Praktikum z ekonometrie

VŠE Praha

Tomáš Formánek

Block 4 – Linear mixed effect models – Outline

Introduction

Linear mixed effect model (LME) – generalization of linear (panel) model

- LMEs & longitudinal data: repeated measurements are performed on each individual unit. Several units are sampled. Number of observations may differ across units (both longitudinal & hierarchical data).

  y_{ti} - observation at time t for i-th individual.
  y_{ij} - ith observation of jth individual (if time aspect secondary).

- LMEs & nesting (hierarchical) data structures: data with two or more groups/levels of observations.

  y_{ij} - observation for i-th company within j-th region.
  y_{ij} - observation for i-th student within j-th class.

  We can group observations at multiple levels:

  y_{tij} - measurement at time period t, in region i within state j.

- Note how indices are ordered (left to right) from individual to highest level of aggregation. (alternative orderings exist in literature).

Introduction

Linear mixed effect model (LME)

- Nested/hierarchical structure of the LME model:
  - Individual units i (Level 1) are nested
  - within j groups (Level 2) with group-specific observation sizes n_j.
- One or more β-coefficients can vary across groups.
- The same nesting/hierarchical framework applies to longitudinal data and their LME-based analysis:
  - Observations at time t (Level 1) are nested
  - within j individual units (Level 2).
  - If appropriate, individual units 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
- Terms "fixed" and "random" have specific meaning for LMEs:
  - A fixed coefficient is an unknown constant to be estimated.
  - A random coefficient varies from "group" to "group".
    By "group", we mean Level 2 aggregation, if data have 2 levels.
    - coefficients vary among schools (Level 2), not within school.
    - coeffs. vary across individuals (Level 2), not over time (Level 1).
- LME models can have some added complexity:
  - Multiple levels of nesting
  - Crossed random effects
  - Correlations between different random coefficients.
- Random coefficients are not estimated, but they can be predicted.

LME model example

- Data:
  London Education Authority Junior School Project dataset,
  - we have 887 students (i) in 48 different schools (j),
  - we want to predict 5th-year math scores.

- We may start by ignoring the school grouping and any possible regressors – we have a trivial model (*single-mean* model):

  $$\texttt{math5}_{ij} = \beta_0 + \varepsilon_{ij}, \quad i = 1, \ldots, n_j, \quad j = 1, \ldots, M, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

  where $M = 48$ and n_j differ among schools, $\texttt{math5}_{ij}$ is the observed math score of i-th student at school j, β_0 is the mean math score across our population (being sampled) and ε_{ij} is the individual deviation from overall mean.

  Population mean math score & the variance of ε are estimated by taking their sample counterparts. Any "school effect" is ignored.

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*)

$$\texttt{math5}_{ij} = \beta_{0j} + \varepsilon_{ij}, \quad i = 1, \ldots, n_j, \quad j = 1, \ldots, M, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

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

- R syntax: `lm(math5 ~ School-1, data=...)`
  $\Rightarrow$ $M = 48$ school-specific intercepts are estimated.
- Using the terminology of LME, β_{0j} are fixed. Hence:
  - **Estimated intercepts only model (refer to) 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:

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

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

  Random effect $u_{0j} = \beta_{0j} - \beta_0$ is the school-specific deviation from overall mean β_0. It can be used to replace the the *fixed effect* β_{0j} :

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

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

- 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*, 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 within the same school share the same random effect u_{0j}, hence are positively "correlated" with ICC $= \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2)$ (see ICC on next slide).
  - 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 coefficient, but we are only interested in estimating σ_u^2.
  - However, upon observed data (and estimated model), we do make predictions using fitted values of $\hat{u}_j$.

LME model: ICC

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

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$$

  - Describes how strongly units in the same group are "correlated".
  - While interpreted as a type of correlation, ICC operates on groups, rather than paired observations.
  - See https://en.wikipedia.org/wiki/Intraclass_correlation for formal definition & relation between ICC and actual correlation.

- Example

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

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

  Here, ICC measures "correlation" between `math5` observations (randomly chosen) within a given school.

LME model with random intercept and fixed slope

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

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

  alternatively:

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

  - Intercept is random, given the u_{0j} element.
  - 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

- 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: $u_{1j} = \beta_{1j} - \beta_1$.

  $$\texttt{math5}_{ij} = (\beta_0 + u_{0j}) + (\beta_1\,\texttt{math3}_{ij} + u_{1j}\,\texttt{math3}_{ij}) + \varepsilon_{ij},$$

  alternatively:

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

  - 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, $\boldsymbol{R} = \sigma_\varepsilon^2 \boldsymbol{I}_n$ is assumed 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 dependent variables (binary, Poisson, etc.).

LME models with multilevel nested effects

Multi-level model example: For 17 years, we follow a total of 86 individual states organized within 9 "global-level" regions (e.g. South America, Europe, Middle East, etc.).

- $\texttt{GDP}_{tij}$ represents individual GDP per capita measurements for:
  t-th time period, e.g. with values ($t = 2000, \dots, 2016$).
  i-th state nested within region j ($i = 1, \dots, M_j$),
  j-th region ($j = 1, \dots, 9$),
- We fit $\texttt{GDP}$ as a function of productivity $\texttt{P}$ and unemployment $\texttt{U}$.
  States are nested in regions, we have 2 levels of random intercepts:
  $u_{0i(j)}$ for each state (within a region),
  v_{0j} for the regions,
  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.

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

  where $i = 1, \dots, 10$ and
  firms are followed over $t = 1, \dots, 20$ years.
  (the usual "*it*" index ordering can be used as well)

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`
- `https://www.r-bloggers.com/2017/12/linear-mixed-effect-models-in-r/`
- Finch, Bolin, Kelley: Multilevel Modeling Using R (2014).