Evaluation und Metaanalyse

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

(5) Linear Mixed Models

Überblick

- Linear Mixed Models sind eine Weiterentwicklung des Allgemeinen Linearen Modells (ALM).
- Durch iterative Schätzverfahren sind Linear Mixed Models in den letzten 50 Jahren sehr populär geworden.
- In **R** sind Linear Mixed Models durch die Verfügbarkeit der Pakete `nlme` und `lme4` sehr verbreitet.
- Klassische Anwendungen von Linear Mixed Models sind Mehrebenen- und Longitudinalanalysen.
- Fixed- und Random-Effects Modelle der Metaanalyse sind spezielle Linear Mixed Models.
- Die Restricted Maximum-Likelihood-Schätzung ist eng mit der Theorie der Linear Mixed Models verwoben.
- Viele weitere Modelle sind Spezialfälle von Linear Mixed Models, z.B. die Bayesianische ALM Schätzung.
- Linear Mixed Models sind die state-of-the-art Inferenzmodelle in vielen Anwendungsfeldern

Wir formulieren zunächst ein allgemeines Linear Mixed Model.

Zur Schätzung eines 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.

Wir betrachten schließlich die Formulierung und Schätzung von Linear Mixed Models mithilfe des **R** Pakets `nlme`.

Modellformulierung

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Bedingter Erwartungswert der Random-Effects

Selbstkontrollfragen

Modellformulierung

Definition (Linear Mixed Model)

Es sei

$$y = X_f \beta_f + X_r \beta_r + \varepsilon, \tag{1}$$

wobei

- y ein n-dimensionaler beobachtbarer Zufallsvektor ist, der *Daten* genannt wird,
- $X_f \in \mathbb{R}^{n \times p}$ eine vorgegeben Matrix ist, die *Fixed-Effects-Designmatrix* genannt wird,
- $\beta_f \in \mathbb{R}^p$ ein unbekannter fester Vektor ist, der *Fixed Effects* genannt wird,
- $X_r \in \mathbb{R}^{n \times q}$ eine vorgegeben Matrix ist, die *Random-Effects-Designmatrix* genannt wird,
- β_r ein q-dimensionaler latenter Zufallsvektor ist der *Random Effects* genannt wird und für den gilt, dass

$$\beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ mit } \Sigma_{\beta_r} \in \mathbb{R}^{q \times q} \text{ p.d.}, \tag{2}$$

- ε ein n-dimensionaler latenter Zufallsvektor ist, der *Zufallsfehler* genannt wird und für den gilt, dass

$$\varepsilon \sim N(0_n, \Sigma_\varepsilon) \text{ mit } \Sigma_\varepsilon \in \mathbb{R}^{n \times n} \text{ p.d. und unabhängig von } \beta_r. \tag{3}$$

Dann heißt (1) *Linear Mixed Model*.

Bemerkungen

- Man bezeichnet Darstellung des Linear Mixed Models in dieser Definition auch als *strukturelle Form* bezeichnen.
- Häufig gelten $\Sigma_{\beta_r} := \sigma^2_{\beta_r} I_q$ mit $\sigma^2_{\beta_r} > 0$ und $\Sigma_\varepsilon := \sigma^2_\varepsilon I_n$ mit $\sigma^2_\varepsilon > 0$.

Modellformulierung

Anwendungsbeispiel

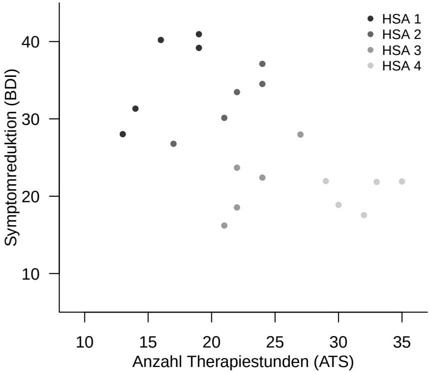
- Studie zum Effekt von Therapiestundenanzahl auf Reduktion der Depressionssymptomatik
- Multizentrendesign mit $k = 4$ Hochschulambulanzen mit jeweils $n_i = 5$ Patient:innen
- Pre-Post-BDI-II Differenz `dBDI` als primäres Ergebnismaß (Positive Werte = Reduktion)
- `HSA`: Hochschulambulanz, `ATS`: Anzahl Therapiestunden, `BDI`: Pre-Post-BDI-II Differenz

```
   HSA ATS BDI
1    1  19  39
2    1  13  28
3    1  14  31
4    1  16  40
5    1  19  41
6    2  17  27
7    2  24  37
8    2  24  35
9    2  22  33
10   2  21  30
11   3  21  16
12   3  22  24
13   3  22  19
14   3  27  28
15   3  24  22
16   4  33  22
17   4  30  19
18   4  32  18
19   4  35  22
20   4  29  22
```

