Kap. Einfache lineare Regression

Geg. Stichprobe $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ quantitative Daten

Bsp. $x_1 = 64$ [%] $y_1 = 2.5$ Bsp. UN-Daten Algerien

Empirische(r) Korrelation(skoeffizient) nach Pearson

$$r_{xy} := \frac{s_{xy}}{s_x \cdot s_y} \quad \text{mit} \quad s_{xy} := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

Wdh.

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$s_x = \sqrt{s_x^2}$$

s_{xy} : empirische Kovarianz

$$s_{xy} = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i \cdot y_i - n \, \bar{x} \cdot \bar{y} \right)$$

Es gilt: 1) Bei linearer Transformation der Daten $x_1, \dots, x_n$, $y_1, \dots, y_n$, d.h. $x_i \mapsto a \cdot x_i + b$ (z.B. °C → °F, $a = \frac{9}{5} = 1{,}8$, $b = 32$) $y_i \mapsto c \cdot y_i + d$, $a \neq 0, c \neq 0$:

$$|r_{ax+b,\, cy+d}| = |r_{xy}|$$

2) $-1 \leq r_{xy} \leq 1$

3) $|r_{xy}| = 1 \iff$ Alle Punkte $(x_1, y_1), \dots, (x_n, y_n)$ liegen auf derselben Geraden, d.h. $y_i = a \cdot x_i + b$ für alle $i = 1, 2, \dots, n$

Bsp. UN-Daten

$$s_{xy} = \frac{1}{35-1} (64 \cdot 2{,}5 + \dots + 21 \cdot 6{,}2 - 35 \cdot 65{,}2 \cdot 2{,}35) \approx -19{,}11$$

$$\bar{x} = \frac{64 + \dots + 21}{35} \approx 65.2 \qquad \bar{y} = \frac{2{,}5 + \dots + 6{,}2}{35} \approx 2{,}39$$

$$s_x^2 = \frac{1}{35-1} (64^2 + \dots + 21^2 - 35 \cdot 65{,}2^2) \approx 336{,}75$$

$$s_y^2 = \frac{1}{35-1} (2{,}5^2 + \dots + 6{,}2^2 - 35 \cdot 2{,}39^2) \approx 1{,}40$$

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} \approx \frac{-19{,}11}{\sqrt{336{,}75} \cdot \sqrt{1{,}40}} \approx -0.879$$ negative Korrelation

Einfache lineare Regression: Regressionsgerade