Kapitel 4 – Berechnung von Kovarianz und Korrelation

Gerald Echterhoff

Die Kovarianz (*cov*) der Merkmale x und y ist der Mittelwert der Produkte korrespondierender Abweichungen von den jeweiligen Mittelwerten von x und y. Die Formel lautet:

$$\mathrm{cov}_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{N}$$

Jede Untersuchungseinheit i liefert ein Messwertpaar (x_i und y_i). Sind beide Werte z. B. weit überdurchschnittlich (oder unterdurchschnittlich), so ergibt sich ein hohes positives Abweichungsprodukt (der Ausdruck im Zähler). Die Summe der Abweichungsprodukte ist ein Maß für den Grad der **gemeinsamen Variation** (der „Ko-Variation“) der beiden Merkmale. Um eine Vergleichbarkeit mit anderen Stichprobengrößen zu gewährleisten, muss die Summe der Abweichungsprodukte zunächst an der Anzahl aller Fälle (*N*) relativiert werden; mathematisch geschieht dies durch die Division der Summe der Abweichungsprodukte durch *N*.

Die Kovarianz hängt zwar nicht mehr von der Stichprobengröße ab, aber immer noch von der Messeinheit der beiden Variablen. Sie ist noch nicht vollständig standardisiert und lässt daher keinen Vergleich zwischen verschiedenen Stichproben zu. Wird diese Standardisierung vorgenommen, so erhält man die Korrelation (*r*). Hierzu wird die Kovarianz durch das Produkt der Standardabweichungen (*SD*) von x und y dividiert. Die Formel für die Korrelation lautet also:

$$r_{xy} = \frac{\mathrm{cov}_{xy}}{SD_x \cdot SD_y}$$

Der Korrelationskoeffizient *r* (auch Produkt-Moment-Korrelation oder Pearson’scher Korrelationskoeffizient genannt) ist somit die standardisierte Kovarianz. Er variiert zwischen –1 (perfekter negativer Zusammenhang), 0 (kein Zusammenhang) und +1 (perfekter positiver Zusammenhang). Für Daten unterhalb von Intervallskalenniveau gibt es entsprechend angepasste Korrelationskoeffizienten (z. B. der Phi-Koeffizient für nominale Daten, s. weiterführende Literatur im Buch).

Aus: Schreier, M., Echterhoff, G., Bauer, J. F., Weydmann, N. & Hussy, W. (2023). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (3. Aufl.). Springer.