

Zusatzaufgabe 1

Mit dieser Zusatzaufgabe können Sie bis zu 2.5 Zusatzpunkte für die Klausur erwerben. Die Zusatzpunkte werden nur bei *bestandener Klausur* angerechnet. Sie können die Aufgaben in einer Gruppe von bis zu drei Personen bearbeiten. Ihre Lösungen sind bis einschließlich Mittwoch, den 7.12.2022, unter Nennung von Namen und Matrikelnummer aller Beteiligten **als kommentiertes R-Skript per E-Mail einzureichen** (martin.arnold@vwl.uni-due.de).

Aufgabe 1

Der Datensatz `stimmung.csv` enthält Beobachtungen des Wohlbefindens von Eltern (`stimmung`, Skala von 1 bis 100) sowie für die Schlafdauer ihres Babys (`schlaf_baby`, Stunden pro Nacht) für 100 Tage.

- (a) Lesen Sie `stimmung.csv` mit einer geeigneten Funktion aus dem Paket `tidyverse` als Objekt `stimmung` in R ein und erzeugen Sie eine statistische Zusammenfassung der Daten.
- (b) Erzeugen Sie eine statistische Zusammenfassung für die Schätzung des linearen Regressionsmodells

$$\text{stimmung}_i = \beta_0 + \beta_1 \text{schlaf_baby}_i + u_i, \quad (1)$$

und nutzen Sie das Paket `ggplot2`, um den geschätzten Regressionszusammenhang gemeinsam mit den Daten grafisch darzustellen.

- (c) Interpretieren Sie das in (b) geschätzte Modell. Beurteilen Sie die Anpassung des Modells an die Daten.
- (d) Nutzen Sie R-Funktionen, um folgende aus der Übung bekannte Eigenschaften der KQ-Schätzung im allgemeinen Modell $y_i = \beta_0 + \beta_1 X_i + u_i$, $i = 1, \dots, n$ für die Schätzung `stimmung_mod` zu validieren:

1. $\sum_{i=1}^n \hat{u}_i = 0$
2. $\sum_{i=1}^n \hat{u}_i X_i = 0$
3. $\frac{1}{n} \sum_{i=1}^n \hat{y}_i = \frac{1}{n} \sum_{i=1}^n y_i$
4. Die geschätzte Regressionsgerade verläuft durch den Punkt (\bar{X}, \bar{Y}) .
Hinweis: nutzen Sie `predict()`.

- (e) Wiederholen Sie die Teilaufgaben (b) und (c) für das Modell

$$\text{stimmung}_i = \beta_1 \text{schlaf_baby}_i + u_i, \quad i = 1, \dots, 100. \quad (2)$$

- (f) Implementieren Sie den Schätzer für β_1 im einfachen Regressionsmodell $y_i = \beta_1 X_i + u_i$, $i = 1, \dots, n$ als R-Funktion `KQOK()`,

```
KQOK <- function(X, Y) {  
  ...  
}
```

und überprüfen Sie Ihre Funktion anhand eines Vergleich mit dem Ergebnis in (e).

Hinweise: Die Paketsammlung `tidyverse` kann mit `install.packages('tidyverse')` installiert werden. Mit `library(tidyverse)` werden sämtliche für die Zusatzaufgabe nötige R-Pakete geladen. Wir empfehlen den DataCamp-Kurs [Introduction to Regression in R](#).