

Statistische Methoden der
Datenanalyse
Beispielsammlung

Übung 8

W. Waltenberger, R. Frühwirth

Institut für Hochenergiephysik
der Österreichischen Akademie der Wissenschaften
A-1050 Wien, Nikolsdorfer Gasse 18

Wintersemester 2017/2018

Übung 8

Beispiel 8.1

Berechnen Sie die Kovarianzmatrix der Residuen $\mathbf{r} = \mathbf{Y} - \hat{\mathbf{Y}}$ des linearen Modells

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{Cov}[\boldsymbol{\varepsilon}] = \mathbf{V}$$

mittels linearer Fehlerfortpflanzung.

Beispiel 8.2

In einem einfachen linearen Regressionsmodell der Form $y_i = \alpha + \beta x_i$, $i = 1, \dots, n$ werden die folgenden Größen definiert:

$$\begin{aligned} \text{Erklärbare Streuung} \quad SS^* &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ \text{Reststreuung,,} \quad SS_R &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ \text{Totale Streuung,} \quad SS_T &= \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

a) Zeigen Sie, dass die folgende **Streuungszerlegung** gilt:

$$SS_T = SS^* + SS_R$$

b) Das **Bestimmtheitsmaß** B der Regression ist definiert durch $B = SS^*/SS_T$. Zeigen Sie, dass gilt:

$$B = r_{xy}^2$$

wo r_{xy} der empirische Korrelationskoeffizient von x und y ist.

Beispiel 8.3

Für 25 Betriebe wurde die Anzahl x der Mitarbeiter und die Höhe der Aufwendungen für Fortbildungskosten y ermittelt.

	x	y
Mittelwert	150	2000
Varianz	625	250000

Die Korrelation zwischen x und y beträgt 0.65. Wie lautet die Gleichung der Regressionsgeraden? Berechnen Sie die Prognose \hat{y} für $x = 200$ und geben Sie den Prognosefehler an.

Beispiel 8.4 (Prog)

Eine Versicherung interessiert sich für die Abhängigkeit der Schadenswahrscheinlichkeit von der Fahrleistung. Folgende Tabelle zeigt Daten einer Erhebung der jährlichen Fahrleistung x (in 1000 km) und der Schadenshäufigkeit y (in Promille):

x	5	10	15	20	25	30	40	50
y	9	10	14	18	22	24	29	29

- Stellen Sie die Punkte in einem Streudiagramm dar und berechnen Sie den empirischen Korrelationskoeffizienten von x und y .
- Ermitteln Sie die Regressionsgerade.
- Welche Schadenshäufigkeit ist für Fahrer mit 35000 km jährlicher Fahrleistung zu erwarten?

Beispiel 8.5 (Prog)

Die folgende Tabelle enthält die Zahl der Sonnenflecken, die zwischen 1970 und 1983 festgestellt wurden sowie die Anzahl der Verkehrstoten durch Autounfälle in diesen Jahren in den USA in tausend. Testen Sie die Hypothese, dass die Zahl der Verkehrstoten nicht von der Zahl der Sonnenflecken beeinflusst wird. Interpretieren Sie Ihr Ergebnis.

Jahr	Sonnenflecken	Verkehrstote
1970	165	54.6
1971	89	53.3
1972	55	56.3
1973	34	49.6
1974	9	47.1
1975	30	45.9
1976	59	48.5
1977	83	50.1
1978	109	52.4
1979	127	52.5
1980	153	53.2
1981	112	51.4
1982	80	46.0
1983	45	44.6

(Quelle: Sheldon M. Ross, Statistik f. Ingenieure und Naturwissenschaftler)