

## KAPITEL 5: REGRESSIONSRECHNUNG

### Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate (MkIQ) ist ein *Verfahren zur Anpassung einer Funktion an eine Punktwolke*. Angewandt wird sie beispielsweise, um eine Gesetzmäßigkeit aus stichprobenartig aufgenommenen Messwerten herzuleiten.

Aufgabe: Suchen einer optimalen Anpassung einer Funktion (Kurve, Fläche) gemäß folgender Gleichungen:

$$y = f(x; \beta_0, \beta_1) \quad (\text{Kurve}), \quad \beta_0 \text{ und } \beta_1 \text{ sind die unbekannt Parameter.}$$

$$z = f(x, y; \beta_0, \beta_1) \quad (\text{Fläche}), \quad \beta_0 \text{ und } \beta_1 \text{ sind die unbekannt Parameter, } x, y \text{ und } z \text{ sind bekannt (Messwerte).}$$

Eine optimale Anpassung liegt vor, wenn die **Summe der Abweichungsquadrate der Messwerte von den Werten der Funktion minimal** ist. Wichtig ist, dass die **Quadrate** der Abweichungen betrachtet werden, denn die Summe der einfachen Abweichungen könnte ja auch dadurch sehr gering sein, dass sich negative und positive Abweichungen aufheben.

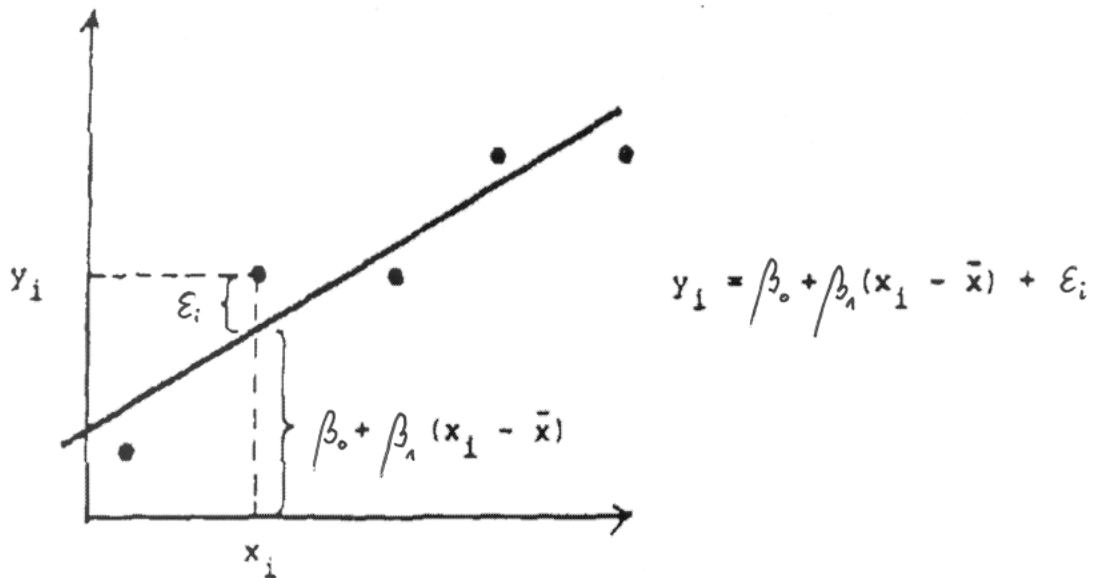
Die einfachen Abweichungen sind die Differenzen aus Messwert und geschätztem Wert:

$\varepsilon_i = y_i - f(x_i; \beta_0, \beta_1)$  bzw.  $\varepsilon_i = y_i - f(x_i, y_i; \beta_0, \beta_1)$ . Wir formulieren also für die Summe der Abweichungsquadrate:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - f(x_i; \beta_0, \beta_1))^2 = \min. \quad (\text{Kurve}) \quad \text{bzw.}$$

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (z_i - f(x_i, y_i; \beta_0, \beta_1))^2 = \min. \quad (\text{Fläche}).$$

Abbildung 105



## Quantifizierung der Abweichungen $\varepsilon_i$

Gegeben sind:  $(x_i, y_i), i = 1, \dots, n$  (Punktwolke).

Funktionsansatz: z.B.  $\hat{y} = \beta_0 + \beta_1 x$  (Gerade).

