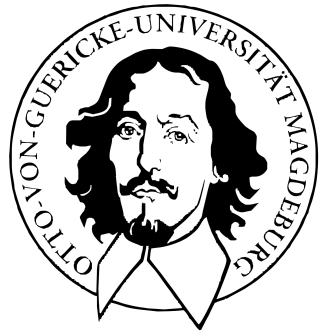


Interventionsforschung

MSc Klinische Psychologie und Psychotherapie

MSc Umweltpsychologie / Mensch-Technik-Interaktion

SoSe 2026

Prof. Dr. Dirk Ostwald

(4) Linear Mixed Models

Überblick

- Das Linear Mixed Model (LMM) ist eine Weiterentwicklung des Allgemeinen Linearen Modells (ALM)
- Durch iterative Schätzverfahren sind LMMs in den letzten 50 Jahren sehr populär geworden
- In **R** sind LMMs durch die Verfügbarkeit der Pakete `nlme` und `lme4` sehr verbreitet
- Klassische Anwendungen von LMMs sind Mehrebenen- und Longitudinalanalysen
- Fixed- und Random-Effects-Modelle der Metaanalyse sind spezielle LMMs
- Viele weitere Modelle sind Spezialfälle von LMMs, z.B. die bayesianische ALM-Schätzung
- LMMs sind die State-of-the-Art-Inferenzmodelle der Psychotherapieforschung (z.B. Detry and Ma (2016))
- In der Umweltpsychologie sind LMMs zunehmend in Gebrauch (z.B. im NudgeProject)

Wir formulieren zunächst das Linear Mixed Model

Zur Schätzung eines Linear Mixed Models betrachten wir dann

- die Generalisierte-Kleinste-Quadrate-Schätzung der Fixed-Effects und ihre Konfidenzintervalle,
- den bedingten Erwartungswert der Random-Effects, sowie
- die Varianzkomponentenschätzung mit Restricted Maximum-Likelihood.

Zur Evaluation eines LMMs betrachten wir dann

- Konfidenzintervalle und p-Werte für Fixed-Effects-Parameter

Anwendungsbeispiel

- Multizentren-Parallelgruppendesign mit Treatmentgruppe und Kontrollgruppe
- Je zwei Gruppen randomisierter Proband:innen an vier verschiedenen HSA-Standorten
- Treatmentfaktor (`TRM`) mit zwei Leveln (`0`: Kontrolle, `1`: Treatment)
- Hochschulambulanzfaktor (`HSA`) mit vier Leveln (`1`: Magdeburg, `2`: Halle, `3`: Dresden, `4`: Marburg)
- 5 Proband:innen pro Gruppe an jedem HSA-Standort
- Primäres Ergebnismaß: BDI-II Differenz
- Random-Intercept-Modell-Analyse

Beispieldatensatz

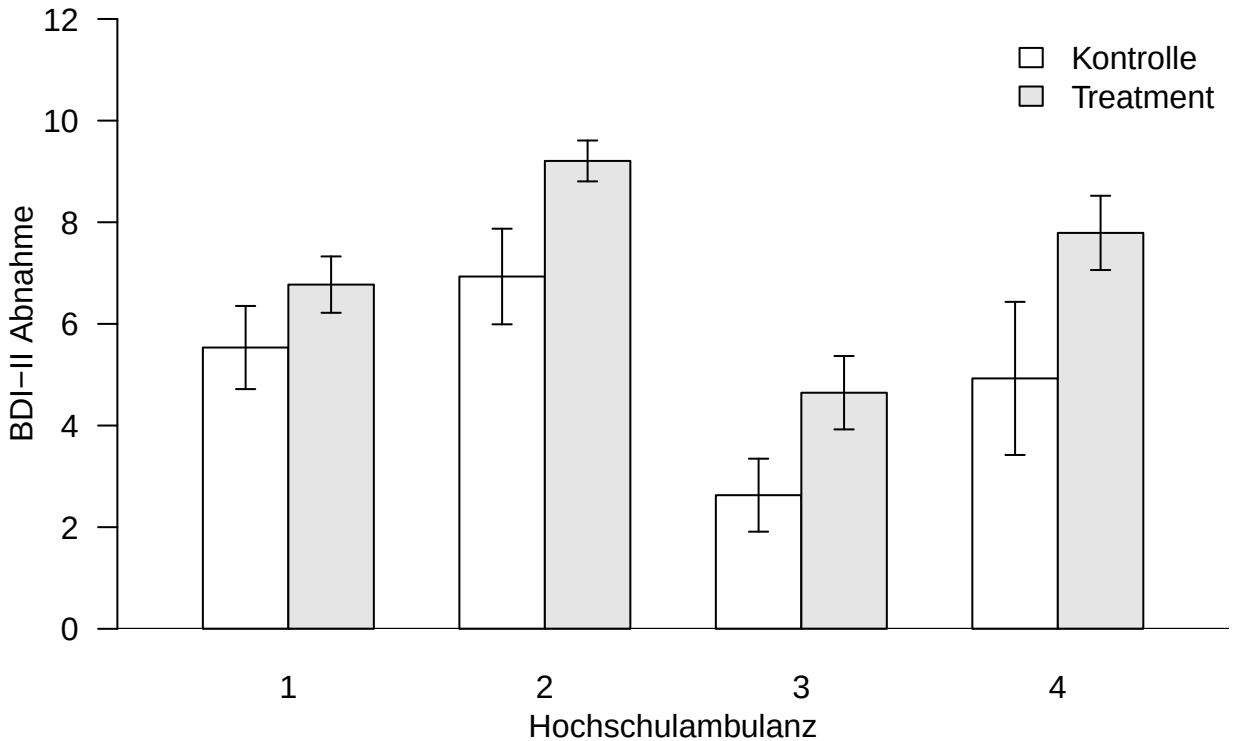
HSA	TRM	BDI
1	0	4.7
1	0	6.2
1	0	4.6
1	0	6.0
1	0	6.2
1	1	7.7
1	1	6.4
1	1	6.6
1	1	6.5
1	1	6.8
2	0	7.0
2	0	5.7
2	0	6.3
2	0	8.1
2	0	7.5
2	1	8.9
2	1	9.8
2	1	9.1
2	1	9.4
2	1	8.8
3	0	1.8
3	0	3.4
3	0	2.1
3	0	3.3
3	0	2.6
3	1	4.7
3	1	4.7
3	1	3.9
3	1	4.2
3	1	5.8
4	0	7.4
4	0	5.0
4	0	4.7
4	0	4.1
4	0	3.5
4	1	7.4
4	1	8.3
4	1	8.3
4	1	6.7
4	1	8.3

Deskriptivstatistik

```r
library(dplyr)                                                    # dplyr für einfache Datengruppierung
D   = read.csv("./4-Daten/mz-anova.csv")                          # Dateneinlesen
DS  = D %>% group_by(HSA, TRM) %>%  summarise(av = mean(BDI, na.rm = TRUE), # Group mean
                                              sd = sd(BDI, na.rm = TRUE),   # Group standard deviation
                                             .groups = "drop")              # Gruppierungsaufhebung
print(DS)
```

```
# A tibble: 8 x 4
    HSA   TRM    av    sd
  <int> <int> <dbl> <dbl>
1     1     0  5.53 0.818
2     1     1  6.77 0.555
3     2     0  6.93 0.941
4     2     1  9.21 0.402
5     3     0  2.63 0.718
6     3     1  4.65 0.722
7     4     0  4.93 1.51
8     4     1  7.79 0.731
```

Visualisierung



Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Definition (Linear Mixed Model in Clusterdarstellung)

Für $i = 1, \ldots, k$ sei

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i, \tag{1}$$

wobei

- y_i ein n_i-dimensionaler beobachtbarer Zufallsvektor ist, der *Daten des $i$ten Clusters* genannt wird,
- $X_i \in \mathbb{R}^{n_i \times p}$ eine vorgegebene Matrix ist, die *Fixed-Effects-Designmatrix des $i$ten Clusters* genannt wird,
- $\beta \in \mathbb{R}^p$ ein unbekannter fester Vektor ist, der *Fixed-Effects-Parameter* genannt wird,
- $Z_i \in \mathbb{R}^{n_i \times q_i}$ eine vorgegebene Matrix ist, die *Random-Effects-Designmatrix des $i$ten Clusters* genannt wird,
- b_i ein q_i-dimensionaler latenter Zufallsvektor ist, der *Random-Effects des $i$ten Clusters* genannt wird, mit

$$b_i \sim N(0_{q_i}, \Sigma_{b_i}) \text{ mit } \Sigma_{b_i} \in \mathbb{R}^{q_i \times q_i} \text{ p.d.}, \tag{2}$$

- ε_i ein n_i-dimensionaler latenter Zufallsvektor ist, der *Zufallsfehler des $i$ten Clusters* genannt wird, mit

$$\varepsilon_i \sim N(0_{n_i}, \Sigma_{\varepsilon_i}) \text{ mit } \Sigma_{\varepsilon_i} \in \mathbb{R}^{n_i \times n_i} \text{ p.d. und unabhängig von } b_j \text{ für } j = 1, \ldots, k. \tag{3}$$

Dann werden die k Gleichungen (1) *Linear Mixed Model in Clusterdarstellung* genannt.

Bemerkungen

- Häufig gelten $\Sigma_{b_i} := \sigma_b^2 I_{q_i}$ mit $\sigma_b^2 > 0$ und $\Sigma_{\varepsilon_i} := \sigma_\varepsilon^2 I_{n_i}$ mit $\sigma_\varepsilon^2 > 0$.
- Die LMM-Clusterdarstellung herrscht in der methodischen Literatur und Anwendungssoftware vor.

Definition (Linear Mixed Model in Kompaktdarstellung)

Gegeben sei die Clusterdarstellung eines Linear Mixed Models und mit $n := \sum_{i=1}^{k} n_i$, $q := \sum_{i=1}^{k} q_i$ seien

$$y := \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix}, \; X := \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}, \; Z := \text{blockdiag}(Z_1, ..., Z_k), \; b := \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}, \; \varepsilon := \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{pmatrix} \tag{4}$$

ein n-dimensionaler beobachtbarer Zufallsvektor, eine $n \times p$ vorgegebene Fixed-Effects-Designmatrix, eine $n \times q$ vorgegebene Random-Effects-Designmatrix, ein q-dimensionaler latenter Zufallsvektor bzw. ein n-dimensionaler latenter Zufallsvektor. Dann wird

$$y = X\beta + Zb + \varepsilon \tag{5}$$

Kompaktdarstellung des Linear Mixed Models genannt und es gelten entsprechend

$$b \sim N(0_q, \Sigma_b) \text{ mit } \Sigma_b \in \mathbb{R}^{q \times q} \text{ p.d. und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \text{ mit } \Sigma_\varepsilon \in \mathbb{R}^{n \times n} \text{ p.d. und unabhängig von } b \tag{6}$$

und mit

$$\Sigma_b := \text{blockdiag}(\Sigma_{b_1}, ..., \Sigma_{b_k}) \text{ und } \Sigma_\varepsilon := \text{blockdiag}(\Sigma_{\varepsilon_1}, ..., \Sigma_{\varepsilon_k}). \tag{7}$$

Bemerkungen

- Wir nehmen durchgängig an, dass X von vollem Spaltenrang ist, d.h. $\text{rg}(X) = p$ gilt.
- Der Betaparametervektor der Kompaktdarstellung entspricht dem Betaparametervektor der Clusterdarstellung.
- In einem LMM hat die Random-Effects-Designmatrix Z immer eine Blockdiagonalmatrixstruktur.

