Institut für Angewandte Mathematik
Sommersemester 2016
Dr. Robert Philipowski, Eva Kopfer, Angelo Profeta

"Einführung in die Statistik"

Übungszettel Nr. 11

Abgabe am 06.07.16

Aufgabe 1 (Simpson Paradoxon) [*5 Punkte*]

Folgende Tabelle zeigt Dauer des Studiums (in Semestern) und Einstiegsgehalt (in 1000 Euro) der Absolventen eines Jahres am Fachbereich Mathematik und Informatik der (hypothetischen) Yule-Universität:

Semester	10	9	11	9	11	12	10	11
Gehalt	35	35	34	36	41	39	40	38

a) Schlägt sich (für diese Absolventen) ein längeres Studium in einem höheren Anfangsgehalt nieder? Bestimmen Sie die Regressionsgerade für Studiendauer gegen Anfangsgehalt.

b) Ändert sich Ihr Befund, wenn Sie zusätzlich erfahren, dass die ersten vier Einträge der Tabelle sich auf die Absolventen des Fachs Informatik, die letzten vier sich auf die Absolventen des Fachs Mathematik beziehen, und Sie dieselbe Regression jeweils innerhalb dieser beiden Gruppen durchführen?

Aufgabe 2 (Lineare Modelle) [*5 Punkte*]

Es seien $\varepsilon_1 \dots, \varepsilon_n$ i.i.d. $\mathcal{N}(0,1)$ verteilt. Außerdem seien $x_1, \dots, x_n$ bekannte reelle Zahlen und $\sigma > 0$ und $\gamma_1, \gamma_2, \gamma_3$ unbekannte reelle Zahlen. Welche der folgenden Modelle sind linear? Geben Sie gegebenenfalls die Designmatrix A an. Für $i \in \{1, \dots, n\}$ sei

$$X_i = \gamma_1 + \gamma_2 x_i^2 + \sigma\varepsilon_i, \tag{1}$$
$$X_i = \gamma_1 + \gamma_2 x_i + \sigma\varepsilon_i^2, \tag{2}$$
$$X_i = e^{\gamma_1} e^{\gamma_2 x_i} x_i^{\gamma_3} e^{\sigma\varepsilon_i}, \tag{3}$$
$$X_i = \gamma_1 + \gamma_1 e^{\gamma_2} x_i + \sigma\varepsilon_i. \tag{4}$$

Finden Sie bei den nicht linearen Modellen, falls möglich, eine Variablentransformation, so dass das transformierte Modell linear ist und geben Sie die dazugehörige Designmatrix A an.

Aufgabe 3 (Autoregressives Modell) [*5 Punkte*]

Zur Beschreibung zeitlicher Entwicklungen mit deterministischer Wachstumstendenz und zufälligen Störungen verwendet man oft das folgende autoregressive Modell:

$$X_k = \gamma X_{k-1} + \sqrt{v}\xi_k, \quad 1 \leq k \leq n.$$

Dabei sind $\gamma \in \mathbb{R}$ und $v > 0$ unbekannte Parameter, $X_0, \ldots, X_n$ die Beobachtungen und $\xi_1, \ldots, \xi_n$ unabhängige zufällige Störungen mit $\mathbb{E}(\xi_k) = 0, \mathrm{Var}(\xi_k) = 1$.

a) Im allgemeinen linearen Modell ist der quadratische Fehler $|Y - C\beta|^2$. Machen Sie einen Ansatz für den quadratischen Fehler, und bestimmen Sie den Kleinste-Quadrate-Schätzer für γ. Ist dieser Schätzer erwartungstreu?

b) Zeigen Sie: Im Fall von standardnormalverteilten $\xi's$ und Startvariable $X_0 = 0$ lautet die Likelihood-Funktion für $\theta = (\gamma, v)$

$$L_\theta = (2\pi v)^{-n/2} \exp\left[-\frac{\sum_{k=1}^n (X_k - \gamma X_{k-1})^2}{2v}\right].$$

Aufgabe 4 (KQ-Schätzer) [*5 Punkte*]

Zeigen Sie, dass der Kleinste-Quadrate-Schätzer im allgemeinen Gauß'schen Regressionsmodell $X = A\gamma + \varepsilon$ mit dem Maximum-Likelihood-Schätzer übereinstimmt.