Statistik 2, Übung 13

Henry Haustein

Aufgabe 1

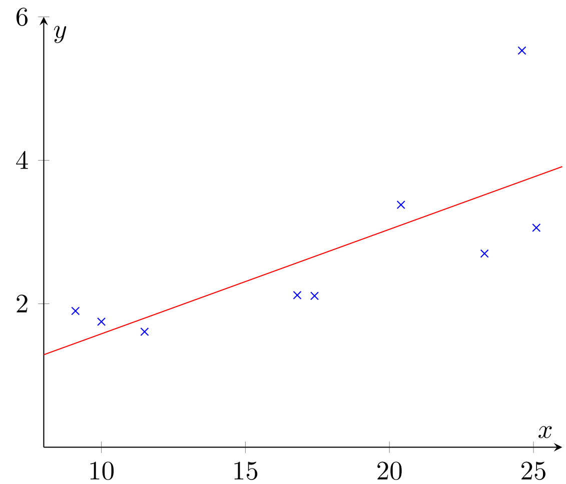
(a) Die Formeln zur Berechnung der Koeffizienten bei der linearen Regression sind gegeben durch

$$\begin{aligned}\hat{\beta}_1 &= \frac{\tilde{s}_{xy}}{\tilde{s}_x^2} \\ &= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2} \\ &= \frac{\frac{1}{9}\cdot 470.341 - 17.5778\cdot 2.6844}{\frac{1}{9}\cdot 3094.28 - 17.5778^2} \\ &= \frac{5.0743}{34.8298} \\ &= 0.1457 \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x} \\ &= 2.6844 - 0.1457\cdot 17.5778 \\ &= 0.124\end{aligned}$$

(b) Mit *Schwerpunkt* ist der Punkt $(\bar{x},\bar{y}) = (17.5778, 2.6844)$ gemeint, die Überprüfung liefert

$$\begin{aligned}\bar{y} &\stackrel{?}{=} \hat{\beta}_0 + \hat{\beta}_1\cdot\bar{x} \\ &\stackrel{?}{=} 0.124 + 0.1457\cdot 17.5778 \\ &\stackrel{?}{=} 2.685 \approx \bar{y}\end{aligned}$$

(c) Graph



(d) Ich komme hier zwar auf eine andere Varianz der Residuen, $\hat{\sigma}^2 = 0.7661$, was auf keine besonders gute Regressionsgerade hindeutet. R^2 ist hier 0.5536.

(e) Wir führen einen rechtsseitigen Test durch:
$H_0 : \hat{\beta}_1 \leq 0.1$
$H_1 : \hat{\beta}_1 > 0.1$
Die Teststatistik ergibt sich zu

$$\begin{aligned} T &= \frac{\hat{\beta}_1 - 0.1}{\sqrt{\hat{\sigma}^2}} \sqrt{n} \sqrt{\mathrm{Var}(x)} \\ &= \frac{0.1457 - 0.1}{\sqrt{0.7661}} \sqrt{9} \sqrt{34.8306} \\ &= 0.9237 \end{aligned}$$

Der kritische Wert ist $t_{n-1;1-\alpha} = t_{8;0.99} = 2.9980$. Die Nullhypothese kann damit nicht abgelehnt werden.

(f) Einsetzen von $x = 0$ liefert $y = 0.124 + 0.1457 \cdot 0 = 0.124$.

(g) Einsetzen von $x = 14.5$ liefert $y = 0.124 + 0.1457 \cdot 14.5 = 2.2361$. Das Konfidenzintervall ist dann gegeben durch

$$\begin{aligned} KI &= \hat{y} \mp \hat{\sigma} \sqrt{1 + \frac{1}{n}\left(1 + \frac{(x - \bar{x})^2}{\mathrm{Var}(x)}\right)} \\ &= 2.2361 \mp \sqrt{0.776} \sqrt{1 + \frac{1}{9}\left(1 + \frac{(14.5 - 17.5778)^2}{34.8306}\right)} \\ &= [0.0250; 4.4472] \end{aligned}$$

Aufgabe 2

Im ersten Modell haben wir

$$\begin{aligned}
\hat{\alpha}_1 &= \frac{\tilde{s}_{xy}}{\tilde{s}_x^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2} \\
&= \frac{\frac{1}{10}\cdot 58.759 - 4.51 \cdot 0.823}{\frac{1}{10}\cdot 415.51 - 4.51^2} \\
&= \frac{2.1642}{21.2109} \\
&= 0.1020 \\
\hat{\alpha}_0 &= \bar{y} - \hat{\alpha}_1 \bar{x} \\
&= 0.823 - 0.1020 \cdot 4.51 \\
&= 0.3630
\end{aligned}$$

Im zweiten Modell haben wir

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\tilde{s}_{xy}}{\tilde{s}_x^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2} \\
&= \frac{\frac{1}{10}\cdot 20.285 - 1.8311 \cdot 0.823}{\frac{1}{10}\cdot 45.1 - 1.8311^2} \\
&= \frac{0.5215}{1.1571} \\
&= 0.4507 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
&= 0.823 - 0.4507 \cdot 1.8311 \\
&= -0.0024
\end{aligned}$$

(a) Das Bestimmtheitsmaß ist definiert durch

$$\begin{aligned}
R^2 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
R_1^2 &= 0.938665 \\
R_2^2 &= 0.9993375
\end{aligned}$$

Das zweite Modell hat eine höhere Erklärkraft.

