

# Lineare Modelle

Laden Sie den Datensatz *Theater.txt* von der Tutoriums Homepage.

Variablenbeschreibung:

**Alter:** Alter der Befragten

**Geschl:** Geschlecht der Befragten (0: männl. / 1: weibl.)

**Gehalt:** Jahresgehalt in Tsd Schweizer Franken

**Kultur:** Ausgaben für Kultur in diesem Jahr

**Theater:** Ausgaben für Theater in diesem Jahr

**Theater\_Vorjahr:** Ausgaben für Theater im Vorjahr

## Aufgabe 1: Lineare Modellbildung

- (i) Stellen Sie die Modellgleichung für ein lineares Modell  $M_0$  auf, das die Ausgaben für Theaterbesuche in diesem Jahr durch alle anderen Kovariablen erklärt. Fitten Sie dieses Modell anschließend in R.
- (ii) Eliminieren Sie alle nicht-signifikanten Variablen aus  $M_0$  und fitten Sie ein neues Modell  $M_1$ .
- (iii) Welches dieser beiden Modelle würden Sie bevorzugen? Begründen Sie.

## Aufgabe 2: Faktorisierung

- (i) Formen Sie die Variable **Geschl** zu einer Dummyvariable **Geschl\_fac** mit Referenzkategorie "männl." um. Wenden Sie die Funktion `summary()` darauf an. Was geschieht?
- (ii) Fitten Sie anschließend ein lineares Modell  $M_{neu}$ , indem Sie die Variable **Theater** durch **Geschl\_fac** und alle anderen Einflussgrößen erklären. Interpretieren Sie das Ergebnis kurz!
- (iii) Welche Werte nehmen die Koeffizientenschätzer an, wenn man "weibl." zur Referenzkategorie macht (kurze Begründung!)? Bestätigen Sie Ihre Vermutung mithilfe von R!
- (iv) Plotten Sie die Variablen **Geschl\_fac** und **Theater** auf geeignete Weise gegeneinander. Interpretieren Sie Ihre Graphik!

## Aufgabe 3: Interaktionen und Transformation von Einflussgrößen

- (i) Erweitern Sie  $M_{neu}$  auf geeignete Weise, um zu überprüfen, ob **Geschl\_fac** und **Theater\_Vorjahr** interagieren.
- (ii) Wie verändert sich der Modellfit, wenn Sie
  - statt der absoluten metrischen Einflussgrößen deren Logarithmen verwenden?
  - zu den absoluten Einflussgrößen auch die Wurzel des Gehaltes hinzunehmen?