Kapitel 1 - Das einfache lineare Regressionsmodell

Einfaches lineares Regressionsmodell

Das einfache lineare Regressionsmodell hat die Form

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$$

für ein festes numerisches x_i und $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. Beachte, dass per Definition gilt $Y_i | x_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$

Kleinste Quadrate (KQ) Schätzer

Wir schätzen die Parameter (β_0, β_1) durch

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1)}{\arg\min} \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 x_i))^2 \tag{1}$$

und nennen $(\hat{\beta}_0, \hat{\beta}_1)$ den KQ-Schätzer von (β_0, β_1) und $\hat{\varepsilon}_i := Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ die Residuen.

Existenz und Berechnung vom KQ Schätzer

Der KQ-Schätzer existiert und ist eindeutig, falls $\sum_{i=1}^{n} (x_i - \overline{x})^2 \neq 0$. Dieser lässt sich berechnen als

$$\begin{aligned} \hat{\beta}_1 &= \frac{s_{xY}}{S_x^2} = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y})}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2} \\ \hat{\beta}_0 &= \overline{Y} - \hat{\beta}_1 \overline{x}. \end{aligned}$$

Durch differenzieren von der Gleichung (1) erhält man $(\hat{\beta}_0, \hat{\beta}_1)$ als Lösung der Normalengleichungen

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$$

$$\sum_{i=1}^{n} \hat{\varepsilon}_i x_i = 0$$

Interpretation der Modellparameter

Wir betrachten die Zufallsvariable $Y = \beta_0 + \beta_1 X$

- Wenn X um eine **Einheit** steigt, dann steigt Y **im Erwartungswert** um β_1 Einheiten.
- Es gilt $\beta_0 = \mathbb{E}(Y | X = 0)$.

Wir betrachten $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$

- Der Parameter σ die erwartete Abweichung der Y_i-Werte von der Regressionsgerade an.