Anwendungsbeispiel

```
# Datengeneration
library(MASS)                                          # multivariate Normalverteilung
library(Matrix)                                        # Blockdiagonalmatrizen
set.seed(0)                                            # Zufallszahlengeneratorzustand
k           = 4                                        # Anzahl Zentren
n_i         = 5                                        # Anzahl Patient:innen pro Zentrum
n           = k*n_i                                    # Gesamtanzahl an Patient:innen
x           = matrix(rep(NaN,n), ncol = k)             # Therapiedauerarray
x[,1]       = round(runif(n_i, min = 10, max = 20))    # Therapiedauer Hochschulambulanz 1
x[,2]       = round(runif(n_i, min = 15, max = 25))    # Therapiedauer Hochschulambulanz 2
x[,3]       = round(runif(n_i, min = 20, max = 30))    # Therapiedauer Hochschulambulanz 3
x[,4]       = round(runif(n_i, min = 25, max = 35))    # Therapiedauer Hochschulambulanz 4
Xi          = array(rep(NaN,n*2), dim = c (n_i,2,k))   # Zentrenspezifische Regressionsmatrizenarray
for(i in 1:k){                                         # Zentreniterationen
    Xi[,,i] = matrix(c(rep(1,n_i),x[,i]), ncol = 2)}   # Zentrumspezifische Regressionmatrix
X_f         = rbind(Xi[,,1],Xi[,,2],Xi[,,3],Xi[,,4])   # Fixed-Effects-Designmatrix
beta_f      = matrix(c(5,1), nrow = 2)                 # Fixed-Effects-Parameter
X_r         = as.matrix(bdiag(Xi[,,1],Xi[,,2],Xi[,,3],Xi[,,4])) # Random-Effects-Designmatrix
beta_r      = matrix(c(15,0,5,0,-5,0,-15,0))           # Random-Effects-Parameter
s_eps       = 15                                       # Varianzkomponente
eps         = mvrnorm(1,rep(0,n), s_eps*diag(n))       # Fehlervektor
y           = X_f %*% beta_f + X_r %*% beta_r + eps    # Datengeneration
HSA         = kronecker(c(1,2,3,4), rep(1,n_i))        # Hochschulambulanzfaktor
D           = data.frame(HSA = HSA, ATS = X_f[,2], BDI = y)   # Dataframe
write.csv(D, "./5_Daten/BDI.csv", row.names = FALSE)   # Speichern
```

Anwendungsbeispiel



Definition (Verteilungsdarstellung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model

$$y = X_f\beta_f + X_r\beta_r + \varepsilon \text{ mit } \beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{4}$$

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

$$\beta_r \sim N(0_q, \Sigma_{\beta_r}) \tag{5}$$

und der bedingten Verteilung

$$y \,|\, \beta_r \sim N(X_f\beta_f + X_r\beta_r, \Sigma_\varepsilon) \tag{6}$$

die Verteilungsdarstellung des Linear Mixed Models

Bemerkungen

- Die Äquivalenz folgt mit dem Theorem zur linear-affinen Transformation multivariate Normalverteilungen.
- Intuitiv beschreibt der Ausdruck $y = X_f\beta_f + X_r\beta_r + \varepsilon$ vor allem 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. Dann gilt für die gemeinsame Verteilung von Daten und Random Effects, dass

$$\begin{pmatrix} \beta_r \\ y \end{pmatrix} \sim N\left(\mu_{\beta_r,y}, \Sigma_{\beta_r,y}\right) \tag{7}$$

mit

$$\mu_{\beta_r,y} := \begin{pmatrix} 0_q \\ X_f\beta_f \end{pmatrix} \in \mathbb{R}^{q+n} \text{ und } \Sigma_{\beta_r,y} := \begin{pmatrix} \Sigma_{\beta_r} & \Sigma_{\beta_r}X_r^T \\ X_r\Sigma_{\beta_r} & X_r\Sigma_{\beta_r}X_r^T + \Sigma_\varepsilon \end{pmatrix} \in \mathbb{R}^{(q+n)\times(q+n)} \tag{8}$$

<u>Beweis</u>

Die gemeinsame Verteilung des Linear Mixed Models ergibt sich direkt durch Anwendung des Theorems zu Gemeinsamen Normalverteilungen auf die Verteilungsdarstellung des Linear Mixed Models.

□

Theorem (Marginale Datenverteilung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model. Dann gilt für die marginale Verteilung der Daten, dass

$$y \sim N\left(\mu_y, \Sigma_y\right) \tag{9}$$

mit

$$\mu_y := X_f \beta_f \in \mathbb{R}^n \text{ und } \Sigma_y := X_r \Sigma_{\beta_r} X_r^T + \Sigma_\varepsilon \in \mathbb{R}^{n \times n} \tag{10}$$

Beweis

Die Aussage ergibt sich direkt aus dem Theorem zur Gemeinsamen Verteilung des Linear Mixed Models und dem Theorem zu Marginalen Normalverteilungen.

$\square$

Bemerkungen

- Linear Mixed Models erlauben es, nicht-sphärische Kovarianzmatrixstrukturen zu modellieren.
- Gilt speziell $\Sigma_{\beta_r} := \sigma^2_{\beta_r} I_q, \sigma^2_{\beta_r} > 0$ und $\Sigma_\varepsilon := \sigma^2_\varepsilon I_n, \sigma^2_\varepsilon > 0$, so folgt

$$y \sim N\left(X_f \beta_f, \sigma^2_{\beta_r} X_r X_r^T + \sigma^2_\varepsilon I_n\right) \tag{11}$$

- Parameter wie $\sigma^2_{\beta_r}$ und σ^2_ε nennt man *Kovarianzkomponenten*.

