



Vorlesung: Lineare Modelle

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

SoSe 2014

- Einführung und Beispiele
- ① Das einfache lineare Regressionsmodell
- Das multiple lineare Regressionsmodell
- Quadratsummenzerlegung und statistische Inferenz im multiplen linearen Regressionsmodell



- Einführung und Beispiele
- ① Das einfache lineare Regressionsmodell
- Das multiple lineare Regressionsmodell
- Quadratsummenzerlegung und statistische Inferenz im multiplen linearen Regressionsmodell

$$Y_i = \underbrace{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}_{x_i' \beta} + \varepsilon_i; \quad i = 1, \dots, n$$

$$x_i' = (1, x_{i1}, \dots, x_{ip})$$

$$Y = X\beta + \varepsilon \tag{2.1}$$

mit

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Interpretation des Modells

- Lineare Abhängigkeit von den Einflussgrößen
- Steigt x_k um eine Einheit, so steigt Y im Erwartungswert um β_k Einheiten, wenn alle anderen **X-Variablen festgehalten werden**
- Linearer Zusammenhang bei Festhalten der übrigen Variablen
- β_k charakterisiert den Einfluss von x_k unter Berücksichtigung der übrigen Variablen („Confounder-Korrektur“)

$$E(\varepsilon_i) = 0$$

$$E(\varepsilon) = \mathbf{0} \tag{2.2}$$

$$V(\varepsilon_i) = \sigma^2 \tag{2.3}$$

$$\{\varepsilon_i \mid i = 1, \dots, n\} \text{ unabh.} \tag{2.4}$$

Aus (2.3), (2.4) folgt: $V(\varepsilon) = \sigma^2 I$

$$\varepsilon_i \sim N(0, \sigma^2) \text{ und (2.4)}$$

$$\varepsilon \sim N(\mathbf{0}, \sigma^2 I) \tag{2.5}$$

Y: Zufallsvektor der Zielgröße
 X: feste Design-Matrix (Matrix der Einflussgrößen)
 β: Vektor der Regressionsparameter
 ε: Störgrößen

KQ-Schätzer

Wir betrachten Modell (2.1). Dann heißt

$$\hat{\beta} = \arg \min_{\beta} \underbrace{(Y - X\beta)'(Y - X\beta)}_{\sum_{i=1}^n (Y_i - x_i' \beta)^2} \tag{2.6}$$

KQ-Schätzer.

$$\hat{\varepsilon}_i := Y_i - x_i' \hat{\beta} \tag{2.7}$$

Es gilt für $(X'X)$ invertierbar: $\hat{\beta}$ **existiert**, ist **eindeutig** und

$$\hat{\beta} = (X'X)^{-1} X'Y. \tag{2.8}$$

Der KQ-Schätzer erfüllt die Normalgleichungen:

$$X' \hat{\varepsilon} = \mathbf{0} \tag{2.9}$$

Produktsummenmatrix

Die Matrix $X'X$ heißt Produktsummenmatrix.

Es gilt:

$$X'X = \begin{pmatrix} n & \sum x_{i1} & \cdots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \cdots & \sum x_{i1}x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_{ip} & \cdots & \cdots & \sum x_{ip}^2 \end{pmatrix} \quad (2.10)$$

Hat-Matrix und Residualmatrix

Sei das Modell (2.1) mit einer Designmatrix X mit $rg(X) = p + 1$ gegeben. Es gilt

$$\hat{Y} = X \underbrace{(X'X)^{-1}X'Y}_{\hat{\beta}} \quad (2.14)$$

$$P := \underbrace{X(X'X)^{-1}X'}_{n \times n} \quad (2.15)$$

$$\hat{\epsilon} = Y - \hat{Y} = \underbrace{(I - P)}_Q Y \quad (2.16)$$

$$Q := I - P \quad (2.17)$$

Eigenschaften des KQ-Schätzers

Sei das Modell (2.1) mit (2.2) gegeben.

