

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

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

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 $\sum_{i=1}^n (x_i - \bar{x})^2 \neq 0$. Dieser lässt sich berechnen als

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

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

$$\begin{aligned} \sum_{i=1}^n \hat{\varepsilon}_i &= 0 \\ \sum_{i=1}^n \hat{\varepsilon}_i x_i &= 0 \end{aligned}$$

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.

Eigenschaften des KQ-Schätzers

Gegeben dem einfachen linearen Modell, gilt für den KQ-Schätzer $(\hat{\beta}_0, \hat{\beta}_1)$

- Erwartungstreue: $\mathbb{E}(\hat{\beta}_0, \hat{\beta}_1) = (\beta_0, \beta_1)$.
- $\mathbb{V}(\hat{\beta}_1) = \frac{\sigma^2}{n S_x^2}$ und $\mathbb{V}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{n S_x^2} \right)$.
- $(\hat{\beta}_0, \hat{\beta}_1)$ ist der maximum-likelihood Schätzer.

Schätzer für σ^2

Gegeben dem einfachen linearen Modell mit $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, gilt

$$\hat{\sigma}^2 := \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2$$

ist ein erwartungstreuer Schätzer von σ^2 und

$$\frac{n-2}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2.$$

Der KQ-Schätzer $(\hat{\beta}_0, \hat{\beta}_1)$ und der Schätzer $\hat{\sigma}^2$ sind stoch.unabhängig.

Konfidenzintervalle für β_0 und β_1

Gegeben dem einfachen linearen Modell mit $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, gilt für $\hat{\beta}_1$ und $\hat{\beta}_0$

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{n-2} \text{ mit } \hat{\sigma}_{\hat{\beta}_1} := \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim t_{n-2} \text{ mit } \hat{\sigma}_{\hat{\beta}_0} := \sqrt{\hat{\sigma}^2 \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}$$

Überschrift