

# Empirische Wirtschaftsforschung

## STAR Projekt

In den späten 1980er Jahren wurde in Tennessee ein Experiment durchgeführt, in dem für 4 Jahre der Effekt der Klassengrösse auf den Lernerfolg von Schülern evaluiert wurde (das sogenannte Student Teacher Achievement Ratio Experiment oder STAR Experiment). Das Experiment verglich die Leistungen von Schülern, die in Klassen von verschiedener Grösse registriert waren. In jeder der teilnehmenden Schulen gab es Klassen von verschiedener Grösse. Innerhalb jeder teilnehmenden Schule wurde nicht nur die Klassengrösse zufällig bestimmt, sondern auch die Kinder zufällig den Klassen zugeteilt. Jedes Jahr mussten die Kinder standardisierte Tests ablegen (sogenanntes SAT). Die im Datensatz `classsize.RData` enthaltenen Variablen sind:

- `id` - ID des Kindes
- `classsize` - Grösse der Klasse, in der das Kind registriert ist
- `tscorek` - Testergebnis des Kindes, gemessen in Punkten
- `sck` - 1 wenn das Kind in einer kleinen Klasse ist; 0 anderfalls
- `boy` - 1 wenn das Kind ein Junge ist; 0 andernfalls
- `freelunk` - 1 wenn das Kind ein kostenloses Mittagessen erhält (Proxy für arme Familienverhältnisse); 0 andernfalls
- `totexpk` - Arbeitserfahrung des Lehrers, gemessen in Jahren

- schidkn - Code der Schule, in dem das Kind ist.

Sie wollen nun den Effekt der Klassengröße auf die Testergebnisse schätzen, weshalb Sie das folgende bivariate Regressionsmodell betrachten:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

Dabei bezeichnet  $y_i$  das Testergebnis,  $x_i$  die Klassengröße und  $u_i$  den Fehlerterm eines Teilnehmers.

1. Laden Sie die Daten “classize.RData” und das R-Skript “R\_Vorlage.R” von Moodle herunter und speichern Sie sie im selben Ordner. Öffnen Sie *R Studio* und öffnen Sie das R-Skript. Laden Sie das Paket “ggplot2” mit dem Befehl `library` (eventuell müssen Sie es vorher mit dem Befehl `install.packages()` installieren) und laden Sie die Daten mit dem Befehl `load`.
2. Verschaffen Sie sich mit den Befehlen `summary()` und `View()` einen ersten Überblick über die Daten. Fallen Ihnen Fehler in den Daten auf? Überprüfen Sie, wie viele Observationen der Datensatz enthält.
3. Berechnen Sie die Mittelwerte und Varianzen von sowie die Kovarianz zwischen `classize` und `tscorek`.
4. Berechnen Sie die geschätzten Koeffizienten  $\hat{\beta}_0$  und  $\hat{\beta}_1$  manuell.
5. Interpretieren Sie den Schätzer für den Achsenabschnitt und den Steigungsparameter.
6. Regressieren Sie die Testergebnisse im Kindergarten auf die Klassengröße mit dem

Befehl `lm()`. Sind die Ergebnisse gleich wie bei manueller Berechnung?

7. Erzielt ein Kind in einer kleineren Klasse notwendigerweise ein besseres Testergebnis?
8. Erklären Sie, was der Fehlerterm  $u_i$  darstellt. Welche Informationen können in diesem enthalten sein?
9. Wie gross ist das R-Quadrat der Regression? Kann die Klassengrösse einen grossen Teil der Varianz der Testergebnisse erklären?
10. Ein Kind besucht eine Klasse mit zehn Schülern - wie hoch ist dessen vorhergesagtes Testergebnis? Verifizieren Sie Ihre Angabe unter der Verwendung des `predict()` Befehls.
11. Diskutieren Sie, ob der gemessene Effekt klein oder gross ist (Hinweis: Benutzen Sie die Standardabweichung der Testergebnisse, um diese Frage zu beantworten).