Übungen zur Vorlesung

Mathematische Statistik

Sommersemester 2013

Institut für Mathematik

Jun.-Prof. Dr. Thorsten Dickhaus
RUD25, Raum 1.203
E-Mail: dickhaus@math.hu-berlin.de
Tel.: 030/2093-5841

Übungen: Mathias Trabs
E-Mail: trabs@math.hu-berlin.de
Tel.: 030/2093-3988

Blatt 6

Abgabe bis Dienstag, 21. Mai 2013, 11:15 Uhr
Jede komplett richtig gelöste Aufgabe ergibt 4 Punkte.
Bitte jede Aufgabe auf einem separaten **Blatt** bearbeiten, Danke!

Aufgaben

21. **Einfache lineare Regression.** Betrachten Sie das Modell

$$\forall 1 \le i \le n : \quad Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i \tag{1}$$

der einfachen linearen Regression mit fixem Design und $x_i \in \mathbb{R}$ für alle $1 \le i \le n$. Dabei seien die Fehlerterme $(\varepsilon_i)_{1\le i\le n}$ stochastisch unabhängig und identisch verteilt mit $\varepsilon_1 \sim \mathcal{N}(0, \sigma^2)$ für unbekanntes $\sigma^2 > 0$. Wir schreiben abkürzend $SSX := \sum_{i=1}^n (x_i - \bar{x}_n)^2$ und setzen $SSX > 0$ voraus.

(a) Berechnen Sie durch analytisches Betrachten der Likelihoodfunktion des Gesamtexperimentes den KQ-Schätzer bzw. den MLE $(\hat{\theta}_0, \hat{\theta}_1)^\top$ für $(\theta_0, \theta_1)^\top$. Nach Aufgabe 22.(b) stimmen diese beiden Schätzer unter Modell (1) überein.

(b) Zeigen Sie:

$$\hat{\theta}_0 \sim \mathcal{N}\left(\theta_0, \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\, SSX}\right), \; \hat{\theta}_1 \sim \mathcal{N}\left(\theta_1, \frac{\sigma^2}{SSX}\right)$$

und leiten Sie daraus asymptotische Konfidenzbereiche für θ_0 und θ_1 ab.

22. **Lineare Regression.**

(a) Lösen Sie Exercise 3.3.1 im Skript.

(b) Betrachten Sie das lineare Regressionsmodell

$$Y_i = \Psi_i^\top \theta^* + \varepsilon_i, \quad 1 \le i \le n \tag{2}$$

und zeigen Sie: Gilt unter Modell (2), dass die Fehlerterme $(\varepsilon_i)_{1\le i\le n}$ stochastisch unabhängig und identisch verteilt sind mit $\varepsilon_1 \sim \mathcal{N}(0, \sigma^2)$, so stimmen der KQ-Schätzer und der MLE für θ^* überein.

23. **Verallgemeinerte Regressionsmodelle.**

    (a) Lösen Sie Exercise 3.7.1 im Skript.

    (b) Lösen Sie Exercise 3.7.2 im Skript.

    (c) Lösen Sie Exercise 3.8.1 im Skript.

24. **Programmieraufgabe.** Die folgende Tabelle enthält Daten über die Anzahl von Stunden, die acht Studierende in einem Zeitraum von drei Wochen vor einer Prüfung zur Vorbereitung aufgewendet haben, sowie ihre Punktzahlen, die sie am Ende in der Prüfung erreicht haben.

Lernzeit in Stunden	20	16	34	23	27	32	18	22
Erreichte Punktzahl	64	61	84	70	88	92	72	77

    Modellieren Sie den Zusammenhang der beiden Merkmale mit der einfachen linearen Regression, siehe Aufgabe 21.

    (a) Geben Sie die Daten in `R` ein.

    (b) Bestimmen Sie die Kleinste-Quadrate–Regressionsgerade des linearen Modells der erreichten Punktzahl in Abhängigkeit von der Lernzeit und tragen Sie diese in ein Streudiagramm der Daten ein. Zeichnen Sie die Residuen mit ein.

    (c) Schätzen Sie in `R` die Regressionskoeffizienten. Führen Sie eine visuelle Residualanalyse durch.

    (d) Prognostizieren Sie mit Hilfe der ermittelten Regressionsgleichung aus Teilaufgabe (c) das Prüfungsergebnis eines / einer Studierenden, der / die sich 30 Stunden für das Lernen Zeit nimmt. Verwenden Sie dazu die Funktion `predict.lm`.