

(ECO-VOI) / KUN

Simple linear regression

→ assume given data points $(x_i, y_i) \quad i=1 \rightarrow n$
 y_i ... dependent variable
 x_i ... independent (explanatory) variable

→ try to approximate y_i through a linear function of x_i
 $y_i \approx g(x_i)$
 $y_i \approx b_1 + b_2 x_i \quad i=1 \rightarrow n$

→ not need to determine b_1 and b_2 that minimize the total approx. error between y_i and $b_1 + b_2 x_i$ is ~~as small~~ small

1 "one
n "n"

→ "differentiate" $S(b_1, b_2) = \sum_{i=1}^n (y_i - (b_1 + b_2 x_i))^2 \rightarrow \min$

1. Ableitung:
 first derivative
 → "times"

$\frac{\partial S}{\partial b_1} = \sum_{i=1}^n (y_i - b_1 - b_2 x_i) \cdot (-1) = 0 \rightarrow \text{KEL}$
 $\Leftrightarrow \sum_{i=1}^n y_i - n b_1 - b_2 \sum_{i=1}^n x_i = 0$
 → directly vereinfacht

$\frac{\partial S}{\partial b_2} = \sum_{i=1}^n (y_i - b_1 - b_2 x_i) \cdot (-x_i) = 0$
 → "y" → "y" b_2

$\frac{\partial S}{\partial b_2} = \sum_{i=1}^n (y_i - b_1 - b_2 x_i) \cdot (-x_i) = 0$
 $\Rightarrow -\sum_{i=1}^n x_i y_i + b_1 \sum_{i=1}^n x_i + b_2 \sum_{i=1}^n x_i^2 = 0$
 → directly vereinfacht
 → **KEL**

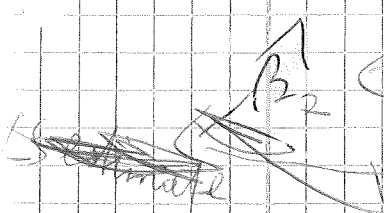
$\Rightarrow b_2 \cdot \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$

$\Leftrightarrow S_{xx} b_2 = S_{xy} \rightarrow$ if not all x_i are equal to the same value, then $S_{xx} > 0$

$\Rightarrow b_2 = \frac{S_{xy}}{S_{xx}}$
 $b_1 = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}$

ASSUMPTION

→ some assumption



- how well the regression line (the linear function in the x 's) explains the data

$$\Rightarrow 0 \leq R^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n u_i^2} \leq 1 \quad \rightarrow \text{"perfect fit"}$$

↙
"best fit possible"

↗
"fit" of the regression line to the data points (usually x -linear)

→ las hier Skript 1.1

→ interpretation of the coefficient β_2

- changing x by 1 unit will change y by the amount of β_2

↙
absolute value

→ assume that β_2 is an unskewed value, not an approximation, and call it β_2

$$\Rightarrow y = \beta_1 + \beta_2 x \rightarrow \text{level-level model}$$

$$\bullet \ln(y) = \beta_1 + \beta_2 x \rightarrow \text{log-level model}$$

↙

⇒ changing x by 1 unit will change y by approx. 100 · β_2 %