MATH3091 Statistical Modelling II
Lecture 6: Linear Mixed Models (LMMs)

Chao Zheng

18 Feb 2022

Recap

In last lecture, we have

- revised linear models with different structures
- introduced the likelihood based inference for linear models (MLE estimation, distribution, model comparison)

In next a few lectures we are going to look at an extension of linear models.

4.1.1 Introduction to LMMs

Restrictions of linear model

In linear models, we generally assume that the error term $\{\epsilon_1, \ldots, \epsilon_n\}$ are i.i.d $\sim N(0, \sigma^2)$, which leads to independent response random variables $\{Y_1, Y_2, ..., Y_n\}$.

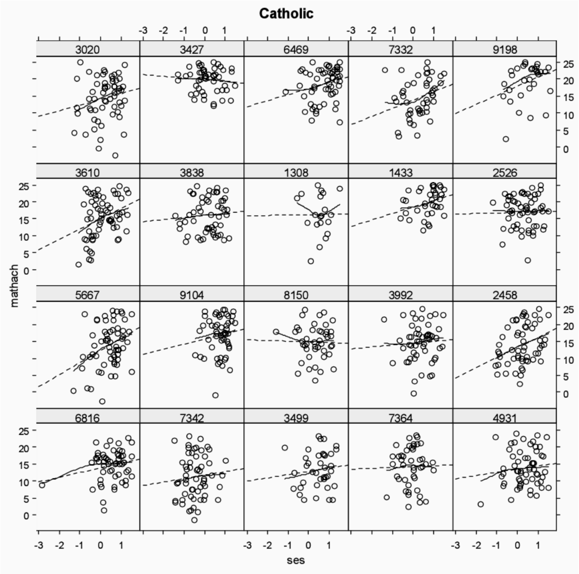
However, in many practical areas such as social, behavioral, and health sciences, correlated/dependent data frequently arise.

Motivating Examample: `mathachieve` dataset

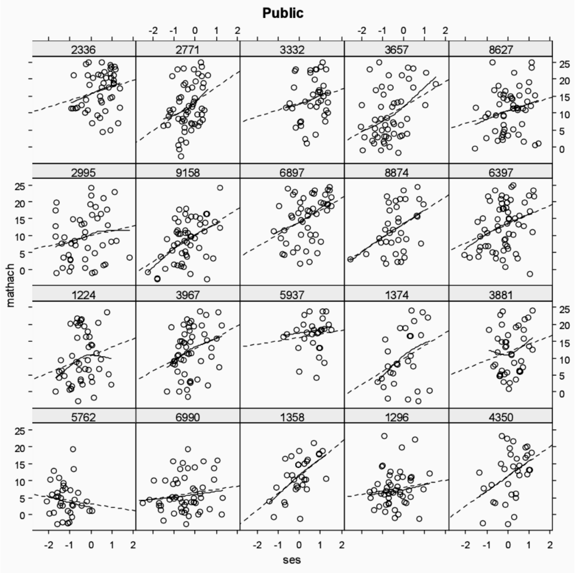
In this dataset, we have scores on a math-achievement test from 7185 high-school students, which are from 160 schools. We want to model the performance of students math-achievement test and the socioeconomic status (SES) of the student's family, and the type of students' school.

It is okay to assume the performance of students across different schools are independent.

However, it is unreasonable to assume that the performance of students in the same school are independent of one-another, since students took the same classes from the same teachers.



Math achievement by socio-economic status for 20 randomly selected Catholic schools. The broken lines give linear least-squares fits.



Math achievement by socio-economic status for 20 randomly selected public schools.

A better model

We will fit a hierarchical linear model to the `math-achievement` data. This model consists of two parts:

(1) within schools, we have the regression of math achievement on the individual-level covariate SES;

(2) across schools, we have for each school estimates the average level of SES, and the type of school

A flexible way to model this type of correlated data is to assume there is a grouped structure.

Cluster data

In clustered data, we have each subject is measured at each data point, and each data point belongs to a cluster (group).

For example: in studies of health services and outcomes, assessments of quality of care are obtained from patients who are grouped by within different clinics.

Longitudinal data

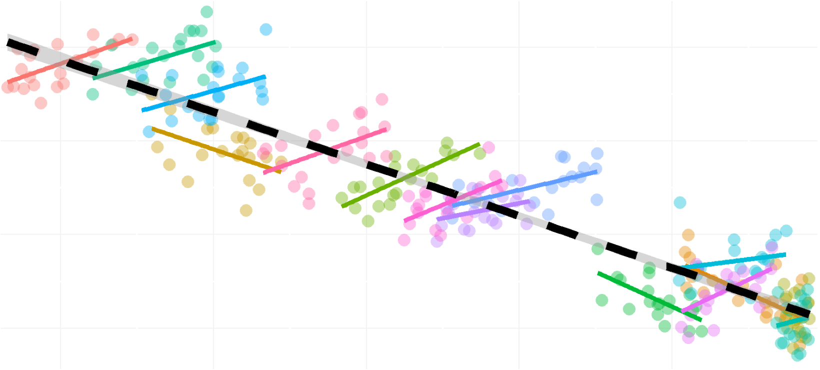
In longitudinal data, we have each subject is measured at different occasions, and each subject (measured at multiple times) forms a group.