Anwendungsbeispiel

- Grundidee: Modellierung der Daten durch ein Parallelgruppendesign-Modell mit zufälligem HSA-Intercept
- Entsprechend des Parallelgruppendesign-Modells ergibt sich für HSA $i = 1, 2, 3, 4$

$$\begin{aligned} y_{i0j} &= \beta_1 + b_i + \varepsilon_{i0j} \qquad \text{mit } \varepsilon_{i0j} \sim N(0, \sigma_\varepsilon^2) \text{ für } j = 1, \ldots, n_{i_0} \\ y_{i1j} &= \beta_1 + \beta_2 + b_i + \varepsilon_{i1j} \text{ mit } \varepsilon_{i1j} \sim N(0, \sigma_\varepsilon^2) \text{ für } j = 1, \ldots, n_{i_1} \end{aligned} \tag{8}$$

und die entsprechende Datenverteilungsform

$$\begin{aligned} y_{i0j} \,|\, b_i &\sim N(\beta_1 + b_i, \sigma_\varepsilon^2) \qquad \text{bedingt u.i.v. für } j = 1, \ldots, n_{i_0} \text{ mit } \beta_1 \in \mathbb{R}, \sigma_\varepsilon^2 > 0 \\ y_{i1j} \,|\, b_i &\sim N(\beta_1 + \beta_2 + b_i, \sigma_\varepsilon^2) \text{ bedingt u.i.v. für } j = 1, \ldots, n_{i_1} \text{ mit } \beta_1, \beta_2 \in \mathbb{R}, \sigma_\varepsilon^2 > 0 \end{aligned} \tag{9}$$

- Insbesondere sollen die HSA-Intercepts b_i hier normalverteilte Zufallsvariablen sein, es soll gelten

$$b_i \sim N(0, \sigma_b^2) \text{ u.i.v. für } i = 1, 2, 3, 4. \tag{10}$$

- β_1 und β_2 sind hier also Fixed-Effects, b_i für $i = 1, 2, 3, 4$ dagegen Random-Effects
- β_1 enkodiert den Erwartungswert der Kontrollgruppe
- β_2 enkodiert den Erwartungswertunterschied zwischen Treatment- und Kontrollgruppe

Anwendungsbeispiel

- Für die Designmatrixform in Clusterdarstellung ergibt sich für $i = 1, 2, 3, 4$ damit

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0_{10}, \sigma^2_\varepsilon I_{10}), \tag{11}$$

ausgeschrieben als

$$\begin{pmatrix} y_{i01} \\ y_{i02} \\ y_{i03} \\ y_{i04} \\ y_{i05} \\ y_{i11} \\ y_{i12} \\ y_{i13} \\ y_{i14} \\ y_{i15} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} b_i + \begin{pmatrix} \varepsilon_{i01} \\ \varepsilon_{i02} \\ \varepsilon_{i03} \\ \varepsilon_{i04} \\ \varepsilon_{i05} \\ \varepsilon_{i11} \\ \varepsilon_{i12} \\ \varepsilon_{i13} \\ \varepsilon_{i14} \\ \varepsilon_{i15} \end{pmatrix}$$

Anwendungsbeispiel

- Weiterhin ergibt sich für das Linear Mixed Model in Kompaktdarstellung

$$y = X\beta + Zb + \varepsilon \text{ mit } \varepsilon \sim N(0_{40}, \sigma_\varepsilon^2 I_{40}) \text{ und } b \sim N(0_4, \sigma_b^2 I_4) \tag{12}$$

ausgeschrieben als

$$\begin{pmatrix} y_{101} \\ y_{102} \\ y_{103} \\ y_{104} \\ y_{105} \\ y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \\ y_{115} \\ y_{201} \\ y_{202} \\ y_{203} \\ y_{204} \\ y_{205} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{215} \\ y_{301} \\ y_{302} \\ y_{303} \\ y_{304} \\ y_{305} \\ y_{311} \\ y_{312} \\ y_{313} \\ y_{314} \\ y_{315} \\ y_{401} \\ y_{402} \\ y_{403} \\ y_{404} \\ y_{405} \\ y_{411} \\ y_{412} \\ y_{413} \\ y_{414} \\ y_{415} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_{101} \\ \varepsilon_{102} \\ \varepsilon_{103} \\ \varepsilon_{104} \\ \varepsilon_{105} \\ \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{113} \\ \varepsilon_{114} \\ \varepsilon_{115} \\ \varepsilon_{201} \\ \varepsilon_{202} \\ \varepsilon_{203} \\ \varepsilon_{204} \\ \varepsilon_{205} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{213} \\ \varepsilon_{214} \\ \varepsilon_{215} \\ \varepsilon_{301} \\ \varepsilon_{302} \\ \varepsilon_{303} \\ \varepsilon_{304} \\ \varepsilon_{305} \\ \varepsilon_{311} \\ \varepsilon_{312} \\ \varepsilon_{313} \\ \varepsilon_{314} \\ \varepsilon_{315} \\ \varepsilon_{401} \\ \varepsilon_{402} \\ \varepsilon_{403} \\ \varepsilon_{404} \\ \varepsilon_{405} \\ \varepsilon_{411} \\ \varepsilon_{412} \\ \varepsilon_{413} \\ \varepsilon_{414} \\ \varepsilon_{415} \end{pmatrix}.$$

Definition (Verteilungsdarstellung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model in Kompaktdarstellung,

$$y = X\beta + Zb + \varepsilon \text{ mit } b \sim N(0_q, \Sigma_b) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{13}$$

Dann nennt man die äquivalente Darstellung dieses Modells mit der marginalen Verteilung

$$b \sim N(0_q, \Sigma_b) \tag{14}$$

und der bedingten Verteilung

$$y \,|\, b \sim N(X\beta + Zb, \Sigma_\varepsilon) \tag{15}$$

die *Verteilungsdarstellung des Linear Mixed Models*.

Bemerkungen

- Die Äquivalenz folgt mit Theorem Linear-affine Transformation.
- Intuitiv beschreibt der Ausdruck $y = X\beta + Zb + \varepsilon$ eine bedingte Verteilung.
- Die Fehlerkovarianzmatrix Σ_ε ist die Kovarianzmatrix dieser bedingten Verteilung.

Theorem (Gemeinsame Verteilung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model in Kompaktdarstellung,

$$y = X\beta + Zb + \varepsilon \text{ mit } b \sim N(0_q, \Sigma_b) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{16}$$

Dann gilt für die gemeinsame Verteilung von Daten und Random-Effects, dass

$$\begin{pmatrix} b \\ y \end{pmatrix} \sim N\left(\mu_{b,y}, \Sigma_{b,y}\right) \tag{17}$$

mit

$$\mu_{b,y} := \begin{pmatrix} 0_q \\ X\beta \end{pmatrix} \in \mathbb{R}^{q+n} \text{ und } \Sigma_{b,y} := \begin{pmatrix} \Sigma_b & \Sigma_b Z^T \\ Z\Sigma_b & Z\Sigma_b Z^T + \Sigma_\varepsilon \end{pmatrix} \in \mathbb{R}^{(q+n)\times(q+n)} \tag{18}$$

Beweis

Die gemeinsame Verteilung des Linear Mixed Models ergibt sich direkt durch Anwendung von Theorem Gemeinsame Normalverteilungen auf die Verteilungsdarstellung des Linear Mixed Models.

□

Theorem (Marginale Datenverteilung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model in Kompaktdarstellung,

$$y = X\beta + Zb + \varepsilon \text{ mit } b \sim N(0_q, \Sigma_b) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{19}$$

Dann gilt für die marginale Verteilung der Daten, dass

$$y \sim N\left(\mu_y, \Sigma_y\right) \tag{20}$$

mit

$$\mu_y := X\beta \in \mathbb{R}^n \text{ und } \Sigma_y := Z\Sigma_b Z^T + \Sigma_\varepsilon \in \mathbb{R}^{n \times n} \tag{21}$$

Beweis

Die Aussage ergibt sich direkt aus dem Theorem zur gemeinsamen Verteilung des LMMs und Theorem Marginale Normalverteilungen. □

Bemerkungen

- LMMs erlauben es, nicht-sphärische Kovarianzmatrixstrukturen zu modellieren.
- Gilt speziell $\Sigma_b := \sigma_b^2 I_q, \sigma_b^2 > 0$ und $\Sigma_\varepsilon := \sigma_\varepsilon^2 I_n, \sigma_\varepsilon^2 > 0$, so folgt

$$y \sim N\left(X\beta, \sigma_b^2 ZZ^T + \sigma_\varepsilon^2 I_n\right) \tag{22}$$

- Parameter wie σ_b^2 und σ_ε^2 nennt man deshalb auch *Kovarianzkomponenten*.
- Wir nehmen durchgängig an, dass marginale Kovarianzmatrizen positiv definit sind.

Definition (Hierarchische Darstellung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model in Kompaktdarstellung,

$$y = X\beta + Zb + \varepsilon \text{ mit } b \sim N(0_q, \Sigma_b) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{23}$$

Dann nennt man die äquivalente Darstellung dieses Modells in der Form

$$\begin{aligned} b &= 0_q + \eta \qquad \text{mit } \eta \sim N(0_q, \Sigma_b) \;\Leftrightarrow\; b \sim N(0_q, \Sigma_b) \\ y &= X\beta + Zb + \varepsilon \;\text{ mit } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \Leftrightarrow y \,|\, b \sim N(X\beta + Zb, \Sigma_\varepsilon) \end{aligned} \tag{24}$$

die *hierarchische Darstellung des Linear Mixed Models*

Bemerkung

- Man nennt diese Darstellung auch ein *Mehrebenenmodell*.
- Es ist leicht, sich LMMs mit mehr als den hier spezifizierten zwei Ebenen vorzustellen.
- Die obigen Verteilungsaussagen gelten natürlich auch für die hierarchische Form des Linear Mixed Models.

Theorem (Populationsparameterdarstellung eines Linear Mixed Models)

Gegeben sei ein Linear Mixed Model der Form

$$y = X\beta + Zb + \varepsilon \text{ mit } b \sim N(0_q, \Sigma_b), \varepsilon \sim N(0_n, \Sigma_\varepsilon) \text{ und } b \text{ und } \varepsilon \text{ unabhängig.} \tag{25}$$

Wenn es eine Matrix $A \in \mathbb{R}^{q \times p}$ gibt, so dass

$$X = ZA \tag{26}$$

gilt, dann ist das Modell äquivalent darstellbar als

$$y = Zu + \varepsilon \text{ mit } u := A\beta + b, u \sim N(A\beta, \Sigma_b), \varepsilon \sim N(0_n, \Sigma_\varepsilon) \text{ und } u \text{ und } \varepsilon \text{ unabhängig.} \tag{27}$$

Man nennt (27) die *Populationsparameterdarstellung des Linear Mixed Models* (25) mit den *Populationsparametern* $\beta \in \mathbb{R}^p$ und $\Sigma_b \in \mathbb{R}^{q \times q}$.

Bemerkungen

- In (27) denkt man sich u aus einer Population mit Erwartungswert $A\beta$ realisiert.
- Der Fixed-Effects-Parameter β parameterisiert dabei explizit den Erwartungswert dieser Population.
- Fixed-Effects-Parameter können also in diesem Fall als Populationserwartungswerte interpretiert werden.
- Random-Effects-Parameter dagegen entsprechen Abweichungen vom Populationserwartungswert.
- Die Fixed-Effects-Designmatrix ist hier eine Linearkombination der Random-Effects-Designmatrix.
- Das Ganze funktioniert genau dann, wenn die Spalten von X im Spaltenraum von Z liegen.
- Demidenko (2013), Kapitel 4.1 bezeichnet Modelle der Form (27) als Linear-Growth-Curve-Modelle
- Die Unabhängigkeit von u und ε folgt aus der Unabhängigkeit von b und ε.

