Chapter 4
Hierarchical Structures Modelling

Hierarchical structures modelling: This will focus on methodology for the analysis of data with complex patterns of variability such as those arising from longitudinal and nested designs: e.g., measurements on subjects over time, or on plots within fields within farms within villages. The goal is to provide students with knowledge and confidence to use hierarchical modelling in their discipline through understanding of statistical ideas and concepts behind hierarchical models set up and estimation.

1. Introduction

Hierarchical linear models and multilevel models are variant terms for what are broadly called linear mixed models (LMM). These models handle data where observations are not independent, correctly modeling correlated error. Uncorrelated error is an important but often violated assumption of statistical procedures in the general linear model family, which includes analysis of variance, correlation, regression, and factor analysis. Violations occur when error terms are not independent but instead cluster by one or more grouping variables. For instance, predicted student test scores and errors in predicting them may cluster by classroom, school, and municipality. When clustering occurs due to a grouping factor (this is the rule, not the exception), then the standard errors computed for prediction parameters will be wrong (eg., wrong b coefficients in regression).
Linear mixed models are a generalization of general linear models to better support analysis of a continuous dependent variable for the following:

1. Random effects: For when the set of values of a categorical predictor variable are seen not as the complete set but rather as a random sample of all values (eg., when the variable "product" has values representing only 30 of a possible 142 brands). Random effects modeling allows the researcher to make inferences over a wider population than is possible with regression or other general linear model (GLM) methods.

2. Hierarchical effects: For when predictor variables are measured at more than one level (eg., reading achievement scores at the student level and

teacher–student ratios at the school level; or sentencing lengths at the offender level, gender of judges at the court level, and budgets of judicial districts at the district level). The researcher can assess the effects of higher levels on the intercepts and coefficients at the lowest level (eg., assess judge-level effects on predictions of sentencing length at the offender level).

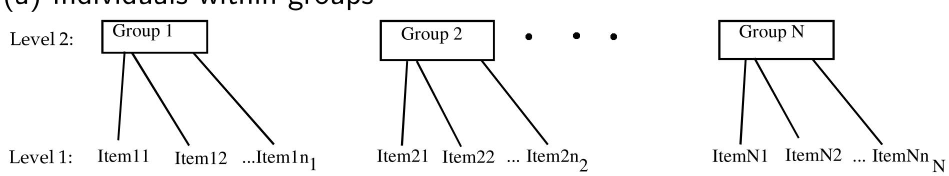
3. Repeated measures: For when observations are correlated rather than independent (eg., before–after studies, time series data, matched-pairs designs). In repeated measures, the lowest level is the observation level (eg., student test scores on multiple occasions), grouped by observation unit (eg., students) such that each unit (student) has multiple data rows, one for each observation occasion.

2. Examples of Hierarchies

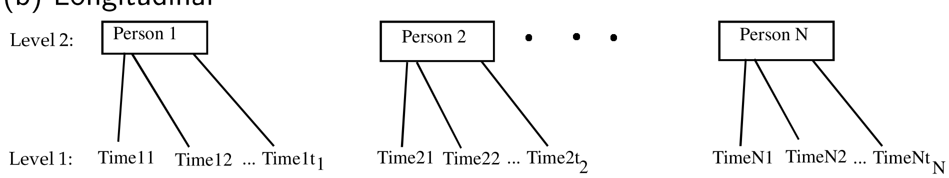
We will look at the following examples;

Example. Some examples of Hierarchies include:

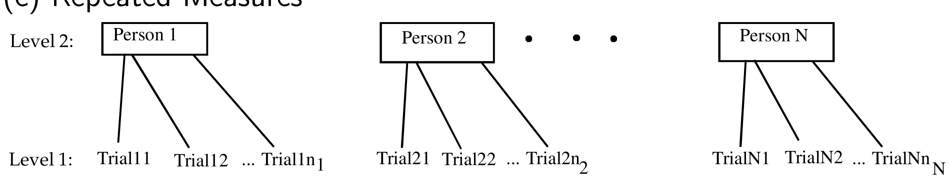
(a) Individuals within groups



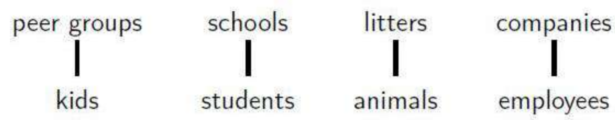
(b) Longitudinal

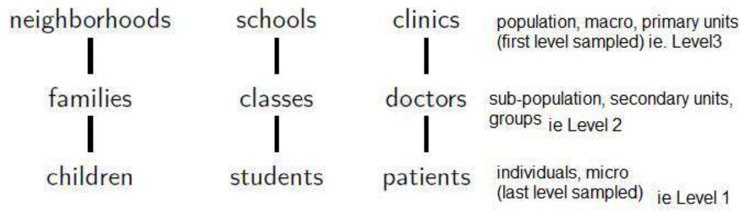


(c) Repeated Measures



More Examples of Hierarchies





The versatility of linear mixed modeling has led to a variety of terms for the models it makes possible. Different disciplines favor one or another label, and different research targets influence the selection of terminology as well. These terms, many of which are discussed later in this chapter, include random intercept modeling, random coefficients modeling, random coefficients regression, random effects modeling, hierarchical linear modeling, multilevel modeling, linear mixed modeling, growth modeling, and longitudinal modeling. Linear mixed models in some disciplines are called "random effects" or "mixed effects" models. In economics, the term "random coefficient regression models" is used. In sociology, "multilevel modeling" is common, alluding to the fact that regression intercepts and slopes at the individual level may be treated as random effects of a higher (eg., organizational) level. And in statistics, the term "covariance components models" is often used, alluding to the fact that in linear mixed models one may decompose the covariance into components attributable to within-groups versus between-groups effects. In spite of many different labels, the commonality is that all adjust observation-level predictions based on the clustering of measures at some higher level or by some grouping variable.

The "linear" in linear mixed modeling has a meaning similar to that in regression: There is an assumption that the predictor terms on the right-hand side of the estimation equation are linearly related to the target term on the left-hand side. Of course, nonlinear terms such as power or log functions may be added to the predictor side (eg., time and time-squared in longitudinal studies). Also, the target variable may be transformed in a nonlinear way (eg., logit link functions). Linear mixed model (LMM) procedures that do the latter are "generalized" linear mixed models. Just as regression and GLM procedures can be extended to "generalized general linear models" (GZLM), multilevel and other LMM procedures can be extended to "generalized linear mixed models" (GLMM).

Linear mixed models for multilevel analysis address hierarchical data, such as when employee data are at level 1, agency data are at level 2, and department data are at level 3. Hierarchical data usually call for LMM implementation. While most multilevel modeling is univariate (one dependent variable), multivariate multilevel modeling for two or more dependent variables is available also. Likewise, models for cross-classified data exist for data that are not strictly hierarchical (eg., as when schools are a lower level and neighborhoods are a higher level, but schools may serve more than one neighborhood).

The researcher undertaking causal modeling using linear mixed modeling should be guided by multilevel theory. That is, hierarchical linear modeling postulates that there are cross-level causal effects. Just as regression models postulate direct effects of independent variables at level 1 on the dependent variable at level 1, so too, multilevel models specify cross-level interaction effects between variables located at different levels. In doing multilevel modeling, the researcher postulates the existence of mediating mechanisms that cause variables at one level to influence variables at another level (eg., school-level funding may positively affect individual-level student performance by way of recruiting superior teachers, made possible by superior financial incentives).

Multilevel modeling tests multilevel theories statistically, simultaneously modeling variables at different levels without necessary recourse to aggregation or disaggregation. Aggregation and disaggregation as used in regression models run the risk of ecological fallacy: What is true at one level need not be true at another level. For instance, aggregated state-level data on race and literacy greatly overestimate the correlation of African American ethnicity with illiteracy because states with many African Americans tend to have higher illiteracy for all races. Individual-level data shows a low correlation of race and illiteracy.

3. Why use Linear Mixed/Hierarchical/Multilevel Modelling?

The central reason, noted above, is that linear mixed models handle random effects, including the effects of grouping of observations under higher entities (eg., grouping of employees by agency, students by school, etc.). Clustering of observations within groups leads to correlated error terms, biased estimates of parameter (eg., regression coefficient) standard errors, and possible substantive mistakes when interpreting the importance of one or another predictor variable. Whenever data are sampled, the sampling unit as a grouping variable may well be a random effect. In a study of the federal bureaucracy, for instance, "agency" might be the sampling unit and error terms may cluster by agency, violating ordinary least squares (OLS) assumptions.

Unlike OLS regression, linear mixed models take into account the fact that over many samples, different b coefficients for effects may be computed, one for each group. Conceptually, mixed models treat b coefficients as random effects drawn from a normal distribution of possible b's, whereas OLS regression treats the b parameters as if they were fixed constants (albeit within a confidence interval). Treating "agency" as a random rather than fixed factor will alter and make more accurate the ensuing parameter estimates. Put another way, the misestimation of standard errors in OLS regression inflates Type 1 error (thinking there is relationship when there is not: false positives), whereas mixed models handle this potential problem. In addition, LMM can handle a random sampling variable like "agencies," even when there are too many agencies to make into dummy variables in OLS regression and still expect reliable coefficients.

In summary, OLS regression and GLM assume error terms are independent and have equal error variances, whereas when data are nested or cross-classified by groups, individual-level observations from the same upper-level group will not be independent but rather will be more similar due to such factors as shared group history and group selection processes. While random effects associated with upper-level random factors do not affect lower-level population means, they do affect the covariance structure of the data. Indeed, adjusting for this is a central point of LMM models and is why linear mixed models are used instead of regression and GLM, which assume independence.

