Karlsruher Institut für Technologie (KIT)

Institut für Wirtschaftstheorie und Statistik (ETS)

Ergänzungen Kapitel 10 DS

Deskriptive Statistik

Dirk Krause

ETS

25.05.2012

Modell der Linearen Regression

In einer Menge von Beobachtungspaaren $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ lässt sich jeder y-Wert y_i beschreiben als:

$$y_i = \hat{y}_i + \varepsilon_i \,, \; i = 1, \ldots, n$$

$\hat{y}_i$: Trendwert, gewonnen aus unterstelltem funktionalen Zusammenhang mit zugehörigem x_i

ε_i: Störgröße

Funktionaler Zusammenhang $f(x) = mx + b$ ist hier linear:

$$\hat{y}_i = f(x_i) = \hat{y}(x_i) = mx_i + b \quad \text{(Modell der Trendwerte)}$$

Für fixe m und b ergeben sich implizit die Störgrößen

$$\varepsilon_i = y_i - \hat{y}_i = y_i - mx_i - b$$

Modell der Linearen Regression

Beobachtungspaare dargestellt als Streudiagramm:

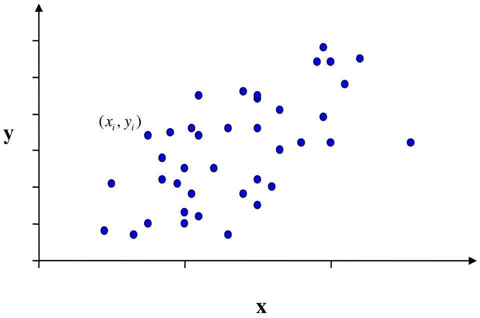

Abbildung: Streudiagramm

Modell der Linearen Regression

Beobachtungspaare dargestellt als Streudiagramm:

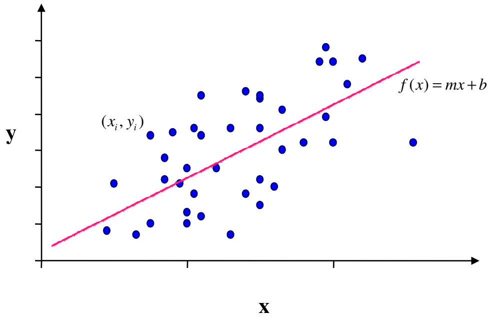

Abbildung: Streudiagramm und funktionaler Zusammenhang

Modell der Linearen Regression

Beobachtungspaare dargestellt als Streudiagramm:

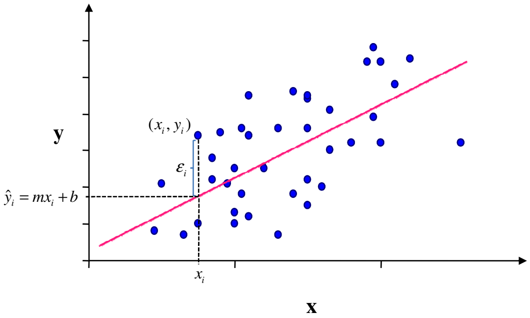

Abbildung: Trendwert und Störgröße

Vertikale Abstandsmessung bzw. Messung der Störgröße.

Modell der Linearen Regression

Anpassungskriterium der optimalen (Regressions-)Geraden:

Minimiere Summe der quadrierten Störgrößen ($\Rightarrow$ Methode der kleinsten Quadrate) über die Wahl der Geradenkoeffizienten m und b.

$$\min_{m,b} \sum_{i=1}^{n} \varepsilon_i^2 \equiv \min_{m,b} \sum_{i=1}^{n} (y_i - mx_i - b)^2$$

Optimale Anpassung der Regressionsgeraden

<u>Optimierungskalkül</u>: Partielle Differentiation der Zielfunktion nach m und b, anschließendes Nullsetzen:

$$\frac{\partial \sum_{i=1}^n (y_i - mx_i - b)^2}{\partial m} = \sum_{i=1}^n 2(y_i - mx_i - b)(-x_i) = -2\sum_{i=1}^n x_i y_i + 2m \sum_{i=1}^n x_i^2 + 2b \sum_{i=1}^n x_i \overset{!}{=} 0 \quad ①$$

$$\frac{\partial \sum_{i=1}^n (y_i - mx_i - b)^2}{\partial b} = \sum_{i=1}^n 2(y_i - mx_i - b)(-1) = -2\sum_{i=1}^n y_i + 2m \sum_{i=1}^n x_i + 2nb \overset{!}{=} 0 \quad ②$$

aus ② :
$$b = \frac{1}{n}\sum_{i=1}^n y_i - m \cdot \frac{1}{n}\sum_{i=1}^n x_i = \bar{y} - m \cdot \bar{x}$$

in ① :
$$-\sum_{i=1}^n x_i y_i + m\sum_{i=1}^n x_i^2 + (\bar{y} - m \cdot \bar{x})\sum_{i=1}^n x_i \overset{!}{=} 0$$

$$\Leftrightarrow m\left(\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i\right) = \sum_{i=1}^n x_i y_i - \bar{y}\sum_{i=1}^n x_i$$

$$\Rightarrow m = \frac{\sum_{i=1}^n x_i y_i - \bar{y}\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i} = \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{y}\bar{x}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2}$$