Beweis

Wir nehmen an, dass es eine Matrix $A \in \mathbb{R}^{q \times p}$ gibt, so dass $X = ZA$. Mit Theorem Linear-affine Transformation gilt für $u := A\beta + b$

$$u \sim N(A\beta, \Sigma_b). \tag{28}$$

Da u nur von b abhängt und b und ε unabhängig sind, sind auch u und ε unabhängig. Weiterhin gilt

$$\begin{aligned}
& y = X\beta + Zb + \varepsilon \quad \text{mit } b \sim N(0_q, \Sigma_b) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \\
\Rightarrow\ & y = ZA\beta + Zb + \varepsilon \text{ mit } b \sim N(0_q, \Sigma_b) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \\
\Rightarrow\ & y = Z(A\beta + b) + \varepsilon \text{ mit } b \sim N(0_q, \Sigma_b) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \\
\Rightarrow\ & y = Zu + \varepsilon \qquad\quad \text{mit } u \sim N(A\beta, \Sigma_b) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon)
\end{aligned} \tag{29}$$

Damit folgt die behauptete Darstellung.

□

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Überblick zur Modellschätzung

Wie bereits gesehen, impliziert das Linear Mixed Model mit

$$\Sigma_b := \sigma_b^2 I_q \text{ und } \Sigma_\varepsilon := \sigma_\varepsilon^2 I_n \tag{30}$$

die gemeinsame Verteilung von Datenvektor und Random-Effects-Vektor

$$\begin{pmatrix} b \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0_q \\ X\beta \end{pmatrix}, \begin{pmatrix} \sigma_b^2 I_q & \sigma_b^2 Z^T \\ \sigma_b^2 Z & \sigma_b^2 Z Z^T + \sigma_\varepsilon^2 I_n \end{pmatrix} \right), \tag{31}$$

sowie die marginale Datenverteilung

$$y \sim N\left(X\beta, \sigma_b^2 Z Z^T + \sigma_\varepsilon^2 I_n\right). \tag{32}$$

Mithilfe der Definitionen des *Varianzkomponentenvektors* θ und des marginalen Datenkovarianzmatrixparameters V_θ

$$\theta := \left(\sigma_b^2, \sigma_\varepsilon^2\right) \text{ und } V_\theta := \sigma_b^2 Z Z^T + \sigma_\varepsilon^2 I_n, \tag{33}$$

wird die marginale Datenverteilung des Linear Mixed Models häufig auch als

$$y \sim N\left(X\beta, V_\theta\right) \tag{34}$$

geschrieben. Allgemein steht θ im Folgenden für die zur jeweiligen Kovarianzstruktur gehörigen Varianz- bzw. Kovarianzkomponenten. Wir nehmen durchgängig an, dass V_θ positiv-definit ist. Weiterhin lassen wir zu, dass die Random-Effects-Kovarianzmatrix Σ_b und die Fehlerkovarianzmatrix Σ_ε beliebige positiv-definite Matrizen sein können, so dass V_θ entsprechend angepasst wird. Das Schätzproblem hat dann drei zentrale Aspekte:

(1) Die Angabe eines Schätzers $\hat{\beta}$ für den Fixed-Effects-Parameter β.
(2) Die Angabe eines Schätzers $\hat{\theta}$ für den Varianzkomponentenparameter θ.
(3) Die Angabe eines Schätzers $\hat{b}$ für den Random-Effects-Parameter b.

Überblick zur Modellschätzung

Die Lösung dieses Problems ist nicht trivial und Gegenstand aktueller Forschung

Generell werden in der Anwendung iterative Verfahren genutzt

Wir beschreiben hier folgenden, der Funktion `lme()` aus dem R-Paket `nlme` nahen Ansatz

(1) Schätzung von β basierend auf der geschätzten Marginalverteilung von y

$\Rightarrow$ V_θ wird durch $V_{\hat{\theta}}$ ersetzt und β durch den *Generalisierten-Kleinste-Quadrate-Schätzer* geschätzt.

(2) Schätzung von θ basierend auf der geschätzten Marginalverteilung von y

$\Rightarrow$ θ wird iterativ mit dem *Restricted-Maximum-Likelihood-Verfahren* geschätzt.

(3) Schätzung von b basierend auf der geschätzten gemeinsamen Verteilung von b und y

$\Rightarrow$ V_θ und β werden durch $V_{\hat{\theta}}$ und $\hat{\beta}$ ersetzt und b durch seinen *bedingten Erwartungswert* geschätzt.

Modellschätzung

Überblick zur Modellschätzung

Iteratives Verfahren zur Linear Mixed Model Parameterschätzung

(0) Initialisierung

- Wahl eines geeigneten Startwerts $\widehat{\theta}^{(0)}$

(1) Für $k = 1, ..., K$

- GLS-Schätzung $\widehat{\beta}^{(k)}$ basierend auf $\widehat{\theta}^{(k-1)}$
- ReML-Schätzung $\widehat{\theta}^{(k)}$ basierend auf $\widehat{\beta}^{(k)}$

(2) Schätzung von $\widehat{b}$ basierend auf $\widehat{\beta}^{(K)}$ und $\widehat{\theta}^{(K)}$

Generalisierte-Kleinste-Quadrate-Schätzung der Fixed-Effects

- Kleinste-Quadrate-Schätzung heißt auf Englisch "Ordinary Least Squares" (OLS).
- Zur Abgrenzung nennen wir den bekannten Betaparameterschätzer im Folgenden "OLS-Schätzer".
- Generalisierte-Kleinste-Quadrate-Schätzung heißt auf Englisch "Generalized Least Squares" (GLS).
- Der GLS-Schätzer ist ein Betaparameterschätzer für das ALM

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \text{ mit } V \neq I_n \tag{35}$$

- Der GLS-Schätzer ist also im Fall nicht-sphärischer Fehlerkovarianzmatrixparameter angezeigt.
- Der GLS-Schätzer stellt sicher, dass T-Statistiken auch im Fall $V \neq I_n$ t-verteilt sind.
- Im Kontext der Fixed-Effects-Schätzung eines LMMs gilt im Hinblick auf obiges ALM speziell

$$X := X, \beta := \beta, \sigma^2 := 1, V := \sigma_b^2 Z Z^T + \sigma_\varepsilon^2 I_n \text{ und somit } \sigma^2 V = V = V_\theta. \tag{36}$$

Definition (Generalisierte-Kleinste-Quadrate-Schätzer)

Gegeben sei ein Allgemeines Lineares Modell der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \tag{37}$$

mit $\sigma^2 > 0$ und einer positiv-definiten Matrix $V \in \mathbb{R}^{n \times n}$. Dann heißt

$$\hat{\beta}_{\text{GLS}} := \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y \tag{38}$$

der *Generalisierte-Kleinste-Quadrate-Schätzer von* β und

$$\hat{\sigma}^2_{\text{GLS}} := \frac{\left(y - X\hat{\beta}_{\text{GLS}}\right)^T V^{-1} \left(y - X\hat{\beta}_{\text{GLS}}\right)}{n - p} \tag{39}$$

der *Generalisierte-Kleinste-Quadrate-Schätzer von* σ^2.

Bemerkungen

- Es muss nicht notwendigerweise $V = I_n$ gelten.
- Die Fehlerkomponenten in ε können unterschiedliche Varianzen haben oder korreliert sein.
- Im Fall $V = I_n$ gilt weiterhin

$$\hat{\beta}_{\text{GLS}} = \left(X^T I_n^{-1} X\right)^{-1} X^T I_n^{-1} y = \left(X^T X\right)^{-1} X^T y =: \hat{\beta}_{\text{OLS}}. \tag{40}$$

Theorem (GLS-Schätzer und OLS-Schätzer)

Gegeben sei ein *untransformiertes ALM* der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \tag{41}$$

mit $\sigma^2 > 0$ und einer positiv-definiten Matrix $V \in \mathbb{R}^{n \times n}$ und es sei $\hat{\beta}_{\text{GLS}}$ der Generalisierte-Kleinste-Quadrate-Schätzer von β. Weiterhin sei $K \in \mathbb{R}^{n \times n}$ eine Matrix mit den Eigenschaften

$$KK^T = V \text{ und } \left(K^{-1}\right)^T K^{-1} = V^{-1} \tag{42}$$

Schließlich sei

$$y^* = X^*\beta + \varepsilon^* \text{ mit } y^* := K^{-1}y, X^* := K^{-1}X, \varepsilon^* := K^{-1}\varepsilon \tag{43}$$

das *transformierte ALM*. Dann gelten

(1) Der GLS-Schätzer des untransformierten ALMs ist der OLS-Schätzer des transformierten ALMs.

(2) Für den Zufallsfehler im transformierten ALM gilt $\varepsilon^* \sim N(0_n, \sigma^2 I_n)$.

Bemerkungen

- Der zu schätzende wahre, aber unbekannte, Parameter β ist in beiden ALMs identisch.
- Im transformierten ALM ist der Fehlerkovarianzmatrixparameter sphärisch, also T-Statistiken t-verteilt.
- Man nennt die Transformation des ALMs durch K auch eine "Whitening-Transformation".
- K mit den geforderten Eigenschaften kann durch die *Cholesky-Zerlegung* von V gewonnen werden.

Beweis

(1) Für den GLS-Schätzer im untransformierten Modell gilt

$$\begin{aligned}\hat{\beta}_{\text{GLS}} &= \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y \\ &= \left(X^T \left(K^{-1}\right)^T K^{-1} X\right)^{-1} X^T \left(K^{-1}\right)^T K^{-1} y \\ &= \left(\left(K^{-1} X\right)^T K^{-1} X\right)^{-1} \left(K^{-1} X\right)^T K^{-1} y \\ &= \left((X^*)^T X^*\right)^{-1} (X^*)^T y^*.\end{aligned} \tag{44}$$

Dies aber entspricht dem OLS-Schätzer im transformierten Modell.

(2) Mit der Tatsache, dass für eine invertierbare Matrix A immer gilt, dass $\left(A^{-1}\right)^T = \left(A^T\right)^{-1}$ und Theorem Linear-affine Transformation ergibt sich

$$\begin{aligned}\varepsilon^* &\sim N\left(K^{-1} 0_n, K^{-1}(\sigma^2 V)(K^{-1})^T\right) \\ &= N\left(0_n, \sigma^2 K^{-1} V (K^{-1})^T\right) \\ &= N\left(0_n, \sigma^2 K^{-1} K K^T (K^{-1})^T\right) \\ &= N\left(0_n, \sigma^2 K^{-1} K (K^{-1} K)^T\right) \\ &= N\left(0_n, \sigma^2 I_n\right).\end{aligned} \tag{45}$$

□

Definition (Fixed-Effects-Parameterschätzer)

Gegeben sei die marginale Datenverteilung eines LMMs mit einem Schätzer $\hat{\theta}$ für θ, also

$$y \sim N(X\beta, V_{\hat{\theta}}). \tag{46}$$

Dann ist der GLS-Schätzer

$$\hat{\beta} = \left(X^T V_{\hat{\theta}}^{-1} X\right)^{-1} X^T V_{\hat{\theta}}^{-1} y \tag{47}$$

ein populärer Schätzer für den Fixed-Effects-Parameter β.