Gesucht werden:  $\beta_0, \beta_1$  mit  $S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - f(x_i; \beta_0, \beta_1))^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min.$

$$\varepsilon_1 = y_1 - \hat{y}_1 = y_1 - \beta_0 - \beta_1 x_1$$

$$\varepsilon_2 = y_2 - \hat{y}_2 = y_2 - \beta_0 - \beta_1 x_2$$

$$\vdots \quad \quad \quad \vdots$$

$$\varepsilon_n = y_n - \hat{y}_n = y_n - \beta_0 - \beta_1 x_n$$

### **Annäherung an eine Gerade beispielhaft für 4 Messwerte:**

$$P_1(2,1) \quad , \quad P_2(5,3) \quad , \quad P_3(8,2) \quad , \quad P_4(11,5)$$

Summe der Abweichungsquadrate:

$$S(\beta_0, \beta_1) = (1 - \beta_0 - \beta_1 \cdot 2)^2 + (3 - \beta_0 - \beta_1 \cdot 5)^2 + (2 - \beta_0 - \beta_1 \cdot 8)^2 + (5 - \beta_0 - \beta_1 \cdot 11)^2$$

Um die Summe der Abweichungsquadrate zu minimieren, müssen wir wie bei der Ermittlung von Extrema bei Funktionen mehrerer unabhängiger Variablen verfahren, nämlich zunächst die ersten partiellen Ableitungen bilden und gleich Null setzen:

$$\begin{aligned} S_{\beta_0} &= 2(y_1 - \beta_0 - \beta_1 x_1) \cdot (-1) + 2(y_2 - \beta_0 - \beta_1 x_2) \cdot (-1) + 2(y_3 - \beta_0 - \beta_1 x_3) \cdot (-1) \\ &\quad + 2(y_4 - \beta_0 - \beta_1 x_4) \cdot (-1) = 0 \end{aligned}$$

$$\begin{aligned} S_{\beta_1} &= 2(y_1 - \beta_0 - \beta_1 x_1) \cdot (-x_1) + 2(y_2 - \beta_0 - \beta_1 x_2) \cdot (-x_2) + 2(y_3 - \beta_0 - \beta_1 x_3) \cdot (-x_3) \\ &\quad + 2(y_4 - \beta_0 - \beta_1 x_4) \cdot (-x_4) = 0 \end{aligned}$$

Die Multiplikation der Gleichungen mit  $\frac{1}{2}$  und Summation ergibt:

$$\begin{aligned} \sum_{i=1}^4 y_i + \sum_{i=1}^4 \beta_0 + \sum_{i=1}^4 \beta_1 x_i &= 0 \quad \text{und} \\ \sum_{i=1}^4 x_i y_i + \sum_{i=1}^4 \beta_0 x_i + \sum_{i=1}^4 \beta_1 x_i^2 &= 0 \end{aligned}$$

Wir formen um:

$$\beta_0 \cdot 4 + \beta_1 \cdot \sum_{i=1}^4 x_i = \sum_{i=1}^4 y_i \quad (4 = \text{Anzahl der Messungen})$$

$$\beta_0 \sum_{i=1}^4 x_i + \beta_1 \sum_{i=1}^4 x_i^2 = \sum_{i=1}^4 x_i y_i$$

Diese Gleichungen werden als *Normalgleichungen* bezeichnet. Sie können verwendet werden, um die Unbekannten  $\beta_0, \beta_1$  zu berechnen.

Für den allgemeinen Fall ergeben sich durch eine analoge Rechnung folgende Normalgleichungen:

$$\beta_0 \cdot n + \beta_1 \cdot \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$\beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

In Matrixschreibweise:

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}, \text{ oder, abgekürzt:}$$

$$X^T X \cdot \vec{\beta} = X^T \vec{y} \quad (\text{Erklärung folgt!})$$

Wir definieren die **Design-Matrix**  $X$ , den **Vektor der abhängigen Variablen**  $\vec{y}$  und den **Fehlervektor**  $\vec{\varepsilon}$  :

$$X := \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}; \quad \vec{y} := \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}; \quad \vec{\varepsilon} := \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

$$y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1$$

Dann ist  $y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2$

$$\vdots \quad \quad \quad \vdots$$

$$y_n = \beta_0 + \beta_1 x_n + \varepsilon_n$$

in Matrixschreibweise:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\vec{y} = X \cdot \vec{\beta} + \vec{\varepsilon}$$

Es ist:

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \cdot \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = X^T \cdot X$$

(2, n)                      (n, 2)                      (2, 2)

und

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix} = X^T \cdot \vec{y}$$

Das System der Normalgleichungen, das wir aus den ersten partiellen Ableitungen (s.o.) erhalten haben, lautet:

$$X^T X \vec{\beta} = X^T \vec{y}$$

Um die Funktion zu bestimmen, die die Abweichungsquadrate minimiert, benötigen wir die Unbekannten  $\beta_i$ , also den Vektor  $\vec{\beta}$ . Dazu formen wir die Normalgleichungen nach den Regeln der Matrizenrechnung um:

$$\vec{\beta} = (X^T X)^{-1} \cdot X^T \vec{y} \quad (\text{Voraussetzung: } |X^T X| \neq 0)$$

und erhalten den Lösungsvektor.

Die Design-Matrix  $X$  kann für viele verschiedene Ansätze (beliebige Anzahl von Messwerten; beliebiger Funktionstyp (Gerade, Parabel, Polynom, Logarithmusfunktion etc.); auch für mehrdimensionale Fälle ( $x, y, z$  und mehr)) konstruiert werden.

