TECHNISCHE UNIVERSITÄT
IN DER KULTURHAUPTSTADT EUROPAS
CHEMNITZ

Professur Psychologie digitaler Lernmedien
Institut für Medienforschung Philosophische Fakultät

Statistik I

Korrelation

Sherlock, Series II, Episode I, A Scandal in Belgravia (2012). BBC One.

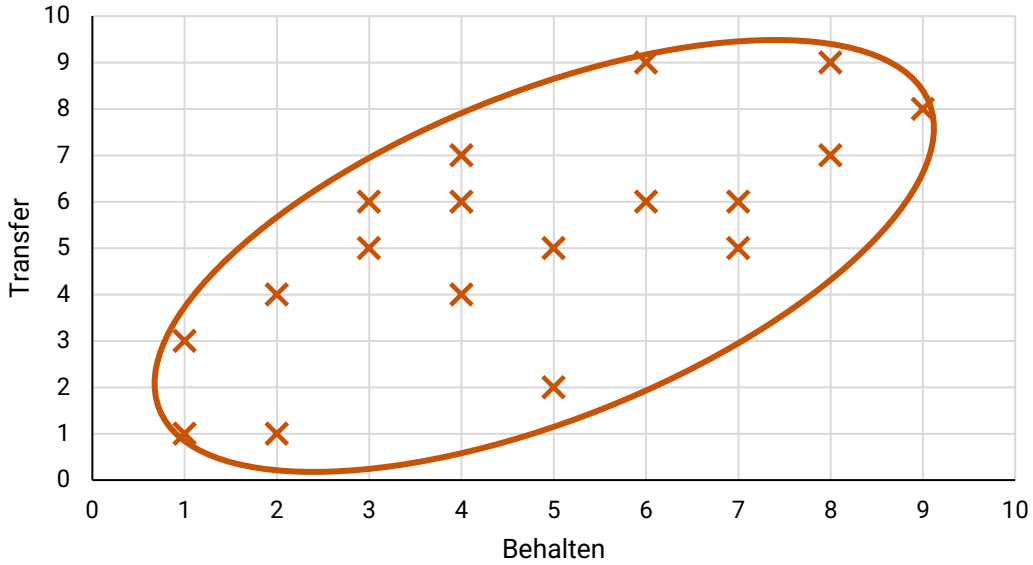
Prof. Dr. Günter Daniel Rey

Überblick

- Kovarianz und Korrelation
- Korrelation und Kausalität
- Fishers *Z*-Transformation
- Signifikanz von Korrelationen
- Partialkorrelation
- Weitere Korrelationstechniken

Einführung

- Zusammenhang zweier Variablen: Die Variablen variieren systematisch miteinander
- Fiktives Beispiel: Zusammenhang zwischen Behaltens- und Transferleistungen



Kovarianz
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Kovarianz und Korrelation quantifizieren den Grad des Zusammenhanges
- Kovarianz zweier Variablen: Durchschnittliches Abweichungsprodukt aller Messwertpaare von ihrem jeweiligen Mittelwert
- Formel (vgl. Formel zur Varianz):

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

x_i	= Wert x der Person i
$\bar{x}$	= Mittelwert von x
y_i	= Wert y der Person i
$\bar{y}$	= Mittelwert von y
n	= Anzahl an Personen

- Kovarianz: Unstandardisiertes Maß für den Grad von Zusammenhängen

Kovarianz

- Beispiel: Berechnung der Kovarianz für den rechts dargestellten Datensatz
- Berechnung:

$$\text{cov}(x, y) = \frac{(9.5 - 6.1) \cdot (9.0 - 6.5) + \ldots + (1.5 - 6.1) \cdot (2.0 - 6.5)}{5 - 1}$$

$$\text{cov}(x, y) = \frac{8.5 + 0.4 + 0.8 + 3.6 + 20.7}{4} = \frac{34}{4} = 8.5$$

- Ergebnis: Die Kovarianz beträgt 8.5

VPN	IQ	Mathe
Sheldon	9.5	9.0
Leonard	6.5	7.5
Howard	4.5	6.0
Rajesh	8.5	8.0
Penny	1.5	2.0
M	6.1	6.5