Implementierung des Fixed-Effects-Parameterschätzers

```
gls = function(y, X, V){
    # Diese Funktion bestimmt den Generalisierte-Kleinste-Quadrate-Schätzer.
    #
    # Inputs
    #   y           : n x 1 Datenvektor
    #   X           : n x p Designmatrix
    #   V           : n x n marginale Datenkovarianzmatrix
    #
    # Outputs
    #   beta_hat    : p x 1 Generalisierte-Kleinste-Quadrate-Schätzer
    # ------------------------------------------------------------------------
    Vi          = solve(V)                                          # Inverse
    beta_hat    = solve(t(X) %*% Vi %*% X) %*% t(X) %*% Vi %*% y    # GLS-Schätzer
    return(beta_hat)                                                # Output
}
```

Motivation von Varianzkomponentenschätzung und Restricted Maximum-Likelihood

Die Maximum-Likelihood-Methode kann zu verzerrten Varianzschätzern führen

Zum Beispiel ist der Maximum-Likelihood-Schätzer des Varianzparameters des Normalverteilungsmodells

$$\hat{\sigma}^2_{\text{ML}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_{\text{ML}})^2 \text{ mit } \hat{\mu}_{\text{ML}} := \frac{1}{n}\sum_{i=1}^{n} y_i =: \bar{y} \tag{48}$$

ist verzerrt und nur asymptotisch erwartungstreu. Speziell gilt mit der Erwartungstreue der Stichprobenvarianz

$$\mathbb{E}\left(\hat{\sigma}^2\right) = \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right) = \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right) = \sigma^2, \tag{49}$$

dass

$$\mathbb{E}\left(\hat{\sigma}^2_{\text{ML}}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_{\text{ML}})^2\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right) = \frac{n-1}{n}\sigma^2. \tag{50}$$

Da $(n-1)/n < 1$ insbesondere bei kleinem n, unterschätzt der Maximum-Likelihood-Schätzer den Wert von σ^2.

Patterson and Thompson (1971) schreiben "The difference between the two methods [ML und ReML] is analogous to the well-known difference between two methods of estimating the variance σ^2 of a normal distribution [wie oben] (...)." und Harville (1977) merkt an "One criticism of the ML approach to the estimation of [σ^2] is that the ML estimator (...) takes no account of the loss in degrees of freedom that results from estimating [μ] (...) These"deficiencies" are eliminated in the restricted Maximum-Likelihood (REML) approach (...)"

$\Rightarrow$ ReML für Varianzparameterschätzung scheint eine gute Idee zu sein.

Referenzen zur Theorie der Restricted-Maximum-Likelihood-Schätzung

Patterson and Thompson (1971)

- Fehlerkontrastmotivation der Restricted-Maximum-Likelihood-Zielfunktion

Harville (1977)

- Übersicht zu Restricted-Maximum-Likelihood-Methoden und numerischer Auswertung

Searle et al. (1992)

- Ausführliche Übersicht zum Problem der Varianzkomponentenschätzung

Bates and DebRoy (2004)

- Integration von Restricted Maximum Likelihood in Penalized Least Squares

Starke and Ostwald (2017)

- Expectation-Maximization und Restricted Maximum Likelihood aus der Perspektive von Variational Inference

Maximum-Likelihood und Restricted Maximum-Likelihood

Betrachtet man die marginale Datenverteilung des LMMs

$$y \sim N(X\beta, V_\theta) \tag{51}$$

so ergibt sich für die Log-Likelihood-Funktion der Varianzkomponenten θ für einen Schätzer $\hat{\beta}$ von β

$$\begin{aligned} \ell(\theta) &= \ln\left((2\pi)^{-\frac{n}{2}} |V_\theta|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y - X\hat{\beta})^T V_\theta^{-1}(y - X\hat{\beta})\right)\right) \\ &= -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|V_\theta| - \frac{1}{2}(y - X\hat{\beta})^T V_\theta^{-1}(y - X\hat{\beta}) \end{aligned} \tag{52}$$

Die (numerische) Maximierung dieser Funktion hinsichtlich θ führt zu einem Maximum-Likelihood-Schätzer von θ,

$$\hat{\theta}_{\mathsf{ML}} := \mathsf{argmax}_\theta \left(-\frac{1}{2}\ln|V_\theta| - \frac{1}{2}(y - X\hat{\beta})^T V_\theta^{-1}(y - X\hat{\beta})\right). \tag{53}$$

Ein zentrales Resultat ist, dass ein Restricted Maximum-Likelihood-Schätzer von θ gegeben ist durch

$$\hat{\theta}_{\mathsf{ReML}} := \mathsf{argmax}_\theta \left(-\frac{1}{2}\ln|V_\theta| - \frac{1}{2}\ln|X^T V_\theta^{-1} X| - \frac{1}{2}(y - X\hat{\beta})^T V_\theta^{-1}(y - X\hat{\beta})\right). \tag{54}$$

Die Zielfunktion der ReML-Methode und der ML-Methode unterscheiden sich also nur hinsichtlich eines Terms.

Im Folgenden wollen wir die Motivation für die Einführung des Terms $-\frac{1}{2}\ln|X^T V_\theta^{-1} X|$ (sehr) grob skizzieren.

Motivation der Restricted-Maximum-Likelihood-Funktion durch Fehlerkontraste

Grundidee des von Patterson and Thompson (1971) formulierten Ansatzes ist es, den Effekt von β aus y herauszurechnen und dann die Likelihood-Funktion der so transformierten Daten hinsichtlich von θ zu maximieren.

Genauer ist das Ziel, den Datenvektor durch eine lineare Transformation mit einer Matrix $M \in \mathbb{R}^{m \times n}$ in einen anderen Vektor z zu transformieren, dessen Erwartungswert für jeden möglichen Wert von β der Nullvektor ist, also

$$z = My \text{ mit } \mathbb{E}(z) = 0_m \text{ für alle } \beta \in \mathbb{R}^p \tag{55}$$

und dann die Log-Likelihood-Funktion von z zu maximieren.

Eine solche Matrix M muss insbesondere die Bedingung

$$MX = 0_{m \times p} \tag{56}$$

erfüllen, denn dann gilt

$$\mathbb{E}(z) = \mathbb{E}(My) = \mathbb{E}(MX\beta) = \mathbb{E}(0_{m \times p}\beta) = 0_m \text{ für alle } \beta \in \mathbb{R}^p. \tag{57}$$

Eine prinzipielle Möglichkeit für die Wahl von M ist die $n \times n$ Matrix

$$M = I_n - P_n \text{ mit } P_n := X(X^T X)^{-1} X^T \in \mathbb{R}^{n \times n} \tag{58}$$

mit der sogenannten *Projektionsmatrix* P.

Motivation der Restricted-Maximum-Likelihood-Funktion durch Fehlerkontraste

Es gilt dann nämlich

$$MX = (I_n - P_n)X = X - P_nX = X - X(X^TX)^{-1}X^TX = X - X = 0_{n\times p}. \tag{59}$$

Nutzt man also diese Matrix M zur Transformation der Daten, ergibt sich

$$z = My = (I_n - P_n)y = y - X(X^TX)^{-1}X^Ty = y - X\hat{\beta} = \hat{\varepsilon} \tag{60}$$

und wir sehen, dass eine solche Matrix M die Daten auf die Residuals, also die Differenz zwischen Daten und Modellvorhersage nach Schätzung der Fixed-Effects projiziert. Die Matrix P_n nennt man dementsprechend auch *Residual-forming matrix* oder *Projektionsmatrix* und die Matrix M *Fehlerkontrastmatrix*. Der Vektor z ist dann der Residualvektor, und ReML wird auch häufig als *Residual Maximum-Likelihood* bezeichnet. Eine Zeile einer solchen Matrix M nennt man *Fehlerkontrast*.

Prinzipiell würde man nun die Log-Likelihood-Funktion von $z \in \mathbb{R}^n$ betrachten. Aufgrund von Theorem Linear-affine Transformation hat z die Verteilung

$$z \sim N(MX\beta, MV_\theta M^T) \tag{61}$$

Die entsprechende Log-Likelihood-Funktion lautet also

$$\ell(\theta) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|MV_\theta M^T| - \frac{1}{2}(My)^T(MV_\theta M^T)^{-1}My. \tag{62}$$

Motivation der Restricted-Maximum-Likelihood-Funktion durch Fehlerkontraste

Leider funktioniert die vorgeschlagene Wahl von M in dieser Form nicht, da $\text{rg}(M) = n - p < n$ und $MV_\theta M^T$ damit singulär ist.

Man wählt daher $n - p$ linear unabhängige Zeilen von M und erhält eine Matrix $K \in \mathbb{R}^{m \times n}$ mit vollem Zeilenrang $m = n - p$.

Dabei gilt weiterhin $\mathbb{E}(z) = \mathbb{E}(Ky) = 0_m$ und man möchte

$$\ell(\theta) = -\frac{m}{2} \ln 2\pi - \frac{1}{2} \ln |KV_\theta K^T| - \frac{1}{2} (Ky)^T (KV_\theta K^T)^{-1} Ky \tag{63}$$

maximieren.

Searle et al. (1992) beweisen nun, dass

$$\ln |KV_\theta K^T| = \ln |V_\theta| + \ln |X^T V_\theta^{-1} X| + c_K \tag{64}$$

und

$$(Ky)^T (KV_\theta K^T)^{-1} Ky = (y - X\hat{\beta})^T V_\theta^{-1} (y - X\hat{\beta}) \tag{65}$$

mit einer nicht von θ abhängigen Konstante c_K. Damit ergibt sich für die Log-Likelihood-Funktion von $z = Ky$ aber, dass

$$\ell(\theta) = -\frac{m}{2} \ln 2\pi - \frac{1}{2} \ln |V_\theta| - \frac{1}{2} \ln |X^T V_\theta^{-1} X| - \frac{1}{2} (y - X\hat{\beta})^T V_\theta^{-1} (y - X\hat{\beta}) + c_K \tag{66}$$

also bis auf additive Konstanten identisch mit der ReML-Zielfunktion ist.

Modellschätzung

Die hier gegebene Darstellung lässt allerdings viele Fragen offen

- Warum sind ReML-Schätzer der Varianzkomponenten unverzerrt (vgl. Foulley (1993))?
- Was sind weitere generelle Eigenschaften der ReML-Schätzer (vgl. Harville (1977))?
- Was genau ist das Problem bei Residualprojektion mit $M = (I_n - P_n)$?
- Wie und wann funktionieren die Beweise von Searle et al. (1992)?

Darüber hinaus ergeben sich zumindest folgende Fragen

- Wie verhält sich die Fehlerkontrastmotivation zur Expectation-Maximization-Motivation (vgl. Laird (1982))?
- Wie verhält sich die Fehlerkontrastmotivation zur bedingten Verteilungsmotivation (vgl. Verbyla (1990))?
- Welche Algorithmen eignen sich zur Maximierung der ReML-Zielfunktion (vgl. Lindstrom and Bates (1990))?
- Wie verhalten sich ReML und Penalized-Least-Squares (vgl. Bates and DebRoy (2004))?

Definition (ReML-Varianzkomponentenschätzer)

Gegeben sei die marginale Datenverteilung eines LMMs basierend auf einem Fixed-Effects-Parameterschätzer $\hat{\beta}$, also

$$y \sim N(X\hat{\beta}, V_\theta) \tag{67}$$

Dann ist der ReML-Schätzer in diesem Modell

$$\hat{\theta}_{\text{ReML}} := \text{argmax}_\theta \left(-\frac{1}{2} \ln |V_\theta| - \frac{1}{2} \ln |X^T V_\theta^{-1} X| - \frac{1}{2} (y - X\hat{\beta})^T V_\theta^{-1} (y - X\hat{\beta}) \right), \tag{68}$$

ein populärer Schätzer für den Varianzkomponentenvektor θ.

