



Psychologische Forschungsmethoden

BSc Philosophie-Neurowissenschaften-Kognition WiSe 2021/22

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(9) Experimentelle Versuchspläne IV

Korrelative Designs

Korrelationskoeffizient

Stichprobenauswahleffekte

Selbstkontrollfragen

Korrelative Designs

Korrelationskoeffizient

Stichprobenauswahleffekte

Selbstkontrollfragen

Definition

Nicht-experimentelle Datenanordnungen, die typischerweise nur den wechselseitigen Zusammenhang (Korrelation) zwischen zwei oder mehreren Variablen betreffen.

Reiß and Sarris (2012)

Bemerkungen

- “Correlation is not causation!”

⇒ Korrelationen können immer durch Drittvariablen kausal bedingt sein.

Ja, aber ...

- Der Begriff der “Kausalität” ist nicht eindeutig definiert.
- Experimentelle Designs werden im Normalfall mit Korrelationen (Regression, ALM) untersucht.
- Kausale Inferenz nutzt auch “nur” probabilistische Modelle, wird wenig angewendet/gelehrt.
- Im Sinne der zeitlichen Präzedenz benutzt kausale Inferenz zum Teil Zeitserienmodelle.
- Das Psychologiestudium sieht keine Auseinandersetzung mit Graphical Models und Differentialgleichungsmodellen vor, die für ein Verständnis zeitgenössischer kausaler Inferenz nötig wäre.

Kausalzusammenhänge sind ein latentes Konstrukt das nur datenanalytisch erschlossen werden kann!

⇒ Kausalzusammenhänge sind sowohl in experimentellen als auch korrelativen Designs latent!

Korrelative Designs

Korrelationskoeffizient

Stichprobenauswahleffekte

Selbstkontrollfragen

Wir bezeichnen einen bivariaten Datensatz mit

$$x = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)). \quad (1)$$

x_i und y_i bezeichnen das erste und zweite Merkmal der i ten Einheit, respektive.

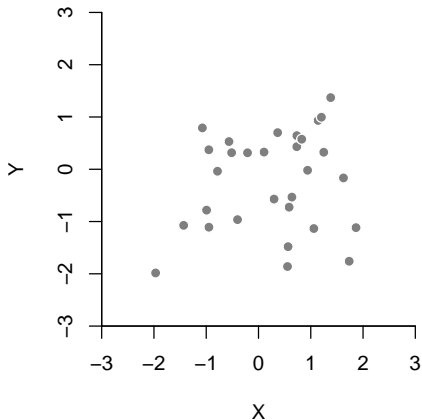
n ist die Anzahl an bivariaten Datenpunkten (x_i, y_i) .

Untenstehende Tabelle zeigt ein Beispiel.

i	1	2	3	4	5	6	7	8	9	10
x_i	3.4	1.5	2.7	4.5	6.1	3.8	2.0	2.5	6.2	9.1
y_i	5.5	7.3	1.1	1.9	4.5	2.3	8.4	8.6	3.9	1.6

Korrelationskoeffizient

Die Darstellung der $(x_i, y_i), i = 1, \dots, n$ in einem Koordinatensystem heißt *Streudiagramm*.



Definition (Empirische Stichprobenkovarianz)

$x = ((x_1, y_1), \dots, (x_n, y_n))$ sei ein bivariater Datensatz. Dann heißt die Zahl

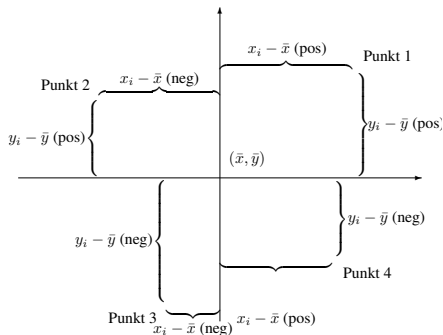
$$C := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (2)$$

empirische Stichprobenkovarianz von x_1, \dots, x_n und y_1, \dots, y_n .

Bemerkungen

- \bar{x} und \bar{y} bezeichnen die Mittelwerte der x_1, \dots, x_n und y_1, \dots, y_n , respektive.
- Der Faktor $1/n$ normiert für die Stichprobengröße.
- Eine Intuition vermittelt untenstehende Abbildung aus Fahrmeir et al. (2016).

Korrelationskoeffizient



	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
Punkt 1 (1. Quadrant)	positiv	positiv	positiv
Punkt 2 (2. Quadrant)	negativ	positiv	negativ
Punkt 3 (3. Quadrant)	negativ	negativ	positiv
Punkt 4 (4. Quadrant)	positiv	negativ	negativ

Abbildung 3.10 und Tabelle 3.9 aus Fahrmeir et al. (2016), Kapitel 3.4.

Definition (Pearson's Stichprobenkorrelationskoeffizient)

$x = ((x_1, y_1), \dots, (x_n, y_n))$ sei ein bivariater Datensatz, C sei die empirische Stichprobenvarianz von x_1, \dots, x_n und y_1, \dots, y_n und S_x bzw. S_y seien die empirischen Stichprobenstandardabweichungen von x_1, \dots, x_n bzw. y_1, \dots, y_n . Dann heißt die Zahl

$$r := \frac{C}{S_x S_y} \quad (3)$$

Pearson's Stichprobenkorrelationskoeffizient oder *empirischer Korrelationskoeffizient*.