(b) Wir testen
$H_0 : \hat{\beta}_0 = 0$
$H_1 : \hat{\beta}_0 \neq 0$

Die Teststatistik ergibt sich zu

$$\begin{aligned} T &= \frac{\hat{\beta}_1 - 0}{\sqrt{\hat{\sigma}^2}\sqrt{1 + \frac{\bar{x}^2}{\mathrm{Var}(X)}}}\sqrt{n} \\ &= \frac{-0.0024 - 0}{\sqrt{0.0001948}\sqrt{1 + \frac{1.8311^2}{1.1569}}}\sqrt{10} \\ &= -0.2796 \end{aligned}$$

Der kritische Wert ist $t_{n-1;1-\frac{\alpha}{2}} = t_{10;0.995} = 3.3554$. Die Nullhypothese kann damit nicht abgelehnt werden.

(c) Wir testen:
$H_0 : \hat{\beta}_1 = 0.4522$
$H_1 : \hat{\beta}_1 \neq 0.4522$
Die Teststatistik ergibt sich zu

$$\begin{aligned} T &= \frac{\hat{\beta}_1 - 0.4522}{\sqrt{\hat{\sigma}^2}}\sqrt{n}\sqrt{\mathrm{Var}(x)} \\ &= \frac{0.0.4507 - 0.4522}{\sqrt{0.0001948}}\sqrt{10}\sqrt{1.1569} \\ &= -0.3469 \end{aligned}$$

Der kritische Wert ist $t_{n-1;1-\frac{\alpha}{2}} = t_{10;0.995} = 3.3554$. Die Nullhypothese kann damit nicht abgelehnt werden.

(d) Einsetzen liefert:

$$\begin{aligned} t_1 &= 0.363 + 0.1020 \cdot 8 \\ &= 1.179 \\ t_2 &= -0.0024 + 0.4507 \cdot \sqrt{8} \\ &= 1.2724 \end{aligned}$$

(e) Umstellen und Einsetzen liefert:

$$\begin{aligned} 5 &= 0.363 + 0.1020 \cdot h_1 \\ h_1 &= \frac{5 - 0.363}{0.1020} \\ &= 45.4608 \\ 5 &= -0.0024 + 0.4507 \cdot \sqrt{h_2} \\ h_2 &= \left(\frac{5 + 0.0024}{0.4507}\right)^2 \\ &= 123.1918 \end{aligned}$$

Aufgabe 3

(a) Wenn man eine Substitution $z = \frac{1}{x}$ macht, so erhält man ein einfaches lineares Modell $a + bz$, von dem wir die Koeffizienten mit den bekannten Formeln bestimmen können. Wir müssen nur daran denken, dass z nie explizit auszurechnen, sondern es in den Formeln einfach nur durch $\frac{1}{x}$ zu ersetzen:

$$\begin{aligned}
\hat{b} &= \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i} - \overline{\frac{1}{x}}\right)(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i} - \overline{\frac{1}{x}}\right)^2} \\
&= \frac{3.5}{\frac{7}{18}} \\
&= 9 \\
\hat{a} &= \bar{y} - \hat{b}\bar{z} \\
&= \bar{y} - \hat{b}\overline{\frac{1}{x}} \\
&= 5.5 - 9 \cdot 0.5 \\
&= 1
\end{aligned}$$

(b) Der Wert a ist so etwas wie eine Fix-Zeit, die unabhängig von der Anzahl der Reinigungskräfte immer anfällt. Hier beträgt a 1 Stunde. Der Wert b ist die Zeit, die eine Reinigungskraft für den gesamten Zug braucht. Mehr Reinigungskräfte brauchen dann logischerweise nur ein Bruchteil der Zeit. Hier ist b 9 Stunden.

(c) Einsetzen liefert: $1 + \frac{9}{9} = 2$ Stunden

(d) Nehmen wir zum Beispiel den Punkt (0,9). Das würde bedeuten, dass 0 Reinigungskräfte den Zug in 9 Stunden reinigen. Diese Selbstreinigung wäre für die Bahn besonders kostengünstig, aber leider reinigen sich die Züge nicht von selbst.
Ein anderer unsinniger Punkt ist (7,0). 7 Reinigungskräfte können also den Zug sofort reinigen und dann direkt den nächsten Zug in nullkommanichts reinigen.