Simulations

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 `sim_data()`, die Daten $(\mathbf{y}, \mathbf{X})$ gemäß des obigen Modells für beliebige Werte von $n, \sigma^2, \beta_0, \beta_1$ und β_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, x1 und x2 enthält.

b) Schreiben Sie eine Funktion `est_model()`, 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.

c) Schreiben Sie eine Funktion `get_MSEs()`, die für beliebige (skalare) Parameter $n, \sigma^2, \beta_0, \beta_1, \beta_2$ und eine beliebige Anzahl an Wiederholungen (n_rep) Daten gemäß des obigen Modells simuliert und den MSE für alle Wiederholungen berechnet. Output der Funktion soll ein Vektor aller MSEs sein.

d) Führen Sie die Funktion aus c) mit verschiedenen Stichprobenumfängen $n \in \{100, 200, \dots, 1000\}$ und festen Parametern $\sigma^2 = 2$, $\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 1.5$ sowie $n_{rep} = 200$ durch und speichern Sie die Ergebnisse als Matrix.

e) Visualisieren Sie Ihre Ergebnisse aus d), z.B. mithilfe von Boxplots, und vergleichen Sie diese.

Hinweis: Es ist hilfreich, wenn Sie jede einzelne Funktion jeweils durch geeignete Funktionsaufrufe testen.

** Durch Lösen dieser Übungsaufgabe kann ein Teil der Prüfungsleistung erbracht werden.*

Abgabetermin: 05.07.2016/06.07.2016