Implementierung des ReML-Varianzkomponentenschätzers

```
llh_reml = function(theta, y, X, Z, beta_hat){
    # Diese Funktion evaluiert die restricted log likelihood
    # Zielfunktion für das Linear Mixed Model.
    #
    # Inputs
    #   theta     : k x 1 Varianzkomponentenvektor
    #   y         : n x 1 Datenvektor
    #   X         : n x p Fixed-Effects-Designmatrix
    #   Z         : n x q  Random-Effects-Designmatrix
    #   beta_hat  : p x 1 Fixed-Effects-Schätzer
    #
    # Outputs
    #   llh_reml    : 1 x 1 Wert der ReML-Zielfunktion
    # ------------------------------------------------------------------------
    n           = nrow(Z)                                          # Datenpunktanzahl
    V           = theta[1]*Z %*% t(Z) + theta[2]*diag(n)           # marginale Datenkovarianzmatrix
    Vi          = solve(V)                                         # Inverse
    R           = y - X%*%beta_hat                                 # Residuals
    T1          = -(1/2)*log(det(V))                               # Erster Term
    T2          = -(1/2)*log(det(t(X) %*% Vi %*% X))               # Zweiter Term
    T3          = -(1/2)*t(R) %*% Vi %*% R                         # Dritter Term
    llh_reml    = T1 + T2 + T3                                     # Restricted Log Likelihood
    return(llh_reml)                                               # Wert der ReML-Zielfunktion
}
```

Implementierung des ReML-Varianzkomponentenschätzers

```
reml  = function(theta, lmm){
    # Diese Funktion ist eine Wrapperfunktion für llh_reml() zum Gebrauch mit
    # der generischen R Optimierungsfunktion optim().
    #
    # Inputs
    #   theta   : k x 1 Varianzkomponentenvektor
    #   lmm     : Liste von LMM-Komponenten
    #
    # Output
    #   reml    : Wert der restricted log likelihood-Funktion
    # ------------------------------------------------------------------------
    y           = lmm$y                                                  # Datenvektor
    X           = lmm$X                                                  # Fixed-Effects-Designmatrix
    Z           = lmm$Z                                                  # Random-Effects-Designmatrix
    beta_hat    = lmm$beta_hat                                           # Fixed-Effects-Schätzer
    l_reml      = llh_reml(theta,y,X,Z,beta_hat)                         # Wert der ReML-Zielfunktion
    return(-l_reml)                                                      # Ausgabeargument
}
```

```
mcov = function(theta, Z){
    # Diese Funktion generiert eine marginale Datenkovarianzmatrix
    # basierend auf einer Random-Effects-Designmatrix und den Varianzkomponenten.
    #
    # Inputs:
    #   theta   : c x 1 Varianzkomponentenvektor
    #   Z       : n x q Random-Effects-Designmatrix
    #
    # Outputs
    #   V_theta : n x n marginale Kovarianzmatrix
    # ------------------------------------------------------------------------
    V_theta = theta[1]*(Z%*%t(Z)) + theta[2]*diag(nrow(Z))                   # marginale Datenkovarianzmatrix
    return(V_theta)                                                      # Ausgabeargument
}
```

Motivation zum Random-Effects-Parameterschätzer

Das Linear Mixed Model impliziert wie gesehen eine gemeinsame Verteilung von Daten und unbeobachtbarem Random-Effects-Vektor b. Ein Standardvorgehen im Bereich der Linear-Mixed-Model-Schätzung ist es, b durch den Erwartungswert der auf den Daten bedingten Verteilung von b zu schätzen, wobei unbekannte Parameterwerte wiederum durch ihre Schätzer ersetzt werden.

Dieses Vorgehen entspricht damit letztlich einer bayesianischen Punktschätzung von b mit marginaler Verteilung ("Prior distribution")

$$b \sim N\left(0_q, \hat{\sigma}_b^2 I_q\right) \tag{69}$$

und bedingter Verteilung ("Likelihood")

$$y \,|\, b \sim N\left(X\hat{\beta} + Zb, \hat{\sigma}_\varepsilon^2 I_n\right) \tag{70}$$

Anwendung von Theorem Bedingte Normalverteilungen auf die hier relevante gemeinsame Verteilung

$$\begin{pmatrix} b \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0_q \\ X\hat{\beta} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_b^2 I_q & \hat{\sigma}_b^2 Z^T \\ \hat{\sigma}_b^2 Z & V_{\hat{\theta}} \end{pmatrix}\right) \tag{71}$$

ergibt dann als Schätzer für den Random-Effects-Parameter den Erwartungswertparameter der Verteilung $b \,|\, y$

$$\hat{b} = \mu_{b|y} = \hat{\sigma}_b^2 Z^T V_{\hat{\theta}}^{-1}(y - X\hat{\beta}). \tag{72}$$

Definition (Random-Effects-Parameterschätzer)

Gegeben sei die gemeinsame Verteilung von Random-Effects und Daten eines LMMs basierend auf einem Fixed-Effects-Parameterschätzer $\hat{\beta}$ und einem Varianzkomponentenschätzer $\hat{\theta} := \left(\hat{\sigma}_b^2, \hat{\sigma}_\varepsilon^2\right)$, also

$$\begin{pmatrix} b \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0_q \\ X\hat{\beta} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_b^2 I_q & \hat{\sigma}_b^2 Z^T \\ \hat{\sigma}_b^2 Z & V_{\hat{\theta}} \end{pmatrix}\right). \tag{73}$$

Dann ist der bedingte Erwartungswert von b in diesem Modell, also

$$\hat{b} = \mu_{b|y} = \hat{\sigma}_b^2 Z^T V_{\hat{\theta}}^{-1}(y - X\hat{\beta}). \tag{74}$$

ein populärer Schätzer für den Random-Effects-Parameter.

Implementierung des Random-Effects-Parameterschätzers

```
rfx = function(lmm){
    # Diese Funktion bestimmt den bedingten Erwartungswert der Random-Effects.
    #   Inputs :
    #     lmm   : R Liste mit Einträgen
    #       $y              : n x 1 Datenvektor
    #       $X              : n x p Fixed-Effects-Designmatrix
    #       $Z              : n x q Random-Effects-Designmatrix
    #       $beta_hat       : p x 1 Fixed-Effects-Parameterschätzer
    #       $s_b_hat        : 1 x 1 Random-Effects-Varianzkomponente
    #       $s_eps_hat      : 1 x 1 Fehlervarianzkomponente
    #   Outputs :
    #     lmm               : R Liste mit zusätzlichen Einträgen
    #       $b_hat          : q x 1 Random-Effects-Parameterschätzer
    # ------------------------------------------------------------------
    y               = lmm$y                                          # Daten
    Z               = lmm$Z                                          # Random-Effects-Designmatrix
    X               = lmm$X                                          # Fixed-Effects-Designmatrix
    beta_hat        = lmm$beta_hat                                   # Fixed-Effects-Parameterschätzer
    s_b_hat         = lmm$s_b_hat                                    # Random-Effects-Varianzkomponentenschätzer
    s_eps_hat       = lmm$s_eps_hat                                  # Fehlervarianzkomponentenschätzer
    theta_hat       = c(s_b_hat,s_eps_hat)                           # Varianzkomponentenschätzer
    V_theta_hat_i   = solve(mcov(theta_hat, Z))                      # Inverser Datenkovarianzmatrixschätzer
    eps_hat         = (y - X %*% beta_hat)                           # Residuals
    lmm$b_hat       = s_b_hat*t(Z) %*% V_theta_hat_i %*% eps_hat     # Random-Effects-Parameterschätzer
    return(lmm)                                                      # Ausgabe
}
```

Modellschätzung

Überblick zur Modellschätzung

Iteratives Verfahren zur Linear Mixed Model Parameterschätzung

(0) Initialisierung

- Wahl eines geeigneten Startwerts $\widehat{\theta}^{(0)}$

(1) Für $k = 1, ..., K$

- GLS-Schätzung $\widehat{\beta}^{(k)}$ basierend auf $\widehat{\theta}^{(k-1)}$
- ReML-Schätzung $\widehat{\theta}^{(k)}$ basierend auf $\widehat{\beta}^{(k)}$

(2) Schätzung von $\widehat{b}$ basierend auf $\widehat{\beta}^{(K)}$ und $\widehat{\theta}^{(K)}$

Modellschätzung

```r
estimate = function(lmm){
    # Diese Funktion schätzt die Parameter eines LMMs.
    #   Inputs :
    #     lmm   : R Liste mit Einträgen
    #       $y              : n x 1 Datenvektor
    #       $X              : n x p Fixed-Effects-Designmatrix
    #       $Z              : n x q Random-Effects-Designmatrix
    #       $c              : 1 x 1 Varianzkomponentenanzahl
    #   Outputs :
    #     lmm   : R Liste mit zusätzlichen Einträgen
    #       $beta_hat       : p x 1 Fixed-Effects-Parameterschätzer
    #       $s_b_hat        : 1 x 1 Random-Effects-Varianzkomponentenschätzer
    #       $s_eps_hat      : 1 x 1 Fehlervarianzkomponentenschätzer
    # ------------------------------------------------------------------------
    y               = lmm$y                                                  # Datenvektor
    X               = lmm$X                                                  # Fixed-Effects-Designmatrix
    Z               = lmm$Z                                                  # Random-Effects-Designmatrix
    c               = lmm$c                                                  # Anzahl Varianzkomponenten
    n               = nrow(X)                                                # Anzahl Datenpunkte
    p               = ncol(X)                                                # Anzahl Fixed-Effects
    q               = ncol(Z)                                                # Anzahl Random-Effects
    K               = 2^3                                                    # maximale Iterationsanzahl
    theta_hat_k     = matrix(rep(NaN, c*K), nrow = c)                        # Varianzkomponentenschätzerarray
    theta_hat_k[,1] = rep(1,c)                                               # Initialisierung
    beta_hat_k      = matrix(rep(NaN, p*K), nrow = p)                        # Fixed-Effects-Schätzerarray
    V_theta_hat_k   = mcov(theta_hat_k[,1], Z)                               # marginale Datenkovarianzmatrix
    beta_hat_k[,1]  = gls(y,X,V_theta_hat_k)                                 # Fixed-Effects-Parameterschätzer
    for (k in 2:K){                                                          # Iterationen
        lmm$beta_hat    = beta_hat_k[, k-1]                                  # Fixed-Effects-Schätzer k-1
        max_l_reml      = optim(par=theta_hat_k[,k-1], fn=reml, lmm=lmm,     # ReML-Varianzkomponentenschätzung
                                method = "L-BFGS-B",                         # L-BFGS-B Algorithmus
                                lower = rep(.Machine$double.eps, c))         # untere Schranke für Varianzkomponenten
        theta_hat_k[,k] = max_l_reml$par                                     # Varianzkomponentenschätzer k
        V_theta_hat_k   = mcov(theta_hat_k[,k], Z)                           # marginale Datenkovarianzmatrix
        beta_hat_k[,k]  = gls(y,X,V_theta_hat_k)}                            # Fixed-Effects-Parameterschätzer
    lmm$beta_hat    = beta_hat_k[,K]                                         # Fixed-Effects-Parameterschätzer
    lmm$s_b_hat     = theta_hat_k[1,K]                                       # Random-Effects-Varianzkomponente
    lmm$s_eps_hat   = theta_hat_k[2,K]                                       # Fehlervarianzkomponente
    lmm             = rfx(lmm)                                               # Random-Effects-Parameterschätzer
    return(lmm)}                                                             # Ausgabe
```

Parameterschätzung im Anwendungsbeispiel

```
D       = read.csv("./4-Daten/mz-anova.csv")                    # Dateneinlesen
n_i     = 4                                                     # Anzahl Zentren
n_j     = 2                                                     # Anzahl Gruppen
n_ij    = 5                                                     # Proband:innen pro Zentrum und Gruppe
n       = n_i*n_j*n_ij                                          # Gesamtanzahl an Proband:innen
X_i     = kronecker(matrix(c(1,1,0,1), ncol = 2), rep(1,n_ij))  # Parallelgruppendesign-Designmatrix
X       = kronecker(rep(1,n_i), X_i)                            # Fixed-Effects-Designmatrix
Z       = kronecker(diag(n_i), rep(1,n_j*n_ij))                 # Random-Effects-Designmatrix
y       = D$BDI                                                 # Daten
lmm     = list(y = y, X = X, Z = Z, c = 2)                      # LMM-Komponenten
lmm     = estimate(lmm)                                         # Modellschätzung
```

```
beta_hat         : 5.01 2.1
b_hat            : 0.1 1.97 -2.36 0.3
sigsqr_b_hat     : 3.26
sigsqr_eps_hat   : 0.77
```