- 1 Der KQ-Schätzer ist **erwartungstreu**:

$$E(\hat{\beta}) = \beta \quad (2.11)$$

- 2 Für die Varianz-Kovarianz-Matrix von $\hat{\beta}$ gilt unter (2.3) und (2.4):

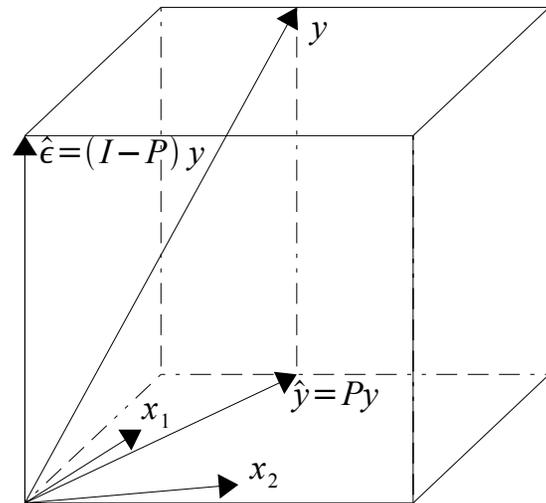
$$V(\hat{\beta}) = \sigma^2(X'X)^{-1} \quad (2.12)$$

- 3 Unter (2.5) gilt:

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}) \quad (2.13)$$

Geometrische Interpretation

Wir betrachten den Vektorraum \mathbb{R}^n . Die Beobachtungen Y und x sind Vektoren. Die KQ - Schätzung ist eine orthogonale Projektion auf den von den x - Vektoren aufgespannten Unterraum:



Eigenschaften von P und Q

P heißt **Hat-Matrix** ($\hat{Y} = PY$), Q **Residualmatrix**. P und Q sind Projektionsmatrizen, und zueinander orthogonal:

$$P' = P, P^2 = P \tag{2.18}$$

$$Q' = Q, Q^2 = Q \tag{2.19}$$

$$PQ = QP = \mathbf{0} \tag{2.20}$$

Erwartungstreue Schätzung von σ^2

Gegeben sei das Modell (2.1) mit (2.2) bis (2.4).

Dann ist:

$$\hat{\sigma}^2 = \frac{1}{n - (p + 1)} \hat{\varepsilon}' \hat{\varepsilon} = \frac{1}{n - (p + 1)} \sum \hat{\varepsilon}_i^2 \tag{2.24}$$

ein erwartungstreuer Schätzer für σ^2 .

Bemerkungen

- Beweis durch Nachrechnen (benutze $(AB)' = B'A'$)
- Bedeutung von $P^2 = P$:
zweimaliges Anwenden der Regression führt zum gleichen Ergebnis
- Bedeutung von $PQ = 0$:
Regression von Residuen liefert $\hat{y} = 0$

Für die Varianz-Kovarianz-Matrizen von \hat{Y} bzw. $\hat{\varepsilon}$ gilt:

$$V(\hat{Y}) = \sigma^2 P \tag{2.21}$$

$$V(\hat{\varepsilon}) = \sigma^2 Q \tag{2.22}$$

$$\text{da } \hat{\varepsilon} = Q\varepsilon \tag{2.23}$$

Beweis

$$\begin{aligned}
E(\hat{\varepsilon}' \hat{\varepsilon}) &= E(\varepsilon' Q' Q \varepsilon) = E\left(\sum_{i=1}^n \sum_{j=1}^n q_{ij} \varepsilon_i \varepsilon_j\right) \\
&= \sum_{i=1}^n E(\varepsilon_i^2) q_{ii} + 0 = \sigma^2 * sp(Q) \\
&= \sigma^2 * (n - sp(P)) = \sigma^2 * (n - (p + 1)) \\
sp(P) &= sp(X(X'X)^{-1}X') = sp((X'X)^{-1}X'X) = sp(I_{p+1})
\end{aligned}$$

Bemerkung:
Für Projektionsmatrizen gilt allgemein: $Sp(P) = rg(P)$

Beispiel: Volumen von Hühnereiern

V : Volumen
d : Durchmesser
r : Radius

