Grundlagen sportwissenschaftlicher Forschung
Deskriptive Statistik 2
Inferenzstatistik 1

Dr. Jan-Peter Brückner

jpbrueckner@email.uni-kiel.de

R.216
Tel. 880 4717

Rückblick: Besonders wichtige Themen

Wissenschaftstheoretischer Hintergrund

- Unterschied wissenschaftliches Wissen – Alltagswissen
- Was ist eine Theorie? Qualitätskriterien von Theorien
- Was ist eine Hypothese? Hypothesenarten
- Grundlegende Aussagen des kritischen Rationalismus
- Kennzeichen und Abgrenzung von quantitativer und qualitativer Forschung

Forschungsablauf und Untersuchungsplanung

- Phasen der empirischen Forschung
- Forschungslogischer Ablauf
- Aufbau eines Forschungsberichts

  → **Wilhelm, A. (2009). *Leitfaden zum Anfertigung wissenschaftlicher Hausarbeiten. Inhaltliche und formale Vorgaben (2. Aufl.)*. Kiel: Institut für Sportwissenschaft**

- Grundlagen von Kausalschlüssen
- Untersuchungspläne: Korrelationsstudie – Quasiexperiment – Experiment (Kennzeichen und Abgrenzung)
- Interne und externe Validität bei verschiedenen Untersuchungsdesigns (s.o.) sowie Feld- und Laboruntersuchungen

Test

- Test, Messen und Skalenniveaus
- Testgütekriterien (Haupt- und Nebengütekriterien)
- Abgrenzung der Hauptgütekriterien
- Bestimmung der Hauptgütekriterien

Deskriptive Statistik

- Maße der zentralen Tendenz und Streuung (Voraussetzung für Anwendung, Verstehen und Interpretieren, Insbesondere: M, Md, SD, r)

Deskriptive Statistik 2

Grundlagen sportwissenschaftlicher Forschung
Deskriptive Statistik 2
Inferenzstatistik 1

Dr. Jan-Peter Brückner

jpbrueckner@email.uni-kiel.de

R.216
Tel. 880 4717

Deskriptive Statistik - Ziele

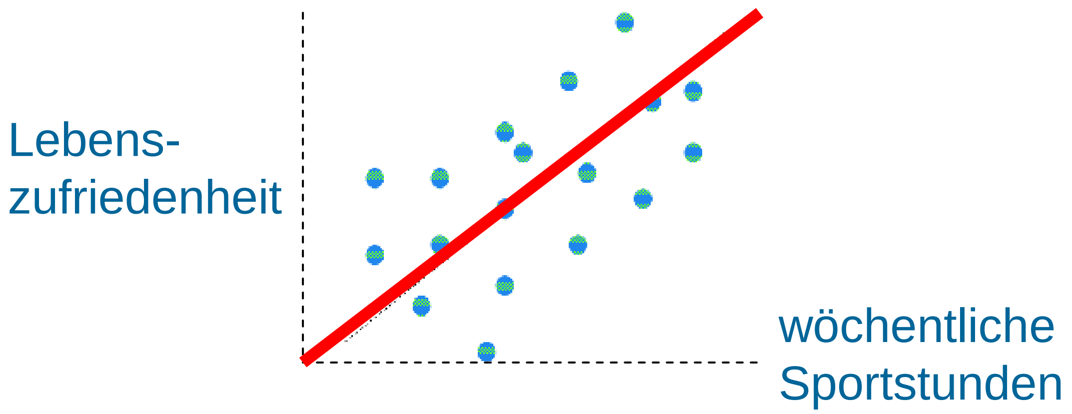
- Beschreiben der Daten
- Zusammenfassen der Daten
- Überblick über die Daten
- Datenanalyse
- Fehlerprüfung/Plausibilitätsprüfung

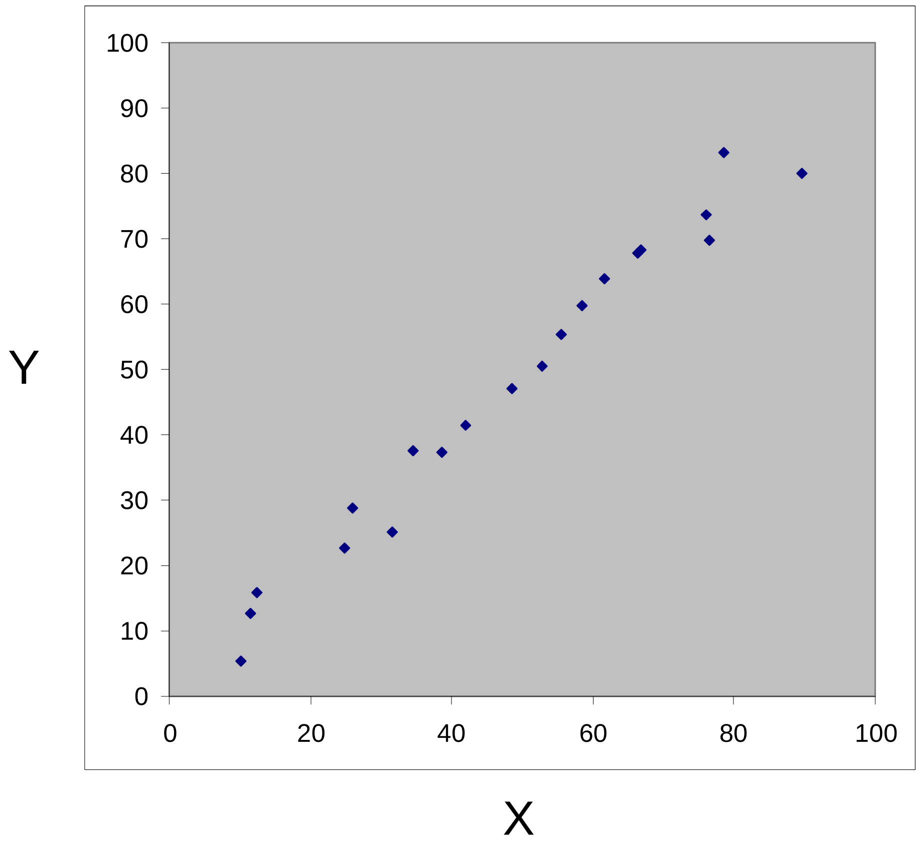
Kovarianz

„Gemeinsame Varianz“ zweier Variablen X und Y:

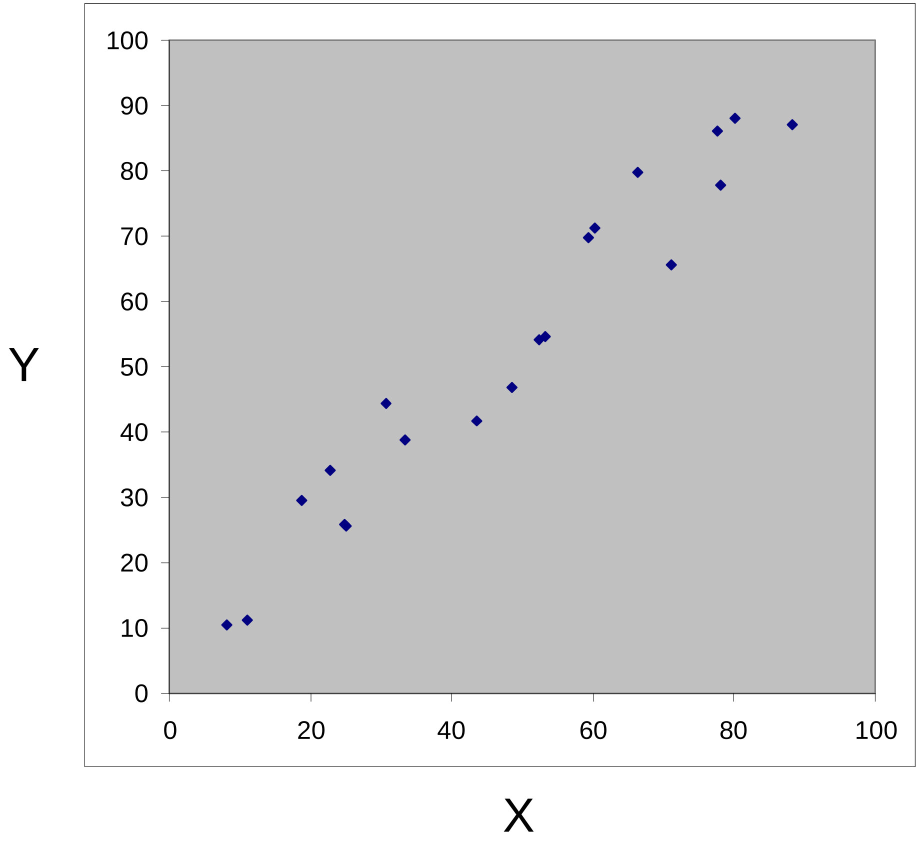
In wiefern ist X hoch ausgeprägt, wenn Y hoch (niedrig) ausgeprägt ist?

Beispiel aus der Einheit „Untersuchungsplanung“:

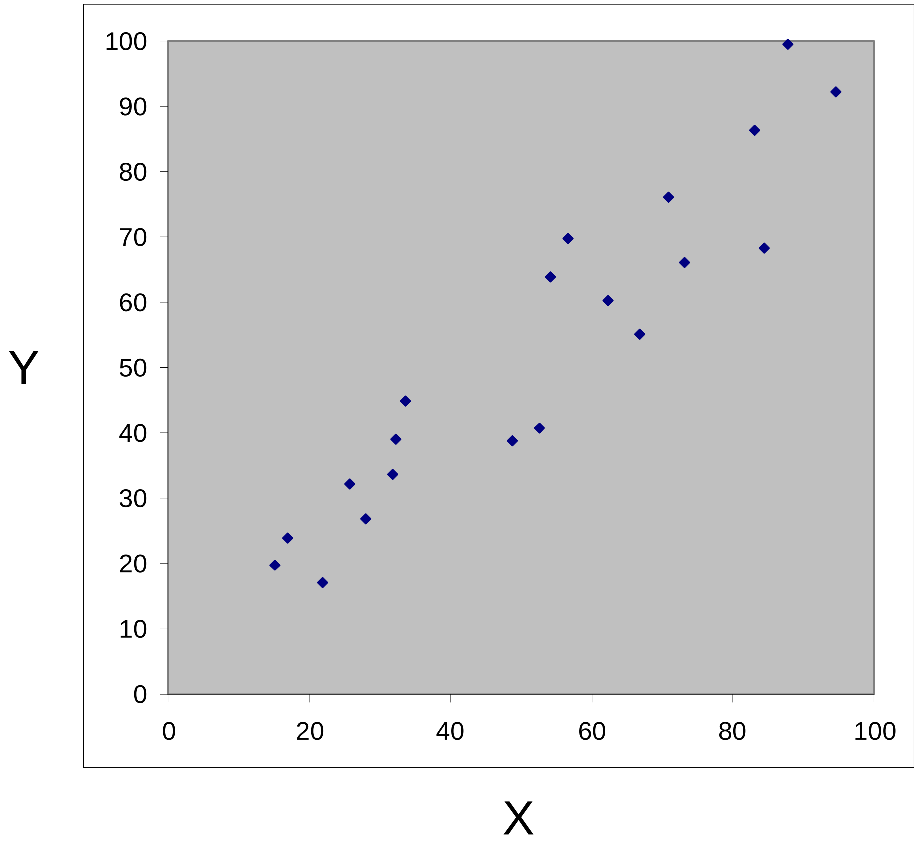




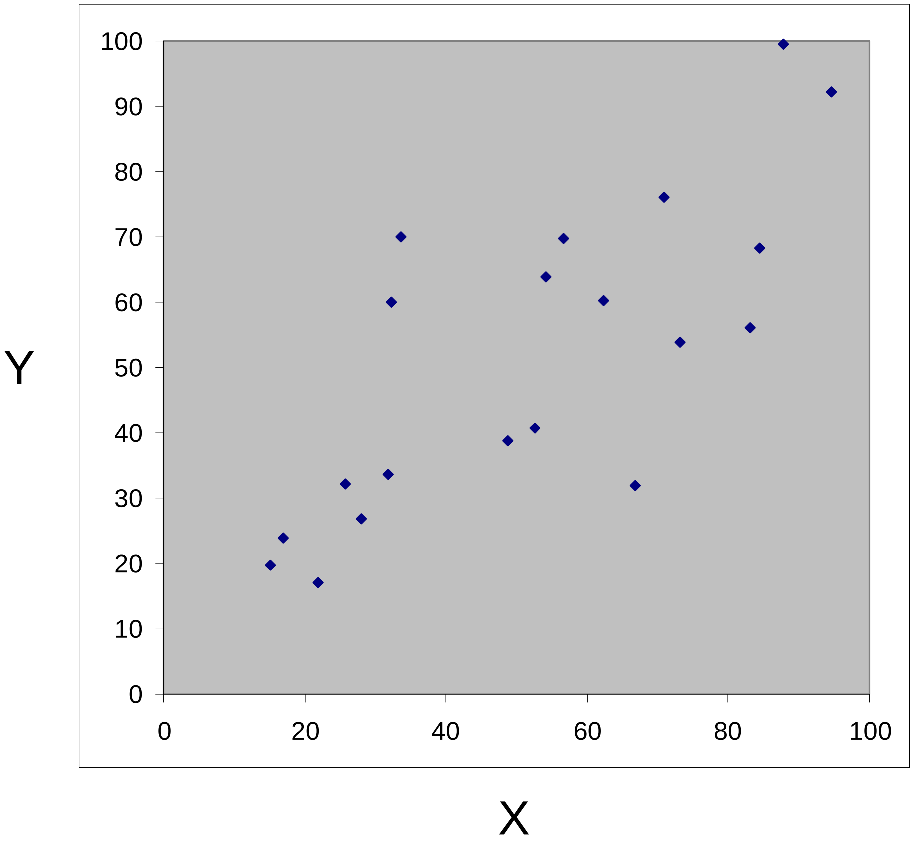
SD_X	13
SD_Y	13
Kov(X,Y)	159,08
$r_{X,Y}$	0,97