Definition (Hierarchische Darstellung des Linear Mixed Models)

Gegeben sei ein Linear Mixed Model

$$y = X_f\beta_f + X_r\beta_r + \varepsilon \text{ mit } \beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{12}$$

Dann nennt man die äquivalente Darstellung dieses Modells in der Form

$$\begin{aligned} \beta_r &= 0_q + \eta && \text{mit } \eta \sim N(0_q, \Sigma_{\beta_r}) \Leftrightarrow \ \beta_r \sim N(0_q, \Sigma_{\beta_r}) \\ y &= X_f\beta_f + X_r\beta_r + \varepsilon && \text{mit } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \ \Leftrightarrow y\,|\,\beta_r \sim N(X_f\beta_f + X_r\beta_r, \Sigma_\varepsilon) \end{aligned} \tag{13}$$

die *hierarchische Darstellung des Linear Mixed Models*

Bemerkung

- Man nennt diese Darstellung auch ein *Mehrebenenmodell*.
- Es ist leicht, sich Linear Mixed Models 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 des Linear Mixed Models)

Mit den Definitionen des Linear Mixed Models und einer Matrix $X \in \mathbb{R}^{q \times p}$ sei

$$y = X_r\beta + \varepsilon \text{ mit } \beta := X\beta_f + \beta_r, X_f := X_r X \text{ und } \beta \sim N(X\beta_f, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \tag{14}$$

Dann ist (14) äquivalent zu (1) und wir nennen (14) die *Populationsparameterdarstellung des Linear Mixed Models* mit *Populationsparameter* $\beta_f \in \mathbb{R}^p$.

Bemerkungen

- In (14) denkt man sich β aus einer Population mit Erwartungswert $X\beta_f$ realisiert.
- Der Fixed-Effects-Pameter β_f parameterisiert dabei explizit der Erwartungswert dieser Population.
- Fixed-Effects-Parameter können also als Populationserwartungswerte interpretiert werden.
- Random-Effects-Parameter dagegen entsprechen Abweichungen vom Populationserwartungswert.
- Wir setzen auch hier die Unabhängigkeit von β und ε implizit voraus.

<u>Beweis</u>

Mit dem Theorem zur linearen Transformation normalverteilter Zufallsvektoren gilt

$$
\begin{aligned}
& y = X_r\beta + \varepsilon \text{ mit } \beta := X\beta_f + \beta_r, X_f := X_rX \text{ und } \beta \sim N(X\beta_f, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \\
\Leftrightarrow\, & y = X_r\beta + \varepsilon \text{ mit } \beta := X\beta_f + \beta_r, X_f := X_rX \text{ und } \beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \\
\Leftrightarrow\, & y = X_r\left(X\beta_f + \beta_r\right) + \varepsilon \text{ mit } X_f := X_rX, \beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \\
\Leftrightarrow\, & y = X_rX\beta_f + X_r\beta_r + \varepsilon \text{ mit } X_f := X_rX, \beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon) \\
\Leftrightarrow\, & y = X_f\beta_f + X_r\beta_r + \varepsilon \text{ mit } \beta_r \sim N(0_q, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon).
\end{aligned}
\tag{15}
$$

Modellformulierung

Beispiel zur Populationsparameterdarstellung des Linear Mixed Models

Für das Random-Effects-Modell der Metaanalyse gilt bekanntlich

$$y_i = \delta_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma_i^2) \text{ und } \delta_i \sim N(\delta, \tau^2) \text{ für } i = 1, ..., n \tag{16}$$

und unter der Annahme der paarweisen Unabhängigkeit aller Zufallsvariablen. In Matrixform entspricht das Random-Effects-Modell der Metaanalyse dann zunächst für $y := (y_1, ..., y_n)^T$ und $\tilde{\delta} := (\delta_1, ..., \delta_n)^T$

$$y = \tilde{\delta} + \varepsilon \text{ mit } \tilde{\delta} \sim N(1_n \delta, \tau^2 I_n) \text{ und } \varepsilon \sim N(0_n, \text{diag}(\sigma_1^2, ..., \sigma_n^2)) \tag{17}$$

Mit den Bezeichnungen der Populationsdarstellung des Linear Mixed Models gilt dann äquivalent für

$$\beta := \tilde{\delta}, X := 1_n, \beta_f := \delta, \Sigma_{\beta_r} := \tau^2 I_n \text{ und } \Sigma_\varepsilon := \text{diag}(\sigma_1^2, ..., \sigma_n^2) \tag{18}$$

die Darstellung

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

Offenbar gilt hier also $X_r := I_n$ und es folgt die Linear Mixed Model Darstellung

$$y = X_r X \beta_f + X_r \beta_r + \varepsilon \text{ mit } \beta_r \sim N(0_n, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{20}$$

also

$$y = 1_n \beta_f + \beta_r + \varepsilon \text{ mit } \beta_r \sim N(0_n, \Sigma_{\beta_r}) \text{ und } \varepsilon \sim N(0_n, \Sigma_\varepsilon). \tag{21}$$

