



1. Übungsblatt

Aufgabe 1	Aufgabe 2	Aufgabe 3	Aufgabe 4	Summe:

Übungsgruppe:

Tutor(in):

Namen:

Aufgabe 1 (Anwendung der linearen Regression, 4 = 3 + 1 Punkte).

In einigen Bundesstaaten der USA wurde in die Mortalität und der Breitengrad erfasst, um die Abhängigkeit der durch Hautkrebs verursachten Todesfälle von der Sonneneinstrahlung zu bestimmen. Folgende Daten liegen vor:

Staat	Delaware	Iowa	Michigan	New Hampshire	Oklahoma	Texas	Wyoming
Mortalität	200	128	117	129	182	229	134
Breite	39	42	44	44	35	31	43

Wir nehmen an, dass die Mortalität Y den linearen Zusammenhang

$$Y_i = a + bX_i + \varepsilon_i,$$

mit dem Breitengrad X besitzt. Hierbei sind $a, b \in \mathbb{R}$ unbekannte Parameter und $\varepsilon_1, \dots, \varepsilon_n$ stochastisch unabhängige Zufallsvariablen („Messfehler“) mit $\mathbb{E}\varepsilon_1 = 0$ und $\text{Var}(\varepsilon_1) = \sigma^2$.

- Berechnen Sie den Kleinst-Quadrate-Schätzer $\hat{\beta} = (\hat{a}, \hat{b})'$ und fertigen Sie eine Skizze an, in der die Datenpunkte sowie die erhaltene Regressionsgerade enthalten sind.
- Bestimmen Sie eine Prognose $\hat{Y} = \hat{a} + \hat{b}\hat{X}$ für die Mortalität in Ohio (Breitengrad $\hat{X} = 40$).

Aufgabe 2 (Überbestimmtes Modell, Identifizierbarkeit, 4 = 1 + 2 + 1 Punkte).

Betrachten Sie das lineare Modell

$$Y_i = \beta_1 v_i + \beta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

wobei $x, v \in \mathbb{R}^n$ deterministische (bekannte) Vektoren, $\beta_1, \beta_2 \in \mathbb{R}$ unbekannte Parameter und $\varepsilon_1, \dots, \varepsilon_n$ stochastisch unabhängige Zufallsvariablen (iid) mit $\mathbb{E}\varepsilon_1 = 0$ und $\text{Var}(\varepsilon_1) = \sigma^2$ sind. Sei $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$ der Kleinst-Quadrate-Schätzer von $\beta = (\beta_1, \beta_2)'$.

(a) Welche Bedingung muss an x, v gestellt werden, damit β identifizierbar ist?

Wir nehmen im Folgenden die Bedingungen von (a) an.

(b) Berechnen Sie $\hat{\beta}$ und bestimmen Sie damit $\text{Var}(\hat{\beta}_2)$.

(c) Zeigen Sie, dass

$$\text{Var}(\hat{\beta}_2) \geq \text{Var}(\hat{\beta}_2^*), \quad (*)$$

wobei $\hat{\beta}_2^*$ der Kleinste-Quadrate-Schätzer im Modell

$$Y_i^* = \beta_2 x_i + \varepsilon_i^*$$

ist. Hierbei sind $\varepsilon_1^*, \dots, \varepsilon_n^*$ iid Zufallsvariablen mit $\varepsilon_1 = 0$ und $\text{Var}(\varepsilon_1) = \sigma^2$. In welchen Fällen gilt in (*) Gleichheit? Interpretieren Sie (*) unter diesem Gesichtspunkt.

Aufgabe 3 (Das Lokationsmodell, 4 = 1 + 2 + 1 Punkte).

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

$$Y_i = \mu + \varepsilon_i, \quad i = 1, \dots, n,$$

wobei $\varepsilon_1, \dots, \varepsilon_n$ iid Zufallsvariablen mit $\mathbb{E}\varepsilon_1 = 0$ und $\text{Var}(\varepsilon_1) = \sigma^2$ seien.

(a) Bestimmen Sie den Kleinste-Quadrate-Schätzer $\hat{\mu}$ für μ und zeigen Sie dessen Konsistenz. *Erinnerung: Sie müssen $\hat{\mu} \xrightarrow{p} \mu$ nachweisen.*

(b) Bestimmen Sie das $c \in \mathbb{R}$, für welches der Schätzer $\tilde{\mu} := c \cdot \bar{Y}_n$ (wobei $\bar{Y}_n := \frac{1}{n} \sum_{i=1}^n Y_i$) den mittleren quadratischen Fehler $\text{MSE}_\mu(\tilde{\mu}) := \mathbb{E}[(\tilde{\mu} - \mu)^2]$ minimiert. Zeigen Sie, dass für diese Wahl von c , $\text{MSE}_\mu(\tilde{\mu}) < \text{MSE}_\mu(\hat{\mu})$ gilt.