Korrelation
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Produkt-Moment-Korrelation nach Pearson gebräuchlichstes Maß für die Stärke des Zusammenhangs zweier (intervallskalierter) Variablen
- Korrelationskoeffizient *r* als standardisiertes (Effektstärke-)Maß für den Zusammenhang zweier Variablen
- Formel:

$$r_{xy} = \frac{\text{cov}_{\text{emp}}}{\text{cov}_{\text{max}}} = \frac{\text{cov}(x, y)}{\hat{\sigma}_x \cdot \hat{\sigma}_y}$$

Cov_{emp}	= Empirische Kovarianz zwischen x und y
Cov_{max}	= Maximale Kovarianz zwischen x und y
$\hat{\sigma}_x$	= Standardabweichung (*SD*) von x
$\hat{\sigma}_y$	= Standardabweichung (*SD*) von y

- Wertebereich von *r* reicht von −1 bis +1
- Wichtig: Korrelationskoeffizient *r* nicht intervallskaliert und nicht als Prozentmaß des Zusammenhanges interpretierbar (i. G. zu r^2)

Korrelation

Wie hoch ist die (gerundete) Korrelation für den rechts dargestellten Datensatz?

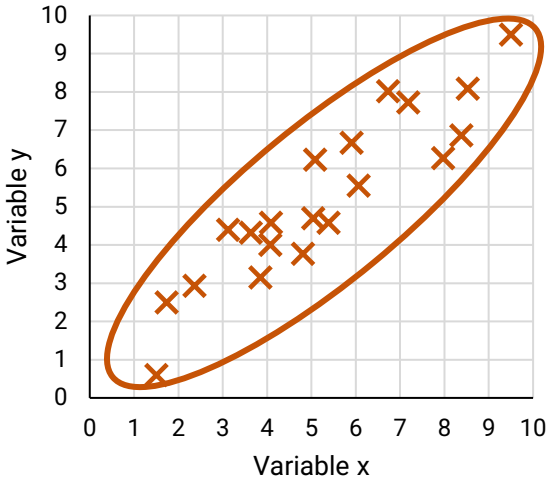
- ☐ A: 0.97
- ☐ B: 1.00
- ☐ C: 0.79
- ☐ D: 0.85

VPN	IQ	Mathe
Sheldon	9.5	9.0
Leonard	6.5	7.5
Howard	4.5	6.0
Rajesh	8.5	8.0
Penny	1.5	2.0
M	6.1	6.5
SD	3.21	2.74

Arten von Zusammenhängen

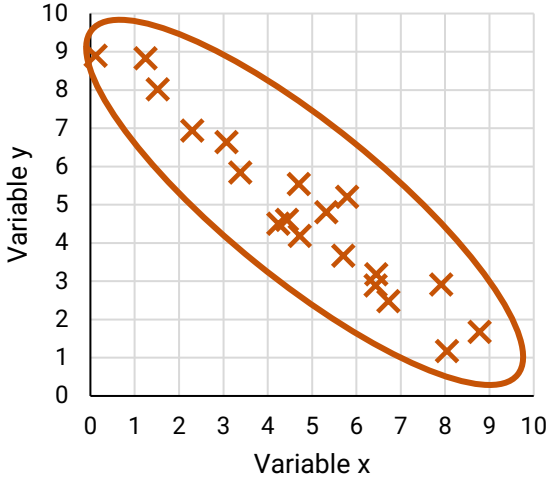
- Beispiele für Zusammenhänge zwischen zwei Variablen x und y:

Hoher positiver Zusammenhang



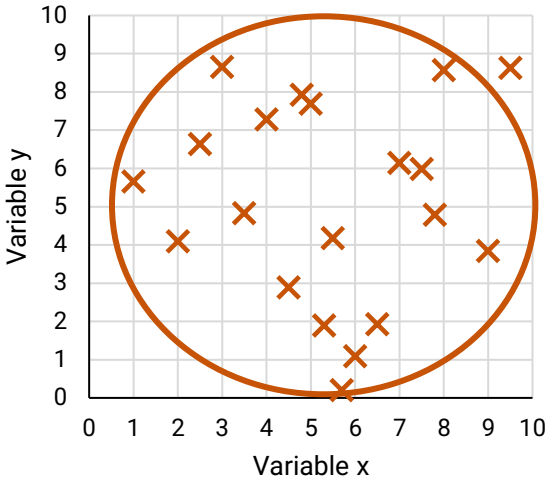
Positive Kovarianz

Hoher negativer Zusammenhang



Negative Kovarianz

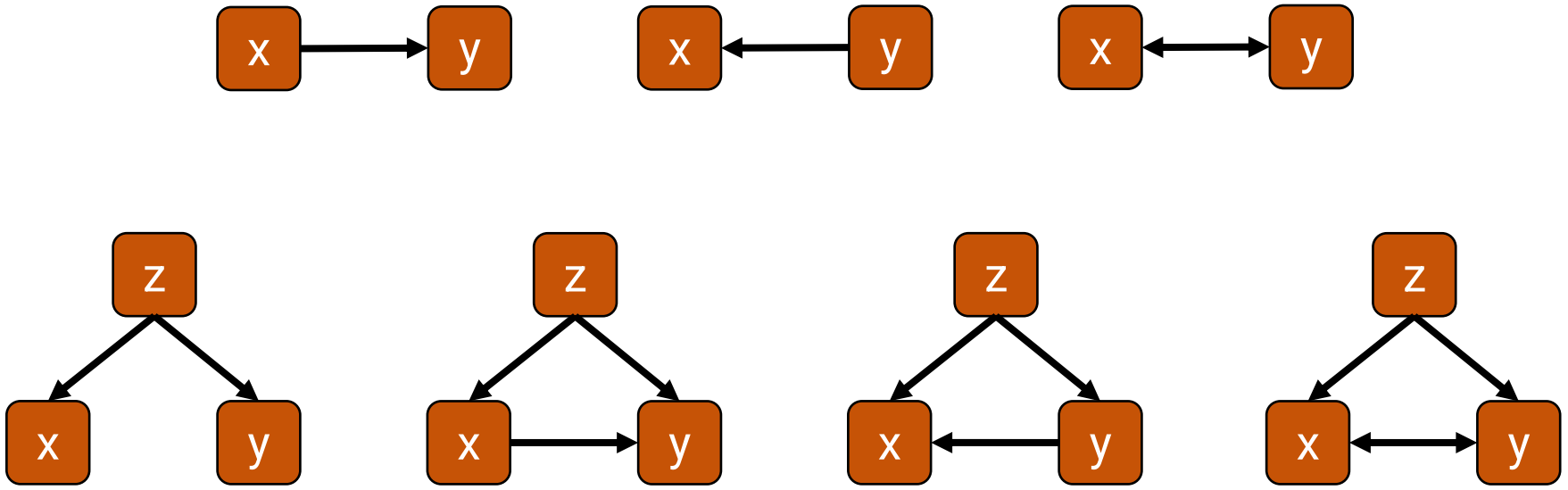
Kein Zusammenhang



Kovarianz von (nahezu) Null

Korrelation und Kausalität (Rey, 2020)

- Wichtig: Korrelation und Kausalität sind nicht identisch
- Mögliche Ursachen für eine Korrelation zwischen den zwei Variablen x und y:



Korrelation und Kausalität (Dubben & Beck-Bornholdt, 2006; Bortz & Schuster, 2010)

- Beispiele für hohe Korrelationen ohne Kausalzusammenhänge
  - Storchenpopulation und Geburtenrate
  - Einsatz von Feuerwehrleuten und Brandschäden
  - Globale Erwärmung und Lebenserwartung
  - Verweildauer im Krankenhaus und späterer Gesundheitszustand (negative Korrelation)
  - Kartoffelkonsum und Stromverbrauch (negative Korrelation)
- Aufdeckung von Scheinzusammenhängen aufgrund von Drittvariablen durch Partialkorrelationen
- Partialkorrelation: Korrelation zwischen Variablen, welche vom Einfluss einer oder mehrerer Drittvariablen statistisch bereinigt wurde