Parameterschätzung im Anwendungsbeispiel mit `lme()` aus `nlme`

```r
library(nlme)                                                  # nlme-R-Paket
D           = read.csv("./4-Daten/mz-anova.csv", header = T)   # Dataframe
D$TRM       = as.factor(D$TRM)                                 # R-Faktor-Kodierung
D$HSA       = as.factor(D$HSA)                                 # R-Faktor-Kodierung
M           = lme(BDI ~ TRM, data = D, random = ~ 1 | HSA)     # LMM-Schätzung
X           = model.matrix(M,D)                                # Fixed-Effects-Designmatrix
Z           = model.matrix(~ M$groups[[1]] - 1)                # Random-Effects-Designmatrix
beta_hat    = M$coefficients$fixed                             # Fixed-Effects-Parameterschätzer
b_hat       = M$coefficients$random$HSA                        # Random-Effects-Parameterschätzer
s_eps_hat   = M$sigma**2                                       # Fehlervarianzkomponentenschätzer
s_b_hat     = diag(getVarCov(M))                               # Random-Effects-Varianzkomponentenschätzer
ci_beta     = intervals(M, which = "fixed", 0.95)              # Konfidenzintervalle
```

```
beta_hat         : 5.01 2.1
b_hat            : 0.1 1.97 -2.36 0.3
sigsqr_b_hat     : 3.26
sigsqr_eps_hat   : 0.77
```

Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Überblick

Im Gegensatz zum ALM gibt es nicht für alle LMM-Parameterschätzer nicht-iterative Lösungen.

$\Rightarrow$ Die frequentistische Verteilungstheorie der LMM-Parameter ist komplexer als im ALM.

Wir sind hier nur an Konfidenzintervallen und p-Werten für die Fixed-Effects-Parameter interessiert.

Dabei unterscheiden wir im Folgenden drei Fälle

	Fixed-Effects-Parameterschätzer	T-Statistik
V_θ bekannt	Normalverteilung	Normalverteilung
$V_\theta = \sigma^2 V$, σ^2 unbekannt	Normalverteilung	Nichtzentrale t-Verteilung
V_θ unbekannt	Normalverteilung für $n \to \infty$	Normalverteilung für $n \to \infty$

Darüber hinaus gibt es weitere Approximationen für $n \ll \infty$

- Die von `nlme` genutzte ANOVA-Containment-Approximation
- Die von `lmerTest` genutzte Satterthwaite-Approximation (vgl. Satterthwaite (1946))
- Die von `pbkrtest` genutzte Kenward-Roger-Approximation (vgl. Kenward and Roger (1997)).
- Die von `lme4` genutzte Profil-Likelihood-Approximation (vgl. Venzon and Moolgavkar (1988)).

Den ANOVA-Containment-Ansatz vertiefen wir in (6) Mehrebenen-Designs.

Theorem (Verteilung von Fixed-Effects-Parameterschätzer und T-Statistik bei V_θ bekannt)

Für bekannte Σ_b und Σ_ε und damit V_θ sei

$$y = X\beta + Zb + \varepsilon \text{ mit } b \sim N(0_q, \Sigma_b), \varepsilon \sim N(0_n, \Sigma_\varepsilon), b \perp \varepsilon \Rightarrow y \sim N(X\beta, V_\theta) \text{ mit } V_\theta = Z\Sigma_b Z^T + \Sigma_\varepsilon \quad (75)$$

das LMM. Dann gilt für den Fixed-Effects-Parameterschätzer

$$\hat{\beta} := \left(X^T V_\theta^{-1} X\right)^{-1} X^T V_\theta^{-1} y, \text{ dass } \hat{\beta} \sim N\left(\beta, \left(X^T V_\theta^{-1} X\right)^{-1}\right). \quad (76)$$

Weiterhin gilt mit einem Kontrastvektor $c \in \mathbb{R}^p$, $c \neq 0$ und einem Parameter $\beta_\circ \in \mathbb{R}^p$ für die T-Statistik

$$T := \frac{c^T\hat{\beta} - c^T\beta_\circ}{\sqrt{c^T \left(X^T V_\theta^{-1} X\right)^{-1} c}}, \text{ dass } T \sim N(\delta, 1) \text{ mit } \delta := \frac{c^T\beta - c^T\beta_\circ}{\sqrt{c^T \left(X^T V_\theta^{-1} X\right)^{-1} c}}. \quad (77)$$

Bemerkungen

- Die T-Statistik wird hier auch als Z-Statistik bezeichnet.
- Die erste Verteilungsaussage ergibt sich direkt aus Theorem Linear-affine Transformation.
- Die zweite Verteilungsaussage ergibt sich in Analogie zur Z-Transformation.
- Den Fall $\beta_\circ := \beta$ nutzt man für die Konstruktion von Konfidenzintervallen.
- Den Fall $\beta := \beta_\circ$ für Hypothesentests und die Berechnung von p-Werten.
- Das Symbol δ bezeichnet hier den Nichtzentralitätsparameter, unten dagegen das Konfidenzniveau.

Theorem (Konfidenzintervalle und p-Werte bei V_θ bekannt)

Gegeben sei ein LMM mit bekannter marginaler Kovarianzmatrix V_θ und es seien $\hat{\beta}$ der Fixed-Effects-Parameterschätzer und T die T-Teststatistik für einen Kontrastvektor $c \in \mathbb{R}^p$, $c \neq 0$. Dann gilt mit der kumulativen Verteilungsfunktion Φ der Standardnormalverteilung,

$$z_\delta := \Phi^{-1}\left(\frac{1+\delta}{2}\right) \tag{78}$$

sowie

$$\lambda_j := \left(\left(X^T V_\theta^{-1} X\right)^{-1}\right)_{jj}, \text{ dem } j\text{ten Diagonalelement von } \left(X^T V_\theta^{-1} X\right)^{-1}, \tag{79}$$

dass für $j = 1, \ldots, p$

$$\kappa_j := \left[\hat{\beta}_j - \sqrt{\lambda_j} z_\delta, \hat{\beta}_j + \sqrt{\lambda_j} z_\delta\right] \tag{80}$$

ein δ-Konfidenzintervall für die $j$te Komponente β_j des Fixed-Effects-Parameters ist. Weiterhin gilt für einen beobachteten Wert t der T-Statistik eines zweiseitigen Hypothesentests mit einfacher Nullhypothese $c^T\beta = c^T\beta_\circ$ und zweiseitiger Alternativhypothese $c^T\beta \neq 0$, dass

$$\text{p-Wert} = 2(1 - \Phi(|t|)) \tag{81}$$

der zur beobachteten Teststatistik t gehörige p-Wert ist.

Theorem (Verteilung von Fixed-Effects-Parameterschätzer und T-Statistik bei $V_\theta = \sigma^2 V$)

Für bekannte Kovarianzstruktur V und unbekannten Varianzparameter σ^2 sei

$$y \sim N(X\beta, \sigma^2 V) \tag{82}$$

die marginale Datenverteilung eines LMM. Dann gilt für den Fixed-Effects-Parameterschätzer

$$\hat{\beta} := \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y, \text{ dass } \hat{\beta} \sim N\left(\beta, \sigma^2 \left(X^T V^{-1} X\right)^{-1}\right). \tag{83}$$

Weiterhin sei der Varianzparameterschätzer gegeben durch

$$\hat{\sigma}^2 := \frac{(y - X\hat{\beta})^T V^{-1} (y - X\hat{\beta})}{n - p} \tag{84}$$

Mit einem Kontrastvektor $c \in \mathbb{R}^p$, $c \neq 0$ und einem Parameter $\beta_\circ \in \mathbb{R}^p$ für die T-Statistik

$$T = \frac{c^T \hat{\beta} - c^T \beta_\circ}{\sqrt{\hat{\sigma}^2 c^T \left(X^T V^{-1} X\right)^{-1} c}}, \text{ dass } T \sim t(\delta, n - p) \text{ mit } \delta = \frac{c^T \beta - c^T \beta_\circ}{\sqrt{\sigma^2 c^T \left(X^T V^{-1} X\right)^{-1} c}}. \tag{85}$$

Bemerkungen

- Das Szenario ist im Wesentlichen analog zur anwendungsorientierten ALM-Theorie.
- In der LMM Anwendung hat man jedoch meist mehrere Varianzskalenparameter
- Die erste Verteilungsaussage ergibt sich direkt aus Theorem Linear-affine Transformation.
- Die zweite Verteilungsaussage ergibt sich in Analogie zur t-Transformation.
- Ausführliche Beweise finden sich zum Beispiel in Rencher and Schaalje (2008).

Theorem (Konfidenzintervalle und p-Werte bei $V_\theta = \sigma^2 V$)

Für ein LMM mit bekannter Kovarianzstruktur V und unbekanntem Varianzparameter σ^2 seien $\hat{\beta}$ der Fixed-Effects-Parameterschätzer, $\hat{\sigma}^2$ der Varianzparameterschätzer und T die T-Statistik für einen Kontrastvektor $c \in \mathbb{R}^p$, $c \neq 0$. Dann gilt mit der kumulativen Verteilungsfunktion $\Psi(\cdot; n-p)$ der t-Verteilung mit $n-p$ Freiheitsgraden,

$$t_\delta := \Psi^{-1}\left(\frac{1+\delta}{2}; n-p\right), \tag{86}$$

sowie

$$\lambda_j := \left(\left(X^T V^{-1} X\right)^{-1}\right)_{jj}, \text{ dem } j\text{ten Diagonalelement von } \left(X^T V^{-1} X\right)^{-1}, \tag{87}$$

dass für $j = 1, \ldots, p$

$$\kappa_j := \left[\hat{\beta}_j - \sqrt{\hat{\sigma}^2 \lambda_j} t_\delta, \hat{\beta}_j + \sqrt{\hat{\sigma}^2 \lambda_j} t_\delta\right] \tag{88}$$

ein δ-Konfidenzintervall für die $j$te Komponente β_j des Fixed-Effects-Parameters ist. Weiterhin gilt für einen beobachteten Wert t der T-Statistik eines zweiseitigen Hypothesentests mit einfacher Nullhypothese $c^T\beta = c^T\beta_\circ$ und zweiseitiger Alternativhypothese $c^T\beta \neq 0$, dass

$$\text{p-Wert} = 2(1 - \Psi(|t|; n-p)) \tag{89}$$

der zur beobachteten Teststatistik t gehörige p-Wert ist.

Bemerkungen

- Das Szenario ist im Wesentlichen analog zur anwendungsorientierten ALM-Theorie.

