BIOS6643 Longitudinal

oftwares

Randy

Department of Biostatistics & Informatics

2.1.1.1 Clustered Data vs. Repeated-Measures and Longitudinal Data

Clustered data

- As data sets in which the dependent variable is **measured once** for **each subject** (the unit of analysis)
- The units of analysis are grouped into, or **nested within**, clusters of units.

Repeated-measures data

- Quite generally as data sets in which the dependent variable is measured **more than once** on the same unit of analysis **across** levels of a repeated-measures factor (or factors).
- The repeated-measures factors may be time or other experimental or observational conditions, as **within-subject** factors.
- **Dropout** of subjects is **not usually a concern** in repeated-measures data, although there may be missing data.

Longitudinal data

- For data sets the dependent variable is measured at **several points** in time for **each unit** of analysis.
- We usually conceptualize longitudinal data as involving **at least two** repeated measurements made over a relatively **long** period of time.
- In contrast to repeated-measures data, **dropout** of subjects is often a **concern** in the analysis of longitudinal data.
- It may be difficult to classify data sets as either longitudinal or repeated-measures data. In the context of analyzing data using *LMMs*, this distinction is *not critical*.
- The important feature of both of these types of data is that the dependent variable is measured more than once for each unit of analysis, with the repeated measures likely to be correlated.

Clustered longitudinal

- Data sets combine features of both clustered and longitudinal data.
- More specifically, the units of analysis are **nested within clusters**, and each unit is measured more than once.

We refer to clustered, repeated-measures, and longitudinal data as **hierarchical data** sets, because the observations can be placed into levels of a hierarchy in the data.

TABLE 2.1: Hierarchical Structures of the Example Data Sets Considered in Chapters 3 through 7

Data Type	**Clustered Data**		**Repeated-Measures/ Longitudinal Data**		
	Two-Level	**Three-Level**	**Repeated-Measures**	**Longitudinal**	**Clustered Longitudinal**
Data set (Chap.)	Rat Pup (Chap. 3)	Classroom (Chap. 4)	Rat Brain (Chap. 5)	Autism (Chap. 6)	Dental Veneer (Chap. 7)
Repeated/ longitudinal measures (t)			Spanned by brain region and treatment	Age in years	Time in months
Subject/unit of analysis (i)	$\boxed{\textit{Rat Pup}}$	$\boxed{\textit{Student}}$	$\boxed{\textit{Rat}}$	$\boxed{\textit{Child}}$	$\boxed{\textit{Tooth}}$
Cluster of units (j)	Litter	Classroom			Patient
Cluster of clusters (k)		School			

Note: Italicized terms in boxes indicate the unit of analysis for each study; (t, i, j, k) indices shown here are used in the model notation presented later in this book.

In particular, we use the index t to denote repeated/longitudinal measurements, the index i to denote subjects or units of analysis, and the index j to denote clusters. The index k is used in models for three-level clustered data to denote "clusters of clusters."

2.1.1.2 Levels of Data

The concept of "levels" of data is based on ideas from the hierarchical linear modeling (HLM) literature *Raudenbush & Bryk, 2002*.

All data sets appropriate for an analysis using LMMs have at least two levels of data.

One notable exception is data sets with crossed random factors (See later lectures).

Levels of data

- Level 1 denotes observations at the **most detailed** level of the data.
    - In a clustered data set, Level 1 represents the **units of analysis** (or **subjects**) in the study.
    - In a repeated-measures or longitudinal data set, Level 1 represents the **repeated measures** made on the same unit of analysis.
    - The **continuous** dependent variable is always measured at Level 1 of the data.
- Level 2 represents the next level of the hierarchy. In clustered data sets, Level 2 observations represent **clusters of units**. In repeated-measures and longitudinal data sets, Level 2 represents the units of analysis.
- Level 3 represents the next level of the hierarchy, and generally refers to clusters of units in clustered longitudinal data sets, or clusters of Level 2 units (**clusters of clusters**) in three-level clustered data sets.

TABLE 2.1: Hierarchical Structures of the Example Data Sets Considered in Chapters 3 through 7

Data Type	**Clustered Data**		**Repeated-Measures/ Longitudinal Data**		
	Two-Level	**Three-Level**	**Repeated-Measures**	**Longitudinal**	**Clustered Longitudinal**
Data set (Chap.)	Rat Pup (Chap. 3)	Classroom (Chap. 4)	Rat Brain (Chap. 5)	Autism (Chap. 6)	Dental Veneer (Chap. 7)
Repeated/ longitudinal measures (t)			Spanned by brain region and treatment	Age in years	Time in months
Subject/unit of analysis (i)	$\boxed{\textit{Rat Pup}}$	$\boxed{\textit{Student}}$	$\boxed{\textit{Rat}}$	$\boxed{\textit{Child}}$	$\boxed{\textit{Tooth}}$
Cluster of units (j)	Litter	Classroom			Patient
Cluster of clusters (k)		School			

Note: Italicized terms in boxes indicate the unit of analysis for each study; (t, i, j, k) indices shown here are used in the model notation presented later in this book.

However, when using the paradigm of levels of data, the distinction between clustered vs. repeated-measures/longitudinal data becomes less obvious.

2.1.2.1 Fixed Factors

The concept of a fixed factor is most commonly used in the setting of a **standard ANOVA or ANCOVA** model.

