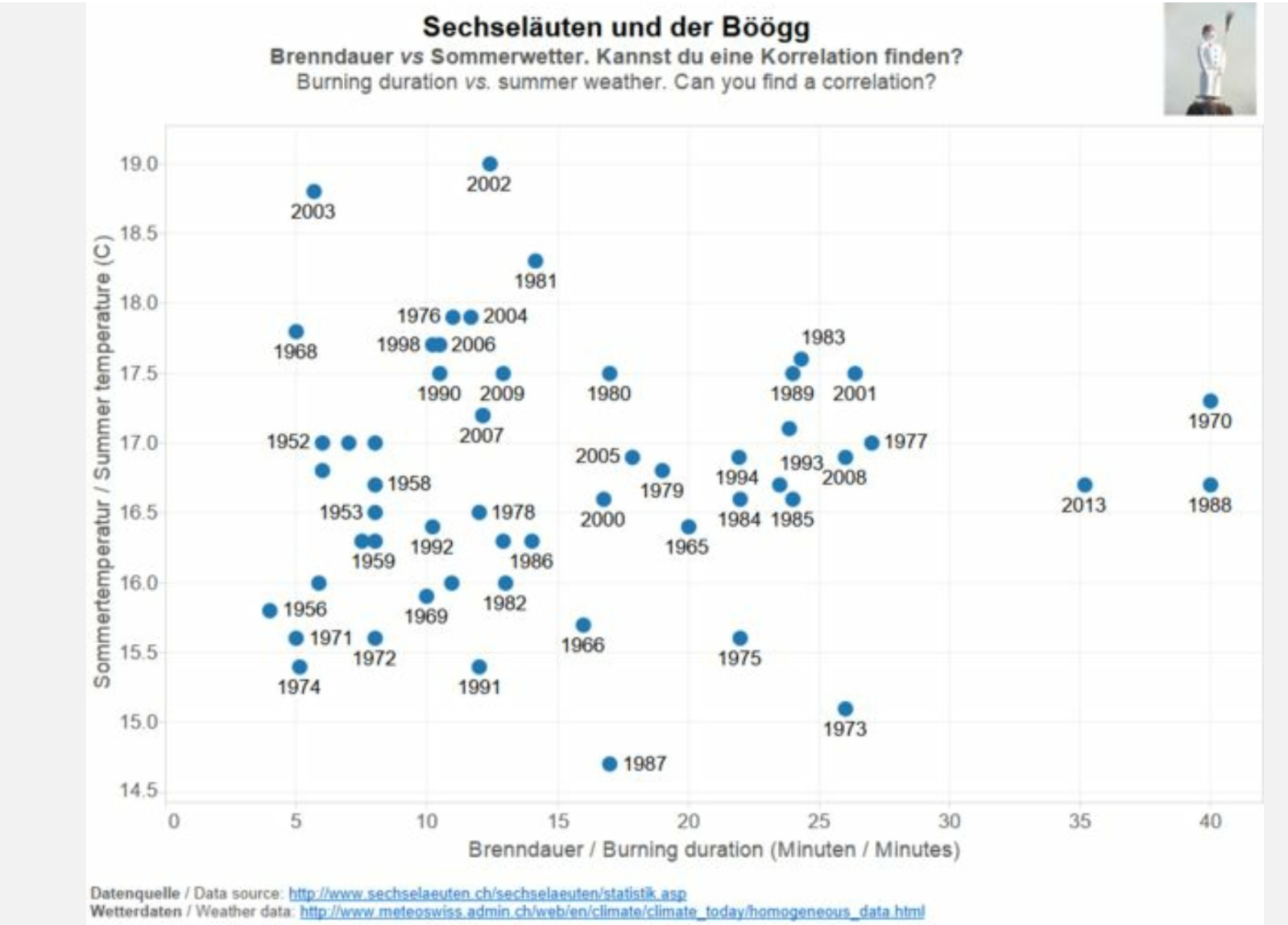
$$r_{uy} = r_{xy}$$

- Die Korrelation ist wie die Kovarianz symmetrisch, d. h.: $r_{xy} = r_{yx}$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Kovarianz vs. Korrelation





Korrelation: 0,08

Quellen: Neue Zürcher Zeitung, `http://www.meteoschweiz.ch`, 2014

- Sind zwei Variablen voneinander unabhängig, ist die Kovarianz sowie die Korrelation zwischen ihnen null.
- Aus einer Kovarianz bzw. Korrelation von null kann man aber umgekehrt nicht einfach die Unabhängigkeit der Variablen folgern, weil der Korrelationskoeffizient nur lineare Zusammenhänge aufdecken kann. Bei einer Korrelation von null kann deshalb immernoch ein nicht-linearer Zusammenhang vorliegen.