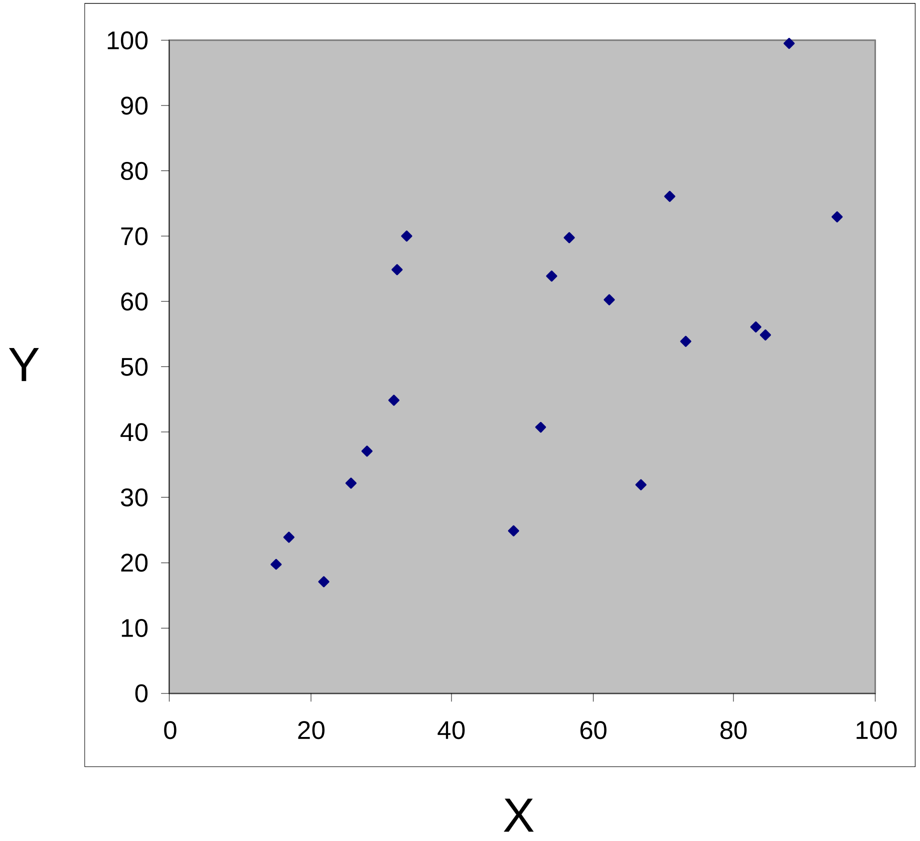
SD_X	12
SD_Y	12
Kov(X,Y)	133,81
$r_{X,Y}$	0,90



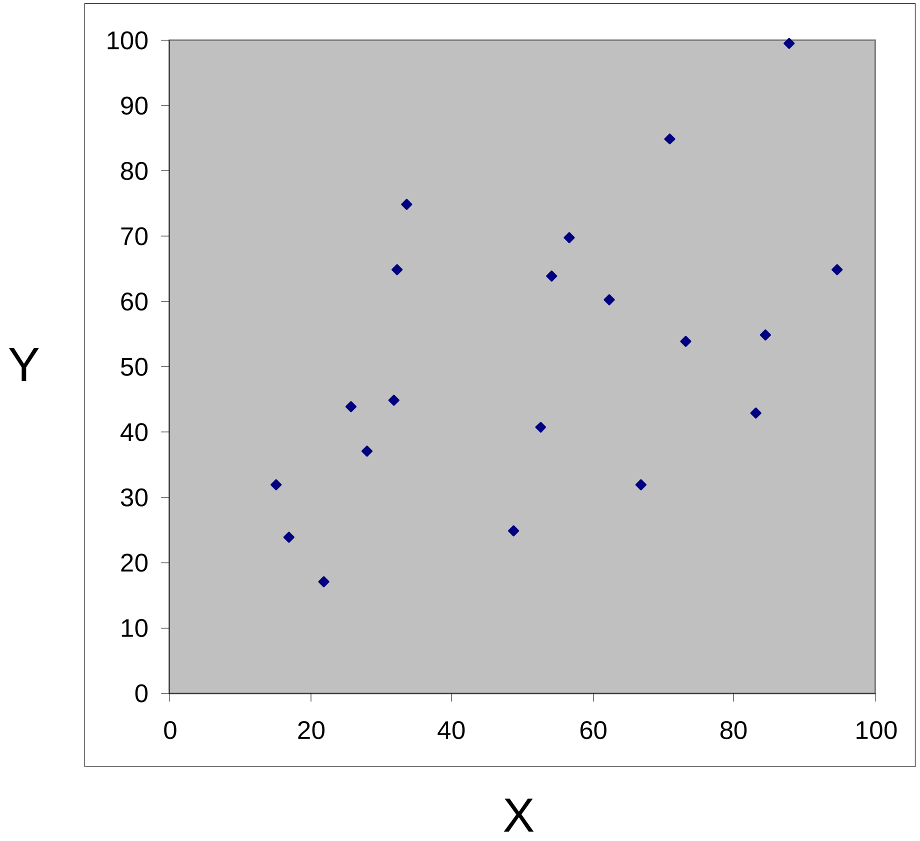
SD_X	12
SD_Y	9
Kov(X,Y)	81,26
$r_{X,Y}$	0,78



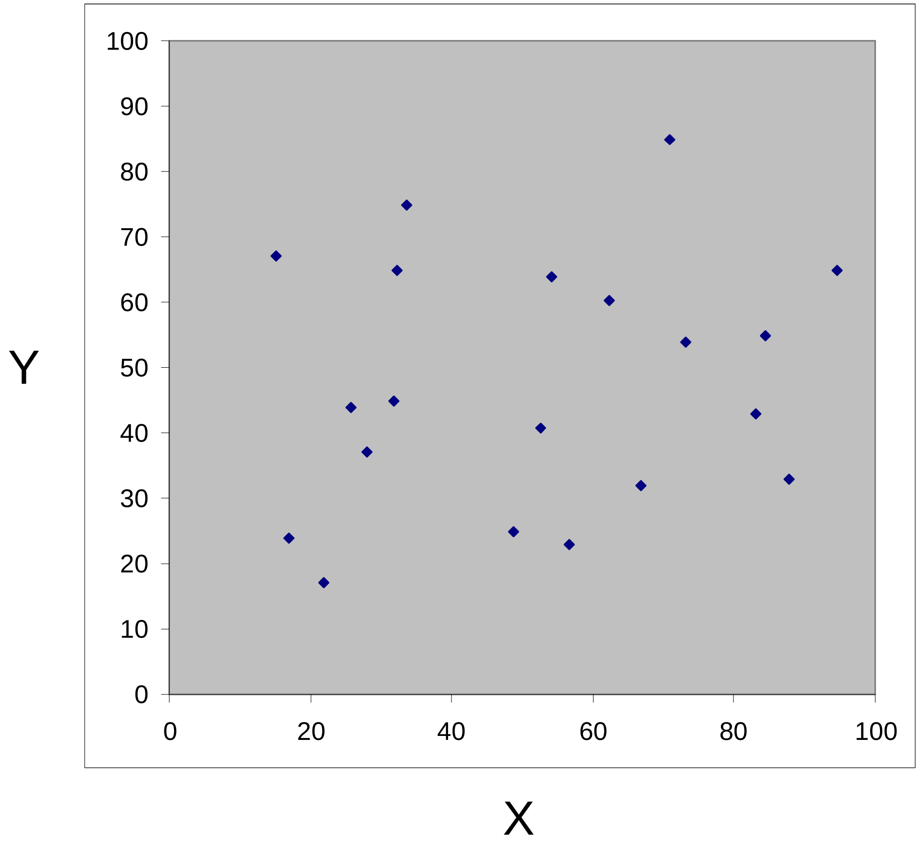
SD_X	12
SD_Y	16
Kov(X,Y)	92,08
$r_{X,Y}$	0,49



