

Korrelation

Kovarianz, Produkt-Moment Korrelation, Rangkorrelation

Der Begriff des (bivariaten) Zusammenhangs: Was bedeutet es, wenn zwei Variablen miteinander zusammenhängen?

Die Ausprägung, die eine Person auf der einen Variable aufweist, gibt zu gewissen Teilen auch Auskunft darüber, welche Ausprägung diese Person auf der anderen Variable erreicht. Beide Variablen variieren dann systematisch miteinander.

Die **Stärke des Zusammenhangs** ist davon abhängig, wie zwingend von der einen Variable auf die andere geschlossen werden kann.

Zusammenhänge können in **zwei Richtungen** vorliegen: positiv oder negativ

- **Positiver Zusammenhang:** Hohe Werte auf der einen Variable entsprechen hohen Werten auf der anderen Variable
- **Negativer Zusammenhang:** Hohe Werte auf der einen Variable entsprechen niedrigen Werten auf der anderen Variable

$$\text{Kovarianz (cov}_{x,y}\text{): } cov_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{n - 1}$$

- Unstandardisiertes Maß, dass den Zusammenhang zweier Variablen erfasst (bivariat)
- Ist in ihrer Berechnung der Varianz nicht unähnlich
- Varianz: Erfasst die Abweichungen einer Variable um ihren Mittelwert **und liefert aufgrund der Quadrierung nur positive Ergebnisse**
- Kovarianz: Gleichgerichtete Abweichungen zweier Variablen von deren Mittelwerten, **kann jedoch auch negative Werte annehmen!**
- Wertespektrum: Betrag der Kovarianz zwischen zwei Variablen kann beliebige Werte zwischen 0 und einer maximalen Kovarianz $|cov_{max}|$ annehmen
- $|cov_{max}|$ ist für positive und negative Zusammenhänge identisch und definiert als das Produkt der beiden Merkmalsstreuungen: $|cov_{max}| = \hat{\sigma}_x * \hat{\sigma}_y$
 - o Die Kovarianz ist positiv, wenn positive Abweichungen vom Mittelwert in x mit positiven in y bzw. negativen Abweichungen in x mit negativen Abweichungen in y einhergehen.
 - o Die Kovarianz ist negativ, wenn positive Abweichungen vom Mittelwert in x mit negativen in y einhergehen und umgekehrt.
 - o Eine Kovarianz von 0 besagt, dass beide Variablen in keinem Zusammenhang zueinander stehen.
- Abweichungen, die in der Formel zur Berechnung der Kovarianz eingehen, **sind abhängig von der Skalierung (Maßstab) der Merkmale.**

Korrelation

Kovarianz, Produkt-Moment Korrelation, Rangkorrelation

Produkt-Moment Korrelation

- Standardisiertes Maß für den Zusammenhang zweier intervallskalierter Variablen
- Gebräuchlichstes Maß für die Stärke eines Zusammenhangs
- Stärke des Zusammenhangs wird mit dem Korrelationskoeffizienten r angegeben.
- Anderer Name für die Korrelation: Pearson-Korrelation
- Berechnung als Standardisierung der Kovarianz
- Die berechnete Kovarianz wird anhand der maximalen Kovarianz relativiert
- Dadurch wird die Kovarianz von der Streuung der Merkmale bereinigt.
- Das resultierende r (Korrelationskoeffizient) ist maßstabsunabhängig

$$r_{xy} = \frac{cov_{emp}}{cov_{max}} = \frac{cov_{x,y}}{\hat{\sigma}_x * \hat{\sigma}_y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{(n - 1) * \hat{\sigma}_x * \hat{\sigma}_y}$$

- Der Korrelationskoeffizient kann nur Werte zwischen ± 1 annehmen
 - o $r = 1$ Perfekt positiver Zusammenhang
 - o $r = -1$ Perfekt negativer Zusammenhang
 - o $r = 0$ Kein Zusammenhang (Nullkorrelation)
 - o $r > 0$ Positiver Zusammenhang
 - o $r < 0$ Negativer Zusammenhang
- Einteilung der Größe des Effekts nach Cohen (1992):
 - o $|.10|$ bis $|.30|$ Kleiner Effekt
 - o $|.30|$ bis $|.50|$ Mittlerer Effekt
 - o Ab $|.50|$ Großer Effekt

Rangkorrelation

- Standardisiertes Maß für den Zusammenhang zweier ordinalskalierter Variablen
- Stellt eine Analogie zur Produkt-Moment-Korrelation dar
- **Anstelle intervallskalierter Messwerte werden Rangplätze von ordinalskalierten Rangreihen verwendet**
- Erfasst, inwieweit zwei Rangreihen systematisch miteinander variieren.
- Stärke des Zusammenhangs wird mit dem Korrelationskoeffizienten r_s angegeben.
- Anderer Name für dieselbe Korrelation: Spearman-Korrelation

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

- d_i ist die Differenz der Ranglistenplätze einer Untersuchungseinheit i bezüglich der Variablen x und y

Korrelation als Signifikanztest:

$$t = \frac{r * \sqrt{N - 2}}{\sqrt{1 - r^2}}; df = N - 2$$

Korrelation impliziert keine Kausalität!