Tutorium 2

Analyse longitudinaler Daten

Prof. Dr. Sonja Greven, Dipl. Stat. Jona Cederbaum,
Alexander Bauer

10. Mai 2016

Übersicht

Nachtrag: Hausman-Test & ICC

Hausman-Test

Hausman-Test: *bzgl. Random effects assumption (REA)*

$$T = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})^T \left(\hat{Var}(\hat{\boldsymbol{\beta}}_{FE}) - \hat{Var}(\hat{\boldsymbol{\beta}}_{RE})\right)^{-1} (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \overset{H_0}{\sim} \chi^2_p$$

- Testidee:
  - $\hat{\beta}_{FE}$ ist konsistent
  - $\hat{\beta}_{RE}$ ist unter REA konsistent (und effizient)

  $\Rightarrow$ Schätzer sollten sich unter REA nicht systematisch unterscheiden
- Testentscheidung:
  - „H_1“: REA ist verletzt
  - „H_0“: Keine gesicherte Aussage bzgl. REA möglich
- **Umsetzung in R:** Keine Implementierung für `lme`/`lmer`/`gam`

  $\Rightarrow$ Deshalb: Test per Hand rechnen

ICC

ICC: *Intra-Class Correlation*

Für ein Modell mit einem Random Intercept und Conditional Independence ($\boldsymbol{\Sigma}_i = \sigma^2 I_{n_i}$) 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?

$\Leftrightarrow$ 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

Schätzung der fixed effects

Grundlagen:

- Schätzung basiert auf **marginalem Modell**:

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

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

Schätzung der fixed effects

ML-Schätzung:

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

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

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

  Best: Kleinste Varianz aller linearen erwartungstreuen Schätzer
  Linear
  Unbiased: Erwartungstreue - $\mathbb{E}(\hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\alpha})) = \boldsymbol{\beta}$
  Estimator

- Problem I: $\hat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\alpha})$ abhängig von unbekanntem $\boldsymbol{\alpha}$
- 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.

  $\Rightarrow$ 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 $L_{ML}(\widehat{\beta}(\boldsymbol{\alpha}), \boldsymbol{\alpha})$:

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

Vorgehen Schätzung der fixed effects:

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

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
  $$\mathbb{E}[(\widetilde{\mathbf{b}} - \mathbf{b})^T(\widetilde{\mathbf{b}} - \mathbf{b})] \geq \mathbb{E}[(\widehat{\mathbf{b}} - \mathbf{b})^T(\widehat{\mathbf{b}} - \mathbf{b})]$$
  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 Popoulationmittels $\mathbf{X}_i\hat{\beta}$ und der Beobachtungen $\mathbf{y}_i$:

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

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

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