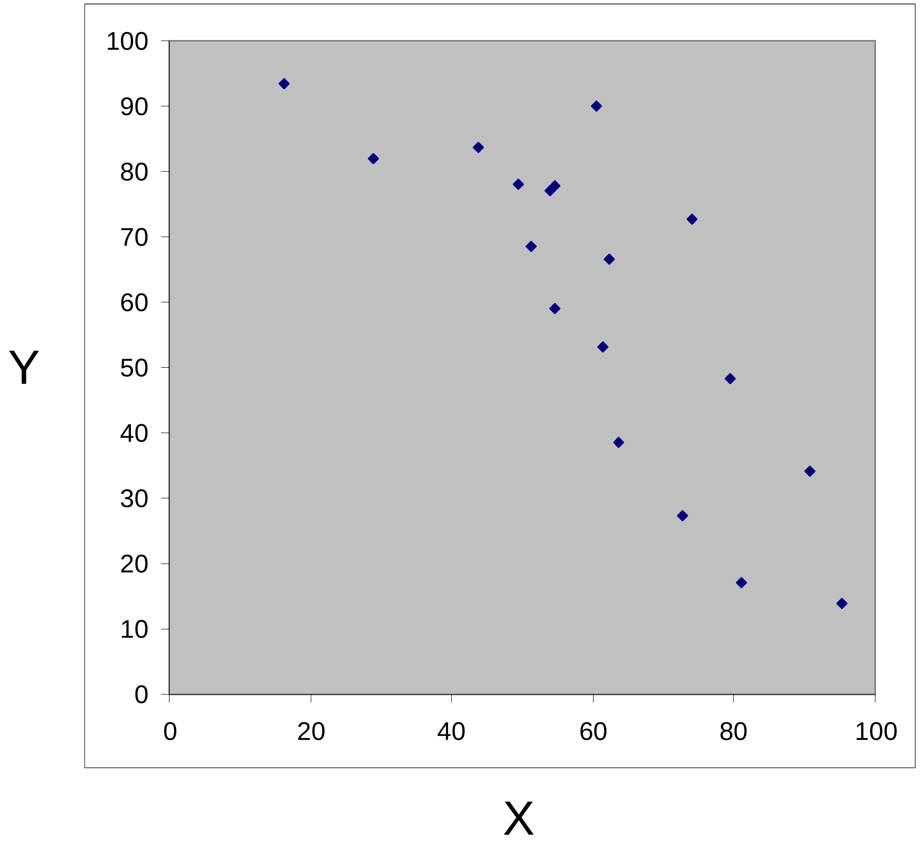
SD_X	12
SD_Y	17
Kov(X,Y)	66,38
$r_{X,Y}$	0,33



SD_X	12
SD_Y	17
Kov(X,Y)	43,12
$r_{X,Y}$	0,21



SD_X	12
SD_Y	19
Kov(X,Y)	-11,52
$r_{X,Y}$	-0,05



SD_X	14
SD_Y	12
Kov(X,Y)	-94,50
$r_{X,Y}$	-0,57

Forschungsablauf im Überblick

Untersuchungsanlass
↓
Problemexploration
↓
Formulieren einer allgemeinen Theorie
↓
Problemeingrenzung
↓
Ableitung einer speziellen Hypothese
↓

Untersuchungs-planung
↓
Operationalisierung der Hypothese
↓
Aufstellen von statistischen Hypothesen
↓
Untersuchungs-durchführung Datenerhebung
↓

Datenauswertung Signifikanztest
↓
Interpretation und Hypothesen-entscheidungen
↓
Theoretische Einordnung
↓
Präsentation/ Publikation

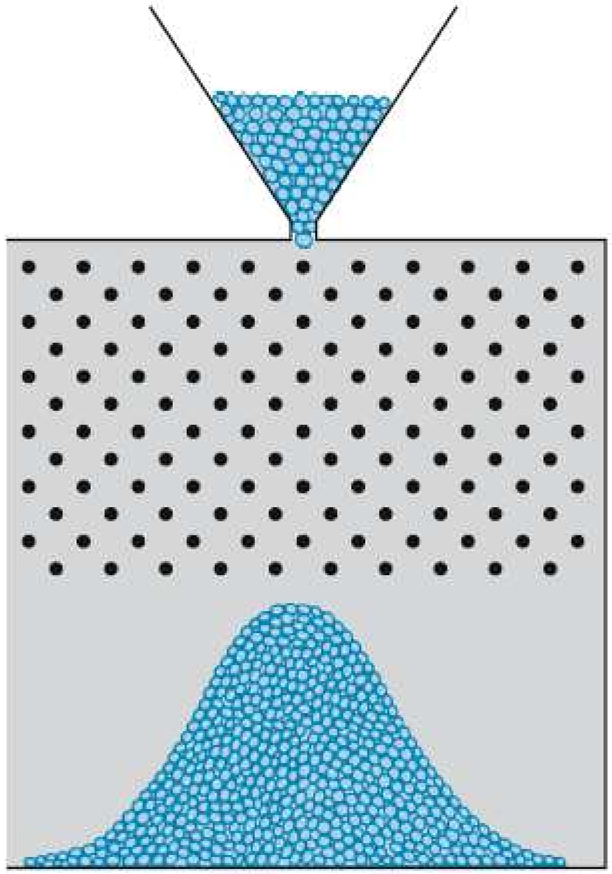
Inferenzstatistik 1

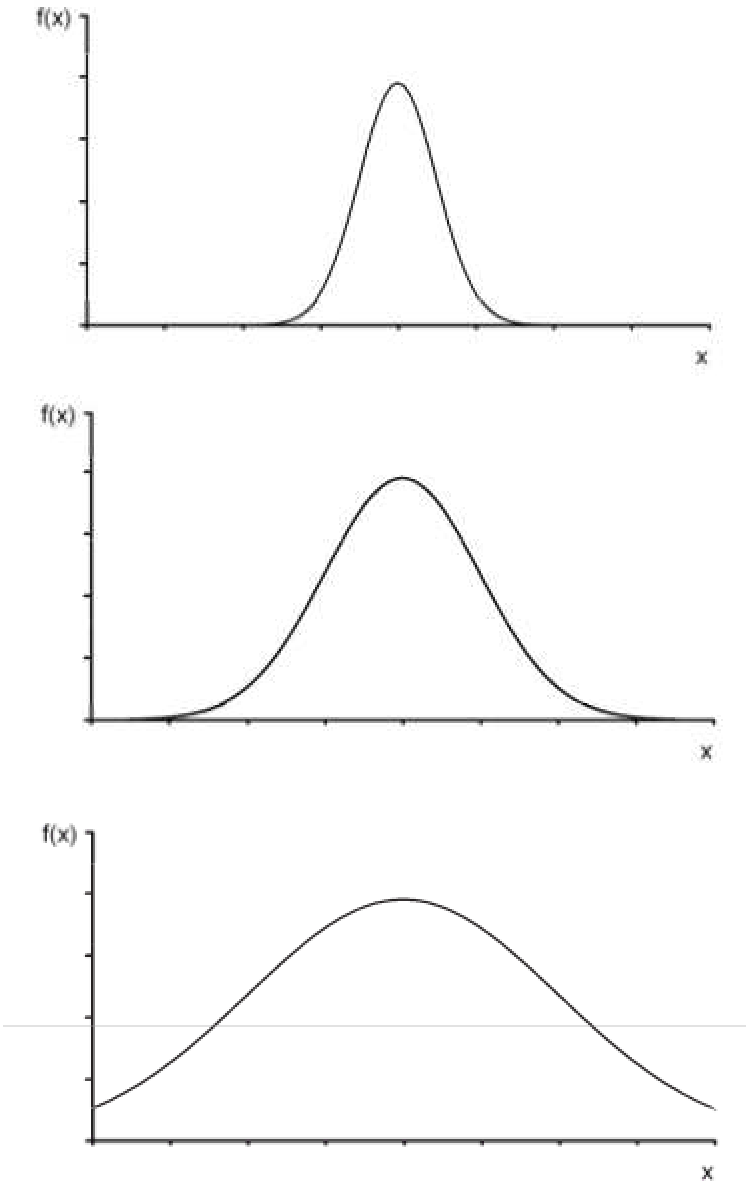
Testen von Zusammenhangshypothesen

Inferenzstatistik

- Schließende Statistik
- Verallgemeinerung:
  Stichprobe → Grundgesamtheit
- Voraussetzung:
  Repräsentative Stichprobe
  z.B. große Zufallsstichprobe

