CDSI Workshop
Introduction to linear mixed models for continuous data

Mila Sun

Department of Epidemiology, Biostatistics and Occupational Health
McGill University

Materials: https://github.com/mila-sun/LMM-workshop

Winter, 2022

COMPUTATIONAL AND
DATA SYSTEMS INITIATIVE

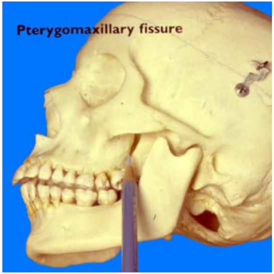
McGill Faculty of Science

Outline

- Motivation example
- Theory/Background
  - Random intercept and random coefficient models
  - Model assumptions
  - Expectation, variance and covariance
- Example: Test scores
  - Data exploration
  - Modelling
  - Assumption checking
  - Interpreting results and visualizing the model
- Extension and remarks
  - Shrinkage in linear mixed model
  - Fixed effects vs random effects
  - How do we choose models?
  - Other package in R

Motivation: dental growth

- Dental growth measurements of the distance (in mm) from the pterygomaxillary fissure to the center of the pituitary gland is obtained from x-rays.



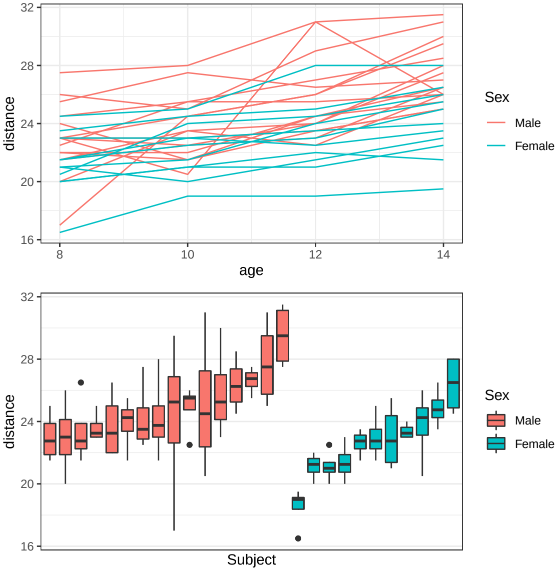
Motivation: dental growth

- 11 girls and 16 boys at the ages of 8, 10, 12 and 14 years.
- Here we have an example of repeated measures or longitudinal data.

Table 1.3 Dental growth data for boys and girls

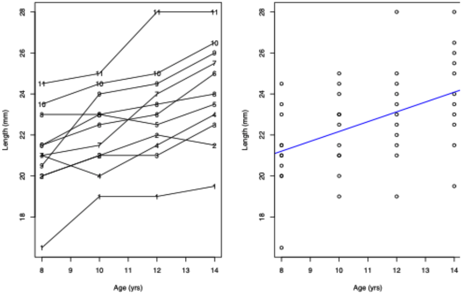
	Age (years)					Age (years)			
Girl	8	10	12	14	Boy	8	10	12	14
1	21.0	20.0	21.5	23.0	1	26.0	25.0	29.0	31.0
2	21.0	21.5	24.0	25.5	2	21.5	22.5	23.0	26.5
3	20.5	24.0	24.5	26.0	3	23.0	22.5	24.0	27.5
4	23.5	24.5	25.0	26.5	4	25.5	27.5	26.5	27.0
5	21.5	23.0	22.5	23.5	5	20.0	23.5	22.5	26.0
6	20.0	21.0	21.0	22.5	6	24.5	25.5	27.0	28.5
7	21.5	22.5	23.0	25.0	7	22.0	22.0	24.5	26.5
8	23.0	23.0	23.5	24.0	8	24.0	21.5	24.5	25.5
9	20.0	21.0	22.0	21.5	9	23.0	20.5	31.0	26.0
10	16.5	19.0	19.0	19.5	10	27.5	28.0	31.0	31.5
11	24.5	25.0	28.0	28.0	11	23.0	23.0	23.5	25.0
					12	21.5	23.5	24.0	28.0
					13	17.0	24.5	26.0	29.5
					14	22.5	25.5	25.5	26.0
					15	23.0	24.5	26.0	30.0
					16	22.0	21.5	23.5	25.0

Motivation: dental growth



Motivation: dental growth

- We distinguish two importantly different types of question:
  - What is the mean difference in length with 1-year difference in age? [marginal]
  - For a specific child, what is the growth curve? [conditional]
- The former is a question about whole populations; the latter is about particular subjects (or types of subjects).



