**Statistik 1**
Prof. Dr. Jan Johannes
Sergio Brenner Miguel
Sommersemester 2019

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

# 02. Übungsblatt

| Aufgabe 1 | Aufgabe 2 | Aufgabe 3 | Aufgabe 4 | Σ |
|---|---|---|---|---|
| | | | | |

## Aufgabe 1 (4 Punkte)

Bei acht Absolvent*innen werden anhand einer Befragung die Anzahl der Mitstudierenden und das Einstiegsgehalt ermittelt:

| Anzahl Mitstudierende $x_i$ | 125 | 100 | 160 | 100 | 185 | 145 | 220 | 185 |
|---|---|---|---|---|---|---|---|---|
| Einstiegsgehalt $Y_i$ | 80 | 100 | 75 | 80 | 110 | 110 | 135 | 130 |

a) Modellieren Sie die Beobachtungen mit Hilfe eines linearen Modells und bestimmen jene Gerade, die den quadratischen Fehler minimiert. Zeichnen Sie die Daten und die berechnete Regressionsgerade in ein geeignetes Koordinatensystem ein.

b) Es stellt sich heraus, dass die ersten Vier ein anderes Fach studiert haben als die anderen Vier. Bestimmen und zeichnen Sie die Regressionsgeraden für beide Studienfächer getrennt.

c) Wie erklären Sie sich die unterschiedlichen Ergebnisse in a) und b)? Formulieren Sie in Kürze Ihre Beobachtungen.

d) Verwenden Sie den Varianzschätzer aus dem Gauß-Markov Satz zur Schätzung der Varianz in den drei Modellen.

## Aufgabe 2 (4 Punkte)

Betrachten Sie für $\mu \in \mathbb{R}$ das Lokationsmodell

$$Y_i = \mu + \varepsilon_i, \qquad i \in [\![1, n]\!]$$

wobei $\varepsilon_1, \dots, \varepsilon_n$ stochastisch unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}[\varepsilon_1] = 0$ und $\mathbb{V}\mathrm{ar}[\varepsilon_1] = \sigma^2 > 0$ sind.

a) Bestimmen Sie den gewöhnlichen Kleinste Quadrate Schätzer $\widehat{\mu}$ für $\mu$ und zeigen Sie dessen schwache Konsistenz, d.h. $\widehat{\mu} \xrightarrow{\mathbb{P}} \mu$.

b) Bestimmen Sie das $c \in \mathbb{R}$, für welches der Schätzer $\tilde{\mu} := c \cdot \bar{Y}_n$, $\bar{Y}_n := \frac{1}{n}\sum_{i=1}^n Y_i$, den mittleren quadratischen Fehler $\mathrm{MSE}_\mu(\tilde{\mu}) := \mathbb{E}[(\tilde{\mu} - \mu)^2]$ minimiert. Zeigen Sie für diese optimale Wahl von $c$, dass $\mathrm{MSE}_\mu(\tilde{\mu}) < \mathrm{MSE}_\mu(\widehat{\mu})$ gilt.
*Hinweis:* Vereinfachen Sie den mittleren quadratischen Fehler und interpretieren sie diesen als Funktion in $c$, die Sie daraufhin minimieren können.

c) Vergleichen Sie die Varianzen der Schätzer $\tilde{\mu}$, mit optimal gewählten $c$ und $\widehat{\mu}$. Steht diese Aussage im Widerspruch zum Gauß-Markov-Theorem ?