Viele biologische, psychologische, …
Variablen sind normalverteilt.





SD↓

SD↑

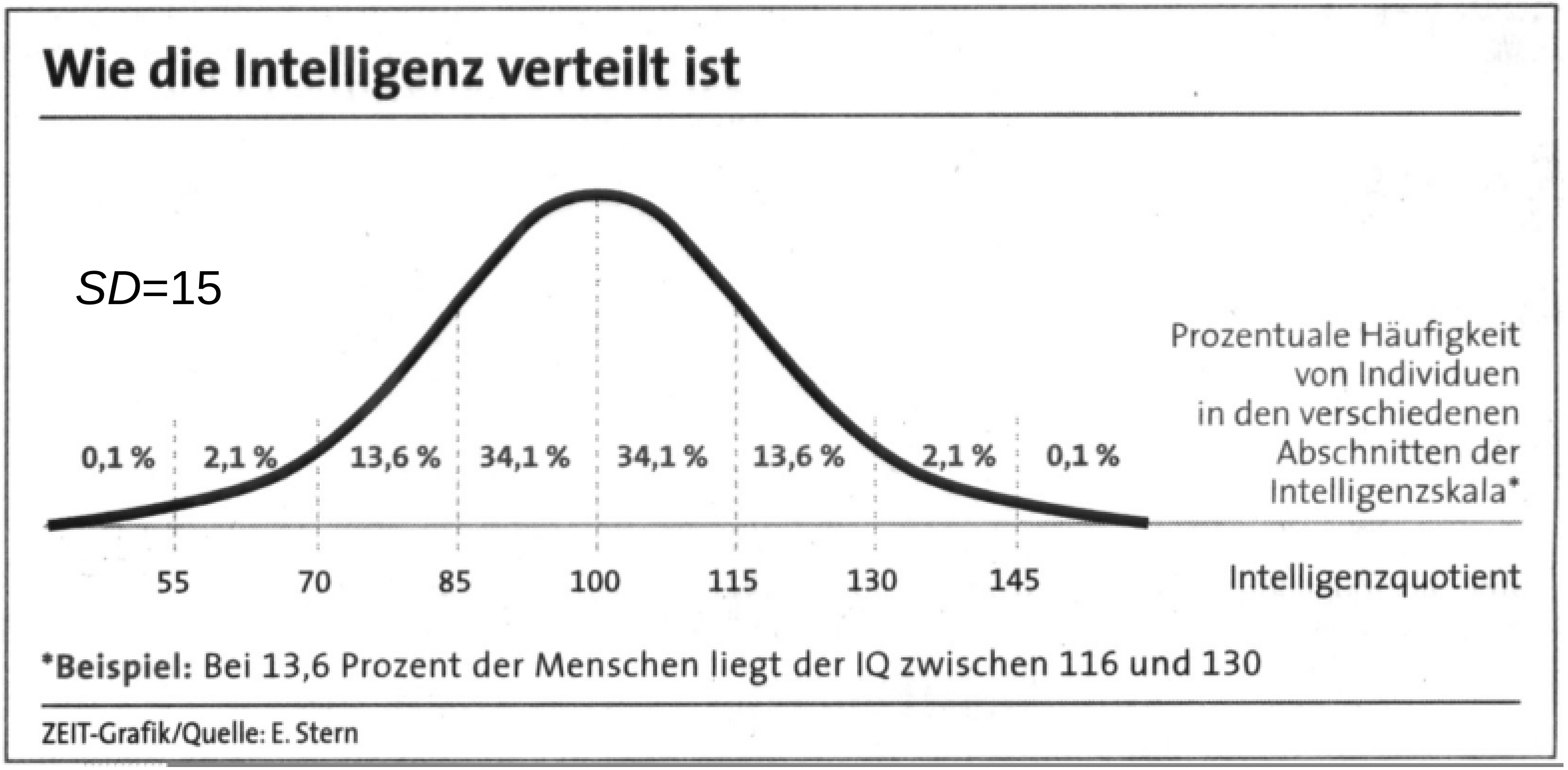
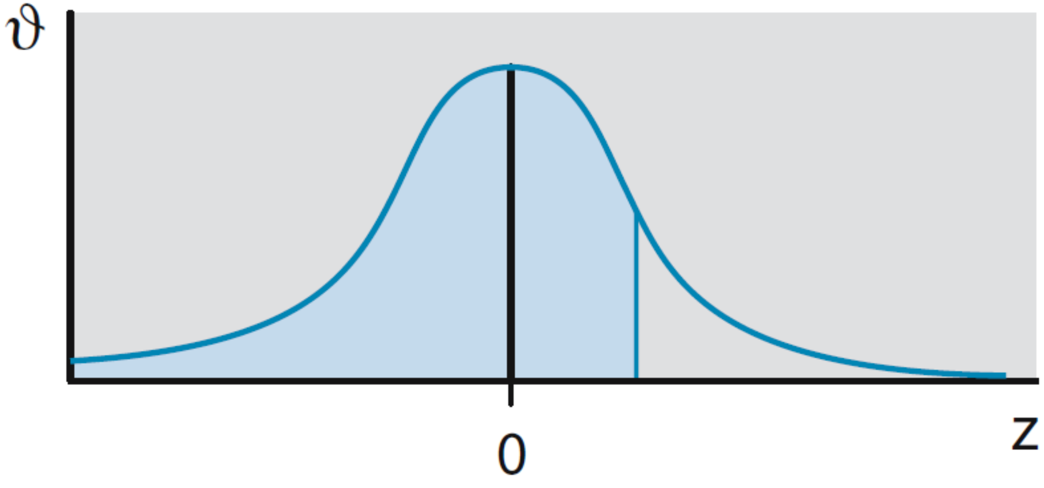

Wie die Intelligenz verteilt ist
SD=15
0,1 %
2,1 %
13,6 %
34,1 %
34,1 %
13,6 %
2,1 %
0,1 %
Prozentuale Häufigkeit von Individuen in den verschiedenen Abschnitten der Intelligenzskala*
55
70
85
100
115
130
145
Intelligenzquotient
*Beispiel: Bei 13,6 Prozent der Menschen liegt der IQ zwischen 116 und 130
ZEIT-Grafik/Quelle: E. Stern