Modellformulierung

Überblick zur Modellschätzung

Wie bereits gesehen, impliziert das Linear Mixed Model mit

$$\Sigma_{\beta_r} := \sigma^2_{\beta_r} I_q \text{ und } \Sigma_\varepsilon := \sigma^2_\varepsilon I_n \tag{22}$$

die gemeinsame Verteilung von Datenvektor und Random-Effects-Vektor

$$\begin{pmatrix} \beta_r \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0_q \\ X_f \beta_f \end{pmatrix}, \begin{pmatrix} \sigma^2_{\beta_r} I_q & \sigma^2_{\beta_r} X_r^T \\ \sigma^2_{\beta_r} X_r & \sigma^2_{\beta_r} X_r X_r^T + \sigma^2_\varepsilon I_n \end{pmatrix} \right), \tag{23}$$

sowie die marginale Datenverteilung

$$y \sim N\left(X_f \beta_f, \sigma^2_{\beta_r} X_r X_r^T + \sigma^2_\varepsilon I_n \right). \tag{24}$$

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

$$\theta := \left(\sigma^2_{\beta_r}, \sigma^2_\varepsilon \right) \text{ und } V_\theta := \sigma^2_{\beta_r} X_r X_r^T + \sigma^2_\varepsilon I_n, \tag{25}$$

wird die marginale Datenverteilung des Linear Mixed Models häufig auch als

$$y \sim N\left(X_f \beta_f, V_\theta \right) \tag{26}$$

geschrieben. Das Schätzproblem für ein Linear Mixed Model hat dann drei zentrale Aspekte:

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

Modellformulierung

Überblick zur Modellschätzung

Die Lösung dieses Problems ist nicht trivial und Gegenstand aktueller Forschung.

Generell werden in der Anwendung iterative Verfahren genutzt und meist folgende Ansätze verfolgt:

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

$\Rightarrow$ V_θ wird durch $V_{\hat{\theta}}$ ersetzt und β_f 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 β_r basierend auf der geschätzten gemeinsamen Verteilung von β_r und y

$\Rightarrow$ V_θ, β_f werden durch $V_{\hat{\theta}}$, $\hat{\beta}_f$ ersetzt und β_r durch seinen *bedingten Erwartungswert* geschätzt.

Überblick zur Modellschätzung

Iteratives Verfahren zur Schätzung der Parameter eines Linear Mixed Models

(0) Initialisierung

- Wahl eines geeigneten Startwerts $\hat{\beta}_f^{(0)}$

(1) Für $k = 1, ..., K$

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

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

Modellformulierung

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Bedingter Erwartungswert der Random-Effects

Selbstkontrollfragen

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Generalisierte Kleinste-Quadrate Schätzung

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

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \text{ mit } V \neq I_n \tag{27}$$

- 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 Linear Mixed Models gilt in Hinblick auf obiges ALM speziell

$$X := X_f, \beta := \beta_f, \sigma^2 := \sigma_\varepsilon^2 \sigma_{\beta_r}^2, V := \frac{1}{\sigma_\varepsilon^2} X_r X_r^T + \frac{1}{\sigma_{\beta_r}^2} I_n \text{ und somit } \sigma^2 V = V_\theta. \tag{28}$$

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Definition (Generalisierte Kleinste-Quadrate Schätzer)

Gegeben sei ein Allgemeines Lineares Modell der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \text{ mit} \tag{29}$$

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

$$\hat{\beta}_{\mathsf{GLS}} := \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y \tag{30}$$

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

$$\hat{\sigma}^2_{\mathsf{GLS}} := \frac{\left(y - X\hat{\beta}_{\mathsf{GLS}}\right)^T V^{-1} \left(y - X\hat{\beta}_{\mathsf{GLS}}\right)}{n-p} \tag{31}$$

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

$$\hat{\beta}_{\mathsf{GLS}} = \left(X^T I_n^{-1} X\right)^{-1} X^T I_n^{-1} y = \left(X^T X\right)^{-1} X^T y =: \hat{\beta}_{\mathsf{OLS}}. \tag{32}$$

Theorem (GLS-Schätzer und OLS-Schätzer)

Gegeben sei ein *untransformiertes ALM* der Form

$$y = X\beta + \varepsilon \text{ mit } \varepsilon \sim N\left(0_n, \sigma^2 V\right) \tag{33}$$

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

$$KK^T = V \text{ und } \left(K^{-1}\right)^T K^{-1} = V^{-1} \tag{34}$$

Schließlich sei

$$y^* = X^*\beta + \varepsilon^* \text{ mit } y^* := K^{-1}y, X^* := K^{-1}X, \varepsilon^* := K^{-1}\varepsilon \tag{35}$$

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, Betaparameter 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.

