Statistik 2, Übung 4

Henry Haustein

Aufgabe 1

Wir wollen den Gesamtfehler $U = \sum_{i=1}^{3} u_i^2$ minimieren. Dabei sind die u_i die Abweichung zwischen dem wahren Wert von p und einer Schätzung $\hat{p}_i$. Die Schätzung für einen Platz am Fenster ist $\hat{p}_1 = \frac{4}{10}$, für einen Platz am Gang $\hat{p}_3 = \frac{5}{10}$ und für den Platz in der Mittel gilt $1 - 2\hat{p}_2 = \frac{1}{10} \Leftrightarrow \hat{p}_2 = \frac{9}{20}$. Mit diesen Informationen können wir eine Funktion für U aufstellen:

$$\begin{aligned} U = \sum_{i=1}^{3} u_i &= \left(p - \frac{4}{10}\right)^2 + \left(p - \frac{5}{10}\right)^2 + \left(p - \frac{9}{20}\right)^2 \\ &= \left(p^2 - \frac{8}{10}p + \frac{16}{10}\right) + \left(p^2 - \frac{10}{10}p + \frac{25}{10}\right) + \left(p^2 - \frac{18}{20}p + \frac{81}{20}\right) \\ &= 3p^2 - \frac{27}{10}p + \frac{163}{20} \end{aligned}$$

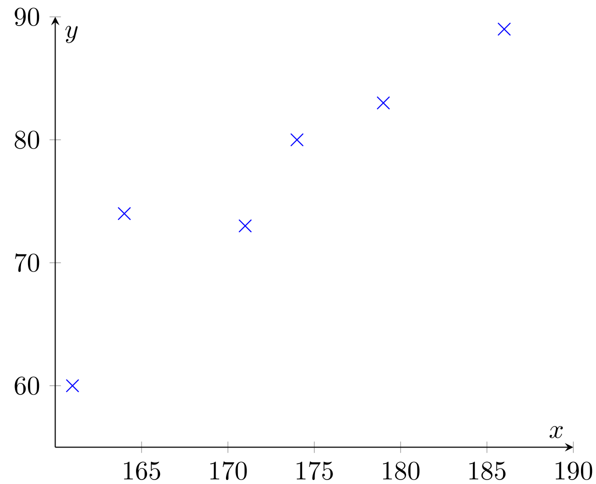
Um ein Minimum von U zu finden, müssen wir U nach p ableiten, also

$$\begin{aligned} \frac{dU}{dp} = 6p - \frac{27}{10} &= 0 \\ 6p &= \frac{27}{10} \\ p &= \frac{27}{60} = 0.45 \end{aligned}$$

Überprüfen wir noch schnell um, welche Art von Extremum es sich handelt: $\frac{d^2U}{dp^2} = 6 > 0 \Rightarrow$ Minimum. Das heißt $p = 0.45$ minimiert den aufsummieren quadratischen Fehler.

Aufgabe 2

(a) Wir wählen den Scatterplot als Darstellungsform:



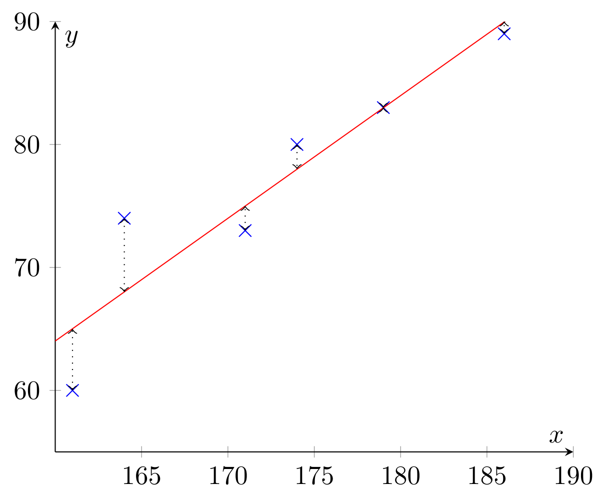
(b) Für ein Modell $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ sind die Koeffizienten gegeben durch

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$
$$\beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}$$

Man kann weiterhin $\bar{x} = 172.5$ und $\bar{y} = 67.5$ berechnen und es ergibt sich

$$y = -95.665 + 0.998x$$

(c) Die Fehler sind die Abweichung zwischen dem mittels Regression geschätzten y und dem gemessenen y.



