



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

Institut für Statistik



# Vorlesung: Lineare Modelle

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

SoSe 2014

- 5 Metrische Einflußgrößen: Polynomiale Regression, Trigonometrische Polynome, Regressionsplines, Transformationen.
- 6 Modelldiagnose
- 7 Variablenselektion
- 0 Das allgemeine lineare Modell: Gewichtete KQ-Methode, Autokorrelierte und heteroskedastische Störterme**
- 0 Das logistische Regressionsmodell
- 10 Das gemischte lineare Regressionsmodell („Linear mixed Model“)

# Das allgemeine lineare Modell

---

Das lineare Modell mit heteroskedastischen Störgrößen ist gegeben durch:

$$Y = X\beta + \varepsilon \quad (8.1)$$

$$\varepsilon \sim N(0, \sigma^2 V) \quad (8.2)$$

$$V = \text{diag}(v_1, v_2, \dots, v_n) \quad (8.3)$$

$V$  bekannte Matrix zur Beschreibung der Varianzstruktur

**Gewichtsmatrix:**  $W = V^{-1/2} = \text{diag}(v_1^{-1/2}, v_2^{-1/2}, \dots, v_n^{-1/2})$

# Der gewichtete KQ-Schätzer

---

$$\hat{\beta}_W := (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (8.4)$$

$$\hat{\sigma}^2 = (\hat{\varepsilon}'V^{-1}\hat{\varepsilon})/(n - p') \quad (8.5)$$

$\hat{\beta}_W$  ist ML-Schätzer und minimiert die **gewichtete Residuenquadratsumme**

$$(Y - X\beta)'V^{-1}(Y - X\beta). \quad (8.6)$$

# Herleitung durch Transformation

---

Das Modell (8.1)-(8.3) lässt sich in ein gewöhnliches lineares Modell transformieren:

$$Y^* := WY \quad (8.7)$$

$$X^* := WX \quad (8.8)$$

$$\varepsilon^* := W\varepsilon \quad (8.9)$$

Dann gilt:

$$Y^* = X^* \beta + \varepsilon^* \quad (8.10)$$

$$\varepsilon^* \sim N(0, \sigma^2 I) \quad (8.11)$$

$\hat{\beta}_W$  ist KQ-Schätzer im transformierten Modell

# Verallgemeinerte KQ-Methode

Das lineare Modell mit allgemeiner Varianzstruktur ist gegeben durch:

$$Y = X\beta + \varepsilon \quad (8.12)$$

$$\varepsilon \sim N(0, \sigma^2 V) \quad (8.13)$$

$V \in R^{n \times n}$ : beliebige bekannte Kovarianzmatrix mit vollem Rang

Dann gibt es eine invertierbare Matrix  $T$  mit

$$TT' = V, \quad W = T^{-1} \text{ Gewichtsmatrix.}$$

Der **verallgemeinerte KQ-Schätzer** ist gegeben durch:

$$\hat{\beta}_W := (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (8.14)$$

$$\hat{\sigma}^2 = (\hat{\varepsilon}'V^{-1}\hat{\varepsilon})/(n - p') \quad (8.15)$$

---

Das Modell (8.12)-(8.13) lässt sich wie oben in ein gewöhnliches lineares Modell transformieren:

$$Y^* := WY \quad (8.16)$$

$$X^* := WX \quad (8.17)$$

$$\varepsilon^* := W\varepsilon \quad (8.18)$$

Dann gilt:

$$Y^* = X^*\beta + \varepsilon^* \quad (8.19)$$

$$\varepsilon^* \sim N(0, \sigma^2 I) \quad (8.20)$$

# Eigenschaften des verallgemeinerten KQ-Schätzers

---

Gegeben sei das Modell (8.12) bis (8.13).

Dann gilt:

$$E(\hat{\beta}_W) = \beta \quad (8.21)$$

$$V(\hat{\beta}_W) = \sigma^2(X'V^{-1}X)^{-1} \quad (8.22)$$

Alle Testverfahren und Quadratsummenzerlegungen lassen sich im Modell  $Y^* = X^*\beta + \varepsilon^*$  betrachten und damit auf den Fall homogener Varianzen zurückführen.



# Allgemeines Gauss-Markov-Theorem

---

Sei das Modell

$$\begin{aligned} Y &= X\beta + \varepsilon, & \text{rg } X &= p' \\ E(\varepsilon) &= 0 \\ V(\varepsilon) &= \sigma^2 V \end{aligned}$$

gegeben.

