Prof. Dr. U. Jensen
H. Höfling

20. Juni 2002

Übungen zu Wirtschaftsstatistik

Sommersemester 2002
Blatt 8
Abgabe am 27.6.2002 vor den Übungen

1. In zehn verschiedenen amerikanischen Wintersportorten wurden während einer gewissen Beobachtungszeit die Anzahl der Besucher registriert. Es wird angenommen, daß diese linear von der Gesamtlänge der zur Verfügung stehenden Pisten sowie der Liftkapazität abhängt.

Skigebiet	Pistenlänge Tkm (x_2)	Liftkapazität TPers. (x_3)	Besucherzahl TPers. (y)
1	10,5	2,2	19,93
2	2,5	1	5,84
3	13,1	3,25	23,7
4	4,0	1,5	9,88
5	14,7	3,8	30,01
6	3,6	1,2	7,24
7	7,1	1,9	11,63
8	22,5	5,6	45,68
9	17,0	4,2	36,48
10	6,4	1,85	12,07

   (a) Erstellen Sie ein lineares Modell mit Normalverteilungsannahme. (1)

   (b) Bestimmen Sie Schätzwerte für die Regressionskoeffizienten β und für σ^2. (3)

   (c) Errechnen Sie das Bestimmtheitsmaß des Modells. (2)

   (d) Testen Sie jeweils getrennt, ob die Koeffizienten für die Pistenlänge und die Liftkapazität 0 sind ($\alpha = 0,05$). Geben Sie für beide Regressionskoeffizienten Konfidenzintervalle an. (3)

   Hinweis: Einige Werte:

$$(X^TX)^{-1} = \begin{pmatrix} 0,69349 & 0,18067 & -0,91529 \\ 0,18067 & 0,1315 & -0,57135 \\ -0,91529 & -0,57135 & 2,53161 \end{pmatrix} \quad X^TY = \begin{pmatrix} 202,46 \\ 2848,847 \\ 717,7075 \end{pmatrix}$$

$$\sum_{i=1}^{n} \hat{\epsilon}_i^2 = 19,137 \quad \sum_{i=1}^{n} y_i^2 = 5742,03$$

2. Eine Fabrik in A-Land hat verschiedene Zulieferbetriebe in A-Land und in B-Land (welches von A-Land durch ein kleines Meer getrennt ist). Es wurde die Dauer gemessen, die zwischen Abgabe der Bestellung und Eintreffen der Ware verging. Hier die ermittelten Daten:

Nummer	Entfernung (x_2) (in Tkm)	Land	Dauer (y) (in d)
1	0,6	A	1,5
2	1,3	A	3
3	0,5	A	1,5
4	1	A	2,5
5	0,8	A	2
6	1,2	B	4
7	0,9	B	3,5
8	1,5	B	4,5
9	1,4	B	4,5
10	1,3	B	4

Es wird zunächst angenommen, daß die für die Fahrtdauer folgender Zusammenhang gilt: Fahrtdauer = β_1 + Entfernung $\cdot\beta_2$.

$$\sum_{i=1}^{n} x_i = 10,5 \quad \sum_{i=1}^{n} y_i = 31 \quad \sum_{i=1}^{n} x_i^2 = 12,09$$
$$\sum_{i=1}^{n} y_i^2 = 108,5 \quad \sum_{i=1}^{n} x_i y_i = 35,85 \quad \sum_{i=1}^{n} \hat{\epsilon}_i^2 = 2,175$$

(a) Stellen Sie das lineare Modell auf (Normalverteilungsannahme). (1)

(b) Bestimmen Sie Schätzer für β und σ^2. (2)

(c) Berechnen Sie das Bestimmtheitsmaß für dieses Modell. (2)

(d) Testen Sie die Hypothese „$\beta_1 = 0$“auf einem 5%-Niveau und geben Sie ein Konfidenzintervall an. (2)

(e) Zeichnen Sie die Regressionsgerade und die Messwerte in einen Graphen. Erstellen Sie zwei Scatterplots für die Residuen, wobei Sie einmal auf der x-Achse die Entfernung, beim anderen Mal das Herkunftsland auftragen. Was ist festzustellen? (3)

Offenbar sind die Residuen bei den Frachten aus B-Land positiv und bei A-Land negativ. Man vermutet deshalb, daß sich aufgrund der Zeit auf der Fähre eine zusätzliche konstante Wartezeit ergibt. Man führt deshalb eine Indikatorvariable (Dummyvariable) in die Regression ein. Für Lieferungen aus A-Land gilt weiterhin:
Fahrtdauer = β_1 + Entfernung $\cdot\beta_2$.
Aus B-Land aber: Fahrtdauer = β_1 + Entfernung $\cdot\beta_2$ + β_3.

(f) Stellen Sie das erweiterte lineare Modell auf (Normalverteilungsannahme). (1)

(g) Berechnen Sie die neuen Regressionskoeffizienten. (2)

(h) Berechnen Sie das Bestimmtheitsmaß. (2)

(i) Testen Sie die Hypothese „$\beta_3 = 0$“auf einem 5%-Niveau und geben Sie ein Konfidenzintervall für β_3 an. (3)

Für das erweiterte Modell gelten die folgenden Zahlenwerte:

$$(X_E^T X_E)^{-1} = \begin{pmatrix} 1,3308 & -1,3462 & 0,3654 \\ -1,3462 & 1,6026 & -0,6731 \\ 0,3654 & -0,6731 & 0,6827 \end{pmatrix} \quad X_E^T Y = \begin{pmatrix} 31 \\ 35,85 \\ 20,5 \end{pmatrix} \quad \sum_{i=1}^{n} \hat{\epsilon}_i^2 = 0,0923$$