Formeln zur linearen Regression:

Definiert man die Summe der Fehlerquadrate einer besten Funktion $z = a + b\,s$ zu

$$Q(a,b) = \sum_{i=1}^{n} (z_i - a - b\ s_i)^2$$

so ergeben sich als partielle Ableitungen nach den Parametern der Gerade die linearen Gleichungen:

$$\frac{\partial Q(a,b)}{\partial b} = 2\ \sum_{i=1}^{n} (z_i - a - b\ s_i)\ (-s_i) = 0\ ,$$

und

$$\frac{\partial Q(a,b)}{\partial a} = 2\ \sum_{i=1}^{n} (z_i - a - b\ s_i)\ (-1) = 0\ .$$

Die Lösung ist

$$b = \frac{n \sum_{i=1}^{n} s_i z_i - \left(\sum_{i=1}^{n} s_i\right)\ \left(\sum_{i=1}^{n} z_i\right)}{n \left(\sum_{i=1}^{n} s_i^2\right) - \left(\sum_{i=1}^{n} s_i\right)^2}$$

und

$$a = \frac{\left(\sum_{i=1}^{n} s_i^2\right) \left(\sum_{i=1}^{n} z_i\right) - \left(\sum_{i=1}^{n} s_i z_i\right)\ \left(\sum_{i=1}^{n} s_i\right)}{n \left(\sum_{i=1}^{n} s_i^2\right) - \left(\sum_{i=1}^{n} s_i\right)^2}$$

Aus den Formeln ergibt sich noch der Korrelationskoeffizient:

$$r_{sz} = \frac{n \left(\sum_{i=1}^{n} s_i\ z_i\right) - \left(\sum_{i=1}^{n} s_i\right)\ \left(\sum_{i=1}^{n} z_i\right)}{\left(n \left(\sum_{i=1}^{n} s_i^2\right) - \left(\sum_{i=1}^{n} s_i\right)^2\right)^{1/2} \left(n \left(\sum_{i=1}^{n} z_i^2\right) - \left(\sum_{i=1}^{n} z_i\right)^2\right)^{1/2}}$$

und die minimale Fehlerquadratsumme:

$$Q(a,b) = \left(\left(\sum_{i=1}^{n} z_i^2\right) - \frac{1}{n} \left(\sum_{i=1}^{n} z_i\right)^2\right)\ \left(1 - r_{sz}^2\right).$$

Mit den Mittelwerten

$$\bar{s} = \frac{1}{n} \; (\sum_{i=1}^{n} s_i)$$

und

$$\bar{z} = \frac{1}{n} \; (\sum_{i=1}^{n} z_i)$$

wird daraus kürzer:

$$b = \left(\sum s_i \; z_i - n \; \bar{s} \; \bar{z}\right) / \left(\; (\sum s_i^2 \;) \; - \; n \; \bar{s}^2\right)$$

$$a = \left(\; (\sum s_i^2) \; \bar{z} - (\sum s_i \; z_i \;) \; \bar{s} \; \right) / \left(\; (\sum s_i^2) \; - \; n \; \bar{s}^2 \; \right)$$

und

$$r_{sz} = \left(\; \sum s_i \; z_i \; - \; n \; \bar{s} \; \bar{z} \; \right) / \left(\; (\sum s_i^2 \; - \; n \; \bar{s}) \; (\sum z_i^2 \; - \; n \; \bar{z}) \; \right)^{1/2}$$

sowie

$$Q(a,b) = \left(\; \sum z_i^2 \; - \; n \; \bar{z}^2 \; \right) \; \left(1 - r_{sz}^2\right)$$

Der Punkt ($\bar{s}$, $\bar{z}$) ist der Schnittpunkt der Regressionsgraden $z = a + b\,s$ mit der invertierten Graden $z = c + d\,s$, die aus $s = A + B\,z$ entsteht: $(s - A)/B = z$, also $c = -A/B$ und $d = 1/B$.

Im **Beispiel** ist

a= -19.05, b= 5.07, A= 4.10, B = 0.19, also $c = -21.57$, und $d = 5.25$ mit ($\bar{s}$, $\bar{z}$) = (14.2, 53.0) . Der Winkel zwischen den Graden ist

$$tan \; \alpha = \frac{m_2 - m_1}{1 + m_1 \; m_2}$$

mit den Anstiegen der beiden Graden als $m_1 = 5.07$, und $m_2 = 5.25$ ergibt sich

$$\alpha = 0.37^o \; .$$