Motivation: dental growth

- Are our data independent? We collected multiple samples from 27 children.
- Even for marginal inferences, it will still be invalid to treat the data as $27 \times 4 = 108$ independent outcomes; we actually have 27 sets of 4 correlated outcomes.
- For either question of interest, ignoring the dependence leads to incorrect standard errors and confidence interval coverage.

Motivation: dental growth

- Are our data independent? We collected multiple samples from 27 children.
- Even for marginal inferences, it will still be invalid to treat the data as $27 \times 4 = 108$ independent outcomes; we actually have 27 sets of 4 correlated outcomes.
- For either question of interest, ignoring the dependence leads to incorrect standard errors and confidence interval coverage.

$\Longrightarrow$ Add a **random effect** for subject. This allows us to resolve this non-independence by assuming a different "baseline" distance value for each subject.

Mixed models

Mixed models (also known as random effects models, hierarchical models, multilevel models, varying coefficient models). Why is the mixed model called a "mixed" model?

The simple linear model

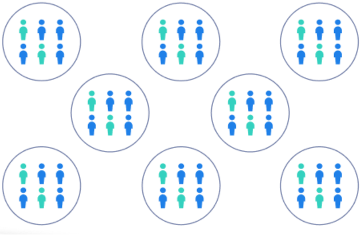
$$y = \beta_0 + \beta_1 x + \varepsilon$$

It is a "fixed-effects-only" model that had one or more fixed effects and a general error term ε which did not have any interesting structure.

The linear mixed model
Basic idea: assume each unit has a regression model characterized by both fixed effects, that are common to all units in the population, and random effects that are unit-specific. "Mixed" effects refer to the combination of both fixed and random effects.

Notations

- i represents the cluster (cluster=subject for longitudinal data)
- j represents the observation within cluster (j = time for longitudinal data)
- Y_{ij} is the outcome for cluster i's observation j
- $Y_i = (Y_{i1}, \ldots, Y_{in_i})$ is the vector of responses for cluster i
- Assume that responses on different clusters are independent, but there is dependence between observations in the same cluster.



Linear mixed models (LMM)

- Consider the random intercept regression model:

$$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + \varepsilon_{ij}$$

- In this model, the intercept is specific for each cluster and can be written as:

$$\beta_{0i} = \beta_0 + u_i$$

- β_0 is the "part" of the intercept that is common to all clusters
- u_i is the "part" of the intercept that is specific to each cluster
- u_i is assumed to be random
- ε_{ij} is the random error
- The traditional linear regression model can be seen as a special case of the mixed model where $u_i = 0$ for all i

Model assumptions

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

- Random errors ε_{ij} are independent of each other:

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

- u_i is the random effect (random intercept) for cluster i, and all u_i are independent of each other:

$$u_i \sim \mathcal{N}(0, \sigma_u^2)$$

- ε_{ij} and u_i are independent of each other

Expectation and variance of the outcome

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

- Expectation of the outcome is the same as in a traditional regression model:

$$\mathrm{E}[Y_{ij}|X_{ij}] = \beta_0 + \beta_1 X_{ij}$$

- Variance of the outcome:

$$\begin{aligned} \mathrm{Var}[Y_{ij}|X_{ij}] &= \mathrm{Var}[u_i] + \mathrm{Var}[\varepsilon_{ij}] \\ &= \sigma_u^2 + \sigma_\varepsilon^2 \end{aligned}$$

- The variability of the observations is due to the variability between clusters and to sources of variations that are not explained by the model (measurement errors).

Why does it model correlated data?

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

- Covariance between two observations from the same cluster:

$$\begin{aligned}
Cov(Y_{ij}, Y_{ik}) &= Cov(u_i + \varepsilon_{ij}, u_i + \varepsilon_{ik}) \\
&= Cov(u_i, u_i) + Cov(u_i, \varepsilon_{ik}) + Cov(\varepsilon_{ij}, u_i) + Cov(\varepsilon_{ij}, \varepsilon_{ik}) \\
&= Var(u_i) + 0 + 0 + 0 \\
&= \sigma_u^2
\end{aligned}$$

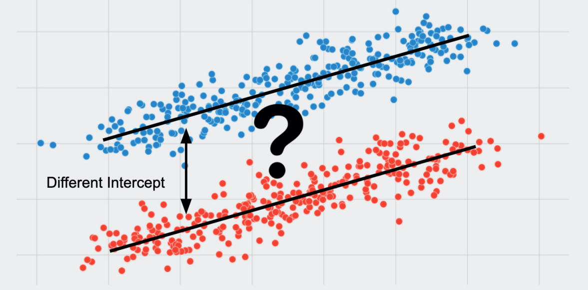
- The correlation between two observations from the same cluster is not zero!

