BIOS 755: Multilevel Models

Alexander McLain

March 30, 2023

- We know that mixed models can be used to analyze longitudinal data.
- Mixed models can also be used to analyze multilevel data.
- Longitudinal data are clustered. Individual subjects, or units, provide multiple observations.
- Multilevel data are also clustered.
  - Randomized trials study patients clustered within practices.
  - In studies of social determinants of health, we may study families of individuals clustered by neighborhood within a community.

Multilevel Data

- Multilevel data can arise from the study design or a natural hierarchy in the target population, or sometimes both.
- In studies of outcomes within families, we measure the outcome status of each family member.
- Interventions will be delivered in particular classrooms or schools.
- In disease mapping studies, we may have the disease status of people and their census tract.
- Other naturally occurring clusters include hospital wards, medical practices, doctors and neighborhoods.

Multilevel Linear Models

- The dominant approach to analysis of multilevel data employs a type of linear mixed effects model known as the hierarchical linear model.
- The correlation induced by clustering is described by random effects at each level of the hierarchy.
- Time may not be an issue; the data can be on separate individuals at one time point.
- Space may be an issue.

Multilevel Linear Models

In a multilevel model, the response is obtained at the first level, but covariates can be measured at any level.

- For example, if we are studying BMI, we can measure individual diets, family attitudes about food and purchasing habits, and community attributes such as the density of fast-food restaurants.
- If we are studying the prevalence of ADHD, the outcome is at the child level and the covariate data may consist of information about the child, the child's parent(s) or guardian(s), the county the child lives in, and the state.

Two-level Linear Models

- Let i index level 1 units and j index level 2 units.
- We assume n_2 level 2 units in the sample. Each of these clusters ($j = 1, 2, \ldots, n_2$) is composed of n_{1j} level 1 units. (In a two-level study of physician practices, we would study n_2 practices, with n_{1j} patients in the jth practice.)
- Let Y_{ij} denote the response for patient i in practice j (this is a little backwards from before).
- Associated with each Y_{ij} is a $1 \times p$ (row) vector of covariates, $\boldsymbol{X}_{ij}$
- Consider the following model for the mean: $E(Y_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}$

Two-level Linear Models

- For example, in a multi-center clinical trial comparing two treatments, we might assume that:

$$E(Y_{ij}) = \beta_0 + \beta_1 Trt_{ij}$$

  where Trt_{ij} is an indicator variable for treatment group.
- The two-level hierarchical linear model assumes that the correlation within practices can be described by a random effect.
- Given that random effect, the outcomes are assumed to be independent.
- Thus, we may assume that

$$Y_{ij} = \beta_0 + \beta_1 Trt_{ij} + b_{0j} + e_{ij}$$

  where $b_{0j} \sim N(0, g)$ and $e_{ij} \sim N(0, \sigma^2)$

Two-level Linear Models

- The model defines two sources of variation which determine the degree of clustering.
- The within-cluster variation ($var(e_{ij}) = \sigma^2$) and
- the between-cluster variation ($var(b_{0j}) = g$).

Features of the Two-level Linear Models

- Thus, we assume that
$$Y_{ij} = \boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{b}_j + e_{ij}$$
with one or more random effects.
- For a given level 2 unit, the random effects are assumed constant across level 1 units.
  - All patients at the same center have the same random effect values.
- The conditional expectation of Y_{ij}, given the level 2 random intercept, is
$$E(Y_{ij}|\boldsymbol{b}_j) = \boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{b}_j$$
- Level 1 observations are assumed to be conditionally independent given the random effects.

Analysis with a Two-level Linear Model

- Suppose that we observe patients i within a given clinic j
- Some questions of interest may be:
  - Is there significant between clinic heterogeneity?
  - Can we quantify the between clinic heterogenity?
  - What is the degree of within clinic correlation?
- We are interested in the following random intercept model

$$Y_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_{0j} + e_{ij}$$

where $var(e_{ij}) = \sigma^2$ and $var(b_j) = g$.

Analysis with a Two-level Linear Model

- Is there significant between clinic heterogeneity?
  - Is $g = 0$? How could we test this?
- Can we quantify the between clinic heterogenity?
  - How could we give a range of intercept values?
- What is the degree of within clinic correlation?
  - Observations for a given clinic have an intraclass correlation (ICC) equal to

$$Corr(Y_{ij}, Y_{kj}) = \frac{g}{g + \sigma^2}.$$

- What sample sizes, i.e., n_2 and n_{1j}, would be needed to answer these questions?