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Beweis

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

$$
\begin{aligned}
\hat{\beta}_{\mathsf{GLS}} &= \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y \\
&= \left(X^T \left(K^{-1}\right)^T K^{-1} X\right)^{-1} X^T \left(K^{-1}\right)^T K^{-1} y \\
&= \left(\left(K^{-1} X\right)^T K^{-1} X\right)^{-1} \left(K^{-1} X\right)^T K^{-1} y \\
&= \left({X^*}^T X^*\right)^{-1} {X^*}^T y^*.
\end{aligned} \tag{36}
$$

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 dem Theorem zur linear-affinen Transformation multivariater Normalverteilungen ergibt sich

$$
\begin{aligned}
\varepsilon^* &\sim N\left(K^{-1} 0_n, K^{-1}(\sigma^2 V) {K^{-1}}^T\right) \\
&= N\left(0_n, \sigma^2 K^{-1} V {K^{-1}}^T\right) \\
&= N\left(0_n, \sigma^2 K^{-1} K K^T {K^{-1}}^T\right) \\
&= N\left(0_n, \sigma^2 K^{-1} K K^T {K^T}^{-1}\right) \\
&= N\left(0_n, \sigma^2 I_n\right).
\end{aligned} \tag{37}
$$

□

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Definition (Fixed-Effects-Parameterschätzer)

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

$$y \sim N(X_f \beta_f, V_{\hat{\theta}}). \tag{38}$$

Dann ist der GLS-Schätzer in diesem Modell

$$\hat{\beta}_f = \left(X_f^T V_{\hat{\theta}}^{-1} X_f\right)^{-1} X_f^T V_{\hat{\theta}}^{-1} y \tag{39}$$

ein populärer Schätzer für den Fixe-Effects-Parameter β_f des Linear Mixed Models.

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

Modellformulierung

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Bedingter Erwartungswert der Random-Effects

Selbstkontrollfragen

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Motivation

Die Maximum-Likelihood Methode kann auf verzerrte Varianzschätzer führen

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

$$\hat{\sigma}^2_{\mathsf{ML}} = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{\mu}_{\mathsf{ML}})^2 \text{ mit } \hat{\mu}_{\mathsf{ML}} := \frac{1}{n}\sum_{i=1}^n y_i =: \bar{y}. \tag{40}$$

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

$$\mathbb{E}\left(\hat{\sigma}^2\right) = \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2\right) = \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^n (y_i - \bar{y})^2\right) = \sigma^2, \tag{41}$$

dass

$$\mathbb{E}\left(\hat{\sigma}^2_{\mathsf{ML}}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n (y_i - \hat{\mu}_{\mathsf{ML}})^2\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^n (y_i - \bar{y}_n)^2\right) = \frac{n-1}{n}\sigma^2 \tag{42}$$

Da $(n-1)/n < 1$ insbesondere bei kleinem n, unterschätzt der Maximum-Likelihood-Schätzer 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.

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Zentrale 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, Casella, and McCulloch (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

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Maximum-Likelihood und Restricted Maximum-Likelihood

Betrachtet man die marginale Datenverteilung des Linear Mixed Models

$$y \sim N(X_f\beta_f, V_\theta) \tag{43}$$

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

$$\begin{aligned}\ell(\theta) &= \ln\left((2\pi)^{-\frac{n}{2}}|V_\theta|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(y-X_f\hat{\beta}_f)^TV_\theta^{-1}(y-X_f\hat{\beta}_f)\right)\right)\\ &= -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|V_\theta| - \frac{1}{2}(y-X_f\hat{\beta}_f)^TV_\theta^{-1}(y-X_f\hat{\beta}_f)\end{aligned} \tag{44}$$

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

$$\hat{\theta}_{\text{ML}} := \text{argmax}_\theta\left(-\frac{1}{2}\ln|V_\theta| - \frac{1}{2}(y-X_f\hat{\beta}_f)^TV_\theta^{-1}(y-X_f\hat{\beta}_f)\right). \tag{45}$$

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

$$\hat{\theta}_{\text{ReML}} := \text{argmax}_\theta\left(-\frac{1}{2}\ln|V_\theta| - \frac{1}{2}\ln|X_f^TV_\theta^{-1}X_f| - \frac{1}{2}(y-X_f\hat{\beta}_f)^TV_\theta^{-1}(y-X_f\hat{\beta}_f)\right). \tag{46}$$

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 $-\frac{1}{2}\ln|X_f^TV_\theta^{-1}X_f|$ (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_f 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_f der Nullvektor ist, also

$$z = My \text{ mit } \mathbb{E}(z) = 0_m \text{ für alle } \beta_f \in \mathbb{R}^p \tag{47}$$

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

Eine solche Matrix M muss insbesondere die Bedingung

$$MX_f = 0_m \tag{48}$$

erfüllen, denn dann gilt

$$\mathbb{E}(z) = \mathbb{E}(My) = \mathbb{E}(MX_f\beta_f) = \mathbb{E}(0_{mp}\beta_f) = 0_m \text{ für alle } \beta_f \in \mathbb{R}^p. \tag{49}$$

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

$$M = I_n - P_n \text{ mit } P_n := X_f(X_f^T X_f)^{-1} X_f^T \in \mathbb{R}^{n \times n} \tag{50}$$

mit der sogenannten *Projektionsmatrix* P.

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Motivation der Restricted Maximum-Likelihood Funktion durch Fehlerkontraste

Es gilt dann nämlich

$$MX_f = (I_n - P_n)X_f = X_f - P_nX_f = X_f - X_f(X_f^TX_f)^{-1}X_f^TX_f = X_f - X_f = 0_{nn}. \tag{51}$$

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

$$z = My = (I_n - P_n)y = y - X_f(X_f^TX_f)^{-1}X_f^Ty = y - X_f\hat{\beta}_f = \hat{\varepsilon} \tag{52}$$

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 sind dann die Residuals und ReML wird auch häufig als *Residual Maximum-Likelihood* bezeichnet. Eine Zeile einer solchen Matrix M nennt man auch *Fehlerkontrast*, die Matrix M daher eine *Fehlerkontrastmatrix*.