(d) $-95.665 + 0.998 \cdot 170 = 74.005$

Aufgabe 3

Um vernünftig mit dem Modell arbeiten zu können, sollten wir es erst in eine passende Form bringen: Wir definieren dazu $\tilde{x}_i = 16x_i - x_i^2$ und erhalten

$$y_i = \kappa + \tilde{x}_i$$

Dieses Modell hat starke Ähnlichkeit mit einem "normalen" linearen Modell, nur dass hier $\beta_0 = \kappa$ und $\beta_1 = 1$ ist. Wir können wieder unsere Formeln zur Bestimmung von $\beta_0 = \kappa$ heranziehen und erhalten

$$\kappa = \beta_0 = \bar{y} - 1 \cdot \bar{\tilde{x}}$$

Die transformierten $\tilde{x}_i$ lauten

x_i	6	7	7.5	7.75	8	8.75	9.25
$\tilde{x}_i$	60	63	63.75	63.9375	64	63.4375	62.4375

$\bar{y} = 8.2857$ und $\bar{\tilde{x}} = 62.9375$. Damit folgt $\kappa = -54.6518$.

Aufgabe 4

Auch hier müssen wir das Modell erst in eine lineare Form bringen, damit wir damit arbeiten können. Wir wenden den natürlichen Logarithmus auf beide Seiten an:

$$\ln(y_i) = \ln(\beta_0) + \beta_1 x_i$$

Mit ein bisschen Umbenennung der Variablen ($\tilde{y}_i = \ln(y_i)$ und $\tilde{\beta}_0 = \ln(\beta_0)$) ergibt sich

$$\tilde{y}_i = \tilde{\beta}_0 + \beta_1 x_i$$

Die Formeln zur Bestimmung von $\tilde{\beta}_0$ und β_1 liefern:

$$\begin{aligned} \tilde{\beta}_0 &= \bar{\tilde{y}} - \beta_1 \bar{x} \\ \beta_1 &= \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \tilde{y}_i - \bar{x}\bar{\tilde{y}}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} \end{aligned}$$

Es sind zum Glück schon alle dieser Summen in der Aufgabenstellung gegeben[1].

$$\begin{aligned} \beta_1 &= \frac{\frac{1}{20} \cdot -249.635 - \bar{x}\bar{\tilde{y}}}{\frac{1}{20}930.76 - \bar{x}^2} \\ &= \frac{-12.48175 - 6.08 \cdot (-1.60657)}{46.538 - 6.08^2} \\ &= \frac{-2.713804}{9.5716} \\ &= -0.2835267 \\ \tilde{\beta}_0 &= \bar{\tilde{y}} - (-0.2835267) \cdot \bar{x} \\ &= -1.60657 - (-0.2835267) \cdot 6.08 \\ &= 0.1172723 \end{aligned}$$

mit $\bar{x} = \frac{1}{20}\sum_{i=1}^{2} 0x_i = \frac{1}{20}121.6 = 6.08$ und $\bar{\tilde{y}} = \frac{1}{20}\sum_{i=1}^{2} 0y_i = \frac{1}{20} \cdot -32.131 = -1.60657$. Da wir β_0 zu $\tilde{\beta}_0$ transformiert haben, müssen wir die Transformation rückgängig machen. Es ergibt sich für $\beta_0 = 1.124426$.

[1] Die Aufgabenstellung hat die Variablen nicht umbenannt, deswegen hier aufpassen: $\log(y_i) = \tilde{y}_i$