Tabelle B (Fortsetzung)

z	Fläche	Ordinate	z	Fläche	Ordinate	z	Fläche	Ordinate
1,20	0,8849	0,1942	1,70	0,9554	0,0940	2,20	0,9861	0,0355
1,21	0,8869	0,1919	1,7					
1,22	0,8888	0,1895	1,7					
1,23	0,8907	0,1872	1,7					
1,24	0,8925	0,1849	1,7					
1,25	0,8944	0,1826	1,7					
1,26	0,8962	0,1804	1,7					
1,27	0,8980	0,1781	1,7					
1,28	0,8997	0,1758	1,7					
1,29	0,9015	0,1736	1,7					
			1,8					
			1,8					
			1,8					
			1,8					
			1,8					
			1,85	0,9678	0,0721	2,35	0,9906	0,0246
			1,86	0,9686	0,0707	2,36	0,9909	0,0246
1,37	0,9147	0,1561	1,87	0,9693	0,0694	2,37	0,9911	0,0241
1,38	0,9162	0,1539	1,88	0,9699	0,0681	2,38	0,9913	0,0235
1,39	0,9177	0,1518	1,89	0,9706	0,0669	2,39	0,9916	0,0229
1,40	0,9192	0,1497	1,90	0,9713	0,0656	2,40	0,9918	0,0224
1,41	0,9207	0,1476	1,91	0,9719	0,0644	2,41	0,9920	0,0219
1,42	0,9222	0,1456	1,92	0,9726	0,0632	2,42	0,9922	0,0213
1,43	0,9236	0,1435	1,93	0,9732	0,0620	2,43	0,9925	0,0208
1,44	0,9251	0,1415	1,94	0,9738	0,0608	2,44	0,9927	0,0203
1,45	0,9265	0,1394	1,95	0,9744	0,0596	2,45	0,9929	0,0198
1,46	0,9279	0,1374	1,96	0,9750	0,0584	2,46	0,9931	0,0194
1,47	0,9292	0,1354	1,97	0,9756	0,0573	2,47	0,9932	0,0189
1,48	0,9306	0,1334	1,98	0,9761	0,0562	2,48	0,9934	0,0184
1,49	0,9319	0,1315	1,99	0,9767	0,0551	2,49	0,9936	0,0180
1,50	0,9332	0,1295	2,00	0,9772	0,0540	2,50	0,9938	0,0175
1,51	0,9345	0,1276	2,01	0,9778	0,0529	2,51	0,9940	0,0171
1,52	0,9357	0,1257	2,02	0,9783	0,0519	2,52	0,9941	0,0167
1,53	0,9370	0,1238	2,03	0,9788	0,0508	2,53	0,9943	0,0163
1,54	0,9382	0,1219	2,04	0,9793	0,0498	2,54	0,9945	0,0158

$$z_i = \frac{x_i - M_X}{SD_X}$$



Auch statistische Kennwerte folgen bei Vorliegen spezifischer Voraussetzungen bestimmten Verteilungen.

z.B.:

- *Normalverteilung*
- *t-Verteilung*
- *F-Verteilung*
- *Chi²-Verteilung*

Statistischer Schluss

- X und Y seien zwei Zufallsvariablen, die bestimmte Voraussetzungen erfüllen
(z.B. X und Y seien intervallskaliert und normalverteilt).
- Annahme:
X und Y sind unkorreliert.
D.h. in der Grundgesamtheit gilt $r_{X,Y}=0$.
In Bezug auf Grundgesamtheiten werden griechische Buchstaben verwendet.
$\rho_{X,Y}=0$
- Diese Annahme wird als **Nullhypothese H_0** bezeichnet.

Statistischer Schluss

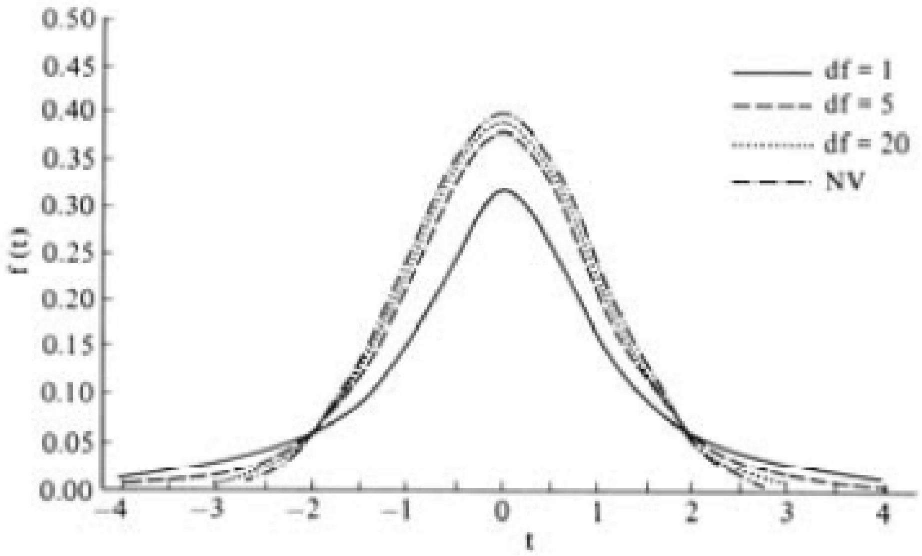
- Ziehen von Zufallsstichproben der Länge n:

  Die Korrelation r beschreibt den Zusammenhang von X und Y *in der jeweiligen Stichprobe*.

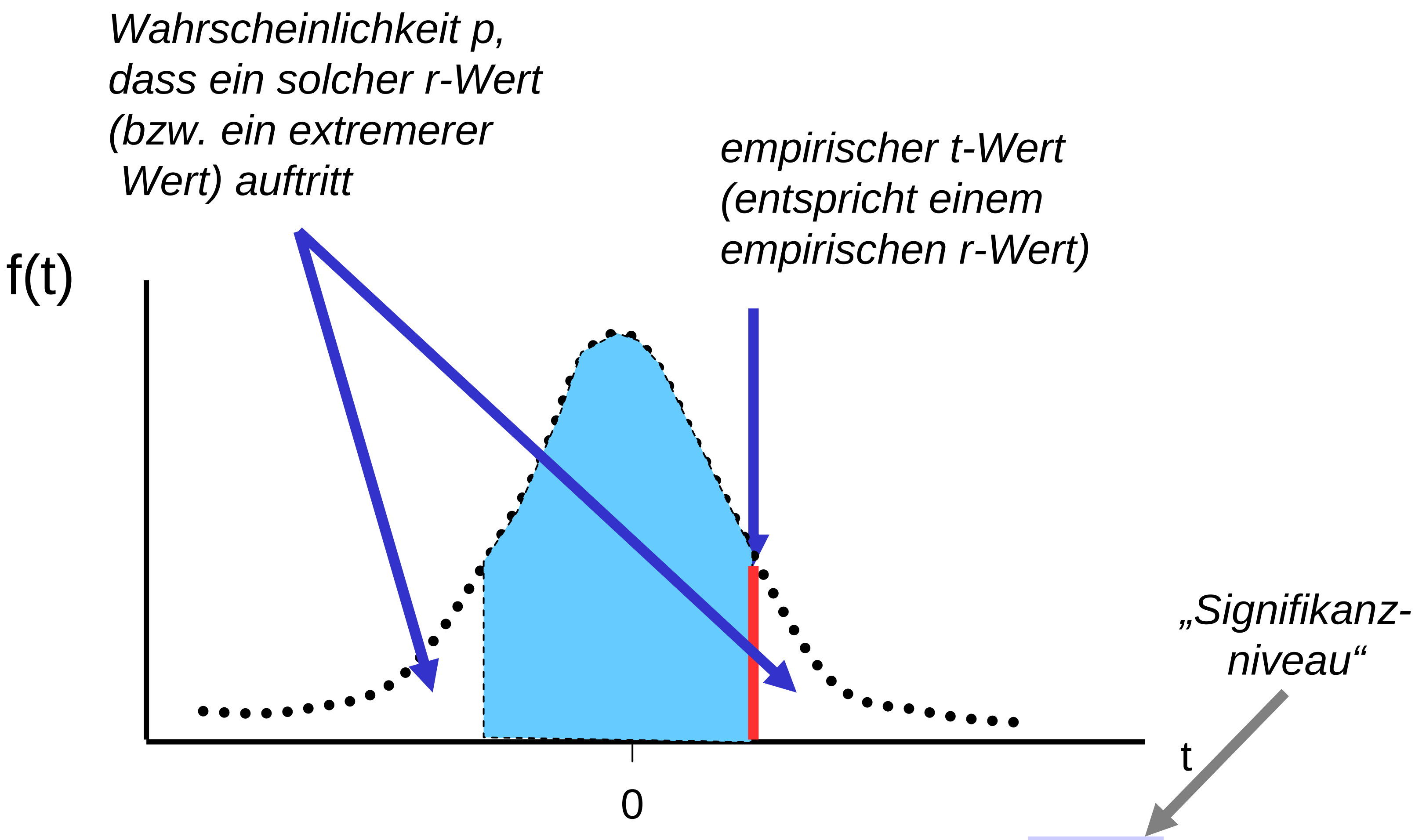
- Auch wenn in der Grundgesamtheit $\rho_{X,Y}=0$ gilt, wird r fast nie genau den Wert Null annehmen!

