

## Simulations I

Sie möchten den Zusammenhang zwischen einer skalaren Zielgröße  $y$  und zwei Einflussgrößen mittels eines linearen Regressionsmodells beschreiben. Dabei gilt:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \quad i = 1, \dots, n, \quad \epsilon_i \sim N(0, \sigma^2)$$

- a) Schreiben Sie eine Funktion, die Daten  $(\mathbf{y}, \mathbf{X})$  gemäß des obigen Modells für beliebige Werte von  $n, \sigma^2, \beta_0, \beta_1$  und  $\beta_2$  erzeugt. Dabei soll gelten:

$$x_{i1} \sim B(0.5) \quad \text{und} \quad x_{i2} \sim N(0, 1), \quad i = 1, \dots, n$$

Die Funktion soll einen Datensatz zurückgeben, der die Spalten  $y$ ,  $x_1$  und  $x_2$  enthält.  
Testen Sie Ihre Funktion durch einen geeigneten Funktionsaufruf.

- b) Schreiben Sie eine Funktion, die ein lineares Modell der Daten aus a) berechnet und den MSE  $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$  der Schätzung zurückgibt.  
Testen Sie Ihre Funktion durch einen geeigneten Funktionsaufruf.

- c) Schreiben Sie eine Funktion, die für eine beliebige Kombination der Parameter  $n, \sigma^2, \beta_0, \beta_1, \beta_2$  und eine beliebige Anzahl an Wiederholungen (`n_repetitions`) Daten gemäß des obigen Modells simuliert und den MSE für alle Wiederholungen berechnet. Output der Funktion soll ein Vektor aller MSEs sein.  
Testen Sie Ihre Funktion durch einen geeigneten Funktionsaufruf.

- d) Schreiben Sie eine Funktion, die als Input jeweils einen Vektor an Parametern für  $n, \sigma^2, \beta_0, \beta_1$  und  $\beta_2$  verlangt und für alle möglichen Parameterkombinationen und eine beliebige Anzahl an Wiederholungen (`n_repetitions`) Daten simuliert und den MSE berechnet.  
Output der Funktion soll eine Liste sein, die zwei Elemente enthält. Das erste Listenelement soll 'parameter\_combinations' heißen und den Datensatz aller untersuchten Parameterkombinationen enthalten. Das zweite Listenelement soll eine Matrix sein, die in jeder Spalte alle Wiederholungen zu einer Parameterkombination enthält.  
Testen Sie Ihre Funktion durch einen geeigneten Funktionsaufruf.

- e) Führen Sie die Funktion aus d) mit von Ihnen sinnvoll gewählten Parametern aus.

- f) Zeichnen Sie für jede Parameterkombination einen Boxplot der MSEs und vergleichen Sie die Ergebnisse.