For example, suppose the unemployment rate remained high for a long period of time during Covid. One can see if the same collection of individuals stay unemployed over the entire period or if they move in and out of unemployment over the time period.

Within group and between group

One importance consequence of grouping is that observations within a group are more similar (correlated) than observations in different groups (independent).

The group structure can be expressed in terms of dependence among the observations within the same group. Such data can also be regarded as hierarchical/multilevel data.



Fixed and random effects

How can we extend the linear model to allow for such grouped dependent structures ?

To incorporate the correlation within groups and the variation between groups, we can extend linear model by introducing random effects in the model and thus obtain Linear Mixed Models (LMMs).

Fixed and random effects

- fixed effect = population-level parameters (β in linear model) that are associated with quantitative variables
- random effect = grouped-level "parameters", whose values are randomly sampled from a population of values being studied

Fixed effects (parameters) are constant across observations, whilst random effects vary (between groups).

To include dependence, we must make, random effect, these grouped-level parameters to be random (not constants).

Fixed effects (parameters) don't change if you re-run an experiment, random effects do.

4.1.2 Basics of Linear Mixed Models

Grouped observations

First, we denote the observations of response variable **within i-th group** as

$$\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i}),$$

where n_i is the number of observations in i-th group. These are observations of (non-independent) random variables

$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i}).$$

Here we let $i = 1, 2 \ldots, m$, where m is the number of groups, and we let $n = \sum_{i=1}^{m} n_i$.

Linear mixed models

A general LMM is as follows:

$$
\begin{aligned}
Y_{ij} =& x_{ij,1}\beta_1 + x_{ij,2}\beta_2 + \cdots + x_{ij,p}\beta_p \\
& + u_{ij,1}\gamma_{i,1} + \cdots + u_{ij,q}\gamma_{i,q} \\
& + \epsilon_{ij} \\
=& \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{u}_{ij}^T\boldsymbol{\gamma}_i + \epsilon_{ij}, \quad j = 1, \dots, n_i; \quad i = 1, \dots, m,
\end{aligned}
$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is parameter vector of fixed effects, and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{iq})^T$ is the vector of random effects in i-th group.

Moreover let $\boldsymbol{\epsilon_i} = (\epsilon_{i1}, , \dots, \epsilon_{in_i})^T$ represents random errors within i-th group, where $\epsilon_{ij} \sim N(0, \sigma^2)$ independently.

We assume independence between each groups, that means $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_m$, $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots, \boldsymbol{\epsilon}_m$ are all independent.

Interpretation - `mathachieve` dataset

For example, in `mathachieve` dataset, we want to analysis the relationship between the response variable:

- `mathach`: the student's score on a math-achievement test

and 3 explanatory variables:

- `cses`, the adjusted socioeconomic status of the student's family;
- `meanses`, the average socioeconomic status for students in each school;
- `sector`, a factor coded Catholic or Public for the type of student's school.

Interpretation - `mathachieve` dataset

The individual-level equation for individual j in school i is

$$\texttt{mathach}_{ij} = \beta_{i0} + \beta_{i1}\texttt{cses}_{ij} + \epsilon_{ij}$$

At the school level, we will entertain the possibility that the school intercepts depend upon sector and upon the average level of SES in the schools:

$$\beta_{i0} = \beta_0 + \beta_1\texttt{meanses}_i + \beta_2\texttt{sector}_i + \gamma_{i0}$$

$$\beta_{i1} = \beta_3 + \gamma_{i1}$$

Rearranging terms we have a LMMs:

$$\begin{aligned}\texttt{mathach}_{ij} =& \beta_0 + \beta_1\texttt{meanses}_i + \beta_2\texttt{sector}_i + \beta_3\texttt{cses}_{ij} \\ &+ \gamma_{i0} + \gamma_{i1}\texttt{cses}_{ij} + \epsilon_{ij}\end{aligned}$$

Random coefficients

As we said $\boldsymbol{\gamma}_i$ is a vector of random coefficients. During this course, unless otherwise specified we assume that

$$\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \boldsymbol{D}),$$

where $\boldsymbol{D}$ is a $q \times q$ covariance matrix of the random effects.

We often assume $\boldsymbol{D} = \boldsymbol{D}_{\boldsymbol{\theta}}$ is parameterized by some unknown parameter $\boldsymbol{\theta}$.

By introducing $\boldsymbol{\gamma}_i, i = 1, \ldots, m$, we include the with-in group correlation into the linear model.

