

# Tutorium 3

## Analyse longitudinaler Daten

Prof. Dr. Sonja Greven, Almond Stöcker,  
Johanna Völkl

Material: Alexander Bauer

20. Juni 2017

- 1 Schätzung der Kovarianz von  $\hat{\beta}$
- 2 Inferenz im LLMM

# Schätzung der Kovarianz von $\hat{\beta}$

1 Schätzung der Kovarianz von  $\hat{\beta}$

2 Inferenz im LLMM

Schätzung der Kovarianz von  $\hat{\beta}$ 

## Grundlagen:

- Schätzung basiert auf **marginalem Modell**:

$$\mathbf{Y}_i \sim \mathcal{N}(\mathbf{X}_i\beta, \mathbf{V}_i(\alpha))$$

$$\text{mit } \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \Sigma_i$$

- Schätzung von  $\hat{\beta}$  per GLS mit  $\mathbf{W} = \mathbf{V}(\alpha)^{-1}$ :

$$\hat{\beta}_{ML}(\alpha) = \{\mathbf{X}^T\mathbf{W}\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y}$$

$$\text{Cov}(\hat{\beta}_{ML}) = \{\mathbf{X}^T\mathbf{W}\mathbf{X}\}^{-1}\{\mathbf{X}^T\mathbf{W} \text{Cov}(\mathbf{Y}) \mathbf{W}\mathbf{X}\}\{\mathbf{X}^T\mathbf{W}\mathbf{X}\}^{-1}$$

# Schätzung der Kovarianz von $\hat{\beta}$

## Schätzung von $Cov(\hat{\beta}_{ML})$ :

- **Fall 1:** Kovarianz  $Cov(\mathbf{Y})$  korrekt spezifiziert  
⇒ Schätzer mit höchster Effizienz:

$$\widehat{Cov}(\hat{\beta}_{ML}) = \{\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X}\}^{-1}$$

- **Fall 2:** Kovarianz  $Cov(\mathbf{Y})$  falsch spezifiziert  
⇒ Obiger Schätzer liefert ungültige Standardfehler für  $\beta$   
(Inferenz folglich auch nicht mehr gültig)  
⇒ Besser: Verwendung eines robusten Schätzers

Schätzung der Kovarianz von  $\hat{\beta}$ 

Schätzung von  $Cov(\hat{\beta}_{ML})$ :

- Robuster Sandwich-Schätzer:

$$\widehat{Cov}(\hat{\beta}_{ML}) = \left\{ \sum_{i=1}^N \mathbf{x}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{x}_i \right\}^{-1} \left\{ \sum_{i=1}^N \mathbf{x}_i^T \widehat{\mathbf{V}}_i^{-1} \widehat{Cov}(\mathbf{Y}_i) \widehat{\mathbf{V}}_i^{-1} \mathbf{x}_i \right\} \left\{ \sum_{i=1}^N \mathbf{x}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{x}_i \right\}^{-1}$$

mit  $\mathbf{V}_i = \mathbf{V}_i(\hat{\alpha})$  und  $\widehat{Cov}(\mathbf{Y}_i) = \mathbf{r}_i \mathbf{r}_i^T = (\mathbf{y}_i - \mathbf{X}_i \hat{\beta})(\mathbf{y}_i - \mathbf{X}_i \hat{\beta})^T$

**Vorteil:** Auch bei Misspezifikation konsistent

**Nachteile:** Wenig effizient; Nicht zu empfehlen bei kleinem  $N$

# Inferenz im LLMM

1 Schätzung der Kovarianz von  $\hat{\beta}$

2 Inferenz im LLMM

# Inferenz im LLMM

## Inferenz für fixed effects:

- Formulierung linearer Hypothesen  $H_0 : \mathbf{L}\beta = \mathbf{0}$

Bsp.:

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_i + \epsilon_{ij}$$

a)  $H_0 : \beta_1 = 0 \quad \Leftrightarrow \quad \mathbf{L} = (0, 1, 0)$

b)  $H_0 : \beta_1 = \beta_2 = 0 \quad \Leftrightarrow \quad \mathbf{L} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

c)  $H_0 : \beta_1 = \beta_2 \quad \Leftrightarrow \quad \mathbf{L} = (0, 1, -1)$

⇒ Testen per Wald-, t-, F- oder LQ-Test

# Inferenz im LLMM

## Inferenz für fixed effects:

$$T = (\hat{\beta} - \beta)^T \mathbf{L}^T \left[ \mathbf{L} \widehat{\text{Cov}}(\hat{\beta}) \mathbf{L}^T \right]^{-1} \mathbf{L} (\hat{\beta} - \beta)$$

- Wald-Test:

$$(\hat{\beta} - \beta) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, (\mathbf{X}^T \mathbf{V}(\hat{\alpha})^{-1} \mathbf{X})^{-1}), \quad T \stackrel{H_{0,a}}{\sim} \chi^2_{\text{rank}(\mathbf{L})}$$

Problem: Unsicherheit von  $\hat{\alpha}$  wird in Schätzung  $\mathbf{V}$  nicht berücksichtigt