Prinzipiell würde man nun die Log-Likelihood Funktion von $z \in \mathbb{R}^n$, das aufgrund des Theorems zur linear-affinen Transformation multivariater Normalverteilungen die Verteilung

$$z \sim N(MX\beta_f, MV_\theta M^T) \tag{53}$$

hat, also

$$\ell(\theta) = \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|MV_\theta M^T| - \frac{1}{2}(My)^T(MV_\theta M^T)^{-1}My. \tag{54}$$

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Motivation der Restricted Maximum-Likelihood Funktion durch Fehlerkontraste

Leider funktioniert die vorgeschlagene Wahl von M in dieser Form nicht, "da rg$(M) = m < n$".

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

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

$$\ell(\theta) = \frac{m}{2}\ln 2\pi - \frac{1}{2}\ln|KV_\theta K^T| - \frac{1}{2}(Ky)^T(KV_\theta K^T)^{-1}Ky \tag{55}$$

maximieren.

Searle, Casella, and McCulloch (1992) beweisen nun, dass

$$\ln|KV_\theta K^T| = \ln|V_\theta| + \ln|X_f V_\theta^{-1} X_f| \tag{56}$$

und

$$(Ky)^T(KV_\theta K^T)^{-1}Ky = y^T P_n y = (y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f) \tag{57}$$

Dies ist intuitiv zumindest unter dem Aspekt, dass K Teil von P_n ist, einsichtig. Damit ergibt sich für die Log-Likelihood-Funktion von $z = Ky$ aber, dass

$$\ell(\theta) = \frac{m}{2}\ln 2\pi - \frac{1}{2}\ln|V_\theta| - \frac{1}{2}\ln|X_f V_\theta^{-1} X_f| - \frac{1}{2}(y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f) \tag{58}$$

also identisch mit der ReML Zielfunktion ist.

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

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, Casella, and McCulloch (1992)?

Darüber hinaus ergeben sich zumindest folgende Fragen

- Was 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 Linear Mixed Models basierend auf einem Fixed-Effects-Parameterschätzer $\hat{\beta}_f$, also

$$y \sim N(X_f\hat{\beta}_f, V_\theta) \tag{59}$$

Dann ist der ReML Schätzer in diesem Modell,

$$\hat{\theta}_{\text{ReML}} := \text{argmax}_\theta \left(-\frac{1}{2}\ln|V_\theta| - \frac{1}{2}\ln|X_f^T V_\theta^{-1} X_f| - \frac{1}{2}(y - X_f\hat{\beta}_f)^T V_\theta^{-1}(y - X_f\hat{\beta}_f)\right), \tag{60}$$

ein populärer Schätzer für den Varianzkomponentenvektor θ.

```
llh_reml = function(theta, y, X_f, X_r, beta_f_hat, Sigma_eps){
    # Diese Funktion evaluiert die negative restricted log likelihood
    # Zielfunktion für  das Random-Effects-Modell der Metanalyse.
    #
    # Inputs
    #   theta       : k x 1 Varianzkomponentenvektor
    #   y           : n x 1 Datenvektor
    #   X_f         : n x p Fixed-Effects-Designmatrix
    #   X_r         : n x q  Random-Effects-Designmatrix
    #   beta_f_hat  : p x 1 Fixed-Effects Schätzer
    #
    # Outputs
    #   llh_reml    : 1 x 1 Wert der ReML Zielfunktion
    # -------------------------------------------------------------------
    n           = nrow(X_r)                                          # Datenpunktanzahl
    V           = theta[1]*X_r %*% t(X_r) + theta[2]*diag(n)         # marginale Datenkovarianzmatrix
    Vi          = solve(V)                                           # Inverse
    R           = y - X_f%*%beta_f_hat                               # Residuals
    T1          = -(1/2)*log(det(V))                                 # Erster Term
    T2          = -(1/2)*log(det(t(X_f) %*% Vi %*% X_f))             # 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
```

