Prof. Dr. U. Stadtmüller
C. Hering

Sommer 2010
29.06.2010
Blatt 11

Angewandte Statistik für Biometrie

(Abgabe: Di., 06.07.2010, 13:15 Uhr, vor den Übungen)

1. (a) Im folgenden seien ε_t, $t \in \mathbb{Z}$ unabhängige und identisch verteilte Zufallsvariablen mit Erwartungswert 0 und Varianz σ^2. Untersuche die folgenden Zeitreihen auf Stationarität und gib im Falle der Stationarität die Autokovarianzfunktion an.
   i. $X_t = \varepsilon_1 \cos(ct) + \varepsilon_2 \sin(ct)$, $t \in \mathbb{Z}$
   ii. $X_t = \varepsilon_t \cos(ct) + \varepsilon_{t-1} \sin(ct)$, $t \in \mathbb{Z}$
   iii. $X_t = \varepsilon_t - X_{t-1}$ mit $X_0 = 0$, $t \in \mathbb{Z}$.

   (b) Gegeben sei die folgende Zeitreihe.

$$X_t = \xi + \sum_{i=1}^{p} \alpha_i X_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \beta_j \varepsilon_{t-j},$$

   wobei ε_t unabhängige und identisch verteilte Zufallsvariablen mit Mittelwert 0 und Varianz σ^2 seien. Nimm an, dass $\sum_{i=1}^{p} \alpha_i \neq 1$ ist, gehe davon aus, dass die Zeitreihe X_t stationär ist und berechne unter dieser Voraussetzung den Mittelwert $\mathbb{E}(X_t)$.

(5 Punkte)

2. Gegeben sei das mixed-effect model aus der Vorlesung, d.h.

$$X_{ij\nu} = \mu + \alpha_i + b_j + c_{ij} + \varepsilon_{ij\nu}, \quad 1 \le i \le k,\ 1 \le j \le l,\ 1 \le \nu \le m,\ n = k \cdot l \cdot m$$

   wobei $\vec{\varepsilon} \sim N(\vec{0}, \sigma_\varepsilon^2 I_n)$, $\vec{b} \sim N(\vec{0}, \sigma_b^2 I_l)$ und $\vec{c} \sim N(0, \Sigma_c)$ unabhängig sein sollen. Für die Kovarianzmatrix von c soll gelten: $\mathrm{Var}(c_{ij}) = \frac{k-1}{k}\sigma_c^2$, $\mathrm{Cov}(c_{i_1 j_1}, c_{i_2 j_2}) = 0$, falls $j_1 \neq j_2$ und $\mathrm{Cov}(c_{i_1 j_1}, c_{i_2 j_2}) = -\frac{1}{k}\sigma_c^2$, falls $i_1 \neq i_2$ und $j_1 = j_2$.
   Zeige, dass dann gilt: $\mathrm{Cov}(\overline{X_{i_1..}}, \overline{X_{i_2..}}) = \frac{1}{l}(\sigma_b^2 - \frac{1}{k}\sigma_c^2)$, falls $i_1 \neq i_2$.

(5 Punkte)

3. Der Datensatz `bryk.dat` enthält die Ergebnisse eines Mathetest unter 7185 Schülern durchgeführt an amerikanischen Highschools im Jahre 1982. Neben der Nummer der Schule, und den Ergbnissen (`mathach`) enthält der Datensatz auch einen sogenannten Sozioökonmischen Faktor (`ses`, 0-zentriert), sowie die Variable (`sector`), die angibt, ob es sich um eine Schule handelt, die vom Staat getragen wird, oder ob die katholische Kirche die Trägerschaft der Schule inne hat. Laden den Datensatz `bryk.dat` von der Veranstaltungshomepage.

   - Passe für jeden Sektor je ein lineares Modell an den Datensatz an. Hierbei soll der Faktor `ses` als zufällige und der Faktor `school` als deterministische Einflussgröße ins Modell einfließen.
   - Erzeuge und veranschauliche mittels zweier plots (einen für jeden Sektor) die Konfidenzintervalle der geschätzten y-Achsenabschnitte und Steigungen für alle Schulen. Interpretiere und kommentiere die resultierenden Graphiken.
   - Erzeuge für die geschätzten y-Achsenabschnitte und Steigung für jeden Sektor jeweils einen Boxplot und vergleiche die Ergebnisse zwischen den Sektoren.

   Hinweis: Das Paket `nlme` mit seinen Befehlen: `lmList` und `intervals` kann verwendet werden.

(5 Punkte)

http://www.uni-ulm.de/mawi/zawa/lehre/sommer2010/asb2010.html