Kleinste-Quadrate-Methode (Least Squares method)

$$\sum_{i=1}^{n} \Big(\underset{\substack{\uparrow \\ \text{beob.}}}{y_i} - \underbrace{(m \cdot x_i + b)}_{\text{unter linearem Modell erwartet}}\Big)^2 =: Q(m, b)$$

innere Ableitung nach b

$$\frac{\partial}{\partial b} Q(m, b) = 2 \sum_{i=1}^{n} \big(y_i - (m x_i + b)\big)^{2-1} (-1) \overset{!}{=} 0 \qquad | : (-2)$$

$$\frac{\partial}{\partial m} Q(m, b) = 2 \sum_{i=1}^{n} \big(y_i - (m \cdot x_i + b)\big)^{2-1} (-x_i) \overset{!}{=} 0 \qquad | : (-2)$$

$$0 \overset{!}{=} \sum_{i=1}^{n} y_i - m \cdot \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} b = n\bar{y} - m \cdot n \cdot \bar{x} - n b \qquad | + nb$$

$$nb = n\bar{y} - n \cdot m \cdot \bar{x} \quad | : n \qquad \hat{b}_{LS} = \hat{b} = \bar{y} - \hat{m}\bar{x}$$

$$= n \cdot (\bar{y} - m \cdot \bar{x})$$

$$\hat{m}_{LS} = \hat{m} = \frac{s_{xy}}{s_x^2} \qquad \left(\text{vgl. } r_{xy} = \frac{s_{xy}}{s_x s_y}\right)$$

Regressionsgerade: $\hat{y} = \hat{m} \cdot x + \hat{b}$ wobei x aus Beobachtungsbereich

Bsp: $x \in [20, 85]$, sonst

unzulässige Extrapolation ($\to$ out-of-sample-Prognose)

Bsp. $\hat{m} = \dfrac{s_{xy}}{s_x^2} = \dfrac{-19{,}17}{336{,}75} \approx -0{,}057$ $\qquad \hat{b} = \bar{y} - \hat{m}\bar{x}$

$= 2{,}39 - (-0{,}057) \cdot 65{,}2$

$\approx 6{,}09$

Prognose: $x = 40$

$$\hat{y} = -0{,}057 \cdot 40 + 6{,}09 \approx 3.8$$

für Yemen $x = 20$: $\hat{y} = -0.057 \cdot 20 + 6.09 \approx 5,\ldots$

Prognosefehler: $\hat{y}_{35} - y_{35} = 5,\ldots - 6{,}2 \approx -1,\ldots$

Bestimmtheitsmaß $\quad R^2 := \dfrac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \dfrac{\cancel{(n-1)} \cdot s_{\hat{y}}^2}{\cancel{(n-1)} \cdot s_y^2} = \dfrac{s_{\hat{y}}^2}{s_y^2}$

emp. Varianz der y-Werte

Anteil der durch die Regression erklärten Varianz $s_{\hat{y}}^2$ (Variation) an der beob. Gesamtvar. s_y^2