Praktikum z ekonometrie - Týden 7
Mixed effect models

VŠE Praha

Tomáš Formánek

Content

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 $1 \leq i \leq n$,
  ε_i are *iid* normally distributed and uncorrelated with regressors.
  β-coefficients are considered fixed unknowns that must be estimated together with σ^2 (the variance of ε).

- In a mixed model, there are multiple groups (or panels or individuals) of observations.

  One or more β-coefficient can vary across groups (panels, individuals).

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 sample of groups to sample of groups.
- LME models can have some added complexity:
    - Correlations between different random coefficients.
    - Multiple levels of nesting withing the groups
- Random coefficients are not estimated, but they can be predicted.

LME model example

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

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

  $$y_{ij} = \beta_0 + \varepsilon_{ij}, \qquad i = 1, \ldots, M, \qquad j = 1, \ldots, n_i, \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

  where y_{ij} is the observed $\texttt{math5}_{ij}$ score of j-th student at school i,
  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_i 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*)

$$y_{ij} = \beta_{0i} + \varepsilon_{ij}, \qquad i = 1, \dots, M, \qquad j = 1, \dots, n_i, \qquad \varepsilon \sim N(0, \sigma^2),$$

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

  - `R` syntax: `lm(math5 ~ 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 example - continued

- *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:

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

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

- Now, the the *random effect* u_i can replace the the *fixed effect* β_{0i}:

  $\beta_{0i} = \beta_0 + u_i \quad \Rightarrow \quad u_i = \beta_{0i} - \beta_0.$

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

LME model example - continued

- The *random effects* model is given as:

$$y_{ij} = \beta_0 + u_i + \varepsilon_{ij}, \qquad u_i \sim N(0, \sigma_u^2), \qquad \varepsilon_{ij} \sim N(0, \sigma^2),$$

  and we assume u_i are *iid* and independent from ε_{ij}.

  - Observations in the same school share the same random effect u_i, hence they are (positively) correlated with $corr = \sigma_u^2 / (\sigma_u^2 + \sigma^2)$.
  - This *random effects* model has three parameters: β_0, σ_u^2 and σ^2. (regardless of M, the number of schools).
  - Note that the *random effect* u_i "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 $\hat{u}_i$.

LME model example - continued

- *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).

  $y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \varepsilon_{ij}$,

  i.e.

  $\texttt{math5}_{ij} = \beta_0 + \beta_1 \texttt{math3}_{ij} + u_i + \varepsilon_{ij}$,

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

LME model example - continued

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

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

  i.e.

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

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

LME model formal definition

- Linear models

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

- can be generalized into LME models

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

  where
  $\boldsymbol{X}$ 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, $\boldsymbol{R} = \sigma^2\boldsymbol{I}$, 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:

- $\texttt{GDP}_{ijt}$ represents individual GDP per capita measurements for:
  i-th regions ($i = 1, \ldots, 9$),
  j-th state nested within region i ($j = 1, \ldots, M_i$),
  t-th yearly measurement at state level ($t = 1990, \ldots, 2006$).
  …say, we follow a total of 48 individual states across 17 years.

- 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, and another due to the state within region (random slopes can be added as well).

- $\texttt{GDP}_{ijt} = \beta_0 + \beta_1 \texttt{P}_{ijt} + \beta_2 \texttt{U}_{ijt} + u_i + v_{j(i)} + \varepsilon_{ijt}.$

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.
- $\mathtt{I}_{it} = \beta_0 + \beta_1 \mathtt{M}_{it} + \beta_2 \mathtt{C}_{it} + u_i + v_t + \varepsilon_{it}$.

  where $i = 1, \ldots, 10$ firms are followed over $t = 1, \ldots, 20$ years.