Analysis with a Two-level Linear Model

- Suppose that we observe patients i within a given clinical practice j.
  - Does the treatment have an effect?
  - Does the type of practice (hospital, private, other) have an impact?
- We are interested in the following random intercept model

$$Y_{ij} = \beta_0 + \beta_1 TRT_{ij} + \beta_2 Type_j + b_{0j} + e_{ij}$$

  where $var(e_{ij}) = \sigma^2$ and $var(b_j) = g$.
- How many different level j's do we need to determine β_2?

The Three-Level Linear Model

- Now consider a three-level longitudinal clinical trial in which
  1. physician practices are randomized to treatment,
  2. patients are nested within practices, and
  3. patients are measured at baseline and at three occasions after treatment.
- Level 1 is occasions, level 2 is patients, and level 3 is practice. Denote the response at the ith observation of the jth patient in the kth clinic by Y_{ijk}.
- Covariates can be measured at any of three levels. Random effects will be at levels 2 and 3.

The Three-Level Linear Model

- The general three-level linear model is written as follows:

$$Y_{ijk} = \boldsymbol{X}_{ijk}\beta + \boldsymbol{Z}^{(3)}_{ijk}\boldsymbol{b}^{(3)}_{k} + \boldsymbol{Z}^{(2)}_{ijk}\boldsymbol{b}^{(2)}_{jk} + e_{ijk}$$

- An example: Let t_{ijk} denote the time from baseline at which Y_{ijk} is obtained. Also, let Trt_{ijk} denote the treatment given to the jth patient at the ith occasion in the kth clinic.
- The treatment may be constant over occasions for a given patient, and is Trt_{jk}.
- Then a hierarchical three-level model for the response is given by

$$Y_{ijk} = \beta_0 + \beta_1 t_{ijk} + \beta_2(Trt_{ij} \times t_{ijk}) + \beta_3 X_{jk} + \beta_4 X_k + b^{(3)}_k + b^{(2)}_{jk} + e_{ijk}$$

The Three-Level Linear Model

- If
$$Var(b_k^{(3)}) = G^{(3)}, \; Var(b_{jk}^{(2)}) = G^{(2)}, \;\; \text{and } Var(e_{ijk}) = \sigma^2$$
and all random effects are assumed to be independent, then
$Var(Y_{ijk}) = G^{(2)} + G^{(3)} + \sigma^2$
- Thus, the observations for a given patient have an intraclass correlation structure, with
$$Corr(Y_{ijk}, Y_{i'jk}) = \frac{G^{(2)} + G^{(3)}}{G^{(2)} + G^{(3)} + \sigma^2}$$
- The expectation is $E(Y_{ijk}) = \beta_0 + \beta_1 t_{ijk} + \beta_2(Trt_{ij} \times t_{ijk}) + \beta_3 X_{jk} + \beta_4 X_k$

Estimation

- For the three-level linear model, the standard distributional assumptions are that:

$$b_k^{(3)} \sim N(0, G^{(3)}),\ b_{jk}^{(2)} \sim N(0, G^{(2)}), \quad \text{and } e_{ijk} \sim N(0, \sigma^2)$$

- As before, we use REML (or ML) to obtain estimates of $G^{(3)}$, $G^{(2)}$, and σ^2.
- This estimation procedure is available in `SAS PROC MIXED` or `lmer` in `R`.
- It is also available in MLwiN and HLM, two stand-alone programs developed for multi-level modeling.

Example one

- In a classic developmental study, ethylene glycol at doses of 0, 750, 1,500, or 3,000 mg/kg/day was administered to 94 pregnant mice.
- The crude results were as follows:

Dose				Weight (gm)	
(mg/kg)	Sqrt(Dose/750)	Dams	Fetuses	Mean	St. Dev.
0	0	25	297	0.97	0.10
750	1	24	276	0.90	0.10
1500	1.4	22	229	0.76	0.11
3000	2	23	226	0.70	0.12

- Based on experience and these data, the investigators modeled the response as linear in $\sqrt{dose}$.