With-in group correlations

Consider the following simple LMMs:

$$Y_{ij} = \beta_0 + \gamma_i + \epsilon_{ij}, \quad i = 1, \dots, m; \quad j = 1, \dots, n_i,$$

where the random effect and the error satisfy:

$$\gamma_i \sim N(0, \sigma_\gamma^2), \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2).$$

It can be shown that within $\{Y_{i1}, Y_{i2}, \dots, Y_{in_i}\}$ is:

$$\begin{aligned} r_{jk}^i = \mathrm{corr}(Y_{ij}, Y_{ik}) =& \frac{\mathrm{Cov}(\gamma_i + \epsilon_{ij}, \gamma_i + \epsilon_{ik})}{\sqrt{\mathrm{Var}(\gamma_i + \epsilon_{ij})\mathrm{Var}(\gamma_i + \epsilon_{ik})}} \\ =& \frac{\sigma_\gamma^2}{\sigma_\gamma^2 + \sigma_\epsilon^2}. \end{aligned}$$

Thus, if $\sigma_\gamma = 0$, there is no with-in group correlation. If $\sigma_\epsilon^2 \ll \sigma_\gamma^2$, the correlation r_{jk}^i becomes very high.

Two sources of randomness

Specifically, by introducing the random effect, we make the data incorporates two sources of randomness: the within-group randomness and the between-group randomness. Thus, it can be interpreted as two-stage hierarchical.

- stage 1: specifies the within-group randomness, which is given by fixing i and letting $j = 1, \ldots, n_i$;
- stage 2: specifies the between-group randomness, which is given by letting $i = 1, \ldots, m$.

LMM Matrix form I

To present the LMMs in a matrix form, let $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{in_i})^T$ be a $n_i \times p$ matrix, and $\boldsymbol{U}_i = (\boldsymbol{u}_{i1}, \boldsymbol{u}_{i2}, \ldots, \boldsymbol{u}_{in_i})^T$ be a $n_i \times q$ matrix. Therefore, we can write

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{X}_i \beta + \boldsymbol{U}_i \gamma_i + \boldsymbol{\epsilon}_i \\ &= \boldsymbol{X}_i \beta + \boldsymbol{\epsilon}_i^*, \qquad\qquad i = 1 \ldots, m, \end{aligned}$$

where $\boldsymbol{\epsilon}_i^* = \boldsymbol{U}_i \gamma_i + \boldsymbol{\epsilon}_i$ can be regarded as random "errors" in i-th group. Note that $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_{n_i})$. It is straightforward to see that

$$\boldsymbol{\epsilon}_i^* \sim N(\mathbf{0}, \boldsymbol{U}_i \boldsymbol{D} \boldsymbol{U}_i^T + \sigma^2 \boldsymbol{I}_{n_i}),$$

which leads to the marginal model:

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i \beta,\ \boldsymbol{U}_i \boldsymbol{D} \boldsymbol{U}_i^T + \sigma^2 \boldsymbol{I}_{n_i}).$$

LMM Matrix form II

If we want to further present all the groups $i = 1, 2, \ldots, m$ in one matrix formula, we can write

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y_1} \\ \vdots \\ \boldsymbol{Y_m} \end{pmatrix} \in \mathbb{R}^{n}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X_1} \\ \vdots \\ \boldsymbol{X_m} \end{pmatrix} \in \mathbb{R}^{n \times p},$$

and

$$\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\gamma_1} \\ \vdots \\ \boldsymbol{\gamma_m} \end{pmatrix} \in \mathbb{R}^{mq}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{\epsilon_m} \end{pmatrix} \in \mathbb{R}^{n \times 1}.$$

Moreover, let

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U_1}, & \mathbf{0}_{n_1 \times q}, & \cdots, & \mathbf{0}_{n_1 \times q} \\ \mathbf{0}_{n_2 \times q}, & \boldsymbol{U_2}, & \cdots, & \mathbf{0}_{n_2 \times q} \\ \vdots & & \ddots & \\ \mathbf{0}_{n_m \times q}, & \mathbf{0}_{n_m \times q} & \cdots & \boldsymbol{U_m} \end{pmatrix} \in \mathbb{R}^{n \times mq}$$

Matrix form II

Therefore, we can write the LMMs for all group as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\epsilon}.$$

Since $\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \boldsymbol{D})$, we have that $\boldsymbol{\gamma} \sim N(\mathbf{0}, \mathcal{G})$, where

$$\mathcal{G} = \begin{pmatrix} \boldsymbol{D} & & & \\ & \boldsymbol{D} & & \\ & & \ddots & \\ & & & \boldsymbol{D} \end{pmatrix} \in \mathbb{R}^{mq \times mq}.$$

Similarly, if we express $\boldsymbol{\epsilon}^* = \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, we have

$$\boldsymbol{\epsilon}^* \sim N(\mathbf{0}, \boldsymbol{V}), \qquad \boldsymbol{V} = \boldsymbol{U}\mathcal{G}\boldsymbol{U}^T + \sigma^2 \boldsymbol{I}_{n,}.$$

As $\boldsymbol{D} = \boldsymbol{D}_{\boldsymbol{\theta}}$, we can write $\mathcal{G}$ and $\boldsymbol{V}$ as $\mathcal{G}_{\boldsymbol{\theta}}$ and $\boldsymbol{V}_{\boldsymbol{\theta}}$, respectively.

Conclusion

- We have introduced the fixed effects and random effects
- We have introduced the linear mixed models (LMMs) for modelling data with dependence
- We have seen LMMs in different matrix forms.