BIO 226: APPLIED LONGITUDINAL ANALYSIS

LECTURE 20

Multilevel Models

Multilevel Models

Until today, this course has focused on the analysis of longitudinal data.

Mixed models can also be used to analyze multilevel data.

Hierarchical or multilevel data arise when there is a *clustered/grouped* structure to the data.

Data of this kind frequently arise in the social, behavioral, and health sciences since individuals can be grouped in so many different ways.

For example, in studies of health services and outcomes, assessments of quality of care are often obtained from patients who are *nested* within different clinics.

Such data can be regarded as hierarchical/multilevel, with patients referred to as the level 1 units and clinics the level 2 units.

In this example there are two levels in the data hierarchy and, by convention, the lowest level of the hierarchy is referred to as level 1.

The term "level", as used in this context, signifies the position of a unit of observation within a hierarchy.

Clustering in multilevel data can be due to a naturally occurring hierarchy in the target population or a consequence of study design (or sometimes both).

Naturally Occurring Data Hierarchies

Studies of nuclear families: observations on the mother, father, and children (level 1 units) nested within families (level 2 units).

Studies of health services/outcomes: observations on patients (level 1 units) nested within clinics (level 2 units).

Studies of education: observations on children (level 1 units) nested within classrooms (level 2 units).

Note: Naturally occurring hierarchical data structures can have more than two levels, e.g., children (level 1 units) nested within classrooms (level 2 units), nested within schools (level 3 units).

Clustering as Consequence of Study Design

Longitudinal Studies: the clusters are composed of the repeated measurements obtained from a single individual at different occasions.

In longitudinal studies the level 1 units are the repeated occasions of measurement and the level 2 units are the subjects.

Cluster-Randomized Clinical Trials: Groups (level 2 units) of individuals (level 1 units), rather than the individuals themselves, are randomly assigned to different treatments or interventions.

Complex Sample Surveys: Many national surveys use multi-stage sampling, e.g., NHANES.

For example, in 1st stage, "primary sampling units" (PSUs) are defined based on counties in the United States. A first-stage random sample of PSUs are selected. In 2nd stage, within each selected PSU, a random sample of census blocks are selected. In 3rd stage, within selected census blocks, a random sample of households are selected.

Resulting data can be regarded as hierarchical, with households being the level 1 units, area segments the level 2 units, and counties the level 3 units.

Finally, clustering can be due to both study design and naturally occurring hierarchies in the target population.

Example: Clinical trials are often conducted in many different centers to ensure sufficient numbers of patients and/or to assess the effectiveness of the treatment in different settings.

Observations from a multi-center longitudinal clinical trial can be regarded as hierarchical data with 3 levels, with repeated measurement occasions (level 1 units) nested within subjects (level 2 units) nested within clinics (level 3 units).

Distinctive Feature of Multilevel Data

Distinctive feature of multilevel data is that they are *clustered*.

A consequence of this clustering is that measurement on units within a cluster are more similar than measurements on units in different clusters.

For example, two children selected at random from the same family are expected to respond more similarly than two children randomly selected from different families.

The clustering can be expressed in terms of correlation among the measurements on units within the same cluster.

Statistical models for hierarchical data must account for the intra-cluster correlation at each level; failure to do so can result in misleading inferences.

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.

Note: 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.

Combining covariates measured at different levels of the hierarchy within a single regression model is central to hierarchical modelling.

We begin by introducing the ideas with the two-level model.

Later we move to the three-level model to illustrate the general approach.

Two-Level Linear Models

Notation:

Let i index level 1 units and j index level 2 units (by convention, the subscripts are ordered from the lowest to the highest level).

We assume n_2 level 2 units in the sample.

Each of these clusters ($j = 1, 2, \cdots, n_2$) is composed of n_{1j} level 1 units.

For example, in a two-level study of physician practices, we would study n_2 practices, with n_{1j} patients in the j^{th} practice.

Let Y_{ij} denote the response for patient i in the j^{th} practice.

Associated with each Y_{ij} is a $1 \times p$ (row) vector of covariates, X_{ij}

Consider the following model for the mean:

$$E(Y_{ij}) = X_{ij}\beta$$

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

$$E(Y_{ij}) = \beta_1 + \beta_2 \mathrm{Trt}_{ij}$$

where Trt_{ij} is an indicator variable for treatment group (or Trt_j if treatment is constant within practice).

