

Tutorium 2

Analyse longitudinaler Daten

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

Material: Alexander Bauer

30. Mai 2017

- 1 ICC
- 2 Schätzung der fixed effects
- 3 Prädiktion der random effects

ICC: *Intra-Class Correlation*

Für ein Modell mit einem Random Intercept und Conditional Independence ($\Sigma_i = \sigma^2 I_{n_i}$) beschreibt der ICC die marginale Korrelation von 2 Beobachtungen am selben Subjekt:

$$ICC = \frac{d^2}{d^2 + \sigma^2} \in [0, 1]$$

Anwendung: Ist Aufnahme eines Random Effects sinnvoll?

⇔ Wird ein hoher Anteil der durch die fixed effects nicht erklärten Varianz ($d^2 + \sigma^2$) durch den Random Effect erklärt?

Schätzung der fixed effects

- 1 ICC
- 2 Schätzung der fixed effects
- 3 Prädiktion der random effects

Schätzung der fixed effects

Grundlagen:

- Schätzung basiert auf **marginalem Modell**:

$$\mathbf{Y}_i \sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i)$$

- $\boldsymbol{\alpha}$: Vektor aller (Ko-)Varianzparameter (aus \mathbf{D} , $\boldsymbol{\Sigma}_i$)
- $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$: Vektor aller zu schätzender Parameter
- $\mathbf{V}_i(\boldsymbol{\alpha}) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$: Kovarianzmatrix im marginalen Modell

Schätzung der fixed effects

ML-Schätzung:

- Minimierung des GLS-Kriteriums mit $\mathbf{W} = \mathbf{V}(\alpha)^{-1}$:

$$(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{W}(\mathbf{y} - \mathbf{X}\beta)$$

- $\hat{\beta}_{ML}(\alpha)$ ist **BLUE**

Best: Kleinste Varianz aller linearen erwartungstreuen Schätzer

Linear

Unbiased: Erwartungstreue - $\mathbb{E}(\hat{\beta}_{ML}(\alpha)) = \beta$

Estimator

- Problem I: $\hat{\beta}_{ML}(\alpha)$ abhängig von unbekanntem α
- Problem II: Bei Varianzparametern führt ML zu Unterschätzung
 \Rightarrow **REML-Schätzung** für Varianzkomponenten

Schätzung der fixed effects

REML-Schätzung:

- **Ziel:** α -Schätzung mit einer Likelihood, die nicht von β abhängt
- **Umsetzung:** Schätzung ohne Bias durch Betrachtung der Likelihood einer linearen Transformation von \mathbf{Y}

Schätzung durch Fehlerkontraste $\mathbf{U} = \mathbf{A}^T \mathbf{Y}$ mit \mathbf{A} einer Zentrierungsmatrix mit

$$\mathbb{E}(\mathbf{A}^T \mathbf{Y}) = \mathbf{0}$$

und $N - p$ linear unabhängigen Spalten.

⇒ Anwendung von ML auf die Likelihood dieser Fehlerkontraste/Residuen, welche unabhängig von β ist

Schätzung der fixed effects

REML-Schätzung im LMM:

- REML-Likelihood basiert auf Profile-Likelihood $L_{ML}(\hat{\beta}(\alpha), \alpha)$:

$$L_{REML}(\alpha) = const \left| \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i(\alpha)^{-1} \mathbf{X}_i \right|^{-1/2} L_{ML}(\hat{\beta}(\alpha), \alpha)$$

Vorgehen Schätzung der fixed effects:

- REML-Schätzung der Varianzkomponenten α
(Möglich, da $L_{REML}(\alpha)$ unabhängig von β !)
- Schätzung von β per GLS: empirischer BLUE $\hat{\beta}(\hat{\alpha}_{REML})$

Prädiktion der random effects

- 1 ICC
- 2 Schätzung der fixed effects
- 3 Prädiktion der random effects

Prädiktion der random effects

Prädiktion:

- \mathbf{b}_i als ZVs nicht schätzbar \Rightarrow „Prädiktion“ der REs
- Schätzung basiert hier **nicht** auf marginaler Likelihood, da REs nur in konditionaler Likelihood als feste Parameter vorkommen!
- Herleitung als empirische Bayes-Schätzer
(alternative Herleitung über Henderson's mixed model equations)
 $\Rightarrow \hat{\mathbf{b}}_i(\theta)$ ist **BLUP**

Best: Kleinster MSE aller linearen erwartungstreuen Prädiktoren

$$\mathbb{E}[(\tilde{\mathbf{b}} - \mathbf{b})^T (\tilde{\mathbf{b}} - \mathbf{b})] \geq \mathbb{E}[(\hat{\mathbf{b}} - \mathbf{b})^T (\hat{\mathbf{b}} - \mathbf{b})]$$

Linear

Unbiased (Erwartungstreu): $\mathbb{E}(\hat{\mathbf{b}}_i(\theta)) = \mathbb{E}(\mathbf{b}_i) = 0$

Prediction

\Rightarrow Praxis: $\hat{\mathbf{b}}_i(\hat{\theta})$ als *empirischer* BLUP

Prädiktion der random effects

Shrinkage-Effekt:

- Prädiktion als gewichtete Summe des Populationmittels $\mathbf{X}_i\hat{\boldsymbol{\beta}}$ und der Beobachtungen \mathbf{y}_i :

$$\begin{aligned}\hat{\mathbf{Y}}_i &= \mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\hat{\mathbf{b}}_i \\ &= \boldsymbol{\Sigma}_i(\hat{\boldsymbol{\alpha}})\mathbf{V}_i(\hat{\boldsymbol{\alpha}})^{-1}\mathbf{X}_i\hat{\boldsymbol{\beta}} + (\mathbf{I}_{n_i} - \boldsymbol{\Sigma}_i(\hat{\boldsymbol{\alpha}})\mathbf{V}_i(\hat{\boldsymbol{\alpha}})^{-1})\mathbf{y}_i\end{aligned}$$

⇒ Shrinkage: Geschätzte random effects aus LMM sind betragsmäßig kleiner als wenn man sie als fixed effects berechnen würde

⇒ Je größer die Residualvarianz $\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\alpha}})$ im Vergleich zur Variabilität zwischen den Individuen $\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T$, desto mehr Gewicht liegt auf dem Populationsmittel