

Aufgabe 1 Häufig wird die Anzahl Zitationen einer wissenschaftlichen Publikation als Indikator für den wissenschaftlichen Erfolg der Autorinnen und Autoren verwendet. Diskutieren Sie die drei Gütekriterien einer Messung an diesem Beispiel! Gehen Sie auf Vor- und Nachteile dieser Messung ein!

Aufgabe 2 Einfache lineare Regression unter Messfehlern

Untersucht werden soll eine Stichprobe $(\xi_1, \eta_1), \dots, (\xi_n, \eta_n)$ eines latenten bivariaten stochastischen Merkmals (ξ, η) , wobei die einfache lineare Regressionsbeziehung

$$\eta_i = \beta_0 + \beta_1 \xi_i + \epsilon_i, \quad i = 1, \dots, n,$$

mit den üblichen Annahmen an die Fehler ϵ_i angenommen werde.

Beobachtet werden die zugehörigen, (potentiell) fehlerbehafteten Daten $(X_1, Y_1) \dots (X_n, Y_n)$, wobei für jedes $i = 1, \dots, n$, gilt:

$$X_i = \xi_i + \delta_i \tag{1}$$

und

$$Y_i = \eta_i + \gamma_i. \tag{2}$$

Dabei seien jeweils die Annahmen des Grundmodells der klassischen Testtheorie erfüllt.

Teil A Illustrieren Sie sich diese Ausgangssituation und die unten betrachteten Spezialfälle

- durch einfache, kurze Simulationsuntersuchungen sowie
- graphisch, indem Sie in einem fiktiven Scatterplot die einzelnen Daten durch Realisationen von Messfehlern der Form

$$\delta_i = \begin{cases} d & \text{mit Wsk } 0.5 \\ -d & \text{mit Wsk } 0.5 \end{cases} \quad \text{bzw.} \quad \gamma_i = \begin{cases} d & \text{mit Wsk } 0.5 \\ -d & \text{mit Wsk } 0.5 \end{cases} \quad i = 1, \dots, n$$

mit festem d „verrauschen“ und dabei d variieren.

Vergleichen Sie für folgende Spezialfälle den auf (ξ_i, η_i) , $i = 1, \dots, n$, beruhenden „Benchmark-Schätzer“ $\hat{\beta}_1$ für β_1 mit dem „naiven Schätzer“ $\hat{\beta}_1^-$, der aus $\hat{\beta}_1$ entsteht, indem man einfach ξ_i durch x_i und η_i durch y_i , $i = 1, \dots, n$, ersetzt.

Teil B (Messfehler nur in abhängigen Variable)

Sei in (1) die Variable $\delta_i \equiv 0$, $i = 1, \dots, n$.

- a) Zeigen Sie, dass dann, falls man ξ_1, \dots, ξ_n als feste Größen auffasst, gilt:

$$\mathbb{E}(\hat{\beta}_1^-) = \mathbb{E}(\hat{\beta}_1). \quad (3)$$

- b) Was bedeutet dies inhaltlich?

- c) Bestätigen Sie (3) durch Einsetzen von Y_i , $i = 1, \dots, n$, in die Regressionsgleichung!

Teil C (Ausschließlicher Kovariablenmessfehler)

Sei nun umgekehrt in (2) die Variable $\gamma_i \equiv 0$, $i = 1, \dots, n$. Für große n gilt ungefähr

$$\hat{\beta}_1^- \approx \left(\frac{\text{Var}(\xi_i)}{\text{Var}(\xi_i) + \text{Var}(\delta_i)} \right) \hat{\beta}_1.$$

- a) Was bedeutet dies inhaltlich?

- b) [Für Hauptfachstudierende der Statistik und der Mathematik ab dem 4. Semester]
Präzisieren Sie obige Aussage und weisen Sie sie nach!

- c) Warum lässt sich die Argumentation aus Teil B c) hier nicht direkt übertragen?