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Hilfsfunktionen zur ReML-Zielfunktionoptimierung

```
reml  = function(theta, lmm){
    # Diese Funktion ist eine Wrapperfunktion für l_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 likelhood Funktion
    # -------------------------------------------------------------
    y           = lmm$y                                         # Datenvektor
    X_f         = lmm$X_f                                       # Fixed-Effects-Designmatrix
    X_r         = lmm$X_r                                       # Random-Effects-Designmatrix
    beta_f_hat  = lmm$beta_f_hat                                # Fixed-Effects Schätzer
    l_reml      = llh_reml(theta,y,X_f,X_r,beta_f_hat)          # Wert der ReML Zielfunktion
    return(-l_reml)}                                            # Ausgabeargument

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

Modellformulierung

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Bedingter Erwartungswert der Random-Effects

Selbstkontrollfragen

Bedingter Erwartungswert der Random Effects

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

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

$$\beta_r \sim N\left(0_q, \hat{\sigma}^2_{\beta_r} I_q\right) \tag{61}$$

und bedingter Verteilung ("Likelihood")

$$y \,|\, \beta_r \sim N\left(X_f \hat{\beta}_f + X_r \beta_r, \hat{\sigma}^2_{\varepsilon} I_n\right) \tag{62}$$

Anwendung des Theorems zur bedingten Normalverteilungen auf die hier relevante gemeinsame Verteilung

$$\begin{pmatrix} \beta_r \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0_q \\ X_f \hat{\beta}_f \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2_{\beta_r} I_q & \hat{\sigma}^2_{\beta_r} X_r^T \\ \hat{\sigma}^2_{\beta_r} X_r & V_{\hat{\theta}} \end{pmatrix}\right) \tag{63}$$

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

$$\hat{\beta}_r = \mu_{\beta_r|y} = \hat{\sigma}^2_{\beta_r} X_r^T V_{\hat{\theta}}^{-1}(y - X_f \hat{\beta}_f). \tag{64}$$

Bedingter Erwartungswert der Random Effects

Definition (Random-Effects-Parameterschätzer)

Gegeben sei die gemeinsame Verteilung von Random-Effects und Daten eines Linear Mixed Models basierend auf einem Fixed-Effects-Parameterschätzer $\hat{\beta}_f$ und einem Varianzkomponentenschätzer $\hat{\theta} := \left(\hat{\sigma}^2_{\beta_r}, sig\hat{m}a^2_{\varepsilon}\right)$, also

$$\begin{pmatrix} \beta_r \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0_q \\ X_f\hat{\beta}_f \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2_{\beta_r} I_q & \hat{\sigma}^2_{\beta_r} X_r^T \\ \hat{\sigma}^2_{\beta_r} X_r & V_{\hat{\theta}} \end{pmatrix}\right). \tag{65}$$

Dann ist bedingte Erwartungswert von β_r in diesem Modell, also

$$\hat{\beta}_r = \mu_{\beta_r|y} = \hat{\sigma}^2_{\beta_r} X_r^T V_{\hat{\theta}}^{-1}(y - X_f\hat{\beta}_f). \tag{66}$$

ein populärer Schätzer für den Random-Effects-Parameter.

Bedingter Erwartungswert der Random Effects

```
rfx = function(lmm){
    # Diese Funktion bestimmt den bedingten Erwartungswert der Random-Effects
    #   Inputs :
    #     lmm   : R Liste mit Einträgen
    #       $y              : n x 1 Datenvektor
    #       $X_f            : n x p Fixed-Effects-Designmatrix
    #       $X_r            : n x q Random-Effects-Designmatrix
    #       $beta_f_hat     : p x 1 Fixed-Effects-Parameterschätzer
    #       $s_beta_r_hat   : 1 x 1 Random-Effects-Varianzkomponente
    #       $s_eps_hat      : 1 x 1 Fehler-Varianzkomponente
    #   Outputs :
    #     lmm               : R Liste mit zusätzlichen Einträgen
    #       $beta_r_hat     : q x 1 Random-Effects-Parameterschätzer
    # -------------------------------------------------------------------
    y               = lmm$y                                             # Daten
    X_r             = lmm$X_r                                           # Random-Effects-Designmatrix
    X_f             = lmm$X_f                                           # Fixed-Effects-Designmatrix
    beta_f_hat      = lmm$beta_f_hat                                    # Fixed-Effects-Parameterschätzer
    s_beta_r_hat    = lmm$s_beta_r_hat                                  # Random-Effects-Varianzkomponentenschätzer
    s_eps_hat       = lmm$s_eps_hat                                     # Fehlervarianzkomponentenschätzer
    theta_hat       = c(s_beta_r_hat,s_eps_hat)                         # Varianzkomponentenschätzer
    V_theta_hat_i   = solve(mcov(theta_hat, X_r))                       # Inverser Datenkovarianzmatrixschätzer
    eps_hat         = (y - X_f %*% beta_f_hat)                          # Residuals
    lmm$beta_r_hat  = s_beta_r_hat*t(X_r) %*% V_theta_hat_i %*% eps_hat} # Random-Effects-Parameterschätzer
```

Iteratives Verfahren zur Schätzung der Parameter eines Linear Mixed Models

(0) Initialisierung

- Wahl eines geeigneten Startwerts $\hat{\beta}_f^{(0)}$

(1) Für $k = 1, ..., K$

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

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

Modellschätzung

Iteratives Verfahren zur Schätzung der Parameter eines Linear Mixed Models

