Übung 2: Klassisches lineares Modell 2

Aufgabe 1

Es wird ein lineares Modell mit sechs zu schätzenden Parametern $\beta_0, \ldots, \beta_5$ betrachtet.

a) Ermitteln Sie den Wert der F-Statistik für die Testsituation

$$H_0 : (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T = \mathbf{0} \quad \text{gegen} \quad H_1 : (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T \neq \mathbf{0}$$

und den Fall, dass $n = 40$ Beobachtungen vorliegen und das Bestimmtheitsmaß $R^2 = 0{,}2$ beträgt. Beurteilen Sie anschließend, ob die Nullhypothese H_0 bei einem Signifikanzniveau von $\alpha = 0{,}05$ zu verwerfen ist.

b) Es liegen nun $n = 400$ Beobachtungen vor. Beurteilen Sie, ob die Nullhypothese H_0 jetzt zu verwerfen ist.

c) Der KQ-Schätzer für $\boldsymbol{\beta}$ sei nun durch $\widehat{\boldsymbol{\beta}} = (2, 2, 3, 3, 4, 1)^T$ gegeben. Berechnen Sie das 95%-Konfidenzintervall für die Linearkombination $\mathbf{c}^T \widehat{\boldsymbol{\beta}}$ mit $\mathbf{c}^T = (1, 2, 1, 4, 5, 3)$ und $\widehat{\sigma}\left(\mathbf{c}^T \widehat{\boldsymbol{\beta}}\right) = 0{,}9$ für den Fall, dass $n = 40$ Beobachtungen vorliegen.

d) Der KQ-Schätzer für $\boldsymbol{\beta}$ sei wieder durch $\widehat{\boldsymbol{\beta}} = (2, 2, 3, 3, 4, 1)^T$ gegeben und $n = 40$. Ferner gelte für den Vektor mit den Werten der erklärenden Variablen $\mathbf{x}_* = (1, 1, 3, 2, 2, 1)^T$ und $\widehat{\sigma}\left(\mathbf{x}_*^T \widehat{\boldsymbol{\beta}}\right) = 0{,}3$. Als Schätzung für den Varianzparameter (mittlerer quadratischer Fehler) σ^2 liegt der Wert $\widehat{\sigma}^2 = 4$ vor. Berechnen Sie mittels diesen Angaben das $(1-\alpha)$-Prognoseintervall für y zum Signifikanzniveau $\alpha = 0{,}05$.

Aufgabe 2

Für ein lineares Modell der Form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

liegen folgende Beobachtungen vor:

y	x_1	x_2
120	3	10
108	5	7
92	-2	3
61	1	-12
198	-5	21
21	-2	-29

a) Prüfen Sie bei einem Signifikanzniveau von $\alpha = 0{,}025$, ob mindestens eine der beiden unabhängigen Variablen einen signifikanten Einfluss auf die Zielvariable hat.

b) Führen Sie einen beidseitigen Test für die Nullhypothese

$$H_0 : \beta_1 = -1 \quad \wedge \quad \beta_2 = 1 \qquad \text{gegen} \qquad \beta_1 \neq -1 \quad \vee \quad \beta_2 \neq 1$$

durch. Kann die Nullhypothese bei einem Signifikanzniveau von $\alpha = 0{,}025$ abgelehnt werden?

Aufgabe 3

Es wird wieder die Situation aus Aufgabe 5 des 1. Übungsblatts betrachtet.

a) Ermitteln Sie eine Schätzung für die Varianz-Kovarianzmatrix der Residuen $\widehat{\boldsymbol{\varepsilon}}$.

b) Berechnen Sie die standardisierten Residuen r_i für $i = 1, 2, 3$ und beurteilen Sie damit und unter Verwendung einer gängigen Daumenregel, ob es sich bei den Beobachtungen y_1, y_2, y_3 um Ausreißer handelt.

c) Beurteilen Sie anhand einer gängigen Daumenregel für die Hebelwerte, ob die Beobachtungen y_1, y_2, y_3 selbst-schätzend sind. Erläutern Sie ferner, ob die Anwendung dieser Daumenregel in diesem Fall sinnvoll ist.

Aufgabe 4

Ein Unternehmen möchte mithilfe eines klassischen linearen Modells untersuchen, ob die Ausgaben für Marketing einen signifikanten Einfluss auf den Umsatz des Unternehmens haben. Für die Daten aus den letzten 6 Jahren ergaben sich folgende Werte:

Umsatz y	Marketingkosten x
10	5
13	7
13	6
5	3
30	12
9	4

a) Bestimmen Sie den KQ-Schätzer $\widehat{\boldsymbol{\beta}}$.

b) Das Unternehmen vermutet, dass sich innerhalb der Daten Ausreißer befinden, die das Ergebnis verzerren könnten. Untersuchen Sie, ob man aufgrund der standardisierten Residuen eine der Beobachtungen als Ausreißer betrachten sollte.

c) Untersuchen Sie die Hebelwerte der einzelnen Beobachtungen. Prüfen Sie mit Hilfe der Cook-Distanz und einer bekannten Daumenregel, ob es sich bei der Beobachtung mit dem größten Hebelwert, um eine einflussreiche Beobachtung handelt?