Hinweis: Für die Bestimmung von c vereinfachen Sie zunächst den Ausdruck $\mathbb{E}[(\tilde{\mu} - \mu)^2]$ und minimieren Sie diesen dann in c mit analytischen Mitteln.

(c) Vergleichen Sie $\text{Var}(\tilde{\mu})$ und $\text{Var}(\hat{\mu})$. Entsteht ein Widerspruch zum Gauß-Markov-Theorem?

Aufgabe 4 (Verteilung des KQ-Schätzers bei normalverteilten Fehlern, 4 = 1 + 1.5 + 1 + 0.5 Punkte).

In dieser Aufgabe betrachten wir ein allgemeines lineares Modell

$$Y = X\beta + \varepsilon,$$

wobei $X = (x_1, \dots, x_n)' \in \mathbb{R}^{n \times k}$ deterministisch und bekannt, $\beta \in \mathbb{R}^{k \times 1}$ der unbekannte Parameter und $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ ein Vektor von iid $\mathcal{N}(0, \sigma^2)$ verteilten Zufallsvariablen sei. Wir nehmen im Folgenden an, dass $\text{Rang}(X) = k$.

(a) Zeigen Sie: Für den Kleinste-Quadrate-Schätzer $\hat{\beta}$ gilt $\hat{\beta} - \beta \sim \mathcal{N}(0, \sigma^2(X'X)^{-1})$.

Hinweis: Nutzen Sie die bekannten Transformationsregeln für multivariate Normalverteilungen: Ist $Z \sim \mathcal{N}(\mu, \Sigma)$ mit $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d \times d}$ positiv semidefinit, dann gilt $AZ + b \sim \mathcal{N}(A\mu + b, A\Sigma A')$ für $A \in \mathbb{R}^{k \times d}$ und $b \in \mathbb{R}^k$.

(b) Ein Schätzer für σ sei durch $\hat{\sigma}^2 := \frac{1}{n-k}(Y - X\hat{\beta})'(Y - X\hat{\beta})$ gegeben. Zeigen Sie, dass $(n-k)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-k}^2$ -verteilt ist und $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$.

Hinweis:

- Zeigen Sie zuerst, dass $(Y - X\hat{\beta}) = (I_n - \Pi_X)\varepsilon$, wobei $\Pi_X := X(X'X)^{-1}X'$.

- Nutzen Sie ohne Beweis, dass $I_n - \Pi_X = S\Lambda S'$, wobei $\Lambda = \text{diag}(1, \dots, 1, 0, \dots, 0)$ mit $n - k$ Einsen und S orthogonal.
 - Definieren Sie $\tilde{\varepsilon} := S'\varepsilon$. Berechnen Sie die Verteilung von $\tilde{\varepsilon}$ und stellen Sie $Y - X\hat{\beta}$ durch $\tilde{\varepsilon}$ dar.
 - Sind Z_1, \dots, Z_r iid $\mathcal{N}(0, 1)$ -verteilt, so gilt $\sum_{i=1}^r Z_i^2 \sim \chi_r^2$
- (c) Zeigen Sie, dass $\hat{\sigma}^2$ und $\hat{\beta}$ stochastisch unabhängig sind.
 Hinweis: Zeigen Sie zuerst, dass $\Pi_X = S\tilde{\Lambda}S'$ eine Diagonalisierung mit S aus (b) besitzt und $\tilde{\Lambda} = \text{diag}(0, \dots, 0, 1, \dots, 1)$ mit k Einsen. Stellen Sie dann $\hat{\beta}$ und $Y - X\hat{\beta}$ durch verschiedene Komponenten von $\tilde{\varepsilon}$ dar.
- (d) Zeigen Sie, dass für beliebiges $a \in \mathbb{R}^k, a \neq 0$, $\frac{a'(\hat{\beta} - \beta)}{\hat{\sigma}(a'(X'X)^{-1}a)^{1/2}} \sim t_{n-k}$ gerade t -verteilt ist.
 Hinweis: Sind U, S unabhängig mit $U \sim \mathcal{N}(0, 1)$, $S^2 \sim \chi_m^2$, dann besitzt $T := \frac{U}{(\frac{1}{m}S^2)^{1/2}}$ eine t -Verteilung mit m Freiheitsgraden (Notation: $T \sim t_m$).

Abgabe:

In Zweiergruppen, bis spätestens Donnerstag, den **12.11.2020, 11:00 Uhr**.

Bitte geben Sie Ihre Lösungsvorschläge im PDF-Format ab. Es reicht, wenn einer der beiden Personen das Dokument hochlädt.

Homepage der Vorlesung:

<https://ssp.math.uni-heidelberg.de/stat-ws2020/index.html>