- Zur Berechnung von Hand kann auch folgende Formel verwendet werden:

$$r = \frac{n \cdot \sum_i x_i \cdot y_i - \left(\sum_i x_i\right) \cdot \left(\sum_i y_i\right)}{\sqrt{\left[n \cdot \sum_i x_i^2 - \left(\sum_i x_i\right)^2\right] \cdot \left[n \cdot \sum_i y_i^2 - \left(\sum_i y_i\right)^2\right]}}$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- Wir berechnen die folgenden Summen für die x und y Variablen:

	x	y	x^2	y^2	$x \cdot y$
	2	1	4	1	2
	1	2	1	4	2
	9	6	81	36	54
	5	4	25	16	20
	3	2	9	4	6
Summen:	20	15	120	61	84

- Nun können wir die Summen bzw. ihre Quadrate in die Formel für die Korrelation einsetzen.
- Man erhält:

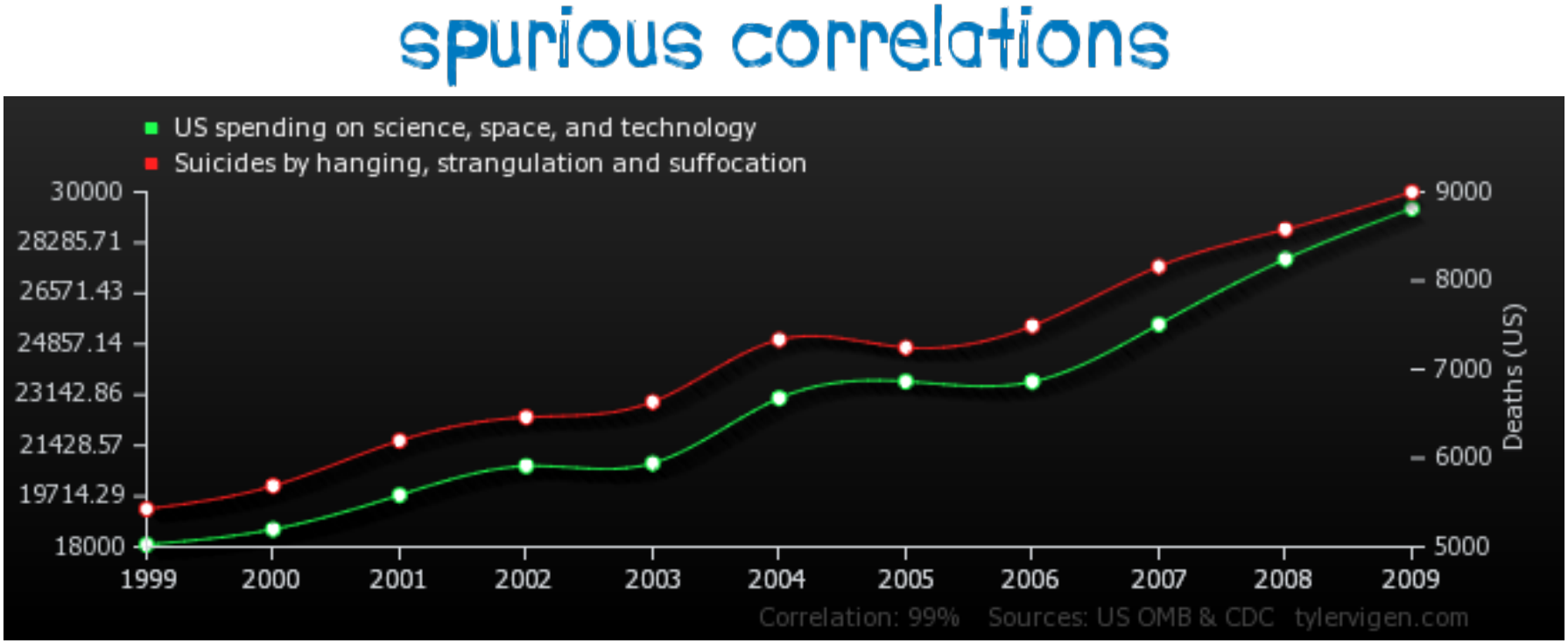
$$r = \frac{5 \cdot 84 - 20 \cdot 15}{\sqrt{(5 \cdot 120 - 400) \cdot (5 \cdot 61 - 225)}} = 0{,}949$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010). Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

