ECO - VO I | KW 1

[1] Simple Linear Regression

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

→ try to approximate y_i through a linear function of x_i

$$\left.\begin{array}{l} y_i \approx g(x_i) \\ y_i \approx b_1 + b_2 x_i \end{array}\right\} i = 1 \to n$$

→ we want to determine b_1 and b_2 that minimize the ~~total~~ approx. error between y_i and $b_1 + b_2 x_i$ is ~~…~~ small

$\frac{1}{n}$ → „one over n"

→ „differentiate"

$$S(b_1, b_2) = \sum_{i=1}^{n} \left(y_i - (b_1 + b_2 x_i)\right)^2 \to \min$$

1. Ableitung: „first derivative"

$$\frac{\partial S}{\partial b_1} \Rightarrow \bar{y} - b_1 - b_2 \bar{x} = 0 \to \boxed{NE\,I}$$ → bereits vereinfacht

$$b_1 = \bar{y} - b_2 \bar{x}$$

· → „times"

$\bar{y}$ → „y bar"

$$\frac{\partial S}{\partial b_2} \Rightarrow \frac{1}{n} \sum_{i=1}^{n} x_i y_i - b_1 \bar{x} - b_2 \frac{1}{n} \sum_{i=1}^{n} x_i^2 = 0$$ → bereits vereinfacht → $\boxed{NE\,II}$

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

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

$$\Rightarrow b_2 = \frac{S_{xy}}{S_{xx}}$$

$$b_1 = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}$$ → same assumption

$\hat{\beta}_2$

~~Estimate~~

b_1 $\hat{\beta}_1$

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

$\Rightarrow 0 \leq R^2 = 1 - \frac{SSR}{SST} \leq 1$ → „perfect fit"

„worst fit possible" „fit" of the regression line to the data points $(x_1, y_1), \ldots (x_n, y_n)$

⟹ lies hier Skript 1.1

⟹ interpretation of the coefficient $\hat{\beta}_2$

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

absolute value

→ assume that $\hat{\beta}_2$ is an concrete value, not an approximation, and call it β_2

⟹ $y = \beta_1 + \beta_2 x$ → level - level model

- $\ln(y) = \beta_1 + \beta_2 x$ → log - level model

%

⟹ changing x by 1 unit will change y by approx $100 \cdot \beta_2$ %