Fishers *Z*-Transformation (z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Problem: Berechnung von Mittelwerten aus Korrelationen aufgrund des fehlenden Intervallskalenniveaus nicht unmittelbar möglich
- Lösung: Fishers *Z*-Transformation (nicht mit der *z*-Standardisierung verwechseln!)
- Berechnungsschritte
  - Transformation der einzelnen Korrelationen in Fishers *Z*-Werte
  - Berechnung des Mittelwertes zu den Fishers *Z*-Werten
  - Rücktransformation dieses Mittelwertes in eine Korrelation
- Berechnung in Excel mittels der Funktionen „FISHER()“ und „FISHERINV()“
- Beispiel: Mittelwert aus $r = .10$ und $r = .90$ ist $r = .66$ und nicht $r = .50$

Signifikanz von Korrelationen (z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Signifikanztest für Korrelationen analog zum *t*-Test
- Formel:

$$t(df) = \frac{r \cdot \sqrt{N-2}}{\sqrt{1-r^2}}$$

r = Korrelation
N = Stichprobenumfang

- Formel für die Freiheitsgrade: df = *N* − 2
- Beispiel: In einer Studie mit 100 Studierenden korrelieren Behalten und Transfer mit *r* = 0.3
- Berechnung: $t(98) = \frac{0.3 \cdot \sqrt{100-2}}{\sqrt{1-0.3^2}} \approx \frac{0.30 \cdot 9.90}{0.95} \approx 3.11$
- Da $t_{emp} = 3.11 \geq t_{krit} = 1.66$ wird H_0 zugunsten der H_1 verworfen, d. h. das Ergebnis ist signifikant; $r = .3$, $t(98) = 3.11$, $p < .01$

Partialkorrelation (Rasch, Friese, Hofmann & Naumann, 2021)

- Partialkorrelation: Korrelation zwischen Variablen, welche vom Einfluss einer oder mehrerer Drittvariablen statistisch bereinigt (herauspartialisiert) wurde
- Formel:

$$r_{xy|z} = \frac{r_{xy} - r_{yz} \cdot r_{xz}}{\sqrt{(1 - r_{yz}^2) \cdot (1 - r_{xz}^2)}}$$

$r_{xy|z}$ = Partialkorrelation der beiden interessierenden Merkmale

r_{xy} = Korrelation nullter Ordnung der beiden interessierenden Merkmale

r_{xz} = Korrelation von x mit der Drittvariablen z

r_{yz} = Korrelation von y mit der Drittvariablen z

- Signifikanztest analog zum *t*-Test
- Formel:

$$t(df) = r_{xy|z} \cdot \frac{\sqrt{N - 2}}{\sqrt{1 - r_{xy|z}^2}}$$

- Formel für die Freiheitsgrade: df = *N* – 3

Weitere Korrelationstechniken (Rasch, Friese, Hofmann & Naumann, 2021)

- Neben Produkt-Moment-Korrelationen (für intervallskalierte Daten): Weitere Verfahren für verschiedene Kombinationen von Skalenniveaus (intervallskaliert, ordinal, nominal)
- Übersicht über die Korrelationstechniken:

	Intervallskala	**Rangskala**	**Nominalskala (dichotom)**
Intervallskala	Produkt-Moment-Korrelation	Rangkorrelation	Punktbiseriale Korrelation
Rangskala		Rangkorrelation	Punktbiseriale Rangkorrelation
Nominalskala (dichotom)			Phi-Koeffizient

Punktbiseriale Korrelation (Rasch, Friese, Hofmann & Naumann, 2021)

- Punktbiseriale Korrelation bestimmt den Zusammenhang zwischen einer *echt* dichotomen und einer intervallskalierten Variable
- Formel: $r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{\hat{\sigma}_y} \cdot \sqrt{\frac{n_0 \cdot n_1}{N^2}}$
- Wertebereich von *r* reicht von −1 bis +1
- Signifikanztest wieder analog zum *t*-Test
- Formel: $t(df) = \frac{r_{pb} \cdot \sqrt{N - 2}}{\sqrt{1 - {r_{pb}}^2}}$

x = dichotome Variable in den Ausprägungen x_0 und x_1 (nicht in der Formel)
y = intervallskalierte Variable
$\bar{y}_0$ = Mittelwert der y-Werte in x_0
$\bar{y}_1$ = Mittelwert der y-Werte in x_1
n_0 = Stichprobengröße in x_0
n_1 = Stichprobengröße in x_1
$N = n_0 + n_1$ (Anzahl aller Untersuchungseinheiten)
$\hat{\sigma}_y$ = geschätzte Populationsstreuung aller y-Werte

