Praktikum z ekonometrie - Týden 7
Mixed effect models

VŠE Praha

Tomáš Formánek

Content

1. Introduction
2. LME model example
3. LME model in matrix form
4. More complex LME models
5. LME models in R

Introduction

Linear mixed model (LME) is a generalization of linear model

- Standard linear model looks like:

  $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$
  where $1 \leq i \leq n$.

- In a mixed model, we often use a nesting (hierarchical) structure: there are multiple groups (or panels or individuals) of observations. For example:

  y_{ij} - observation for i-th individual within j-th group.

  Also, we can group the observations at multiple levels:

  y_{tij} - time period t, country i, region j (group of countries).

- Please note that indices are ordered (left to right) from individual to highest level of aggregation. (alternative notations/orderings exist in literature).

Introduction

Linear mixed model (LME) is a generalization of linear model

- y_{ij} - e.g. observation for i-th individual within j-th group.

  $y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$

  where $1 \leq i \leq n_j$ and the number of individuals may differ across groups j.

- One or more β-coefficients can vary across groups.

- Nested/hierarchical structure of the LME model:
  - Individuals i (Level 1) are nested
  - within j groups (Level 2).

Introduction

Longitudinal data and LME models

- Longitudinal data: series of measurements are performed on each individual (say, over time). Several individuals are sampled.

  y_{ij} - i-th observation of the j-th individual.

  $y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij}$

  where $1 \leq i \leq n_j$: the number of observations i can differ across individuals j.

- One or more β-coefficients can vary across individuals.

- The same nesting/hierarchical framework applies to longitudinal data and their LME-based analysis:
  - Observations i (Level 1) are nested
  - within j individuals (Level 2).
  - If appropriate, individuals can be nested in groups (Level 3) ...

Introduction

- Mixed models are called “mixed”, because the β-coefficients are a mix of fixed parameters and random variables
  - The terms “fixed” and “random” are being used in the statistics-biostatistics sense.
  - A fixed coefficient is an unknown constant to be estimated.
  - A random coefficient is one which varies from group to group (longitudinal data: by individuals…at Level 2).
- LME models can have some added complexity:
  - Correlations between different random coefficients.
  - Multiple levels of nesting
- Random coefficients are not estimated, but they can be predicted.

LME model example

- Data:
  London Education Authority Junior School Project dataset,
  48 different schools (j) and 887 different students (i).
  We predict 5th-year math scores.

- We may start by simply ignoring the school grouping (*single-mean* model):

$$y_{ij} = \beta_0 \ + \ \varepsilon_{ij}, \quad i = 1, \ldots, n_j, \quad j = 1, \ldots, M, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

  where y_{ij} is the observed $\texttt{math5}_{ij}$ score of i-th student at school j,
  β_0 is the mean math score across our population (being sampled).
  ε_{ij} is the individual deviation from overall mean.

  In our sample, $M = 48$ and n_j may differ among schools.

  Population mean math score & the variance of ε are estimated by taking their sample counterparts.

LME model example - continued

- The school effect (differences among schools) may be incorporated in the model by allowing the mean of each school to be represented by a separate parameter (*fixed effect*)

$$y_{ij} = \beta_{0j} \ + \ \varepsilon_{ij}, \quad i = 1, \ldots, n_j, \quad j = 1, \ldots, M, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where β_{0j} is the school-specific mean math score.
ε_{ij} is the individual deviation from the school-specific mean.

  - `R` syntax: `lm(math5` $\sim$ `School-1, data=...)`
  - $M = 48$ school-specific intercepts are estimated,
  - **Estimated intercepts only model the specific sample** of schools, while -usually- the main interest is in the population from which the sample was drawn.
  - Regression does not provide an estimate of the between-school variability, which is also of central interest.

LME model with random intercept

- *Random effects* model can solve the above problems by treating the school effects as random variations around a population mean.

