universität
innsbruck
Institut für Geographie

Everything hangs together: Lineare Regression

716408 | Sozialwiss. Methoden – How 2 do Things with Numbers

KMH

SS 22 (updated: 2022-06-07)

Wozu Regression?

Exkurs: Unterschied Regression & Korrelation

- **Korrelation:** Stärke eines monotonen Zusammenhangs
  - Korrelation kann von einem 3. (nicht berücksichtigten) Faktor ausgehen → Scheinkorrelation
  - BSP: Korrelation zwischen Geburtenrate und Storchenpopulation in einem Bezugsraum
  - Prognose i.e.S. nicht möglich
- **Regression:** Beschreibung einer Ursache-Wirkungs-Beziehung
  - Formalisierung eines Zusammenhangs als (z.B. lineare) Gleichung
    → Wie genau wirkt ein Faktor auf einen anderen?
  - Möglichkeit zur **Prognose**

Lineare Regression im Detail

- **Abhängige Variable** („Regressand“): zu erklärende Variable
- **Unabhängige Variable** („Regressor“): erklärende Variable(n)
- Ursache-Wirkungs-Beziehung als lineare Gleichung darstellen:

$$f_x = \alpha_0 + \alpha_1 x$$

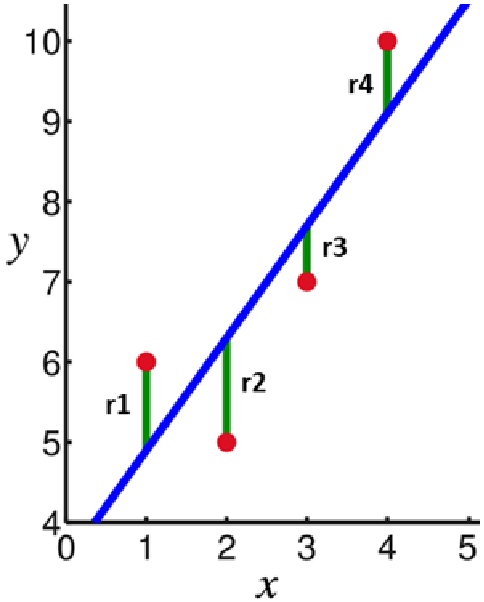
- Parameterschätzung über Methode der kleinsten Fehlerquadrate

$$\min_{\alpha_0,\alpha_1} \sum_{i=1} r_i^2$$

$$r_1 = \alpha_0 + \alpha_1 x_1 - y_1$$

$$\dots$$

$$r_n = \alpha_0 + \alpha_1 x_n - y_n$$

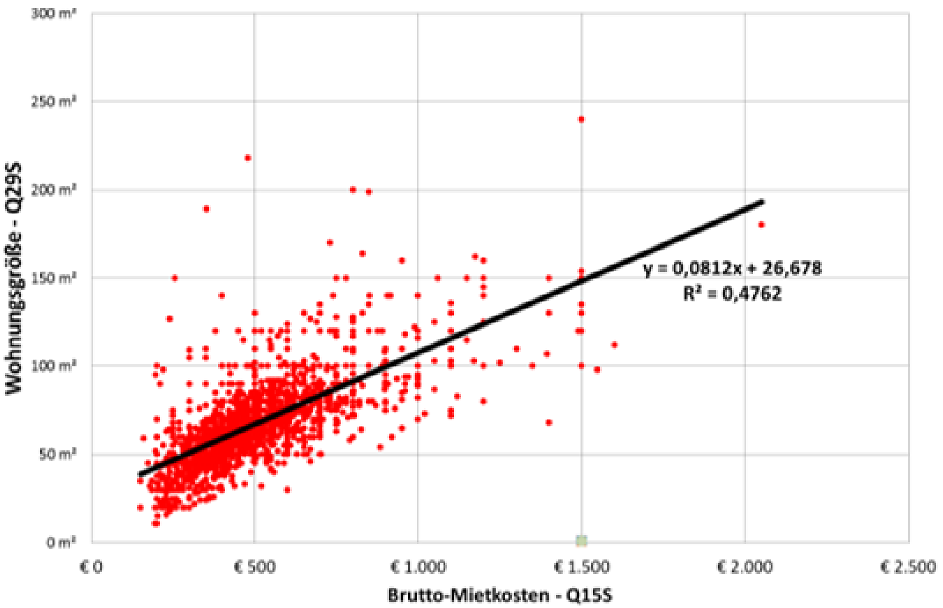


Lineare Regression im Detail

$$f_x = \alpha_0 + \alpha_1 x$$

$$\alpha_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\alpha_0 = \overline{y} - \alpha_1 \overline{x}$$



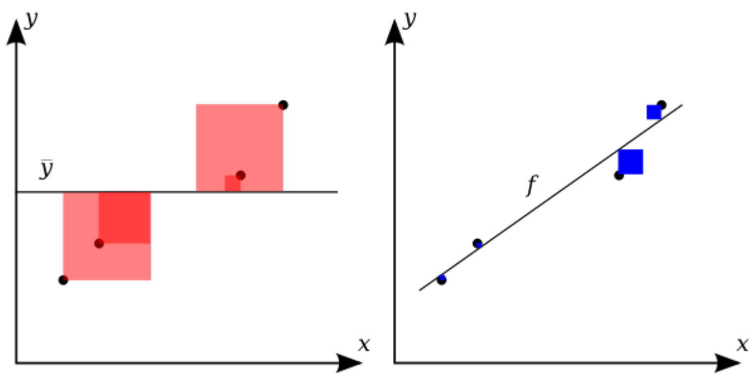
@ Güte einer gefunden Regressionslösung

- **Bestimmtheitsmaß R^2:**
  - Durch Regression erklärte Varianz der abhängigen Variable
  - Wertbereich: 0 bis 1 (= vollständige Varianzerklärung)

$$SS_{\text{res}} = \sum_i (y_i - f_i)^2$$

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$



(Orzetto, Wikimedia, CC BY-SA)