- Auch wenn in der Grundgesamtheit $\rho_{X,Y}=0$ gilt, verteilt sich der Stichprobenkennwert r um den Wert Null nach einer bestimmten Verteilung.
  (Genaugenommen: Die r-Werte werden in t-Werte transformiert und folgen der t-Verteilung mit n-2 Freiheitsgraden.)

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


0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
f (t)
–4
–3
–2
–1
0
1
2
3
4
t
df = 1
df = 5
df = 20
NV

Hypothesentest (Signifikanzprüfung)



Wenn p sehr klein ist (kleiner ist als ein vorher festgelegter α-Wert)
ist das Ergebnis so unwahrscheinlich, dass ich nicht an sein eintreten glaube!

Hypothesenentscheidung

$p < .05$ (bzw. $p < .01$)

→ Verwerfen der Annahme, dass die Nullhypothese gilt!

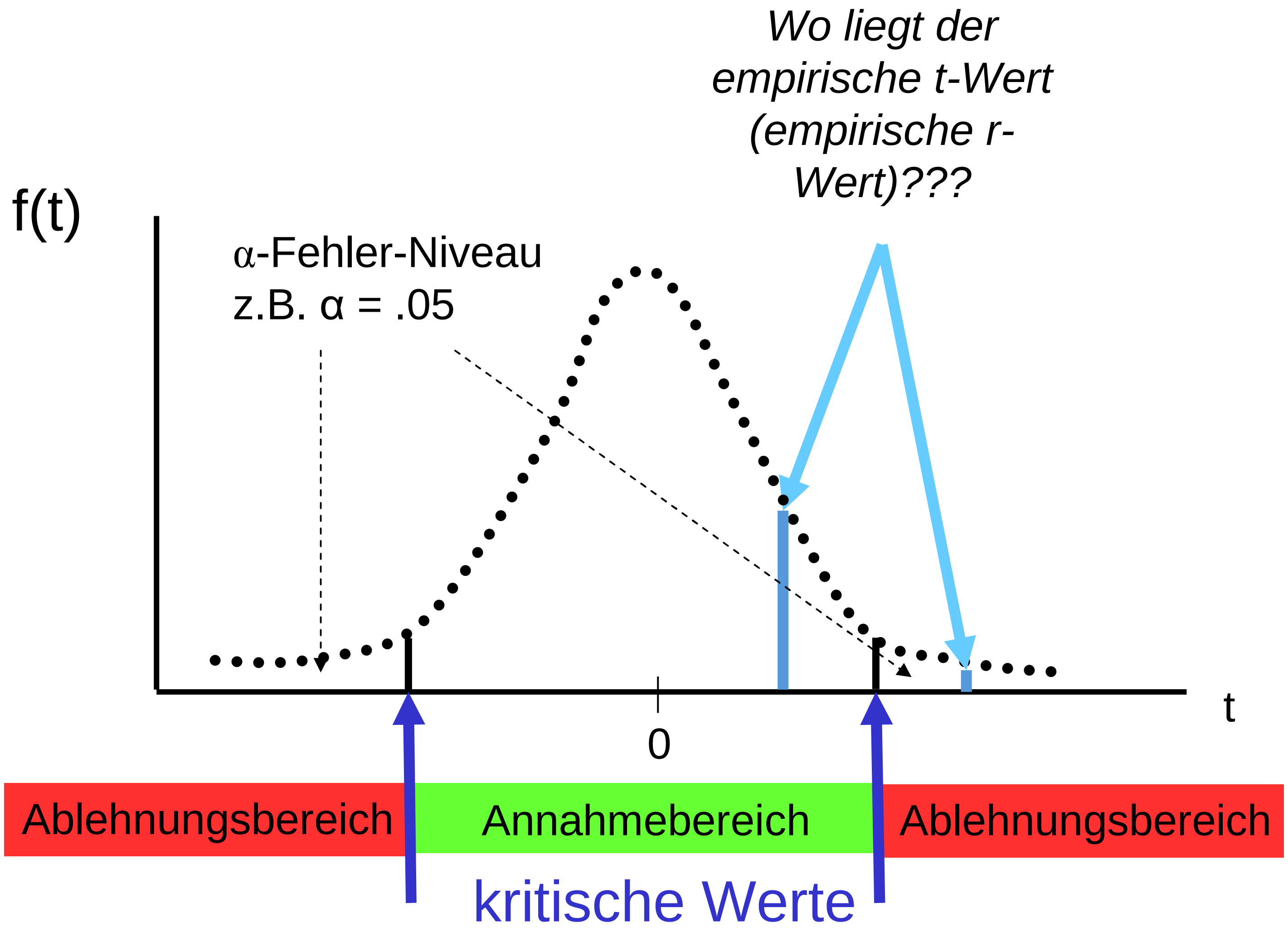
H_0:
X und Y sind unkorreliert. ($\rho_{X,Y}=0$)

→ Das Gegenteil gilt:

H_1:
X und Y korrelieren. ($\rho_{X,Y}\neq 0$)

→ In der Grundgesamtheit besteht ein Zusammenhang zwischen den Variablen X und Y.