```
estimate = function(lmm){
    # Diese Funktion schätzt die Parameter eines Linear Mixed Models.
    #   Inputs :
    #     lmm    : R Liste mit Einträgen
    #       $y               : n x 1 Datenvektor
    #       $X_f             : n x p Fixed-Effects-Designmatrix
    #       $X_r             : n x q Random-Effects-Designmatrix
    #       $c               : 1 x 1 Varianzkomponentenanzahl
    #   Outputs :
    #     lmm    : R Liste mit zusätzlichen Einträgen
    #       $beta_f_hat      : p x 1 Fixed-Effects-Parameterschätzer
    #       $s_beta_r_hat    : 1 x 1 Random-Effects-Varianzschätzer
    #       $s_eps_hat       : 1 x 1 Datenvarianzschätzer
    # -----------------------------------------------------------------------
    y                       = lmm$y                                           # Datenvektor
    X_f                     = lmm$X_f                                         # Fixed-Effects-Designmatrix
    X_r                     = lmm$X_r                                         # Random-Effects-Designmatrix
    c                       = lmm$c                                           # Anzahl Varianzkomponenten
    n                       = nrow(X_f)                                       # Anzahl Datenpunkte
    p                       = ncol(X_f)                                       # Anzahl Fixed-Effects
    q                       = ncol(X_r)                                       # Anzahl Random-Effects
    K                       = 2^3                                             # maximale Iterationsanzahl
    theta_hat_k             = matrix(rep(NaN, c*K), nrow = c)                 # Varianzkomponentenschätzerarray
    theta_hat_k[,1]         = rep(1,c)                                        # Initialisierung
    beta_f_hat_k            = matrix(rep(NaN, p*K), nrow = p)                 # Fixed-Effects-Schätzerarray
    V_theta_hat_k           = mcov(theta_hat_k[,1], X_r)                      # marginale Datenkovarianzmatrix
    beta_f_hat_k[,1]        = gls(y,X_f,V_theta_hat_k)                        # Fixed-Effects-Parameterschätzer
    for (k in 2:K){                                                           # Iterationen
        lmm$beta_f_hat      = beta_f_hat_k[, k-1]                             # Fixed-Effects-Schätzer k-1
        max_l_reml          = optim(par=theta_hat_k[,k-1],fn=reml,lmm=lmm)    # ReML-Varianzkomponentschätzung
        theta_hat_k[,k]     = max_l_reml$par                                  # Varianzkomponentenschätzer k
        V_theta_hat_k       = mcov(theta_hat_k[,k], X_r)                      # marginale Datenkovarianzmatrix
        beta_f_hat_k[,k]    = gls(y,X_f,V_theta_hat_k)}                       # Fixed-Effects-Parameterschätzer
    lmm$beta_f_hat          = beta_f_hat_k[,K]                                # Fixed-Effects-Parameterschätzer
    lmm$s_beta_r_hat        = theta_hat_k[1,K]                                # Random-Effects-Varianzkomponente
    lmm$s_eps_hat           = theta_hat_k[2,K]                                # Fehler-Varianzkomponente
    lmm$beta_r_hat          = rfx(lmm)                                        # Random-Effects-Parameterschätzer
    return(lmm)}                                                              # Ausgabe
```

Modellschätzung

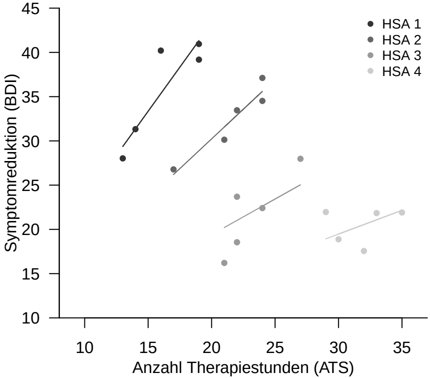
Iteratives Verfahren zur Schätzung der Parameter eines Linear Mixed Models

```
library(Matrix)                                                          # Blockdiagonalmatrizen
D           = read.csv("./5_Daten/BDI.csv")                              # Dateneinlesen
n           = nrow(D)                                                    # Anzahl Datenpunkte
k           = 4                                                          # Anzahl Zentren
n_i         = n/k                                                        # Anzahl Datenpunkte pro Zenturm
x           = matrix(rep(NaN,n), ncol = k)                               # Therapiedauerarray
for(i in 1:k){                                                           # Zentreniterationen
    x[,i] = D$ATS[D$HSA == i]}                                           # Anzahl Therapiestunden pro Zentrum
Xi          = array(rep(NaN,n*2), dim = c (n_i,2,k))                     # Zentrenspezifische Regressionsmatrizenarray
for(i in 1:k){                                                           # Zentreniterationen
    Xi[,,i] = matrix(c(rep(1,n_i),x[,i]), ncol = 2)}                     # Zentrumspezifische Regressionmatrix
X_f         = rbind(Xi[,,1],Xi[,,2],Xi[,,3],Xi[,,4])                     # Fixed-Effects-Designmatrix
X_r         = as.matrix(bdiag(Xi[,,1],Xi[,,2],Xi[,,3],Xi[,,4]))          # Random-Effects-Designmatrix
y           = D$BDI                                                      # Daten
lmm         = list(y = y, X_f = X_f, X_r = X_r, c = 2)                   # LMM Komponenten
lmm         = estimate(lmm)                                              # Modellschätzung
```

```
beta_f_hat        : 3.37 1.17
beta_r_hat        : 0.07 0.82 0.01 0.17 -0.07 -0.36 0 -0.63
sigsqr_beta_r_hat : 0.42
sigsqr_eps_hat    : 6.21
```

Iteratives Verfahren zur Schätzung der Parameter eines Linear Mixed Models

$$\hat{y} = X_f\hat{\beta}_f + X_r\hat{\beta}_r$$



Modellformulierung

Generalisierte Kleinste-Quadrate Schätzung der Fixed-Effects

Varianzkomponentenschätzung mit Restricted Maximum-Likelihood

Bedingter Erwartungswert der Random-Effects

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Linear Mixed Models wieder.

2. Geben Sie das Theorem zur Marginalen Datenverteilung des Linear Mixed Models wieder.

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.

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.

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.

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.

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

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.