Theorem (Verteilung von Fixed-Effects-Parameterschätzer und T-Statistik bei V_θ unbekannt)

Für unbekannten Varianzkomponentenvektor θ sei

$$y \sim N(X\beta, V_\theta) \tag{90}$$

die marginale Datenverteilung eines LMMs und sei $\hat{\theta}$ ein ReML-Schätzer von θ. Dann gilt für den Fixed-Effects-Parameterschätzer

$$\hat{\beta} := \left(X^T V_{\hat{\theta}}^{-1} X\right)^{-1} X^T V_{\hat{\theta}}^{-1} y, \text{ dass } \hat{\beta} \overset{a}{\sim} N\left(\beta, \left(X^T V_{\hat{\theta}}^{-1} X\right)^{-1}\right) \text{ für } n \to \infty. \tag{91}$$

Weiterhin gilt mit einem Kontrastvektor $c \in \mathbb{R}^p$, $c \neq 0$ und einem Parameter $\beta_\circ \in \mathbb{R}^p$ für die T-Statistik

$$T := \frac{c^T\hat{\beta} - c^T\beta_\circ}{\sqrt{c^T\left(X^T V_{\hat{\theta}}^{-1} X\right)^{-1} c}}, \text{ dass } T \overset{a}{\sim} N(\delta, 1) \text{ mit } \delta := \frac{c^T\beta - c^T\beta_\circ}{\sqrt{c^T\left(X^T V_{\hat{\theta}}^{-1} X\right)^{-1} c}} \text{ für } n \to \infty. \tag{92}$$

Bemerkungen

- Es handelt sich um eine asymptotische Wald-Approximation.
- Die Matrix $V_{\hat{\theta}}$ ist zufällig, weil $\hat{\theta}$ aus den Daten geschätzt wird.
- Beweise z.B. in Harville (1977), Cressie and Lahiri (1993) und Richardson and Welsh (1994).
- Zur zusätzlichen Unsicherheit durch geschätzte Varianzkomponenten siehe auch Kackar and Harville (1984).

Theorem (Konfidenzintervalle und p-Werte bei V_θ unbekannt)

Für ein LMM mit unbekannter marginaler Kovarianzmatrix V_θ seien $\hat{\theta}$ ein ReML-Schätzer von θ, $\hat{\beta}$ der Fixed-Effects-Parameterschätzer und T die T-Teststatistik für einen Kontrastvektor $c \in \mathbb{R}^p$, $c \neq 0$. Dann gilt für $n \to \infty$ mit der kumulativen Verteilungsfunktion Φ der Standardnormalverteilung,

$$z_\delta := \Phi^{-1}\left(\frac{1+\delta}{2}\right) \tag{93}$$

sowie

$$\lambda_j := \left(\left(X^T V_{\hat{\theta}}^{-1} X\right)^{-1}\right)_{jj}, \text{ dem } j\text{ten Diagonalelement von } \left(X^T V_{\hat{\theta}}^{-1} X\right)^{-1}, \tag{94}$$

dass für $j = 1, \ldots, p$

$$\kappa_j := \left[\hat{\beta}_j - \sqrt{\lambda_j} z_\delta, \hat{\beta}_j + \sqrt{\lambda_j} z_\delta\right] \tag{95}$$

ein approximatives δ-Konfidenzintervall für die $j$te Komponente β_j des Fixed-Effects-Parameters ist. Weiterhin gilt für einen beobachteten Wert t der T-Statistik eines zweiseitigen Wald-Tests mit einfacher Nullhypothese $c^T\beta = c^T\beta_\circ$ und zweiseitiger Alternativhypothese $c^T\beta \neq 0$, dass

$$\text{p-Wert} = 2(1 - \Phi(|t|)) \tag{96}$$

der zur beobachteten Teststatistik t gehörige approximative p-Wert ist.

Modellevaluation

```r
evaluate = function(lmm){

    # Diese Funktion evaluiert ein geschätztes LMM basierend auf der
    # asymptotischen Wald-Approximation.
    # Inputs:
    #  .lmm    : R Liste mit Einträgen
    #     $y          : n x 1 Datenvektor
    #     $X          : n x p Fixed-Effects-Designmatrix
    #     $Z          : n x q Random-Effects-Designmatrix
    #     $beta_hat   : p x 1 Fixed-Effects-Parameterschätzer
    #     $s_b_hat    : 1 x 1 Random-Effects-Varianzkomponentenschätzer
    #     $s_eps_hat  : 1 x 1 Fehlervarianzkomponentenschätzer
    # Outputs
    #  .lmm    : R Liste mit zusätzlichen Einträgen
    #     $C_beta_hat : p x p Fixed-Effects-Kovarianzmatrixschätzer
    #     $se_beta_hat: p x 1 Fixed-Effects-Standardfehlerschätzer
    #     $kappa_u    : p x 1 untere Fixed-Effects-Wald-KI-Grenzen
    #     $kappa_o    : p x 1 obere Fixed-Effects-Wald-KI-Grenzen
    #     $pvals      : p x 1 Fixed-Effects-p-Werte
    # ------------------------------------------------------------
    y               = lmm$y                                      # Daten
    Z               = lmm$Z                                      # Random-Effects-Designmatrix
    X               = lmm$X                                      # Fixed-Effects-Designmatrix
    beta_hat        = lmm$beta_hat                               # Fixed-Effects-Parameterschätzer
    s_b_hat         = lmm$s_b_hat                                # Random-Effects-Varianzkomponentenschätzer
    s_eps_hat       = lmm$s_eps_hat                              # Fehlervarianzkomponentenschätzer
    theta_hat       = c(s_b_hat,s_eps_hat)                       # Varianzkomponentenschätzer
    V_hat_i         = solve(mcov(theta_hat, Z))                  # inverser Datenkovarianzmatrixschätzer
    lmm$C_beta_hat  = solve(t(X) %*% V_hat_i %*% X)              # Fixed-Effects-Kovarianzmatrixschätzer
    lmm$se_beta_hat = sqrt(diag(lmm$C_beta_hat))                 # Fixed-Effects-Standardfehlerschätzer
    lmm$delta       = 0.95                                       # Konfidenzniveau
    lmm$zdelta      = qnorm((1 + lmm$delta)/2)                   # Wald-KI (Rencher (2008) S. 491)
    lmm$kappa_u     = lmm$beta_hat - lmm$zdelta*lmm$se_beta_hat  # untere KI-Grenzen
    lmm$kappa_o     = lmm$beta_hat + lmm$zdelta*lmm$se_beta_hat  # obere KI-Grenzen
    lmm$t           = beta_hat/lmm$se_beta_hat                   # T-Statistiken
    lmm$pvals       = 2*(1 - pnorm(abs(lmm$t)))                  # p-Werte
    return(lmm)
}
```

Modellevaluation

Approximative Fixed-Effects-Inferenz im Anwendungsbeispiel

```r
lmm         = evaluate(lmm)                                    # Modellevaluation
E           = data.frame(                                      # Ausgabeformatierung
"Value"     = lmm$beta_hat,                                    # Parameterschätzer
"Std.Error" = lmm$se_beta_hat,                                 # Standardfehler
"lower"     = lmm$kappa_u,                                     # untere KI-Grenzen
"upper"     = lmm$kappa_o,                                     # obere KI-Grenzen
"T"         = lmm$t,                                           # T-Statistiken
"pval"      = lmm$pvals,                                       # p-Werte
row.names   = c("beta_1_hat", "beta_2_hat"))                   # Zeilenbeschriftung
print(E, digits = 3)                                           # Ausgabe
```

```
           Value Std.Error lower upper    T     pval
beta_1_hat  5.01     0.924  3.20  6.82 5.42 5.95e-08
beta_2_hat  2.10     0.277  1.56  2.64 7.57 3.62e-14
```

Fixed-Effects-Inferenz im Anwendungsbeispiel mit `lme()` aus `nlme`

```
library(nlme)                                                    # nlme-R-Paket
D           = read.csv("./4-Daten/mz-anova.csv", header = T)     # Dataframe
D$TRM       = as.factor(D$TRM)                                   # R-Faktor-Kodierung
D$HSA       = as.factor(D$HSA)                                   # R-Faktor-Kodierung
M           = lme(BDI ~ TRM, data = D, random = ~ 1 | HSA)       # LMM-Schätzung
t_beta      = summary(M)$tTable                                  # Fixed-Effects-Tabelle
ci_beta     = intervals(M, which = "fixed", 0.95)$fixed          # Konfidenzintervalle
E           = data.frame(                                        # Ausgabeformatierung
"Value"     = t_beta[, "Value"],                                 # Parameterschätzer
"Std.Error" = t_beta[, "Std.Error"],                             # Standardfehler
"lower"     = ci_beta[, "lower"],                                # untere KI-Grenzen
"upper"     = ci_beta[, "upper"],                                # obere KI-Grenzen
"T"         = t_beta[, "t-value"],                               # T-Statistiken
"pval"      = t_beta[, "p-value"])                               # p-Werte
print(E, digits = 3)                                             # Ausgabe
```

```
            Value Std.Error lower upper    T     pval
(Intercept)  5.01     0.924  3.13  6.88 5.42 4.49e-06
TRM1         2.10     0.277  1.54  2.66 7.57 7.05e-09
```

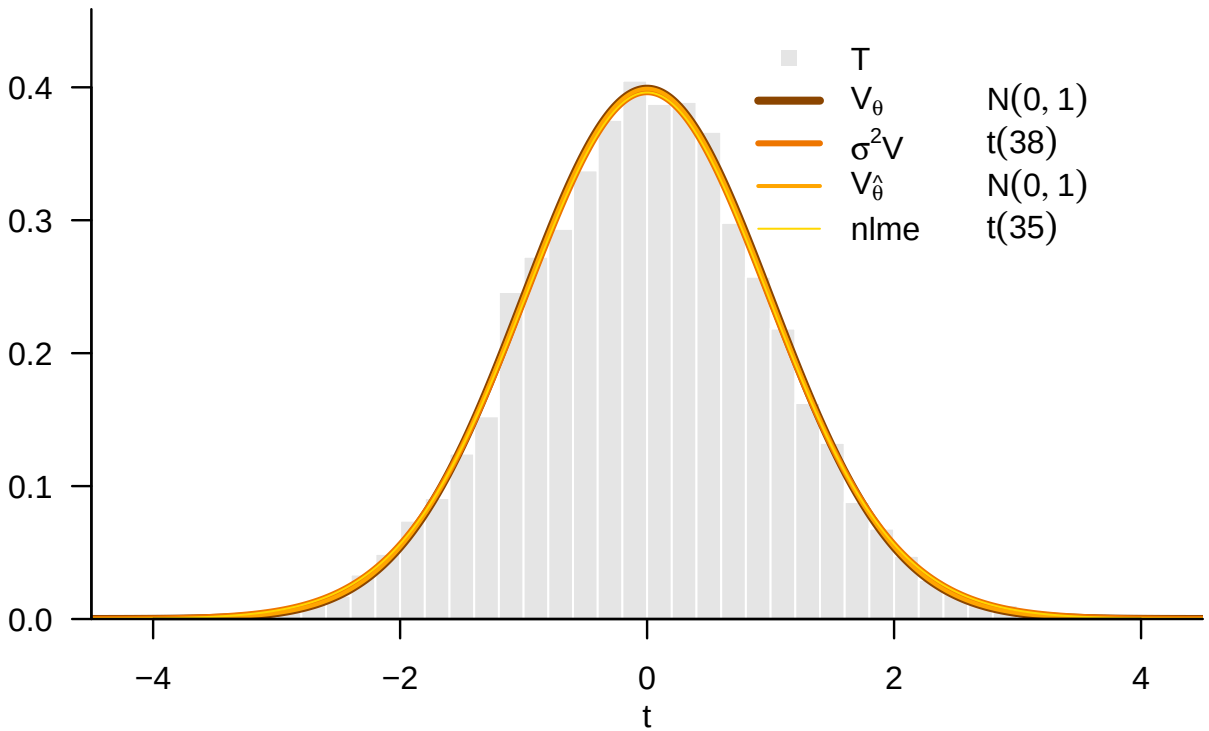
- Offenbar nutzt `lme()` **nicht** die hier diskutierte asymptotische Approximation.
- `lme()` nutzt stattdessen eine ANOVA containment-basierte Approximation (vgl. (6) Mehrebenen-Designs).