Dann ist  $\hat{\beta}_W$  unter den erwartungstreuen linearen Schätzern derjenige mit der kleinsten Varianz:

$\hat{\beta}_W$  ist BLUE-Schätzer (**b**est **l**inear **u**nbiased **e**stimator).

# Beispiele für Varianzstrukturen

---

- AR(1) (allgemeine Zeitreihenstruktur)
- Longitudinale Daten (Mehrdimensionale Zeitreihen) Blockdiagonale Struktur
- Symmetrische Struktur (gemischte Modelle)

# Weitere Schätzstrategien

---

Im Allgemeinen müssen die Parameter der Varianzstruktur geschätzt werden.

Dazu gibt es verschiedene Verfahren:

- ML
- REML (Restricted Maximum Likelihood)
- Robuste Varianzschätzung mit „Working correlation“

# ML und REML-Schätzung I

---

Sei das Modell

$$\begin{aligned} Y &= X\beta + \varepsilon, & \text{rg } X &= p' \\ E(\varepsilon) &= 0 \\ V(\varepsilon) &= \sigma^2 V(\vartheta) \end{aligned}$$

gegeben. Der Parameter(vektor)  $\vartheta$  ist zu schätzen.

Als Log-Likelihood ergibt sich (von additiven Konstanten abgesehen):

$$l(\beta, \vartheta) = -\frac{1}{2} (\ln |V(\vartheta)| + (\mathbf{Y} - X\beta)' V^{-1}(\vartheta) (\mathbf{Y} - X\beta)) \quad (8.23)$$

# ML und REML-Schätzung II

---

Ist  $\vartheta$  bekannt, so ist der MLE von  $\beta$  bedingt auf  $\vartheta$  (gewichteter KQ-Schätzer:)

$$\hat{\beta}(\vartheta) = \left( X' V(\vartheta)^{-1} X \right)^{-1} X' V^{-1}(\vartheta) Y. \quad (8.24)$$

Einsetzen liefert die Profil-Log-Likelihood:

$$l(\vartheta) = -\frac{1}{2} \left( \ln |V(\vartheta)| + (\mathbf{Y} - X\beta(\vartheta))' V^{-1}(\vartheta) (\mathbf{Y} - X\beta(\vartheta)) \right) \quad (8.25)$$

# ML und REML-Schätzer

---

Maximieren von (10.22) bezüglich  $\vartheta$  liefert ML-Schätzer. Da dieser nicht erwartungstreu ist, verwendet man häufig den sogenannten restringierten ML-Schätzer:

Dieser maximiert

$$L_R(\vartheta) = l(\vartheta) - \frac{1}{2} \ln |X' V(\vartheta)^{-1} X| \quad (8.26)$$

Im einfachen linearen Modell entspricht der REML-Schätzer dem erwartungstreuen Schätzer von  $\sigma^2$ .

# Inferenz bezüglich von $\beta$

---

Es gilt:  $\hat{\beta}(\vartheta)$  einer multivariaten Normalverteilung mit Erwartungswert  $\beta$  und Kovarianzmatrix

$$\text{var}(\hat{\beta}) = (X'V^{-1}X)^{-1} \quad (8.27)$$

Da  $V$  unbekannt ist, wird es durch den (RE)ML-Schätzer  $V(\hat{\vartheta})$  ersetzt.

Zur Konstruktion von Konfidenzintervallen und entsprechenden Tests nimmt man an, dass  $\beta$  asymptotisch normalverteilt ist. Für spezielle Modelle ist dies bewiesen, aber eine allgemeingültige asymptotische Normalverteilungsaussage ist nicht nachgewiesen.

Da die Varianzmatrix  $V$  nur geschätzt wird, werden in der Praxis deshalb häufig approximative  $t$ -Tests und entsprechende Konfidenzintervalle benutzt, die die Verteilung von  $(\hat{\beta}_j - \beta_j)/s.\hat{e}(\hat{\beta}_j)$  durch eine  $t$ -Verteilung approximieren und die zugehörigen Freiheitsgrade geeignet schätzen.

# Beispiele

---

- Wildzeitreihen:  
Varianzstruktur: Unabhängigkeit der Einzelzeitreihen, aber AR(1)-Struktur für jede einzelne Zeitreihe
- Schuldaten mit Korrelation innerhalb einer Klasse



