Die Regressionsgerade

In Reinhard Ramls Dokumenten ist die Notation für die Regressionsgerade

$$\hat{Y} = kX + d,$$

in der Vorlesung zur Psychologischen Methodenlehre und Statistik wird diese Gerade durch

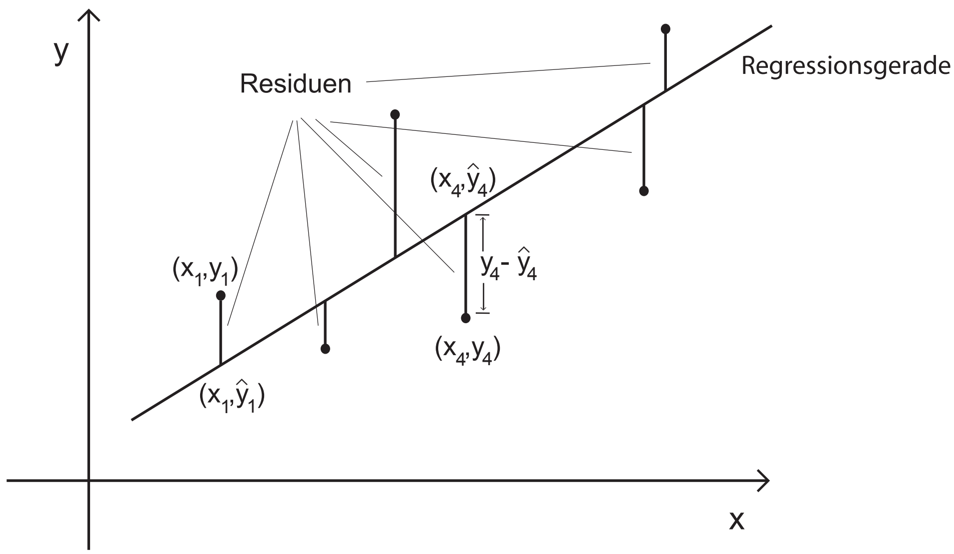
$$\hat{Y} = bX + a$$

bezeichnet.

Wie wir bereits wissen berechnet sich die Regressionsgerade durch die Minimierung der Residuen (**Methode der kleinsten Quadrate**), d.h.

$$Q(a, b) = \sum_i (y_i - \hat{y}_i)^2 = \sum_i (y_i - bx_i - a_i) = \text{minimal}.$$

Dazu eine grafische Veranschaulichung dieser Methode:



Wie kann man die Residuen minimieren? Auf die genaue Vorgangsweise wird in einem Unterkapitel der mathematischen Hintergründe von matheonline eingegangen: Lokale Minima

Die Idee ist folgende: $Q(a, b)$ ist eine Funktion der Unbekannten a und b. Wir suchen das *Minimum* dieser Funktion. Dazu wird $Q(a, b)$ abgeleitet (jeweils nach a und nach b) und gleich 0 gesetzt. Wir bekommen also 2 Gleichungen mit 2 Unbekannten, aus denen a und b bestimmt werden.

Zur Bestimmung von a leitet man die Quadratsumme der Residuen nach a ab, d.h.

$$\frac{\partial Q}{\partial a} = \sum_{i=1}^{n} \frac{\partial (y_i - bx_i - a)^2}{\partial a} = \sum_{i=1}^{n} \frac{\partial ((y_i - bx_i)^2 - 2(y_i - bx_i)a + a^2)^2}{\partial a}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}(0 - 2(y_i - bx_i) + 2a) \\
&= \sum_{i=1}^{n} -2(y_i - bx_i) + 2a \\
&= 0
\end{aligned}
$$

d.h.

$$
\begin{aligned}
2\underbrace{\sum_{i=1}^{n} a}_{na} &= 2\underbrace{\sum_{i=1}^{n} y_i}_{n\bar{y}} - b\underbrace{\sum_{i=1}^{n} x_i}_{n\bar{x}} \\
na &= n\bar{y} - bn\bar{x}.
\end{aligned}
$$

Wir bekommen

$$\boxed{a = \bar{y} - b\bar{x},}$$

wobei $\bar{x}$ der Mittelwert der x_i und $\bar{y}$ der Mittelwert der y_i ist. Zur Bestimmung von b leitet man die Quadratsumme der Residuen nach b ab, d.h.

$$\frac{\partial Q}{\partial b} = \sum_i \frac{\partial (y_i - bx_i - a)^2}{\partial b}$$

Man erhält (diesmal ohne Anführung der Zwischenschritte)

$$\boxed{b = \frac{\overbrace{\frac{1}{n}\sum_{1}^{n} x_i y_i - \bar{x}\bar{y}}^{=Cov(X,Y)}}{s_X^2}.}$$

b wird auch der *Regressionskoeffizient von Y auf X* genannt

Eine nach der Methode der kleinsten-Quadrat Schätzer schneidet stets den Punkt $(\bar{x}, \bar{y})$.

a und b nennt man auch die **Kleinste-Quadrate Schätzer**.

Wenn umgekehrt X als Funktion von Y betrachtet wird, dann ist die Regressionsgerade von X auf Y berechnet. Man vertauscht einfach X mit Y bzw. die x_i mit den y_i. Die Kalkulationen erfolgen analog.