The two-level hierarchical linear model assumes that the correlation within practices can be described by a random effect.

Thus, we assume that

$$Y_{ij} = X_{ij}\beta + b_j + \epsilon_{ij}$$

Or, more generally,

$$Y_{ij} = X_{ij}\beta + Z_{ij}b_j + \epsilon_{ij}$$

with more than 1 random effect.

Features of the Two-Level Linear Model

1. Model defines two sources of variation. Magnitudes of within- and between-cluster variation determine degree of clustering/correlation.

2. For a given level 2 unit, random effects are assumed constant across level 1 units.

3. Conditional expectation of Y_{ij}, given identity of the level 2 group, is

$$X_{ij}\beta + Z_{ij}b_j$$

4. Level 1 observations are assumed to be conditionally independent given the random effects.

The two-level model is identical to the linear mixed model with intraclass correlation structure for repeated measurements (albeit with reversal of subscripting!).

Three-Level Linear Models

Next, 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.

Let Y_{ijk} denote response at the i^{th} observation of the j^{th} patient in the k^{th} practice.

Covariates can be measured at any of three levels. However, we now introduce random effects to represent clustering at both levels 2 and 3.

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

$$Y_{ijk} = X_{ijk}\beta + Z^{(3)}_{ijk}b^{(3)}_{k} + Z^{(2)}_{ijk}b^{(2)}_{jk} + \epsilon_{ijk}$$

Example: Three-Level Model for the Multi-Level Longitudinal Clinical Trial

Let t_{ijk} denote the time from baseline at which Y_{ijk} is obtained.

Also, let Trt_{ij} denote the treatment given to the j^{th} patient at the i^{th} occasion.
The treatment may be constant over occasions for a given patient (Trt_j).

A hierarchical three-level model for the response is given by

$$Y_{ijk} = \beta_1 + \beta_2 t_{ijk} + \beta_3(\text{Trt}_j \times t_{ijk}) + b_k^{(3)} + b_{jk}^{(2)} + \epsilon_{ijk}$$

This model assumes a common intercept and separate linear trends over time in the two treatment groups.

If

$$\text{Var}(b_k^{(3)}) = G^{(3)}, \text{Var}(b_{jk}^{(2)}) = G^{(2)}, \text{ and } \text{Var}(\epsilon_{ijk}) = \sigma^2,$$

and all random effects are assumed to be independent, then

$$\text{Var}(Y_{ijk}) = G^{(2)} + G^{(3)} + \sigma^2$$

and the covariance between two observations from the same patient is

$$G^{(2)} + G^{(3)}$$

Thus, the observations for a given patient have an intraclass correlation structure, with

$$\text{Corr}(Y_{ijk}, Y_{i'jk}) = \frac{G^{(2)} + G^{(3)}}{G^{(2)} + G^{(3)} + \sigma^2}.$$

Because this is a linear mixed model,

$$E(Y_{ijk}) = \beta_1 + \beta_2 t_{ijk} + \beta_3(\text{Trt}_{ij} \times t_{ijk})$$

Estimation

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

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

Given these assumptions, estimation of the model parameters is relatively straightforward. The GLS estimate of β is given by

$$\beta = \Big\{ \sum_{k=1}^{n_3} (X_k' V_k^{-1} X_k) \Big\}^{-1} \sum_{k=1}^{n_3} (X_k' V_k^{-1} Y_k)$$

where Y_k is a column vector of length $\sum_{j=1}^{n_{2k}} n_{1jk}$, the number of observations in the k^{th} cluster. X_k is the corresponding matrix of covariates, and V_k is the covariance matrix of Y_k.

Estimation (Continued)

As before, we use REML (or ML) to obtain estimates of $G^{(3)}$, $G^{(2)}$, and σ^2.

Once these estimates are obtained, we can estimate the covariance matrices, V_k, and substitute those estimates into the expression for the GLS estimator.

This estimation procedure is available in PROC MIXED in SAS.

It is also available in MLwiN and HLM, two stand-alone programs developed for multilevel modeling.

Case Study 1: Developmental Toxicity Study of Ethylene Glycol

Developmental toxicity studies of laboratory animals play a crucial role in the testing and regulation of chemicals.

Exposure to developmental toxicants typically causes a variety of adverse effects, such as fetal malformations and reduced fetal weight at term.