Why does it model correlated data? - comments

- Under the random intercept model, $Cov(Y_{ij}, Y_{ik}) = \sigma_u^2$, so this model *does* describe the correlation between observations in the same cluster
- Note: under a traditional linear regression model, we have $u_i = 0$, which implies that observations within clusters are independent
- We can also verify that $Cov(Y_{ij}, Y_{i'k}) = 0$ when $i \neq i'$ (i.e., covariance between clusters is zero)
- This particular "random intercept" model has correlation that is constant for all pairs of observations within clusters.

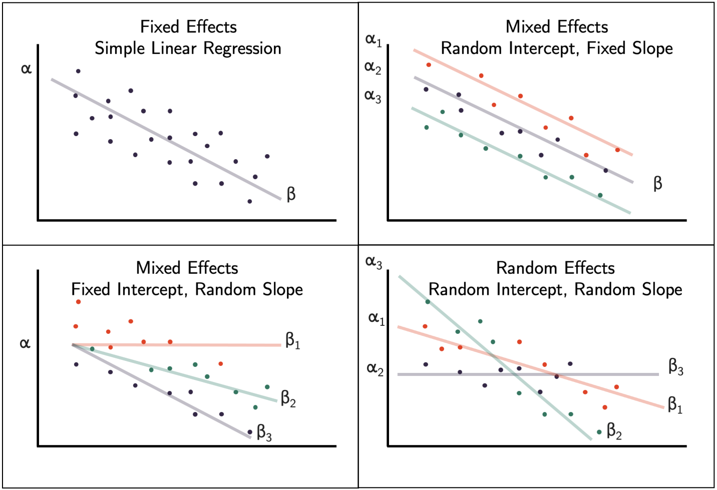
Random intercept versus random coefficient models

- The model $y_{ij} = \beta_0 + u_i + \beta_1 x_{ij} + \varepsilon_{ij}$ is called the "random intercept" model (also called varying intercept model)



- What if we also include a cluster-specific slope $\beta_{1i} = \beta_1 + u_{1i}$? That is, what if we allow additional coefficients to vary?

Random intercept versus random coefficient models



Random coefficient model

- The model allows the slope β_1 to vary across clusters:

$$\begin{aligned} Y_{ij} &= \beta_{0i} + \beta_{1i} X_{ij} + \varepsilon_{ij} \\ &= (\beta_0 + u_{0i}) + (\beta_1 + u_{1i}) X_{ij} + \varepsilon_{ij} \end{aligned}$$

- Assumptions of the model are:
  - ε_{ij}'s are all independent $\mathcal{N}(0, \sigma_\varepsilon^2)$
  - u_{0i}'s are all independent $\mathcal{N}(0, \sigma_{u_0}^2)$
  - u_{1i}'s are all independent $\mathcal{N}(0, \sigma_{u_1}^2)$
  - the ε's are independent from the u_{1i}'s and from the u_{0i}'s
  - u_{1i} and u_{0i} can be set to be independent from each other, or allowed to be correlated

Expectation, variance and covariance

$$Y_{ij} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i})X_{ij} + \varepsilon_{ij}$$

- **Expectation**:

$$\mathrm{E}[Y_{ij}|X_{ij}] = \beta_0 + \beta_1 X_{ij}$$

- **Variance** (in the case where u_{0i} and u_{1i} are independent - more complicated otherwise):

$$\mathrm{Var}[Y_{ij}|X_{ij}] = \sigma^2_{u_0} + \sigma^2_{u_1} X^2_{ij} + \sigma^2_{\varepsilon}$$

- **Covariance within clusters** (in the case where u_{0i} and u_{1i} are independent - more complicated otherwise):

$$Cov(Y_{ij}, Y_{ik}) = \sigma^2_{u_0} + X_{ij} X_{ik} \sigma^2_{u_1}$$

- **Covariance between clusters**:

$$Cov(Y_{ij}, Y_{i'k}) = 0 \text{ if } i \neq i'$$

Parameters of the model

$$Y_{ij} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i})X_{ij} + \varepsilon_{ij}$$

- The coefficients β_0 and β_1 are called fixed coefficients (fixed effects): they are common across all clusters
- The coefficients u_{0i}, u_{1i} are called random coefficients (random effects): they vary across the clusters
- The parameters of the model (to be estimated) are:
  - All fixed coefficients: β_0, β_1;
  - The variance of all random effects and errors: $\sigma^2_{u_0}$, $\sigma^2_{u_1}$, σ^2_{ε};
  - The covariance of the random effects, if we assume/allow them: $Cov(u_{i0}, u_{i1})$.

