

12. Übungsblatt zur Vorlesung
Schließende Statistik WS 2022/23

Aufgabe 44

Man gehe davon aus, dass sich die Abhängigkeit des systolischen Blutdrucks y_i vom Lebensalter x_i eines Menschen durch das einfache lineare Regressionsmodell

$$y_i = \beta_1 + \beta_2 x_i + u_i \quad \text{mit} \quad u_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \quad i = 1, \dots, n$$

erklären lässt.

In einer medizinischen Studie wurden bei $n = 62$ Personen die Merkmale Alter x_i und systolischer Blutdruck y_i erhoben und daraus für die Durchführung einer einfachen linearen Regressionsanalyse bereits die folgenden Zwischenwerte errechnet:

$$\begin{aligned} \sum_{i=1}^{62} y_i &= 9193; & \sum_{i=1}^{62} y_i^2 &= 1384977; & \sum_{i=1}^{62} x_i &= 2882; \\ \sum_{i=1}^{62} x_i^2 &= 148292; & \sum_{i=1}^{62} x_i y_i &= 441517 \end{aligned}$$

- Schätzen Sie β_1 und β_2 mit Hilfe der Kleinst-Quadrate-Methode.
- Berechnen Sie das zugehörige Bestimmtheitsmaß R^2 .
- Schätzen Sie die Varianz σ^2 mit einer erwartungstreuen Schätzfunktion und berechnen Sie $\widehat{\sigma^2_{\hat{\beta}_1}}$ sowie $\widehat{\sigma^2_{\hat{\beta}_2}}$.
- Testen Sie zum Signifikanzniveau $\alpha = 0.001$ (!), ob das Alter einer Person einen signifikanten Einfluss auf den systolischen Blutdruck hat.
- Geben Sie ein zweiseitiges Konfidenzintervall zum Niveau $1 - \alpha = 0.95$ für β_1 an.
- Testen Sie zum Signifikanzniveau $\alpha = 0.05$, ob sich β_1 signifikant von 100 unterscheidet.
- Steigt der systolische Blutdruck eher mit zunehmendem oder eher mit abnehmendem Alter? Begründen Sie Ihre Antwort.
- Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für den systolischen Blutdruck einer 50-jährigen Person an.
- Geben Sie ein Prognoseintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für den Erwartungswert des systolischen Blutdrucks einer 50-jährigen Person an.

Aufgabe 45

In einer einfachen Stichprobe von $n = 42$ britischen Haushalten wurden die wöchentlichen Ausgaben für Kleidung (y) und das jeweilige Haushaltseinkommen (x) erhoben (jeweils in £). Die Schätzung des einfachen linearen Regressionsmodells

$$y_i = \beta_1 + \beta_2 x_i + u_i \quad \text{mit} \quad u_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \quad i = 1, \dots, 42$$

mit der Statistik-Software R produzierte folgende Ausgabe:

```
Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-19.796  -6.174  -3.158   5.663  28.739

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.38913     3.83720   0.101  0.91973
x             0.07383     0.02645   2.791  0.00801 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.26 on 40 degrees of freedom
Multiple R-squared:  0.163,    Adjusted R-squared:  0.1421
F-statistic:  7.79 on 1 and 40 DF,  p-value: 0.008012
```

- Geben Sie die realisierten Kleinst-Quadrate-Schätzwerte für β_1 und β_2 an.
- Geben Sie den realisierten Schätzwert für σ^2 an.
- Welcher Anteil der Gesamtvarianz der Ausgaben für Kleidung wird durch das lineare Modell erklärt?
- Entscheiden Sie mit Hilfe des zugehörigen p -Werts zum Signifikanzniveau $\alpha = 0.10$, ob β_1 signifikant von Null verschieden ist.
- Geben Sie ein Konfidenzintervall zur Vertrauenswahrscheinlichkeit $1 - \alpha = 0.95$ für β_1 an.
- Überprüfen Sie mit einem geeigneten Test zum Niveau $\alpha = 0.01$, ob das Haushaltseinkommen einen signifikant positiven Einfluss auf die Ausgaben für Kleidung hat.
- Welche wöchentlichen Ausgaben für Kleidung (in £) prognostiziert das Modell für einen Haushalt mit einem Haushaltseinkommen von 150 £?