In a typical developmental toxicity experiment, laboratory animals are assigned to increasing doses of a chemical or test substance.

Consider an analysis of data from a development toxicity study of ethylene glycol (EG).

Ethylene glycol is used as an antifreeze, as a solvent in the paint and plastics industries, and in the formulation of various types of inks.

In a study of laboratory mice conducted through the National Toxicology Program (NTP), EG was administered at doses of 0, 750, 1500, or 3000 mg/kg/day to 94 pregnant mice (dams) beginning just after implantation.

Following sacrifice, fetal weight and evidence of malformations were recorded for each live fetus.

In our analysis, we focus on the effects of dose on fetal weight.

Summary statistics (ignoring clustering in the data) for fetal weight for the 94 litters (composed of a total of 1028 live fetuses) are presented in the table on next slide.

Fetal weight decreases monotonically with increasing dose, with the average weight ranging from 0.97g in the control group to 0.70g in the group administered the highest dose.

The decrease in fetal weight is not linear in increasing dose, but is approximately linear in increasing $\sqrt{\text{dose}}$.

Table 1: Descriptive statistics on fetal weight.

Dose				Weight (g)	
(mg/kg)	$\sqrt{\text{Dose}/750}$	Dams	Fetuses	Mean	St. Deviation†
0	0	25	297	0.972	0.098
750	1	24	276	0.877	0.104
1500	1.4	22	229	0.764	0.107
3000	2	23	226	0.704	0.124

†Calculated ignoring clustering.

Because the observations are clustered within dam, the analysis must take account of clustering.

If it does not, the apparent sample size for comparisons between doses will be exaggerated.

To fit a two-level model that is linear in sqrt(dose),

$$Y_{ij} = \beta_1 + \beta_2 \sqrt{\text{dose}/750} + b_j + \epsilon_{ij},$$

we can use the following commands:


```
DATA toxicity;
  INFILE 'c:\bio226\datasets\ethyleneglycol.txt';
  INPUT id dose weight mal;
  newdose=sqrt(dose/750);
  RUN;

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

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

$$\begin{aligned} \widehat{\rho} &= \widehat{\sigma}_2^2/(\widehat{\sigma}_2^2 + \widehat{\sigma}_1^2) \\ &= 0.73/(0.73 + 0.56) \\ &= 0.57 \end{aligned}$$

The estimated decrease in weight, comparing the highest dose to 0 dose, is 0.27 (0.22, 0.33).

The model-based and empirical (sandwich) standard errors are very similar (not shown), indicating that the random effects structure is adequate.

It is also easy to test for linearity on the square root scale, though we have data at only four doses.

Case Study 2: The Television, School, and Family Smoking Prevention and Cessation Program

A randomized study with a 2 by 2 factorial design:
- 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.

Descriptive Statistics

			Pre-THKS		Post-THKS	
CC	TV	n	Mean	Std Dev	Mean	Std Dev
No	No	421	2.15	1.18	2.36	1.30
No	Yes	416	2.09	1.29	2.54	1.44
Yes	No	380	2.05	1.29	2.97	1.40
Yes	Yes	383	1.98	1.29	2.82	1.31

The mean value of Pre-THKS does not differ significantly among treatment groups.

Three-Level Model

Model the adjusted change in THKS scores as function of main effects of CC and TV and the CC $\times$ TV interaction:

$$Y_{ijk} = \beta_1 + \beta_2 \text{Pre-THKS} + \beta_3 \text{CC} + \beta_4 \text{TV} + \beta_5 \text{CC} \times \text{TV} + b_k^{(3)} + b_{jk}^{(2)} + \epsilon_{ijk}.$$

In a slightly modified notation, assume

$$\begin{aligned} \epsilon_{ijk} &\sim N(0, \sigma_1^2) \\ b_{jk}^{(2)} &\sim N(0, \sigma_2^2) \\ b_k^{(3)} &\sim N(0, \sigma_3^2) \end{aligned}$$

This is the standard hierarchical (or multilevel) linear model with school and classroom effects modelled by incorporating random effects at levels 3 and 2, respectively (level 1 units are the children).

PROC MIXED in SAS

```
DATA tvandcc;
  INFILE 'c:\bio226\datasets\tv.txt';
  INPUT sid cid cc tv baseline THKS;
RUN;

