

Stochastik III

(Abgabe: Do., 16.12.2010, 14:15 Uhr, vor den Übungen)

1. Für $A \in \mathbb{R}^{n \times m}$, $\text{rg}(A) = m \leq n$ und $\beta \in \mathbb{R}^m$ sei das lineare Modell $X = A\beta + \varepsilon$ mit $\mathbb{E}(\varepsilon) = 0$ und $\text{Cov}(\varepsilon) = \sigma^2 C$, mit einer symmetrischen, positiv definiten Matrix C gegeben. Man berechne den Least-Squares-Schätzer $\hat{\beta}$ von β bezüglich der Norm $\|x\|_C = \sqrt{x^T C^{-1} x}$. Außerdem zeige man, dass der Schätzer $\hat{\beta}$ erwartungstreu ist und berechne seine Kovarianzmatrix.

(4 Punkte)

2. Der Datensatz `galapagos.dat` enthält Daten von 29 verschiedenen Galapagos Inseln. Zu jeder Insel ist jeweils die Anzahl von verschiedenen Schildkrötenarten (`species`), die Größe der Insel in km^2 (`area`), der höchste Punkt der Insel in m (`elevation`), der Abstand zur nächstgelegenen Insel in km (`nearest`), der Abstand zur Insel Santa Cruz in km (`scruz`) sowie die Größe der Nachbarinsel in km^2 (`adjacent`) gegeben.

- (a) Setze in \mathbb{R} ein lineares Modell auf, welches die Anzahl der vorkommenden Schildkrötenarten als abhängige Variable hat.
- (b) Gib die Schätzung des Parametervektors an.
- (c) Mache eine Vorhersage für die Anzahl an Schildkrötenarten auf der Insel Santa Maria, welche eine Fläche von 170.92 km^2 , eine Höhe von 640 m, einen Abstand von 2.6 km zur nächsten Insel (diese hat eine Größe von 0.1 km^2) und einen Abstand von 49.2 km zur Insel Santa Cruz hat.

(2 + 1 + 2 Punkte)

3. Gegeben sei das sogenannte Cobb-Douglas Modell. Dieses wird verwendet um das Produktionsvolumen einer Volkswirtschaft zu beschreiben.

$$Y_i = e^{\beta_1} \cdot K_i^{\beta_2} \cdot L_i^{\beta_3} \cdot \varepsilon_i,$$

wobei Y das Bruttosozialprodukt, K den Kapitaleinsatz und L den Arbeitseinsatz angebe. Die Residuen ε_i sind iid mit $\mathbb{E}(\varepsilon_i) = 1$ und $\text{Var}(\varepsilon_i) = \sigma^2$.

- (a) Transformiere das Cobb-Douglas Modell so, dass ein lineares Modell entsteht.
- (b) Der Datensatz `prod.dat` enthält die Angaben des Bruttosozialprodukts (in Mrd. Euro), des Kapitaleinsatzes (in Mrd. Euro) und des Arbeitseinsatzes (in Tsd. Erwerbstätigen) in Deutschland für die Jahre 1991-2008. Passe das lineare Modell aus (a) auf diese Daten an. Gib die Residuen, sowie die geschätzten Parameter an.
- (c) Im Jahr 2009 gab es in Deutschland 40,172 Mio. Erwerbstätige und der Kapitaleinsatz betrug 11 904,45 Mrd. Euro. Schätze den Wert des Bruttosozialproduktes mit Hilfe des Modells und vergleiche ihn mit dem tatsächlichen Wert von 2430,94 Mrd. Euro.

(1 + 2 + 2 Punkte)

4. Der Datensatz `auto.dat` enthält die Bremswege (in m) bei bestimmten Geschwindigkeiten (in km/h) eines bestimmten Autos. Der Bremsweg hängt bekanntermaßen von der Geschwindigkeit ab. Passe die folgenden linearen Modelle in `R` auf die Daten an:

(i) $s = \beta_1 + \beta_2 v$

(ii) $s = \beta_1 + \beta_2 v + \beta_3 v^2$

(iii) $s = \beta_1 v + \beta_2 v^2$

Erstelle in ein Schaubild (z.B. mittels des Befehls `par`) jeweils ein Streudiagramm der Daten, welches jeweils zusätzlich die geschätzte Regressionsfunktion enthält. Berechne außerdem für die verschiedenen Ansätze das Bestimmtheitsmaß $B = 1 - \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{x^T x - n \bar{x}^2}$, welches anschaulich angibt wieviel Prozent der Gesamtvariation durch das gefittete Modell erklärt wird. Entscheide basierend auf B welches Modell am Besten geeignet ist.

(6 Punkte)

<http://www.uni-ulm.de/mawi/mawi-stukom/lanzinger/stochiiws10.html>