SAS Code

- Because the observations are clustered within dam, the analysis must take account of clustering. If it does not, the sample size for comparisons between doses will be exaggerated.
- To fit a two-level model that is linear in $\texttt{newdose} = \sqrt{dose}$, use the SAS code

```
proc mixed data=toxicity;
class id;
model weight = newdose / solution chisq;
random intercept / subject=id g;
run;
```

Results

Variable	Estimate	SE	Z
Fixed Effects			
Intercept	0.98	0.02	61.3
Newdose	-0.13	0.01	-10.9
Random Effects			
Level 2 Variance ($\sigma_2^2 \times 100$)	0.73	0.12	6.1
Level 1 Variance ($\sigma_1^2 \times 100$)	0.56	0.03	21.6

Results

- The estimate of σ_2^2 indicates significant clustering of weights within litter. The estimated within-litter correlation is

$$\begin{aligned} \hat{\rho} &= \frac{\sigma_2^2}{\sigma_2^2 + \sigma_1^2} \\ &= \frac{0.73}{0.73 + 0.56} \\ &= 0.57 \end{aligned}$$

- The estimated decrease in weight, comparing the highest dose to 0 dose, is 0.27 (0.22, 0.33).
- Could we rank here?

Example Two

- The Television, School, and Family Smoking Prevention and Cessation Program.
- A randomized study with a 2 by 2 factorial design (None, CC, TV, CC & TV):

Factor 1: A school-based social-resistance curriculum (CC)
Factor 2: A television-based prevention program (TV)

- We report results for 1,600 seventh graders from 135 classes in 28 schools in Los Angeles.
- The response variable, the tobacco and health knowledge scale (THKS), was administered before and after the intervention.
- We consider a linear model for post-intervention THKS, with baseline THKS as a covariate.

Example Two

- Descriptive Statistics:

			Pre-THKS		THKS	
CC	TV	n	Mean	Std Dev	Mean	Std Dev
No	No	421	2.15	1.18	2.34	1.09
No	Yes	380	2.05	1.29	2.82	1.09
Yes	No	416	2.09	1.29	2.48	1.14
Yes	Yes	383	1.98	1.29	2.74	1.07

Example Two

- The model:

$$Y_{ijk} = \beta_0 + \beta_1 \text{Pre-THKS} + \beta_2 CC + \beta_3 TV + \beta_4 CC \times TV + b_k^{(3)} + b_{jk}^{(2)} + e_{ijk}$$

  where we list fixed and random effects on separate lines for clarity.
- Then the random effects and the error are assumed as:

$$b_k^{(3)} \sim N(0, \sigma_3^2),\ b_{jk}^{(2)} \sim N(0, \sigma_2^2),\ e_{ijk} \sim N(0, \sigma_1^2),$$

- This is the standard hierarchical (or multi-level) linear model with random effects at each level to introduce correlation within clusters.
- Fixed effects have both main effects and interactions for CC and TV.

SAS Code

```
proc mixed data=tvandcc covtest;
class sid cid;
model y = pre-THKS cc tv cc*tv / s;
random intercept / subject=sid g;
random intercept / subject=cid(sid) g;
run;
```

- The nesting `cid(sid)` is more computationally efficient
- Note that if `cid` is not unique (i.e., `cid` $= 1$ for the first class in each school) then nesting is absolutely needed.

Results

Variable	Estimate	Standard Error	Z
Fixed Effects			
Intercept	1.70	0.13	13.6
Pre-THKS	0.31	0.03	11.8
CC	0.64	0.16	4.0
TV	0.18	0.16	1.2
CC $\times$ TV	-0.33	0.22	-1.5
Random Effects			
Level 3 Variance (σ_3^2)	0.04	0.03	1.5
Level 2 Variance (σ_2^2)	0.07	0.03	2.3
Level 1 Variance (σ_1^2)	1.60	0.06	27.1

Comments on Results

- Fixed effects:
  - Pre-THKS is an important predictor of knowledge after the intervention.
  - CC had a clear effect on knowledge, but TV did not.
- Random effects:
  - There is relatively little clustering as measured by the small values for the level 2 and 3 variances as compared to the level 1 variance.
  - The variability among classrooms is almost twice as large as the variability among schools.
  - The correlation among children in the same classroom is $(0.04 + 0.07)/(0.04 + 0.07 + 1.60) = 0.06$.

Comments on Results

- The multi-level model can be generalized to an arbitrary number of levels.
- Generalized Linear Mixed Effects models have also been developed for the analysis of binary outcomes and counts in the multi-level setting.
- Multilevel modeling can be difficult:
  - A covariate can operate at different levels
  - It is not always clear how to combine covariates within a single model
  - Though hierarchical linear models with random effects are appealing, the extension to generalized linear models raises difficult problems of interpretation.
  - Marginal models and mixed-effects models can give quite different results in the non-linear setting

Nesting

- Nesting random effects gets very challenging computationally, even for linear mixed models.
- To help with the computations, use

```
random school;
random class(school);
repeated / subject=person(class*school) type=cs;
```

- Instead of

```
random school;
random class(school);
random person(class*school);
```