

Stochastik III - Übungsblatt 4

Abgabe: 05. 12. 2012 vor Beginn der Übung

Aufgabe 1 (4 Punkte)

Gegeben sei das lineare Modell $Y = X\beta + \varepsilon$ mit $\mathbb{E}\varepsilon = o$, $\text{Cov}(\varepsilon) = \sigma^2 I_n$ und $\text{rg}(X) = m$. Sei $\hat{\beta}$ der Kleinste-Quadrate-Schätzer für β und $\hat{\varepsilon} = Y - \hat{Y}$ mit $\hat{Y} = X\hat{\beta}$ der Residuenvektor. Zeigen Sie, dass $\text{Cov}(\hat{\beta}, \hat{\varepsilon}) = \mathbb{E} \left((\hat{\beta} - \mathbb{E}\hat{\beta})(\hat{\varepsilon} - \mathbb{E}\hat{\varepsilon})^\top \right) = o$.

Aufgabe 2 (4 Punkte)

Betrachten Sie den Spezialfall $m = 2$ des multiplen linearen Regressionsmodells, d.h. $Y = X\beta + \varepsilon$ mit

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Zeigen Sie, dass dann der Kleinste-Quadrate-Schätzer $(\hat{\beta}_1, \hat{\beta}_2)^\top = (X^\top X)^{-1} X^\top Y$ dem aus Stochastik I bekannten Schätzer im einfachen linearen Regressionsmodell entspricht, d.h.

$$\hat{\beta}_2 = \frac{s_{xy}^2}{s_{xx}^2}, \quad \hat{\beta}_1 = \bar{y}_n - \hat{\beta}_2 \bar{x}_n,$$

wobei \bar{x}_n bzw. \bar{y}_n die Stichprobenmittel bezeichnen, d.h.

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{bzw.} \quad \bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i,$$

und die Stichprobenvarianz s_{xx}^2 bzw. die Stichprobenkovarianz s_{xy}^2 gegeben sind durch

$$s_{xx}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \quad \text{bzw.} \quad s_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n).$$

Aufgabe 3 (3 + 2 + 2 Punkte) Das Produktionsvolumen der USA zwischen 1932 und 1953 lässt sich mit Hilfe der Cobb-Douglas Produktionsfunktion

$$Y_t = e^{\beta_1} \cdot K_t^{\beta_2} \cdot A_t^{\beta_3} \cdot \varepsilon_t, \quad t \in \{1, \dots, 22\}$$

mit unbekanntenen Konstanten $\beta_1, \beta_2, \beta_3$ beschreiben. Dabei bezeichnet Y die Produktion (in Mrd. Dollar), K den Kapitaleinsatz (in Mrd. Dollar) und A den Arbeitseinsatz (in Mill. Arbeitskräften). Auf der Homepage der Vorlesung finden Sie die Datei *production.dat*, die folgende Daten enthält.

Jahr	t	Produktion	Kapital	Arbeit
1932	1	60.3	297.1	39.3
1933	2	58.2	290.1	39.6
1934	3	64.4	285.4	42.7
1935	4	75.4	287.8	44.2
1936	5	85.0	282.1	47.1
1937	6	92.7	300.3	48.2
1938	7	85.4	301.4	46.4
1939	8	92.3	305.6	47.8
1940	9	101.2	313.3	49.6
1941	10	113.3	327.4	54.1
1942	11	107.8	339.0	59.1
1943	12	105.2	347.1	64.9
1944	13	107.1	353.5	66.0
1945	14	108.8	354.1	64.4
1946	15	131.5	359.4	58.9
1947	16	130.9	359.3	59.3
1948	17	134.7	365.2	60.2
1949	18	129.1	363.2	58.7
1950	19	147.8	373.7	60.0
1951	20	152.1	386.0	63.8
1952	21	154.3	396.5	64.9
1953	22	159.9	408.0	66.0

- (a) Führen Sie den Modellansatz in ein geeignetes lineares Modell über und bestimmen Sie den Kleinste-Quadrate-Schätzer für $\beta = (\beta_1, \beta_2, \beta_3)$.
- (b) Geben Sie die geschätzten Werte \hat{Y}_t sowie die Residuen $\hat{\varepsilon}_t$ an.
- (c) Erstellen Sie ein Schaubild, das die tatsächliche Entwicklung des Produktionsvolumens Y über die Jahre 1932 – 1953 darstellt, sowie die Verlaufskurve der geschätzten Daten.
Hinweis: Mit dem Befehl `lines()` kann ein weiterer Polygonzug zu einem Schaubild hinzugefügt werden.

Aufgabe 4 (5 + 2 Punkte) Der Bremsweg y eines PKWs (in m) hängt von der Geschwindigkeit v (in km/h) ab. Für einen bestimmten Fahrzeugtyp wurde bei 20 Bremsversuchen der Bremsweg in Abhängigkeit von der Geschwindigkeit gemessen. Drei Modelle zur Beschreibung des Zusammenhangs zwischen v und y sollen verglichen werden:

- i) $y = \beta_1 + \beta_2 v$,
- ii) $y = \beta_1 + \beta_2 v + \beta_3 v^2$ und
- iii) $y = \beta_1 v + \beta_2 v^2$.

Die gemessenen Werte sind in den Dateien `distance.txt` und `velocity.txt` enthalten, die Sie auf der Homepage der Vorlesung finden können.

- a) Bestimmen Sie für jedes Modell den Kleinsten-Quadrate-Schätzer für die Parameter β_i und erstellen Sie je ein Streudiagramm der Daten, welches die geschätzte Regressionsfunktion enthalten soll. Lassen Sie sich die drei Grafiken zum besseren Vergleich auf derselben Grafikoberfläche anzeigen.
Hinweis: Der Befehl `par()` ermöglicht die Kontrolle über die Grafikoberfläche, z.B. `par(mfrow = c(2,2))`.

- b) Berechnen Sie für jeden Ansatz die sogenannte Bestimmtheitsmasszahl

$$R^2 = 1 - \frac{\hat{\varepsilon}^\top \hat{\varepsilon}}{y^\top y - n\bar{y}^2}.$$

Entscheiden Sie mit Hilfe der Grösse R^2 , welcher Ansatz der Beste zu sein scheint.