Lineare Regression und Korrelation

Ansatz: $y = b_0 + b_1 x \qquad \mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$b_1 = \frac{\sigma_{xy}}{\sigma_{x^2}} = \frac{\sum_{i=1}^{n}(x_i y_i) - n \cdot \mu_x \cdot \mu_y}{\sum_{i=1}^{n}(x_i^2) - n \cdot \mu_x^2} \quad \begin{matrix}\cdot\frac{1}{n} \\ = \\ \cdot\frac{1}{n}\end{matrix} \quad \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i y_i) - \mu_x \cdot \mu_y}{\frac{1}{n}\sum_{i=1}^{n}(x_i^2) - \mu_x^2}$$

$$b_1 = \frac{\sigma_{xy}}{\sigma_{x^2}} = \frac{Kov(X,Y)}{V(X)}$$

$$b_0 = \mu_y - b_1 \cdot \mu_x$$

Korrelation nach Brevais-Pearson:

$$r = \frac{\sum_{i=1}^{n}(x_i - \mu_x)\cdot(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i - \mu_x)^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \mu_y)^2}} = \frac{\sum_{i=1}^{n}(x_i y_i) - n \cdot \mu_x \cdot \mu_y}{\sqrt{\sum_{i=1}^{n}(x_i^2) - n \cdot \mu_x^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i^2) - n \cdot \mu_y^2}}$$

$$r \underset{\frac{1}{n}}{\overset{\cdot\frac{1}{n}}{=}} \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i) - \mu_x \cdot \mu_y}{\frac{1}{n}\cdot\sqrt{\sum_{i=1}^{n}(x_i^2) - n \cdot \mu_x^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i^2) - n \cdot \mu_y^2}} = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i) - \mu_x \cdot \mu_y}{\sqrt{\frac{1}{n}}\cdot\sqrt{\frac{1}{n}}\cdot\sqrt{\sum_{i=1}^{n}(x_i^2) - n \cdot \mu_x^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i^2) - n \cdot \mu_y^2}}$$

$$r = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i) - \mu_x \cdot \mu_y}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i^2) - \mu_x^2} \cdot \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i^2) - \mu_y^2}} = \frac{Kov(X,Y)}{S(X)\cdot S(Y)} = \frac{Kov(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

$$r = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i) - \mu_x \cdot \mu_y}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i^2) - \mu_x^2} \cdot \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i^2) - \mu_y^2}} = \frac{Kov(X,Y)}{S(X)\cdot S(Y)} = \frac{\sigma(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

Infos zum Korrelationskoeffizienten.

Maß für den Zusammenhang zwischen den Werten zweier Merkmale

$$r = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i)-\mu_x\cdot\mu_y}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i^{\,2})-\mu_x^{\,2}}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i^{\,2})-\mu_y^{\,2}}} = \frac{Kov(X,Y)}{S(X)\cdot S(Y)} = \frac{\sigma(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

$r = [-1;1]$ möglicher Wertebereich

$r = [-1;-0,7[$ sehr starke negative Korrelation

$r = [-0,7;-0,3[$ starke negative Korrelation

$r = [-0,3;0,3[$ keine Korrelation (Punktwolke)

$r = [0,3;0,7[$ starke positive Korrelation

$r = [0,7;1]$ sehr starke positive Korrelation

$$(x_1 + x_2 + \ldots + x_6) = 30$$

$$(y_1 + y_2 + \ldots + y_6) = 36$$

$$(x_1 + x_2 + \ldots + x_6)\cdot(y_1 + y_2 + \ldots + y_6) \neq x_1y_1 + x_2y_2 + \ldots + x_6y_6$$

Kovarianz:

$$Cov(X,Y) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\mu_x)\cdot(y_i-\mu_y) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i)-\mu_x\cdot\mu_y$$

$$V(X) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\mu_x)^2 = \left[\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i)^2\right]-\mu_x^{\,2}$$