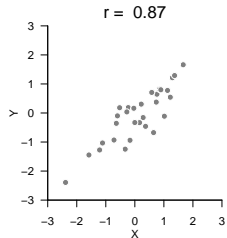
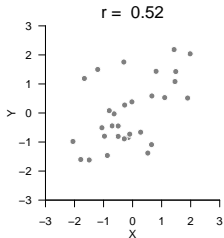
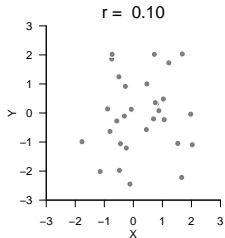
Bemerkungen

- Ohne Beweis halten wir fest, dass gilt $-1 \leq r \leq 1$.
- Ohne Beweis halten wir fest, dass r die Stärke des linearen Zusammenhangs von x und y misst.
- Korrelationsstärken werden in etwa nach

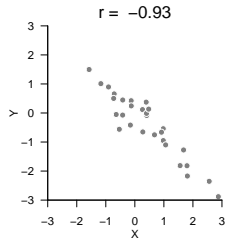
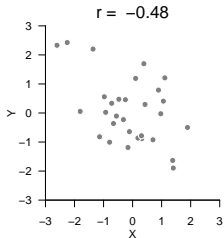
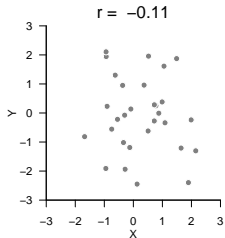
$$|r| < 0.5, 0.5 \leq |r| \leq 0.8, |r| > 0.8 \quad (4)$$

als "schwache," "mittlere," oder "starke" Korrelation bezeichnet

Korrelationsstärken und Streudiagramme



Korrelationsstärken und Streudiagramme



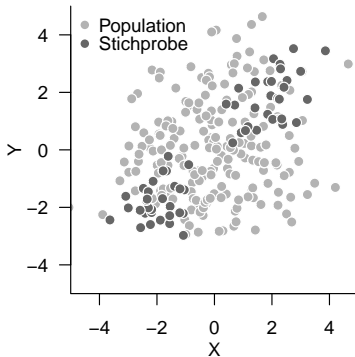
Korrelative Designs

Korrelationskoeffizient

Stichprobenauswahleffekte

Selbstkontrollfragen

Stichprobe aus Extremgruppen

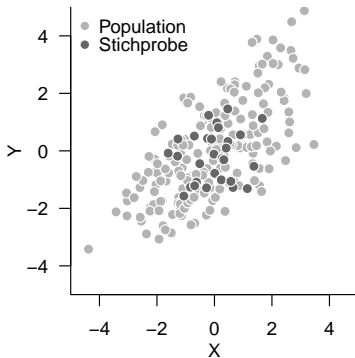


> Korrelation basierend auf Gesamtpopulationsdaten = 0.357

> Korrelation basierend auf Stichprobendaten = 0.919

Stichprobenauswahleffekte

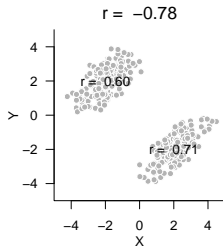
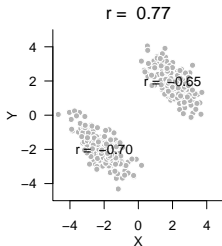
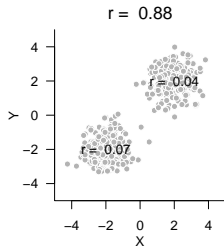
Stichprobe mit zu kleiner Streubreite



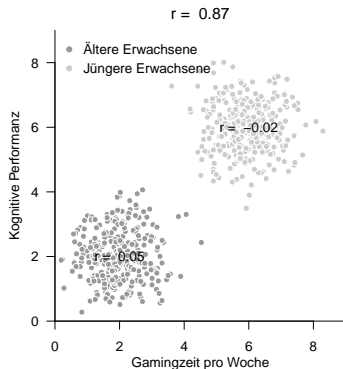
> Korrelation basierend auf Gesamtpopulationsdaten = 0.685

> Korrelation basierend auf Stichprobendaten = 0.186

Bedingte vs. unbedingte Korrelationen (Simpson's Paradox)



Bedingte vs. unbedingte Korrelationen (Simpson's Paradox)



⇒ Durch Drittvariable Alter (Jung, Alt) induzierte Korrelation!

Korrelative Designs

Korrelationskoeffizient

Stichprobenauswahleffekte

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren Sie den Begriff des korrelativen Designs nach Reiß und Sarris (2012).
2. Definieren Sie die Empirische Stichprobenkovarianz.
3. Definieren Sie Pearson's Stichprobenkorrelationskoeffizient.
4. Beschreiben Sie Streudiagramme bivariater Datensätze bei schwachen und stark positiven bzw. negativen Korrelationen.
5. Erläutern Sie die Implikationen einer Stichprobenauswahl aus Extremgruppen.
6. Erläutern Sie die Implikationen einer Stichprobe mit zu kleiner Streubreite.
7. Erläutern Sie die Begriffe der bedingten und der unbedingten Korrelationen.
8. Diskutieren Sie ein Beispiel für eine durch eine Drittvariable induzierte Korrelation.

References

- Fahrmeir, Ludwig, Christian Heumann, Rita Künstler, Iris Pigeot, and Gerhard Tutz. 2016. *Statistik*. Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-50372-0>.
- Reiß, Siegbert, and Viktor Sarris. 2012. *Experimentelle Psychologie: von der Theorie zur Praxis*. Pearson Studium Psychologie. München: Pearson.