- *Fixed effects* model can be reparametrized as:

  $y_{ij} = \beta_{0j} + \varepsilon_{ij}$

  $y_{ij} = \beta_0 + (\beta_{0j} - \beta_0) + \varepsilon_{ij},$

  now, the *random effect* u_{0j} can replace the the *fixed effect* β_{0j} :

  $\beta_{0j} = \beta_0 + u_{0j} \quad \Rightarrow \quad u_{0j} = \beta_{0j} - \beta_0$. Hence:

  $y_{ij} = \beta_0 + u_{0j} + \varepsilon_{ij}.$

- u_{0j} is the school-specific deviation from overall mean β_0.
  u_{0j} is a random variable, specific for the j-th school, with zero mean and unknown variance σ_u^2.
  u_{0j} is a *random effect*, because it is associated with the particular sample units (schools are selected at random from the population).

LME model with random intercept

- The *random effects* model is given as:

$$y_{ij} = \beta_0 + u_{0j} + \varepsilon_{ij}, \qquad u_{0j} \sim N(0, \sigma_u^2), \qquad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2),$$

  and we assume u_{0j} are *iid* and independent from ε_{ij}.

  - Observations in the same school share the same random effect u_{0j}, hence they are (positively) correlated with $corr = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2)$.
  - This *random effects* model has three parameters: β_0, σ_u^2 and σ_ε^2. (regardless of M, the number of schools).
  - Note that the *random effect* u_{0j} "looks like" a parameter, but we are interested in estimating σ_u^2.
  - However, upon observed data (and estimated model), we do make predictions for $\hat{u}_j$.

LME model: ICC

- **ICC:** Intra class correlation (in a LME regression model)

$$\text{ICC} = \frac{\text{Intercept variance}}{\text{Intercept variance} + \text{Residual variance}}$$

  - ICC: Proportion of variance in the outcome variable that occurs between “groups” (schools) to the total variability present.
  - Correlation between two “individuals” (students) randomly selected from the same “group” (school).

- Example 1

  $\texttt{math5}_{ij} = \beta_0 + \beta_1\,\texttt{math3}_{ij} + u_{0j} + \varepsilon_{ij}$,

  where $\sigma_u^2 = \text{var}(u_{0j})$ and $\sigma_\varepsilon^2 = \text{var}(\varepsilon_{ij})$.

  Here, ICC $= \frac{\sigma_u^2}{\sigma_u^2+\sigma_\varepsilon^2}$ measures correlation between `math5` observations (randomly chosen) within a given school.

LME model: ICC

- **ICC:** Intra class correlation (in a LME regression model)

$$\text{ICC} = \frac{\text{Intercept variance}}{\text{Intercept variance} + \text{Residual variance}}$$

- Example 2

  In a longitudinal study, y_{ij} measures the i-th response of the j-th individual.

  $y_{ij} = \beta_0 + \boldsymbol{x}_{ij}\boldsymbol{\beta} + u_{0j} + \varepsilon_{ij},$

  Here, ICC measures correlation between y_{ij} observations for a given individual.

- ICC is interpreted as the correlation between two **appropriately defined** observations from the same cluster/group (individual in a longitudinal study).

LME model with random intercept

- *Random effects* model with random intercept:

  Exogenous regressors are also used in LMEs (like in LRMs).
  For example, `math5` grades depend on `math3` (3^{rd} year grades).

  $y_{ij} = \beta_0 + \beta_1 \, x_{ij} + u_{0j} + \varepsilon_{ij}$,

  i.e.
  $\texttt{math5}_{ij} = \beta_0 + \beta_1 \, \texttt{math3}_{ij} + u_{0j} + \varepsilon_{ij}$,

  - Intercept is random.
  - Slope of the regression line for each school is fixed at β_1.
    ... `math3` has a *fixed effect*.

LME model with random intercept and slope

- *Random effects* model with random intercept and slope:

  If teaching is different from school to school, it would make sense to have different slopes for each of the schools.