**Aufgabe 3 (Frisch-Waugh Theorem; 4 Punkte)**

Sei $Y \circledcirc \mathcal{L}(X\beta, \sigma^2 \mathrm{Id}_n), \beta \in \mathbb{R}^p, \sigma > 0$ durch ein gewöhnliches lineares Modell mit $X = (X_1, X_2)$ und $\beta^t = (\beta_1^t, \beta_2^t)$ für $\beta_i \in \mathbb{R}^{p_i}, p_i > 0$, und $X_i \in \mathbb{R}^{n \times p_i}, i = 1, 2$, sowie $n \geq p_1 + p_2$ adäquat beschrieben. Die Matrix $X$ besitze den Rang $\mathrm{rg}[(X_1, X_2)] = p_1 + p_2$ und die Koordinaten der Störgröße $\varepsilon := Y - X\beta$ seien unabhängig und identisch verteilt. Wir untersuchen den gewöhnlichen Kleinste Quadrate Schätzer $\widehat{\beta}^t = (\widehat{\beta}_1^t, \widehat{\beta}_2^t)$ von $\beta^t = (\beta_1^t, \beta_2^t)$. Dazu sei $\Pi_{\mathcal{R}(X_1)}$ die Orthogonalprojektion auf $\mathcal{R}(X_1) = \{X_1 b : b \in \mathbb{R}^{p_1}\}$ und $\widetilde{X}_2 := \Pi_{\mathcal{R}(X_1)^\perp}(X_2)$ sowie $\widetilde{Y} = \Pi_{\mathcal{R}(X_1)^\perp}(Y)$. Zeigen Sie die folgenden Aussagen

a) $(\widetilde{X}_2^t \widetilde{X}_2)$ ist invertierbar;

b) $\widehat{\beta}_2 = (\widetilde{X}_2^t \widetilde{X}_2)^{-1} \widetilde{X}_2^t Y$;

c) $\widehat{\beta}_2$ ändert sich nicht, wenn $Y$ durch $\widetilde{Y}$ersetzt wird, so dass $\widehat{\beta}_2 = (\widetilde{X}_2^t \widetilde{X}_2)^{-1} \widetilde{X}_2^t \widetilde{Y}$ gilt.

**Aufgabe 4 (4 Punkte)**

Betrachten Sie das folgende lineare Modell

$$Y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \varepsilon_i, \qquad i \in [\![1, n]\!].$$

wobei $x_{\bullet,1}, x_{\bullet,2} \in \mathbb{R}^n$, $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$ unbekannt und $\varepsilon_1, \dots \varepsilon_n$ unabhängig, identisch verteilte Zufallsvariablen mit $\mathbb{E}[\varepsilon_1] = 0$ und $\mathbb{V}\mathrm{ar}[\varepsilon_1] = \sigma^2 > 0$.

a) Unter welcher Voraussetzung an die Vektoren $x_{\bullet,1}, x_{\bullet,2}$ ist der gewöhnliche kleinste Quadrate Schätzer wohldefiniert? Nehmen Sie von nun an, dass diese Bedingung erfüllt ist. Bestimmen Sie nun den gewöhnlichen kleinste Quadrate Schätzer $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$ und berechnen Sie die Varianz von $\widehat{\beta}_2$.

b) Betrachten Sie zusätzlich das folgende lineare Modell

$$Y_i^* = \beta_2 x_{i,2} + \varepsilon^*, \qquad i \in [\![1, n]\!],$$

wobei $\varepsilon_1^*, \dots, \varepsilon_n^*$ stochastisch unabhängige Zufallsvariablen mit $\mathbb{E}[\varepsilon_1^*] = 0$ und $\mathbb{V}\mathrm{ar}[\varepsilon_1^*] = \sigma^2 > 0$. Bestimmen Sie in diesem Modell den gewöhnlichen kleinste Quadrate Schätzer $\widehat{\beta}_2^*$ von $\beta_2$ und zeigen Sie, dass $\mathbb{V}\mathrm{ar}[\widehat{\beta}_2] \geq \mathbb{V}\mathrm{ar}[\widehat{\beta}_2^*]$. Überlegen Sie sich, in welchen Fällen die Gleichheit $\mathbb{V}\mathrm{ar}[\widehat{\beta}_2] = \mathbb{V}\mathrm{ar}[\widehat{\beta}_2^*]$ gelten würde.

---

**Abgabe:**
In Zweiergruppen, bis spätestens Dienstag, den **06. Mai 2019, 11:15 Uhr**.
(Die Übungszettelkästen sind im 1. OG, INF 205, vor dem Dekanat.)

**Homepage:**
https://sip.math.uni-heidelberg.de/vl/st1-ss19/