Formeln zur Berechnung der beiden *Regressionsgeraden*

Gegeben seien n Datenpaare x_i und y_i :

1.Fall: Gesucht sind 2 Regressionsgeraden der Form $y = m \cdot x + b$

Zunächst muss der Schwerpunkt $\boxed{S(\bar{x}/\bar{y}) \quad \text{mit} \quad \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \quad \text{und} \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i}$ der „Punktwolke" berechnet werden !

Ergebnisformel (Lineare Regression) - Minimierung vertikaler Entfernungen :

$$y = m \cdot x + b \quad \text{mit} \quad m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{(\sum_{i=1}^{n} x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y}}{(\sum_{i=1}^{n} x_i^2) - n \cdot \bar{x}^2} \quad \text{und} \quad b = \bar{y} - m \cdot \bar{x}$$

Eine 2. Regressionsgerade durch Minimierung horizontaler Entfernungen lässt sich finden:

$$y = m_2 \cdot x + b_2 \quad \text{mit} \quad m_2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})} = \frac{(\sum_{i=1}^{n} y_i^2) - n \cdot \bar{y}^2}{(\sum_{i=1}^{n} x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y}} \quad \text{und} \quad b_2 = \bar{y} - m_2 \cdot \bar{x}$$

Korrelationskoeffizient r nach PEARSON (Linear) :

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = \frac{(\sum_{i=1}^{n} x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{[(\sum_{i=1}^{n} x_i^2) - n \cdot \bar{x}^2] \cdot [(\sum_{i=1}^{n} y_i^2) - n \cdot \bar{y}^2]}} = \sqrt{\frac{m}{m_2}}$$

Anmerkung:

Der in den obigen Formeln enthaltene Term $\frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$ heißt **Kovarianz** !

Die Kovarianz ist ein Maß für den Zusammenhang zwischen den Merkmalen x und y .

<u>2.Fall:</u> **Der Spezialfall b = 0 (y = m·x) :**

Die beiden Geraden verlaufen dann durch den **Ursprung (0/0)** !

<u>Vertikal:</u> $$\boxed{y = \frac{\sum x_i \cdot y_i}{\sum x_i^2} \cdot x}$$

<u>Horizontal:</u> $$\boxed{y = \frac{\sum y_i^2}{\sum x_i \cdot y_i} \cdot x}$$

<u>Korrelationskoeffizient:</u> $$\boxed{r = \frac{\sum x_i \cdot y_i}{\sqrt{\sum x_i^2 \cdot \sum y_i^2}}}$$