A fixed factor

As a categorical or classification variable, for which the investigator has included all levels (or conditions) that are of interest in the study.

Fixed factors might include **qualitative** covariates, such as gender; **classification** variables implied by a **survey sampling design**, such as region or stratum, or by a **study design**, such as the treatment method in a randomized clinical trial; or **ordinal classification** variables in an observational study, such as age group.

Levels of a fixed factor are chosen so that they represent specific conditions, and they can be used to define contrasts (or sets of contrasts) of interest in the research study.

2.1.2.2 Random Factors

A random factor

As a classification variable with levels that can be thought of as being randomly sampled from a population of levels being studied.

All possible levels of the random factor are not present in the data set, but it is the researcher's intention to make inferences about the entire population of levels.

The classification variables that identify the Level 2 and Level 3 units in both clustered and repeated-measures/longitudinal data sets are often considered to be random factors.

Random factors are considered in an analysis so that variation in the dependent variable across levels of the random factors can be assessed, and the results of the data analysis can be generalized to a greater population of levels of the random factor.

2.1.2.3 Fixed Factors vs. Random Factors

In contrast to the levels of fixed factors, the levels of random factors do not represent conditions chosen specifically to meet the objectives of the study.

However, depending on the goals of the study, the same factor may be considered either as a fixed factor or a random factor, as we note in the following paragraph.

In the Dental Veneer data analyzed, the dependent variable (Gingival Crevicular Fluid, or GCF) is measured repeatedly on selected teeth within a given patient, and the teeth are numbered according to their location in the mouth. In our analysis, we assume that the teeth measured within a given patient represent a random sample of all teeth within the patient, which allows us to generalize the results of the analysis to the larger hypothetical "population" of "teeth within patients." In other words, we consider "tooth within patient" to be a random factor. If the research had been focused on the specific differences between the selected teeth considered in the study, we might have treated "tooth within patient" as a fixed factor. In this latter case, inferences would have only been possible for the selected teeth in the study, and not for all teeth within each patient.

2.1.2.4 Fixed Effects vs. Random Effects

Fixed effects

Fixed effects, called regression coefficients or fixed-effect parameters, describe the relationships between the dependent variable and predictor variables (i.e., fixed factors or continuous covariates) for an entire population of units of analysis, or for a relatively small umber of subpopulations defined by levels of a fixed factor.
Fixed effects may describe contrasts or differences between levels of a fixed factor (e.g., between males and females) in terms of mean responses for the continuous dependent variable, or they may describe the relationship of a continuous covariate with the dependent variable.
Fixed effects are assumed to be unknown fixed quantities in an LMM, and we estimate them based on our analysis of the data collected in a given research study.

Random effects

Random effects are random values associated with the levels of a random factor (or factors) in an LMM.

These values, which are specific to a given level of a random factor, usually represent random devia_{ti}ons from the rela_{ti}onships described by fixed effects.

For example, random effects associated with the levels of a random factor can enter an LMM as random intercepts (represen_{ti}ng random devia_{ti}ons for a given subject or cluster from the overall fixed intercept), or as random coefficients (represen_{ti}ng random devia_{ti}ons for a given subject or cluster from the overall fixed effects) in the model.

In contrast to fixed effects, random effects are represented as random variables in an LMM.

TABLE 2.3: Examples of the Interpretation of Fixed and Random Effects in an LMM Based on the Autism Data Analyzed in Chapter 6

Effect Type	Predictor Variables Associated with Each Effect	Effect Applies to	Possible Interpretation of Effects
Fixed	Variable corresponding to the intercept (i.e., equal to 1 for all observations)	Entire population	Mean of the dependent variable when all covariates are equal to zero
	AGE	Entire population	Fixed slope for AGE (i.e., expected change in the dependent variable for a 1-year increase in AGE)
	SICDEGP1, SICDEGP2 (indicators for baseline expressive language groups; reference level is SICDEGP3)	Entire population within each subgroup of SICDEGP	Contrasts for different levels of SICDEGP (i.e., mean differences in the dependent variable for children in Level 1 and Level 2 of SICDEGP, relative to Level 3)
Random	Variable corresponding to the intercept	CHILDID (individual child)	Child-specific random deviation from the fixed intercept
	AGE	CHILDID (individual child)	Child-specific random deviation from the fixed slope for AGE

2.1.2.5 Nested vs. Crossed Factors and Their Corresponding Effects

When a particular level of a factor (random or fixed) can only be measured within a single level of another factor and not across multiple levels, the levels of the first factor are said to be nested within levels of the second factor. The effects of the nested factor on the response are known as nested effects. For example, in the Classroom data set analyzed in Chapter 4, both schools and classrooms within schools were randomly sampled. Levels of classroom (one random factor) are nested within levels of school (another random factor), because each classroom can appear within only one school.

When a given level of a factor (random or fixed) can be measured across multiple levels of another factor, one factor is said to be crossed with another, and the effects of these factors on the dependent variable are known as crossed effects. For example, in the analysis of the Rat Pup data in Chapter 3, we consider two crossed fixed factors: TREATMENT and SEX. Specifically, levels of TREATMENT are crossed with the levels of SEX, because both male and female rat pups are studied for each level of treatment.