- Vom Vorliegen einer Korrelation in einer Beobachtungsstudie kann man nicht auf einen kausalen Zusammenhang schließen!
- Aus zwei Gründen:
  - Die Richtung des Zusammenhangs – was ist die Ursache, was die Wirkung – lässt sich nicht an der Korrelation ablesen. ($\Rightarrow$ längsschnittliches Design)
  - Drittvariablen (sog. konfundierte Variablen) können den Zusammenhang verursachen. ($\Rightarrow$ randomisiertes, kontrolliertes Experiment)

Beispiele aus der Zeitung

- Menschen mit einer bestimmten Hirnstruktur sind häufiger alkoholabhängig
  - naive Interpretation: aufgrund der Hirnstruktur werden sie alkoholabhängig
  - auch möglich: aufgrund des Alkoholmissbrauchs verändert sich die Hirnstruktur
- Kinder, die oft vorgelesen bekommen, kriegen als Erwachsene bessere Jobs
  - naive Interpretation: Vorlesen macht erfolgreich
  - Konfundierung: Kinder, deren Eltern hohen Bildungsstand haben, bekommen mehr vorgelesen aber erhalten auch bessere Schulbildung, Nachhilfe, Klavierunterricht, Sprachreisen etc., die sich wiederum positiv auf die Jobchancen auswirken



Korrelation: 0.992082

Quelle: `http://www.tylervigen.com/`, 2014

Ein-Stichprobentest für $H_0 : \rho = 0$

- Wir wollen die Nullhypothese $H_0 : \rho = 0$ überprüfen, d.h. dass in der Population kein linearer Zusammenhang zwischen den Variablen besteht.
- Ob eine empirisch ermittelte Korrelation r mit der $H_0 : \rho = 0$ zu vereinbaren ist, lässt sich mit folgender Prüfgröße testen:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

  mit $df = n - 2$
- Hypothesen:
  $H_1 : \rho > 0$
  $H_1 : \rho < 0$
  $H_1 : \rho \neq 0$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Beispiel

- In einer Untersuchung mit $n = 18$ Versuchspersonen ergibt sich zwischen den Variablen „Umfang des Wortschatzes" und „Erreichte Punktzahl in einem Rechtschreib-Test" eine Korrelation von $r = 0{,}62$.
- Die zweiseitige Alternative $\rho \neq 0$ soll auf einem Signifikanzniveau von 1% geprüft werden.
- Erster Schritt: Prüfgröße berechnen:

$$t = \frac{0{,}62 \cdot \sqrt{18-2}}{\sqrt{1-0{,}62^2}} = 3{,}161$$

- Zweiter Schritt: Kritischer Wert aus der t-Tabelle
  Fall c) $t < t_{\alpha/2}(n-2)$ oder $t > t_{1-\alpha/2}(n-2)$ bzw.
  $|\,t\,| > t_{1-\alpha/2}(n-2)$
  $t_{0.995}(16) = 2{,}921$
- Kontrolle: $3{,}161 > 2{,}921 \Rightarrow$ Nullhypothese ablehnen!
- Es liegt ein signifikanter Zusammenhang zwischen Wortschatz und Rechtschreib-Test vor.

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Wiederholung: Signifikanz vs. Effektstärke

Wiederholung: Signifikanz vs. Effektstärke

- Wenn ein Zusammenhang signifikant ist, bedeutet das nicht automatisch, dass es sich um einen starken Zusammenhang handelt.
- Wie bereits besprochen, hängt die Power eines statistischen Tests nicht nur von der Effektstärke (z.B. Größe des Unterschieds, Stärke des Zusammenhangs) ab, sondern vor allem auch von der Stichprobengröße.
- Größere Stichproben erlauben bessere Schätzung $\Rightarrow$ Tests haben höhere Power.

Unterscheidung

- deskriptive (beschreibende) Statistik

  Ziel: Daten aus einer Stichprobe beschreiben

  Wert des Korrelationskoeffizienten aus einer Stichprobe, z.B. $r = 0{,}8$

- induktive (schließende) Statistik

  Ziel: von einer Stichprobe auf (andere Stichproben aus der) Grundgesamtheit schließen

  Kann ein Wert dieser Größe nur durch die Zusammensetzung der Zufalls-Stichprobe zustande kommen, auch wenn in der Grundgesamtheit gar kein Zusammenhang vorliegt? D.h.: $H_0 : \rho = 0$

- signifikant = „überzufällig"

  D.h. wenn in der Grundgesamtheit gar kein Zusammenhang vorliegen würde, wäre es sehr unwahrscheinlich*, dass ein Wert dieser Größe nur durch die Zusammensetzung der Zufalls-Stichprobe zustande kommt.