Bedingung ist jedoch, dass das Modell additiv ist und  $\vec{\beta}$  in linearer Form bzw. Position steht, z.B.

beim Polynom:  $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k$ .

Allgemein wäre die Design-Matrix für den Polynomausgleich:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & x_2^3 & \dots & x_2^k \\ 1 & x_3 & x_3^2 & x_3^3 & \dots & x_3^k \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \dots & x_n^k \end{bmatrix}$$

Nochmals ein einfaches Beispiel:

Durch die drei Punkte:

$$(x_1; y_1) = (2; 1)$$

$$(x_2; y_2) = (4; 1)$$

$$(x_3; y_3) = (4; 3)$$

ist eine „Ausgleichsgerade“ zu legen, d. h. eine Gerade  $\hat{y} = \beta_0 + \beta_1 x$ ,  
so dass die Summe der Quadrate der Fehler  $(y_i - \hat{y}_i)$ ,  $i = 1, 2, 3$ , minimal ist.  
( $\hat{y}_i$  sei  $\beta_0 + \beta_1 x_i$ , also der mittels der Geraden abgelesene  $y$ -Wert zu  $x_i$ ).  
Gesucht sind  $\beta_0$  und  $\beta_1$ .

Lösungsformel („Gaußsche Normalgleichung“):

$$X^T \cdot X \cdot \vec{\beta} = X^T \cdot \vec{y}$$

Darin ist  $\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$  der gesuchte Koeffizientenvektor für die Geradengleichung,

$$X := \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \text{ und } \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad (x_i; y_i) \text{ die gegebenen Punkte } (i = 1, 2, \dots, n).$$

$X$  ist die „Design-Matrix“.

Anwendung der Formel an unserem Beispiel:

$$X := \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 4 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

$$X^T \cdot X = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 10 \\ 10 & 36 \end{bmatrix},$$

$$X^T \cdot \vec{y} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 18 \end{bmatrix}$$

$$\text{Normalgleichung somit: } \begin{bmatrix} 3 & 10 \\ 10 & 36 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 5 \\ 18 \end{bmatrix}.$$

Als lineares Gleichungssystem umgeschrieben:

$$3\beta_0 + 10\beta_1 = 5$$

$$10\beta_0 + 36\beta_1 = 18$$

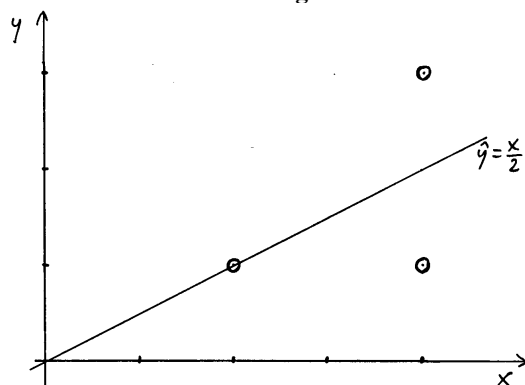
Lösungsschema nach Gauß-Jordan:

$$\begin{array}{cc|c}
 3 & 10 & 5 \\
 5 & 18 & 9 \\
 \hline
 1 & \frac{10}{3} & \frac{5}{3} \\
 5 & 18 & 9 \\
 \hline
 1 & \frac{1}{3} & \frac{5}{3} \\
 0 & 18 - \frac{50}{3} & 9 - \frac{25}{3} \\
 \hline
 1 & \frac{10}{3} & \frac{5}{3} \\
 0 & \frac{4}{3} & \frac{2}{3} \\
 \hline
 1 & \frac{10}{3} & \frac{5}{3} \\
 0 & 1 & \frac{1}{2} \\
 \hline
 1 & 0 & 0 \\
 0 & 1 & \frac{1}{2}
 \end{array}
 \Rightarrow
 \begin{array}{l}
 1 \cdot \beta_0 = 0 \\
 1 \cdot \beta_1 = \frac{1}{2}
 \end{array}
 \Rightarrow
 \begin{array}{l}
 \beta_0 = 0 \\
 \beta_1 = \frac{1}{2}
 \end{array}$$

Die Gleichung der gesuchten Ausgleichsgeraden lautet also:

$$\hat{y} = 0 + \frac{1}{2} \cdot x = \frac{x}{2} \quad (\text{Abb. 106}).$$

Abbildung 106



Das sollte man nach dem Besuch von Vorlesung und Übung beherrschen:

- Was ist die „Methode der kleinsten Quadrate“?
- Warum verwendet man die Quadrate der Abweichungen?
- Wie ist die Design-Matrix aufgebaut?
- Wie lauten die Normalgleichungen?
- Anwendung der Methode der kleinsten Quadrate zur optimalen Anpassung einer Gerade oder einer Polynomfunktion an eine Menge von Messwertepaaren (Punktwolke)