PROC MIXED DATA=tvandcc COVTEST;
  CLASS sid cid;
  MODEL thks = baseline cc tv cc*tv / S;
  RANDOM INTERCEPT / SUBJECT=sid  G ;
  RANDOM INTERCEPT / SUBJECT=cid  G ;
RUN;
```

Table 2: Fixed effects estimates for the THKS scores.

Parameter	Estimate	SE	Z
Intercept	1.702	0.1254	13.57
Pre-Intervention THKS	0.305	0.0259	11.79
CC	0.641	0.1609	3.99
TV	0.182	0.1572	1.16
CC $\times$ TV	-0.331	0.2245	-1.47

Table 3: Random effects estimates for the THKS scores.

Parameter	Estimate	SE	Z
Level 3 Variance:			
σ_3^2	0.039	0.0253	1.52
Level 2 Variance:			
σ_2^2	0.065	0.0286	2.26
Level 1 Variance:			
σ_1^2	1.602	0.0591	27.10

Consider REML estimates of the three sources of variability.

Comparing their relative magnitudes, there is variability at both classroom and school levels, with almost twice as much variability among classrooms within a school as among schools themselves.

Correlation among THKS scores for classmates (or children within same classroom within same school) is approximately 0.061 (or $\frac{0.039+0.065}{0.039+0.06+1.602}$).

Correlation among THKS scores for children from different classrooms within same school is approximately 0.023 (or $\frac{0.039}{0.039+0.06+1.602}$).

Next, consider REML estimates of fixed effects for the interventions.

When compared to their SEs, indicate that neither mass-media intervention (TV) nor its interaction with social-resistance classroom curriculum (CC) have an impact on adjusted changes in THKS scores from baseline.

There is a significant effect of the social-resistance classroom curriculum, with children assigned to the social-resistance curriculum showing increased knowledge about tobacco and health.

The estimate of the main effect of CC, in the model that excludes the CC $\times$ TV interaction, is 0.47 (SE = 0.113, $p < 0.0001$).

The intra-cluster correlations at both the school and classroom levels are relatively small.

It is very tempting to regard this as an indication that the clustering in these data is inconsequential.

However, such a conclusion would be erroneous.

Although intra-cluster correlations are relatively small, they have an impact on inferences concerning the effects of the intervention conditions.

To illustrate this, consider analysis that ignores clustering in the data:

$$Y_{ijk} = \beta_1 + \beta_2 \text{Pre-THKS} + \beta_3 \text{CC} + \beta_4 \text{TV} + \beta_5 \text{CC} \times \text{TV} + e_{ijk},$$

The results of fitting this model to the THKS scores are presented in Table 63 and the estimates of the fixed effects are similar to those reported in Table 61.

However, SEs (assuming no clustering) are misleadingly small for intervention effects and lead to substantively different conclusions about effects of intervention conditions.

This highlights an important lesson: the impact of clustering depends on both the magnitude of the intra-cluster correlation and the cluster size.

For the data from the TVSFP, the cluster sizes vary from 1–13 classrooms within a school and from 2–28 students within a classroom.

With relatively large cluster sizes, even very modest intra-cluster correlation can have a discernible impact on inferences.

Table 4: Fixed effects estimates from analysis that ignores clustering in the THKS scores.

Parameter	Estimate	SE	Z
Intercept	1.661	0.0844	19.69
Pre-Intervention THKS	0.325	0.0258	12.58
CC	0.641	0.0921	6.95
TV	0.199	0.0900	2.21
CC $\times$ TV	-0.322	0.1302	-2.47

Generalizations

The multilevel model can be generalized to an arbitrary number of levels.

Generalized linear mixed effects models (GLMMs) have also been developed for the analysis of binary outcomes and counts in the multilevel setting (see FLW, Chapter 17).

Cautionary Remarks – 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.

- As discussed earlier, marginal models and mixed-effects models can give quite different results in the non-linear setting

Summary

Despite certain complexities, multilevel models are now widely used.

In both designed experiments and studies of effects of family/community factors on health, multilevel models provide a usually effective approach to data analysis that accounts for correlations induced by clustering.

Multilevel models are, in one sense, no different than longitudinal models.

Unlike logistic regression and survival analysis, where concept of regression analysis can be applied quite robustly and with few choices, longitudinal and multilevel analysis require more careful thought about the choice and meaning of models.

This is both their challenge and their reward.

BIO 226: THE END