- Formel für die Freiheitsgrade: df = *N* − 2

Rangkorrelation (Rasch, Friese, Hofmann & Naumann, 2021)

- Rangkorrelation nach Spearman bestimmt den Zusammenhang zwischen zwei ordinalskalierten Variablen
- Analogie zur Produkt-Moment-Korrelation: Anstelle intervall-skalierter Messwerte werden die jeweiligen Rangplätze eingesetzt
- Formel:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{N \cdot (N^2 - 1)}$$

d_i = Differenz der Rangplätze einer Untersuchungseinheit i bezüglich der Variablen x und y
N = Anzahl aller Untersuchungseinheiten

- Wertebereich von *r* reicht wieder von −1 bis +1
- Signifikanztest erneut analog zum *t*-Test (sofern $n \geq 30$)
- Formel:

$$t(df) = \frac{r_s \cdot \sqrt{\mathrm{N} - 2}}{\sqrt{1 - {r_s}^2}}$$

- Formel für die Freiheitsgrade: df = *N* − 2

Beispiele für Korrelationen in Fachzeitschriften

The data were analyzed by means of a 2 × 2 ANOVA with the learner's gender and the speaker's gender as between-factors. For the analysis of problem-solving performance, two additional control variables were included, namely the "Abiturnote" (i.e., final high school grade point average) and intrinsic motivation, resulting in a 2 × 2 ANCOVA. Both covariates showed a significant correlation with problem-solving performance (Abiturnote: $r = -.36$, $P = .001$, whereby better school grades were associated with better learning outcomes; intrinsic motivation: $r = .26$, $P = .02$, whereby higher intrinsic motivation was associated with better learning outcomes), but were independent from each other ($r = -.01$, $P = .94$). The results of the experiment are shown in Table 1.

Quelle: Linek, Gerjets und Scheiter (2010)

Table 6. Correlations between indices of game performance, pre-test and learning outcome

	2	3	4	5	6
(1) Level Reached	.99***	.49**	.41*	.43*	.18
(2) Unique Maths Tasks		.55**	.38*	.44*	.20
(3) All Maths Tasks			-.23	.37*	.25
(4) Accuracy				.06	-.14
(5) Pre-test					-.01
(6) Gain					

Note. * = p<0.05, ** = p<0.01, *** = p<0.001 (two-tailed test of significance).

Quelle: Habgood und Ainsworth (2011)

Zusammenfassung

- Kovarianz als unstandardisiertes und Korrelation als standardisiertes Maß zur Quantifizierung des Zusammenhanges zweier Variablen
- Korrelation und Kausalität sind nicht identisch
- Signifikanztest für Korrelationen analog zum *t*-Test
- Partialkorrelation als Korrelation zwischen Variablen, welche vom Einfluss einer oder mehrerer Drittvariablen statistisch bereinigt wurde
- Punktbiseriale Korrelation und Rangkorrelation als weitere Korrelationstechniken

Prüfungsliteratur

- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2021). *Quantitative Methoden 1: Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften* (5. Aufl.). Heidelberg: Springer.
  - Merkmalszusammenhänge (S. 87–104)

Weiterführende Literatur I

- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
  - Korrelation (S. 153–182)
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5. Aufl.). Weinheim: Beltz.
  - Zusammenhänge zwischen zwei Variablen: Korrelations- und Assoziationsmaße (S. 529–587)
- Leonhart, R. (2022). *Lehrbuch Statistik. Einstieg und Vertiefung* (5. Auflage). Bern: Huber.
  - Korrelation und Regression (S. 261–378)

Weiterführende Literatur II

- Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (3. Aufl.). München: Pearson.
  - Korrelation (S. 207–244)
- Rey, G. D. (2020). *Methoden der Entwicklungspsychologie. Datenerhebung und Datenauswertung* (3., überarbeitete Auflage). Norderstedt BoD.
  - Korrelation (S. 62–66)
- Dubben, H.-H., & Beck-Bornholdt, H.-P. (2006). *Der Hund, der Eier legt. Erkennen von Fehlinformation durch Querdenken*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.