Modellevaluation

Simulation der Verteilung der T-Statistik für $\beta_2 = 0$ im Anwendungsbeispiel

```r
library(MASS)                                                  # multivariate Normalverteilung
set.seed(0)                                                    # Zufallszahlengeneratorzustand
nsim        = 1e4                                              # Anzahl Simulationen
k           = 4                                                # Anzahl Zentren
p           = 2                                                # Anzahl Gruppen
n_i         = 10                                               # Proband:innen pro Zentrum
n_ip        = n_i/p                                            # Proband:innen pro Zentrum und Gruppe
n           = k*n_i                                            # Gesamtanzahl an Proband:innen
X_i         = kronecker(matrix(c(1,1,0,1), ncol = 2), rep(1,n_ip))  # Parallelgruppendesign-Designmatrix
X           = kronecker(rep(1,k), X_i)                         # Fixed-Effects-Designmatrix
Z           = kronecker(diag(k), rep(1,n_i))                   # Random-Effects-Designmatrix
TRM         = kronecker(rep(1,k), kronecker(c(0,1), rep(1,n_ip)))  # Treatmentfaktor
HSA         = kronecker(c(1:k), rep(1,n_i))                    # Hochschulambulanzfaktor
beta        = matrix(c(5,0), nrow = p)                         # Fixed-Effects-Parameter unter H_0
s_b         = 1                                                # Random-Effects-Varianzkomponente
s_eps       = 1                                                # Fehlervarianzkomponente
V_theta     = s_b*Z %*% t(Z) + s_eps*diag(n)                   # marginale Datenkovarianzmatrix
nu_gls      = n - p                                            # GLS-Freiheitsgrade
Tee         = rep(NA, nsim)                                    # T-Statistiken
for(s in 1:nsim){                                              # Simulationsiterationen
    y       = as.numeric(mvrnorm(1, X %*% beta, V_theta))      # Daten unter H_0
    D       = data.frame(HSA = factor(HSA),                    # Simulationsdaten
                         TRM = factor(TRM, levels = c(0,1)),   # Treatmentfaktor als Faktor
                         BDI = y)                              # Datenvektor
    M       = lme(BDI ~ TRM, data = D, random = ~ 1 | HSA)     # LMM-Schätzung
    T       = summary(M)$tTable                                # Fixed-Effects-Tabelle
    Tee[s]  = T[grep("^TRM", rownames(T))[1], "t-value"]}      # Treatment-T-Statistik
nu_nlme     = T[grep("^TRM", rownames(T))[1], "DF"]            # nlme-Freiheitsgrade
```

Verteilung der Treatment-T-Statistik für $\beta_2 = 0$



Modellformulierung

Modellschätzung

Modellevaluation

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des LMMs in Clusterdarstellung wieder.
2. Geben Sie die Definition des LMMs in Kompaktdarstellung wieder.
3. Geben Sie das Theorem zur Marginalen Datenverteilung des LMMs wieder.
4. Geben Sie die Definition der hierarchischen Darstellung des LMMs wieder.
5. Erläutern Sie die Bedeutung des Theorems zur Populationsparameterdarstellung eines Linear Mixed Models.
6. Skizzieren Sie ein iteratives Verfahren zur LMM-Parameterschätzung.
7. Geben Sie die Definition des generalisierten Kleinste-Quadrate-Schätzers wieder.
8. Erläutern Sie den Zusammenhang von GLS- und OLS-Schätzern.
9. Geben Sie die Definition des Fixed-Effects-Parameterschätzers im LMM wieder.
10. Geben Sie die Definition des ReML-Varianzkomponentenschätzers im LMM wieder.
11. Geben Sie die Definition des Random-Effects-Parameterschätzers im LMM wieder.
12. Geben Sie das Theorem zur Verteilung des Fixed-Effects-Parameterschätzers und der T-Statistik bei V_θ unbekannt wieder.
13. Geben Sie das Theorem zu Konfidenzintervallen und p-Werten bei V_θ unbekannt wieder.

Anhang

Theorem (Linear-affine Transformation)

$x \sim N(\mu_x, \Sigma_x)$ sei ein n-dimensionaler normalverteilter Zufallsvektor und es sei

$$z := Ax + b \text{ mit } A \in \mathbb{R}^{m \times n} \text{ und } b \in \mathbb{R}^m. \tag{97}$$

Dann gilt

$$z \sim N(\mu_z, \Sigma_z) \text{ mit } \mu_z = A\mu_x + b \in \mathbb{R}^m \text{ und } \Sigma_z = A\Sigma_x A^T \in \mathbb{R}^{m \times m}. \tag{98}$$

Bemerkung

- Jede linear-affine Transformation eines normalverteilten Zufallsvektors ist wiederum normalverteilt.

Theorem (Gemeinsame Normalverteilungen)

$x \sim N(\mu_x, \Sigma_{xx})$ sei ein m-dimensionaler normalverteilter Zufallsvektor, $A \in \mathbb{R}^{n \times m}$ sei eine Matrix, $b \in \mathbb{R}^n$ sei ein Vektor und y sei ein n-dimensionaler, auf x bedingt normalverteilter Zufallsvektor mit

$$y \,|\, x \sim N(Ax + b, \Sigma_{yy}) \text{ mit } \Sigma_{yy} \in \mathbb{R}^{n \times n}. \tag{99}$$

Dann ist der $m + n$-dimensionale Zufallsvektor $(x^T, y^T)^T$ normalverteilt mit

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\mu_{x,y}, \Sigma_{x,y}\right) \tag{100}$$

und insbesondere

$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ A\mu_x + b \end{pmatrix} \text{ und } \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xx}A^T \\ A\Sigma_{xx} & \Sigma_{yy} + A\Sigma_{xx}A^T \end{pmatrix}. \tag{101}$$

Bemerkungen

- Eine marginale und eine bedingte multivariate Normalverteilung induzieren eine gemeinsame Normalverteilung.
- Die Parameter der gemeinsamen Verteilung ergeben sich aus den Parametern der induzierenden Verteilungen.

Theorem (Marginale Normalverteilungen)

Es sei $n := k + l$ und $x = (x_1, ..., x_n)^T$ sei ein n-dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter

$$\mu = \begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix} \in \mathbb{R}^n \tag{102}$$

mit $\mu_y \in \mathbb{R}^k$ und $\mu_z \in \mathbb{R}^l$ und Kovarianzmatrixparameter

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{pmatrix} \in \mathbb{R}^{n \times n}, \tag{103}$$

mit $\Sigma_{yy} \in \mathbb{R}^{k \times k}$, $\Sigma_{yz} \in \mathbb{R}^{k \times l}$, $\Sigma_{zy} \in \mathbb{R}^{l \times k}$ und $\Sigma_{zz} \in \mathbb{R}^{l \times l}$. Dann sind $y := (x_1, ..., x_k)^T$ und $z := (x_{k+1}, ..., x_n)^T$ k- und l-dimensionale normalverteilte Zufallsvektoren, respektive, und es gilt

$$y \sim N(\mu_y, \Sigma_{yy}) \text{ und } z \sim N(\mu_z, \Sigma_{zz}). \tag{104}$$

Bemerkung

- Die Marginalverteilungen einer multivariaten Normalverteilung sind wiederum Normalverteilungen.

Theorem (Bedingte Normalverteilungen)

$(x^T, y^T)^T$ sei ein $m+n$-dimensionaler normalverteilter Zufallsvektor mit

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \right), \tag{105}$$

mit $x, \mu_x \in \mathbb{R}^m$, $y, \mu_y \in \mathbb{R}^n$, $\Sigma_{xx} \in \mathbb{R}^{m \times m}$, $\Sigma_{xy} \in \mathbb{R}^{m \times n}$, $\Sigma_{yx} \in \mathbb{R}^{n \times m}$ und $\Sigma_{yy} \in \mathbb{R}^{n \times n}$. Dann ist die bedingte Verteilung von x gegeben y eine m-dimensionale Normalverteilung,

$$x \,|\, y \sim N(\mu_{x|y}, \Sigma_{x|y}), \tag{106}$$

mit

$$\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y) \in \mathbb{R}^m \tag{107}$$

und

$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \in \mathbb{R}^{m \times m}. \tag{108}$$

Bemerkung

- Bedingte Verteilungen multivariater Normalverteilungen sind wiederum Normalverteilungen.

Referenzen I

Bates, Douglas M, and Saikat DebRoy. 2004. "Linear Mixed Models and Penalized Least Squares." *Journal of Multivariate Analysis* 91 (1): 1–17. https://doi.org/10.1016/j.jmva.2004.04.013.

Cressie, N., and S. N. Lahiri. 1993. "The Asymptotic Distribution of REML Estimators." *Journal of Multivariate Analysis* 45 (2): 217–33. https://doi.org/10.1006/jmva.1993.1034.

Demidenko, Eugene. 2013. *Mixed Models: Theory and Applications with R*. 2. ed. Wiley Series in Probability and Statistics. Wiley.

Detry, Michelle A., and Yan Ma. 2016. "Analyzing Repeated Measurements Using Mixed Models." *JAMA* 315 (4): 407. https://doi.org/10.1001/jama.2015.19394.

Foulley, JL. 1993. *A Simple Argument Showing How to Derive Restricted Maximum Likelihood*.

Harville, David A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72 (358): 320. https://doi.org/10.2307/2286796.

Kackar, Raghu N., and David A. Harville. 1984. "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models." *Journal of the American Statistical Association* 79 (388): 853–62. https://doi.org/10.1080/01621459.1984.10477102.

Kenward, Michael G., and James H. Roger. 1997. "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood." *Biometrics* 53 (3): 983. https://doi.org/10.2307/2533558.

Laird, Nan M. 1982. "Computation of Variance Components Using the Em Algorithm." *Journal of Statistical Computation and Simulation* 14 (3-4): 295–303. https://doi.org/10.1080/00949658208810550.

Lindstrom, Mary J., and Douglas M. Bates. 1990. "Nonlinear Mixed Effects Models for Repeated Measures Data." *Biometrics* 46 (3): 673. https://doi.org/10.2307/2532087.

Referenzen II

Patterson, H. D., and R. Thompson. 1971. "Recovery of Inter-Block Information When Block Sizes Are Unequal." *Biometrika* 58 (3): 545–54. https://doi.org/10.1093/biomet/58.3.545.

Rencher, Alvin C., and G. Bruce Schaalje. 2008. *Linear Models in Statistics*. 2nd ed. Wiley-Interscience.

Richardson, A. M., and A. H. Welsh. 1994. "Asymptotic Properties of Restricted Maximum Likelihood (ReML) Estimates for Hierarchical Mixed Linear Models." *Australian Journal of Statistics* 36 (1): 31–43. https://doi.org/10.1111/j.1467-842X.1994.tb00636.x.

Satterthwaite, F. E. 1946. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2 (6): 110. https://doi.org/10.2307/3002019.

Searle, Shayle, George Casella, and Charles E. McCulloch. 1992. *Variance Components*. Wiley Series in Probability and Mathematical Statistics. Wiley.

Starke, Ludger, and Dirk Ostwald. 2017. "Variational Bayesian Parameter Estimation Techniques for the General Linear Model." *Frontiers in Neuroscience* 11 (September). https://doi.org/10.3389/fnins.2017.00504.

Venzon, D. J., and S. H. Moolgavkar. 1988. "A Method for Computing Profile-Likelihood-Based Confidence Intervals." *Applied Statistics* 37 (1): 87. https://doi.org/10.2307/2347496.

Verbyla, A. P. 1990. "A Conditional Derivation of Residual Maximum Likelihood." *Australian Journal of Statistics* 32 (2): 227–30. https://doi.org/10.1111/j.1467-842X.1990.tb01015.x.