- t-Test & F-Test:

$$\text{t-Test (nur für skalare Hypothesen): } \frac{\hat{\beta}_k - \beta_k}{\widehat{SE}(\hat{\beta}_k)} \stackrel{H_{0,a}}{\sim} t_{\hat{df}}$$

$$\text{F-Test: } \frac{T}{\text{rank}(\mathbf{L})} \stackrel{H_{0,a}}{\sim} F_{\text{rank}(\mathbf{L}), \hat{df}}$$

$\hat{df}$  wird aus Daten geschätzt

**Anmerkung:** Benutzung robuster Kovarianzen in allen Tests möglich

# Inferenz im LLMM

## Inferenz für fixed effects:

- LQ-Test für genestete Hypothesen:

$$\lambda_N = -2 \ln \left[ \frac{L_{ML}(\hat{\theta}_{ML,0})}{L_{ML}(\hat{\theta}_{ML})} \right] \stackrel{H_0, a}{\sim} \chi_{p_{diff}}^2$$

LQ-Test bei kleinem  $N$  besser als Wald-Test

**Voraussetzung:** Gleiche random effects in den beiden Modellen

**Beachten:**

REML-Likelihoods abhängig von  $\mathbf{X}$  (durch Fehlerkontraste)

⇒ Likelihoods nicht vergleichbar bei unterschiedlichen fixed effects

⇒ LQ-Test zum Testen von fixed effects nur bei ML-Schätzung!

# Inferenz im LLMM

## Inferenz für Varianzparameter:

$$\text{Cov}(\mathbf{b}_i) = \mathbf{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$$

Vergleich dreier Modelle:

- $M_0$ : ohne random effects,  $d_{11} = d_{12} = d_{22} = 0$
- $M_1$ : nur mit random intercept,  $d_{12} = d_{22} = 0$
- $M_2$ : mit (korrelierten) random intercept und slope

Vergleich  $M_2$  vs.  $M_1$  durch  $H_0 : d_{12} = d_{22} = 0$

Vergleich  $M_1$  vs.  $M_0$  durch  $H_0 : d_{11} = 0$

# Inferenz im LLMM

## Inferenz für Varianzparameter:

- LQ-Test:

$$\lambda_N = -2 \ln \left[ \frac{L_{ML/REML}(\hat{\theta}_{ML/REML,0})}{L_{ML/REML}(\hat{\theta}_{ML/REML})} \right]$$

**Voraussetzung:** Gleiche fixed effects in den beiden Modellen

### Beachten:

Varianzparameter befinden sich unter  $H_0$  am Rand des Parameterraums  $[0, \infty)$

$\Rightarrow \lambda_N$  folgt unter  $H_0$  einer Mischverteilung aus einer  $\chi^2_{q_{diff}-1}$ - und einer  $\chi^2_{q_{diff}}$ -Verteilung

# Inferenz im LLMM

## Parametrischer Bootstrap: als Alternative zu Mischverteilungsansatz

- 1) Fitte beide Modelle und berechne LQ-Teststatistik  $t_{obs}$ .
- 2) Für  $b = 1, \dots, B$ :
  - a) Simuliere neue y-Werte
  - b) Fitte beide Modelle erneut auf den neuen y-Werten und berechne die neue LQ-Teststatistik  $t^b$ . Die Werte  $t^b$  dienen als empirische Verteilung von  $t_{obs}$  unter  $H_0$ .
- 3) Schätze den p-Wert:

$$p_{Boot} = \frac{\sum_{b=1}^B I_{t^b > t_{obs}} + 1}{B + 1}$$

# Inferenz im LLMM

## Parametrischer Bootstrap: als Alternative zu Mischverteilungsansatz

- 1) Fitte beide Modelle und berechne LQ-Teststatistik  $t_{obs}$ .  
Notation:  $m_0$  ist das Modell unter  $H_0$ .
- 2) Für  $b = 1, \dots, B$ :
  - a) Simuliere neue y-Werte  $\Rightarrow$  2 Möglichkeiten
    - i) Prognose mit  $\hat{\beta}$  und  $\hat{\mathbf{b}}_i$  aus  $m_0$ , Simulation der  $\epsilon_{ij}$  aus  $\mathcal{N}(\mathbf{0}, \hat{\Sigma}_i)$
    - ii) Prognose mit  $\hat{\beta}$  aus  $m_0$ , Simulation sowohl der  $\epsilon_{ij}$  aus  $\mathcal{N}(\mathbf{0}, \hat{\Sigma}_i)$  als auch der  $\hat{\mathbf{b}}_i$  aus  $\mathcal{N}(\mathbf{0}, \mathbf{D})$   
(In R: Selbst programmieren für lme, simulate.merMod() für lmer)
  - b) Fitte beide Modelle mit neuen y-Werten und berechne Teststatistik  $t^b$
- 3) Schätze den p-Wert:

$$p_{Boot} = \frac{\sum_{b=1}^B I_{t^b > t_{obs}} + 1}{B + 1}$$

# Inferenz im LLMM

## Zusammenfassung:

- Korrekte Schätzung der Kovarianz von  $\hat{\beta}$  zentral für Inferenz
- Testen linearer Hypothesen bzgl. der fixed effects -  $\mathbf{L}\beta = \mathbf{0}$ : (Wald-), t- oder F-Test
- Voraussetzungen LQ-Test:
  - Bei genesteten fixed effects: Gleiche random effects & ML
  - Bei genesteten random effects: Gleiche fixed effects
- In R: Verwendung von `anova.lme` (für `lme`-Modelle)