  Instead of *fixed effects* (using interaction terms `math3:School`), we use random slopes: $u_{1j} = \beta_{1j} - \beta_1$.

  $y_{ij} = \beta_0 + u_{0j} + \beta_1 x_{ij} + u_{1j} x_{ij} + \varepsilon_{ij}$,

  i.e.

  $$\texttt{math5}_{ij} = \underbrace{\beta_0 + \beta_1 \texttt{math3}_{ij}}_{\textit{fixed}} + \underbrace{u_{0j} + u_{1j}\,\texttt{math3}_{ij}}_{\textit{random}} + \varepsilon_{ij},$$

  - We can test whether this extra complexity is justified.
  - u_{0j} and u_{1j} are often correlated, their independence can be tested.
  - Fitted values of $\texttt{math5}_{ij}$ can be produced, along with $\hat{u}_{0j}$ and $\hat{u}_{1j}$.

LME model in matrix form

- Linear models

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0},\, \sigma_{\varepsilon}^{2}\boldsymbol{I}),$$

- can be generalized into LME models

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon} \qquad \boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{G}) \qquad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{R}),$$

where (for balanced panels):
$\boldsymbol{X}$ is a $(n \times k)$ matrix, k is the number of *fixed effects*,
$\boldsymbol{Z}$ is a $(n \times p)$ matrix, p is the number of *random effects*,
$\boldsymbol{G}$ is a $(p \times p)$ variance-covariance matrix of the *random effects*,
$\boldsymbol{R}$ is a $(n \times n)$ variance-covariance matrix of errors.

Independence between $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ is assumed,
Often, $\boldsymbol{R} = \sigma_{\varepsilon}^{2}\boldsymbol{I}$, can be generalized for group-wise correlations,
$\boldsymbol{G}$ is diagonal if *random effects* are mutually independent.

More complex LME models - brief outline

Different types of LME models exist:

- LME models with (multilevel) nested effects,
- LME models with crossed effects,
- Complex behavior of the error term in LME models can be addressed.
- LME models with non-Gaussian (Binary, Poisson, etc.).

LME models with (multilevel) nested effects

Multi-level model example: we follow a total of 48 individual states within 9 regions and across 17 years.

- $\texttt{GDP}_{tij}$ represents individual GDP per capita measurements for:
  t-th time period, e.g. with values $(t = 1990, \ldots, 2006)$.
  i-th state nested within region j $(i = 1, \ldots, M_j)$,
  j-th region $(j = 1, \ldots, 9)$,
- We fit `GDP` as a function of productivity `P` and unemployment `U`.
  We treat states as nested within regions, so we have 2 levels of random intercepts: one due to the regions (v_{0j}), and another due to the state within region (random slopes can be added as well).
- $\texttt{GDP}_{tij} = \beta_0 + \beta_1 \, \texttt{P}_{tij} + \beta_2 \, \texttt{U}_{tij} + u_{0i(j)} + v_{0j} + \varepsilon_{tij}.$

LME models with crossed random effects

Crossed *random effects* example:

- Grunfeld (1958) analyzed data on 10 large U.S. corporations, collected annually from 1935 to 1954 to investigate how investment `I` depends on market value `M` and capital stock `C`.
- Here, we want *random effects* for a given firm and year.

  However, we want the year effect to be the same across all firms, i.e. not nested within firms.
- $\mathtt{I}_{it} = \beta_0 + \beta_1 \mathtt{M}_{it} + \beta_2 \mathtt{C}_{it} + u_{0i} + v_{0t} + \varepsilon_{it}$.

  where $i = 1, \dots, 10$ and
  firms are followed over $t = 1, \dots, 20$ years.
  (note the usual "it" index ordering is used here)

LME models in R

- `{lme4}` package
  `https://www.jstatsoft.org/article/view/v067i01/0`
- `{nlme}` package
  `https://cran.r-project.org/web/packages/nlme/nlme.pdf`
- Finch, Bolin, Kelley: Multilevel Modeling Using R (2014).