  Man würde also erwarten, auch in den meisten anderen Stichproben aus derselben Grundgesamtheit einen Zusammenhang zu finden.

- signifikant heißt aber nicht notwendigerweise: starker, inhaltlich relevanter Zusammenhang

  * d.h. Wahrscheinlichkeit kleiner als Signifikanzniveau α

Signifikanz vs. Effektstärke

zweiseitiger Test für $H_0 : \rho = 0$ auf 5%-Niveau

- Stärke des Zusammenhangs: $r = 0{,}8$
  Stichprobengröße: $n = 100$
  $\Rightarrow$ Zusammenhang signifikant
- Stärke des Zusammenhangs: $r = 0{,}8$
  Stichprobengröße: $n = 5$
  $\Rightarrow$ Zusammenhang nicht signifikant
- Stärke des Zusammenhangs: $r = 0{,}08$
  Stichprobengröße: $n = 100$
  $\Rightarrow$ Zusammenhang nicht signifikant
- Stärke des Zusammenhangs: $r = 0{,}08$
  Stichprobengröße: $n = 5000$
  $\Rightarrow$ Zusammenhang signifikant

Rangkorrelation

Rang-Korrelation nach Spearman

Der Korrelationskoeffizient nach Bravais-Pearson r:

- setzt ein metrisches Skalenniveau voraus,
- unterschätzt den Zusammenhang von x und y, wenn dieser nicht-linear ist,
- reagiert empfindlich auf Ausreißer.

Der Rang-Korrelationskoeffizient nach Spearman r_s

- wird auch bei ordinalskalierten Daten angewendet,
- spricht auch auf nicht-lineare monotone Zusammenhänge an,
- ist robust gegen Ausreißer.

$r_s = +1$ bedeutet, dass y mit steigendem x streng monoton steigt.
$r_s = -1$ bedeutet, dass y mit steigendem x streng monoton fällt.
$r_s = 0$ bedeutet, dass kein monotoner Zusammenhang besteht.

Rang-Korrelation nach Spearman

Die Rang-Korrelation r_s ist definiert als die Korrelation der Rangplätze der Messwerte, d.h.:

- Rang-Korrelationskoeffizient nach Spearman r_s = Korrelationskoeffizient nach Bravais-Pearson r angewendet auf die Ränge:

$$r_s = \frac{n \cdot \sum_i rg(x_i) \cdot rg(y_i) - \left(\sum_i rg(x_i) \right) \cdot \left(\sum_i rg(y_i) \right)}{\sqrt{\left[n \cdot \sum_i rg(x_i)^2 - \left(\sum_i rg(x_i) \right)^2 \right] \cdot \left[n \cdot \sum_i rg(y_i)^2 - \left(\sum_i rg(y_i) \right)^2 \right]}}$$

  Dabei ist $rg(x_i)$ der Rang von x_i und $rg(y_i)$ der Rang von y_i (die Ränge werden für die x-Werte und die y-Werte getrennt vergeben).

- Vereinfachte Berechnung (nur verwenden sofern keine Bindungen vorliegen):

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} (rg(x_i) - rg(y_i))^2}{n \cdot (n^2 - 1)}$$

Beispiel

Messwerte	x_i	33	15	17	11	40
Rangplätze	$rg(x_i)$	4	2	3	1	5

Messwerte	y_i	47	32	29	28	56
Rangplätze	$rg(y_i)$	4	3	2	1	5

Damit ergibt sich für die Rang-Korrelation:

$$r_s = 1 - \frac{6 \cdot [0^2 + (-1)^2 + 1^2 + 0^2 + 0^2]}{5 \cdot (25 - 1)} = 0.90$$

©2010, Springer-Verlag Berlin Heidelberg. Aus J. Bortz & C. Schuster (2010).
Statistik für Human- und Sozialwissenschaftler. Springer: Heidelberg. www.lehrbuch-psychologie.de

Bindungen

- Treten mehrere gleiche Werte auf, spricht man von *Bindungen* (“ties”).
- Die Rangvergabe bei Bindungen wird mit Durchschnittsbildung über die in Frage kommenden Ränge gelöst, d. h. es wird jedem der Werte der durchschnittliche Rang zugeordnet.
- Beispiel:

Messwerte	x_i	1.09	2.17	2.68	2.17	4.5	3.02
Hilfszählung		1	$\underline{2}$	4	$\underline{3}$	6	5
Rangplätze	$rg(x_i)$	1	2,5	4	2,5	6	5

Beim Vorliegen von Bindungen stimmt die vereinfachte Berechnung nur näherungsweise, deshalb sollte die ausführliche Formel verwendet werden.