Hypothesentest



Kritische r-Werte

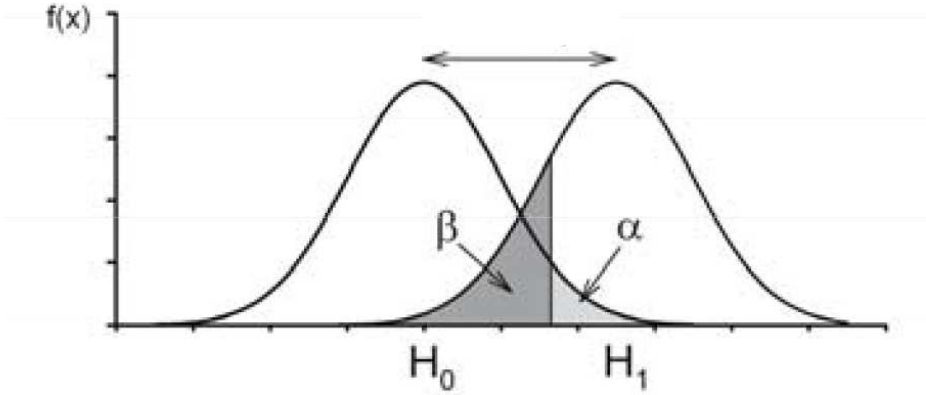
Tabelle 10.5. Betrag der Korrelation, der für den Stichprobenumfang n und das Signifikanzniveau α überschritten werden muss, um die $H_0 : \varrho = 0$ ablehnen zu können (zweiseitiger Test)

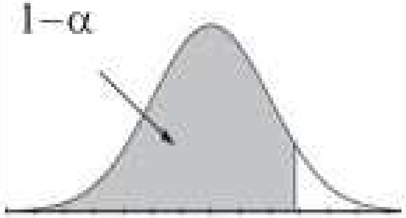
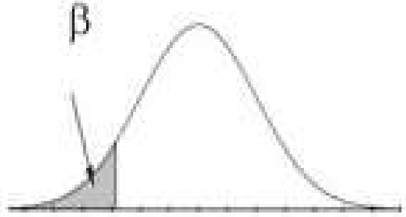
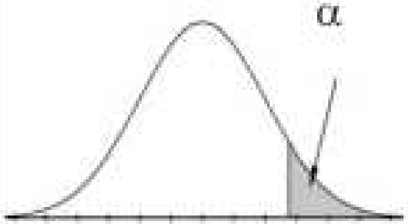
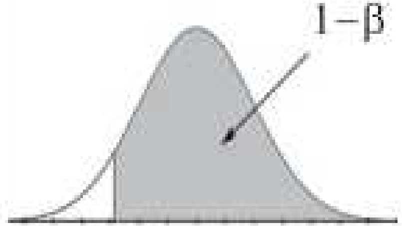
n	$\alpha = 0{,}05$	$\alpha = 0{,}01$
3	0,997	1,000
4	0,950	0,990
5	0,878	0,959
6	0,811	0,917
7	0,754	0,875
8	0,707	0,834
9	0,666	0,798
10	0,632	0,765
11	0,602	0,735
12	0,576	0,708
13	0,553	0,684
14	0,532	0,661
15	0,514	0,641
16	0,497	0,623
17	0,482	0,606
18	0,468	0,590
19	0,456	0,575
20	0,444	0,561
21	0,433	0,549
22	0,423	0,537
23	0,413	0,526
24	0,404	0,515
25	0,396	0,505
26	0,388	0,496
27	0,381	0,487
28	0,374	0,479
29	0,367	0,471
30	0,361	0,463
40	0,312	0,403
60	0,254	0,330
80	0,220	0,286
100	0,197	0,256
120	0,179	0,234

(Bortz & Schuster, 2010, S. 163)

Fehler bei Hypothesenentscheidungen

	In Wirklichkeit gilt die H_0	In Wirklichkeit gilt die H_1
Entscheidung zugunsten der H_0	Richtige Entscheidung	**β-Fehler**
Entscheidung zugunsten der H_1	**α-Fehler**	Richtige Entscheidung



		Realer, aber unbekannter Sachverhalt **POPULATION**	
		H_0	**H_1**
STICHPROBE	**H_0**	$p(H_0/\boldsymbol{H_0})$ $1-\alpha$ 	$p(H_0/\boldsymbol{H_1})$ β 
	H_1	$p(H_1/\boldsymbol{H_0})$ α 	$p(H_1/\boldsymbol{H_1})$ $1-\beta$  Teststärke (power)

Forschungsablauf im Überblick

Untersuchungsanlass
↓
Problemexploration
↓
Formulieren einer allgemeinen Theorie
↓
Problemeingrenzung
↓
Ableitung einer speziellen Hypothese
↓
Untersuchungs-planung
↓
Operationalisierung der Hypothese
↓
Aufstellen von statistischen Hypothesen
↓
Untersuchungs-durchführung Datenerhebung
↓
Datenauswertung Signifikanztest
↓
Interpretation und Hypothesen-entscheidungen
↓
Theoretische Einordnung
↓
Präsentation/ Publikation

Bewertung der Effektstärke

(nur bei Entscheidung für H_1)

$r = 0$	*kein Zusammenhang*
$0 < \lvert r \rvert < 0.4$	*niedriger Zusammenhang*
$0.4 < \lvert r \rvert < 0.7$	*mittlere Zusammenhang*
$0.7 < \lvert r \rvert < 1$	*hoher Zusammenhang*
$r = +/-1$	*vollständiger Zusammenhang*

(Willimczik, 1997)

Beispiel 1

Correlations

		Sporthäufigkeit	Zufriedenheit
Sporthäufigkeit	Pearson Correlation	1	,425**
	Sig. (2-tailed)		,001
	N	62	54
Zufriedenheit	Pearson Correlation	,425**	1
	Sig. (2-tailed)	,001	
	N	54	58

**. Correlation is significant at the 0.01 level (2-tailed).

Beispiel 2

Correlations

		Sporthäufigkeit	aerobeAusdauer
Sporthäufigkeit	Pearson Correlation	1	,210
	Sig. (2-tailed)		,148
	N	62	49
aerobeAusdauer	Pearson Correlation	,210	1
	Sig. (2-tailed)	,148	
	N	49	57

Beispiel 3

Correlations

		aerobe Ausdauer	Maximalkraft	Kraftausdauer
aerobeAusdauer	Pearson Correlation	1	,293*	,490**
	Sig. (2-tailed)		,031	,000
	N	57	54	54
Maximalkraft	Pearson Correlation	,293*	1	,351**
	Sig. (2-tailed)	,031		,004
	N	54	69	66
Kraftausdauer	Pearson Correlation	,490**	,351**	1
	Sig. (2-tailed)	,000	,004	
	N	54	66	69

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Aufgabe

Correlations

		Maximalkraft	Beweglichkeit	Koordinative Fähigkeiten	aerobe Ausdauer
Maximalkraft	Pearson Correlation	1	,100	,259	,293
	Sig. (2-tailed)		,439	,046	,031
	N	69	62	60	54
Beweglichkeit	Pearson Correlation	,100	1	,085	,240
	Sig. (2-tailed)	,439		,539	,097
	N	62	63	55	49
KoordinativeFähigkeiten	Pearson Correlation	,259	,085	1	,280
	Sig. (2-tailed)	,046	,539		,040
	N	60	55	62	54
aerobeAusdauer	Pearson Correlation	,293	,240	,280	1
	Sig. (2-tailed)	,031	,097	,040	
	N	54	49	54	57

Interpretiere den Computerausdruck:
Zugunsten welcher Hypothese fällt die Hypothesenentscheidung aus, wenn auf dem α-Niveau von 0.05 bzw. 0.01 getestet wird? Zwischen welchen Fähigkeiten liegen jeweils signifikante Zusammenhänge vor?

Abgabe bis Di, 12.06., 10:15 h