

Aufgabe 1:

Laden Sie den Datensatz *swiss* aus dem Paket *datasets* und alle weiteren Pakete, die Sie für die Bearbeitung der Aufgaben benötigen. Ziel der Aufgabe ist es, den gegebenen Datensatz deskriptiv zu analysieren und anschließend eine multiple Regression durchzuführen

- (a) Betrachten Sie zunächst alle Variablen im Datensatz und analysieren Sie diese hinsichtlich Ihrer Lage und Streuung. Lassen sich irgendwelche Auffälligkeiten feststellen?
- (b) Stellen Sie als nächstes die Variablen eindimensional graphisch dar. Nutzen Sie hierfür die unterschiedlichen Funktionen, die Sie in den Tutorials rund um den *tidyverse*-Approach kennengelernt haben.
- (c) Führen Sie im Anschluss eine Zusammenhangsanalyse durch und untermauern Sie Ihre Ergebnisse graphisch.
- (d) Nun nutzen Sie die Funktion *lm*, um eine lineare Regressionsanalyse durchzuführen. Bei einer Regressionsanalyse spielt die Auswahl der Variablen eine wesentliche Rolle. Belesen Sie sich hierzu und verwenden Sie eine gängige Methode, um eine Variablenselektion durchzuführen. Als abhängige Variable legen Sie *Fertility* fest. Alle anderen Variablen kommen als erklärende Variablen in Frage. Nachdem Sie sich für ein Modell entschieden haben, nutzen Sie die Funktion *summary*, um sich einen Überblick über das Ergebnis zu verschaffen. Danach verwenden Sie die Funktion *plot*, um die Residuen genauer zu analysieren. Interpretieren Sie die Ergebnisse der Regressions- und Residualanalyse. Weisen die Ergebnisse auf eine Verletzung der Annahmen einer linearen Regression auf Basis der Kleinsten-Quadrate Methode hin?
- (e) Transformieren Sie die abhängige Variable, indem Sie sie logarithmieren. Führen Sie dann die gleiche Analyse wie in Aufgabenteil (d) durch. Gibt es Veränderungen im Vergleich zu Aufgabenteil (d)? Interpretieren Sie die neu erhaltenen Ergebnisse des log-level Modells.
- (f) Bisher fand Ihre Analyse auf Basis der Kleinsten-Quadrate Methode statt. Verwenden Sie nun die Funktion *lmrob* aus dem Paket *robustbase* und schätzen Sie Ihr level-level Modell erneut. Lesen Sie vor Beginn der Auswertung die Hilfe der Funktion *lmrob*. Auf Basis welcher Methode wird nun das Modell geschätzt? Verändert die Verwendung der robusten Methode die Ergebnisse aus Aufgabenteil (d)?