Example: Test scores

Download it here: `https://github.com/mila-sun/LMM-workshop/blob/main/LMM_workshop.pdf`

Shrinkage in linear mixed model

- The estimate of the effect by school is smaller than when we fit a separate linear model to the school's data.
- This is called **shrinkage** in linear mixed models: the school level estimates are shrunk towards the mean slope.
- The less data we have from a given school, the more the shrinkage.

Shrinkage in linear mixed model

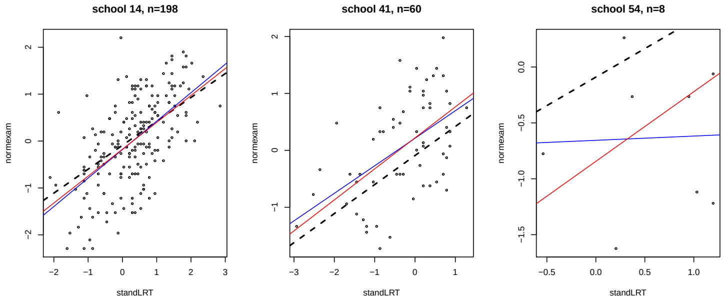


Figure: Simple linear model on all the data (black dashed line), the separate linear model to each school data (blue line), and the linear mixed model (the red line). In all three schools' models, the linear mixed model estimates are shrunk towards the grand mean (black dashed line) estimates.

Fixed effects vs random effects ?

Can include the clustering unit as a fixed effect or a random effect

- As a fixed effect:
  - The clustering unit is included as a dummy variables in a standard regression model
  - Effects of other independent variables are assumed to be the same across the clustering units
- As a random effect:
  - The clustering unit is NOT included as a dummy variable
  - We use a mixed model
  - Effects of other independent variables may vary across the clustering unit.

Fixed effects vs random effects ?

How to decide ??

- **Fixed effects**:
  - All possible values are represented in the study
  - No distribution is estimated
  - If the number of clustering units is small
- **Random effects**:
  - Used when data are clustered (for example, repeated observation of participants or students within schools)
  - When we have a sufficient number of clusters to enable estimation of the random effect (variance)
  - When we are not interested in the "effects" of the clusters themselves.
  - A distribution is estimated
- Can sometimes be tricky to decide which is which

How do we choose models?

- How do we choose the fixed coefficients: same as usual (science, confounding, effect modification, significance (be careful!))
- Which coefficients should we allow to vary?
  - AIC and/or BIC
  - LRT test
  - Conceptual issues - when the structure is important, may want to keep random effects on scientific grounds

Fit a random intercept model: R with nlme

```
# Fit a mixed model with random intercept
> library(nlme)
> m2b<-lme(normexam~standLRT,random=~1|school, data=Exam)
> summary(m2b)
```

Random intercept model: R output with `nlme`

```
> library(nlme)
> m2b<-lme(normexam~standLRT,random=~1|school, data=Exam)
>
> summary(m2b)
Linear mixed-effects model fit by REML
 Data: exam
       AIC      BIC    logLik
  9376.765 9401.998 -4684.383

Random effects:
 Formula: ~1 | school
        (Intercept)  Residual
StdDev:   0.3063315 0.7522402

Fixed effects: normexam ~ standLRT
                Value  Std.Error   DF  t-value p-value
(Intercept) 0.0023228 0.04035436 3993  0.05756  0.9541
standLRT    0.5633069 0.01246796 3993 45.18035  0.0000
 Correlation:
         (Intr)
standLRT 0.008

Standardized Within-Group Residuals:
        Min          Q1          Med          Q3         Max
-3.71658031 -0.63015069  0.02939999  0.68490540  3.26725600

Number of Observations: 4059
Number of Groups: 65
```

- **Random intercept variance**: $\hat{\sigma}_u^2 = 0.306^2 = 0.093$
- **Error variance**: $\hat{\sigma}_\varepsilon^2 = 0.752^2 = 0.565$
- **Fixed effects**: $\hat{\beta}_0 = 0.002$ $(se = 0.040)$, $\hat{\beta}_1 = 0.563$ $(se = 0.012)$
- $Corr(\hat{\beta}_0, \hat{\beta}_1) = 0.008$

Fit a random coefficient model: R with nlme

```
# Fit a mixed model with random intercept
> library(nlme)
> m2b<-lme(normexam~standLRT,random=~standLRT|school, data=Exam)
> summary(m2b)
```