4. Types of Linear Mixed Models

Linear mixed modeling supports a very wide variety of models, too extensive to enumerate here. As mentioned above, different disciplines and authors have employed differing labels for specific types of models, adding to the seeming complexity of the subject. In this section, the most common types of models are defined, using

the most widely applied labels.
The "types" refer to various combinations of what is being predicted and what is doing the predicting. In ordinary regression, the researcher normally is predicting a level 1 (typically individual subject level) dependent variable such as "employee performance score" from one or more level 1 independent variables (eg., from "employee education"). In the multilevel world of linear mixed modeling, however, there are other possibilities. Let level 2 be defined by the grouping variable "agency" and a level 2 variable such as "mean agency education," with the multilevel theory being that the presence of more highly educated employees in an agency has a synergistic effect at the level of the individual. The level 2 grouping variable may have an effect on the intercept (mean score) at level 1 and/or on the b coefficient (slope) of education at level 1. Likewise, the level 2 predictor, mean agency education, may have an effect on the level 1 intercept and the level 1 slope.
There are three broad classes of models: fixed effects, random effects, and mixed. Most models treated in this book are mixed, hence the term "linear mixed modeling."

1. Fixed effects models. Linear mixed modeling is mostly about models involving random effects as well as fixed effects. In mixed models, effects that impact the intercept (representing the mean of the dependent variable when other predictors are zero) are modeled as fixed effects. However, purely fixed effects models such as ordinary regression models may be fitted also. These are models with only fixed factors and optional fixed covariates as predictors. An example would be a study of employee performance score by education, controlling for gender. Most models for analysis of variance, regression, and GLM are fixed effects models. These are the most common type of model in social science. Compared to an OLS regression model, a fixed effects model implemented in LMM will generate very similar if not identical estimates with similar (but not identical) output tables.

2. Random effects models. Random effects models are those with one or more random factors and optional covariates as predictors. Effects that influence the covariance structure are modeled as random factors. Sampling variables (eg., state, where individuals are sampled within a sample of states; subject,

where a sample of subjects have repeated measures over time) are random factors, as is any grouping variable where the clustering of effects creates correlated error. An example would be a study of employee performance score at level 1 by agency at level 2, controlling for salary level at level 1. Score would be the dependent variable, agency the random factor (assuming only a random sample of agencies were studied), and salary the covariate. The level 1 intercept of score may be modeled as a random effect of agency at level 2. Likewise, the level 1 slope of employee education might be modeled as a random effect of agency. If only the intercept is modeled, it is a random intercept model. If the slope is modeled as well, it is a random coefficients model. Some authors use the term "hierarchical linear model" to refer to random effects models in which both intercepts and slopes are modeled.

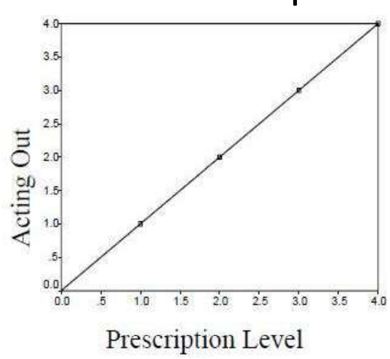
3. Mixed models. Mixed models, naturally, are ones with both fixed and random effects. A given effect may be both fixed and random if it contributes to both the intercept and the covariance structure for the model. Predictors at any level are typically included as fixed effects. For instance, covariates at level 2 are normally included as fixed effect variables. Slopes of variables at lower levels may be random effects of higher-level variables. Grouping variables (eg., school, agency) at any level are random factors.

Hierarchical linear models (HLM) are a type of mixed model with hierarchical data—that is, where nested data exist at more than one level (eg., student-level data and school-level data, with students nested within schools). In explaining a dependent variable, HLM models focus on differences between groups (eg., schools) in relation to differences within groups (eg., among students within schools). While it is possible to construct one-level models in linear mixed modeling, most use of LMM can be seen as one or another form of HLM, so the two terms are often used synonymously in spite of nuanced differences. **Random intercept models** are models where only the intercept of the level 1 dependent variable is modeled as an effect of the level 2 grouping variable and possibly other level 1 or level 2 (or higher) covariates. Random coefficients models are ones where the coefficient(s) of lower-level predictor(s) is/are modeled as well. There are several major types of random intercept and random coefficient models (See John Fox, Linear Mixed Models, 2002). The **null model**, also called the "unconditional model" or a "one-way ANOVA with random effects," is a type of random intercept model that predicts the level 1 intercept of the dependent variable as a random effect of the level 2 grouping variable, with no other predictors at level 1 or 2 in a two-level model. For instance, differences in mean performance scores may be analyzed in terms of the random effect of agency at level 2. The researcher is testing to see if there is an agency effect. The null model is used to calculate the intraclass correlation coefficient (ICC), which is a test of the need for mixed modeling. The null model also serves as a "baseline model" for purposes of comparison with later, more complex models. Note that a model is "conditional" by the presence of predictors at level 1 or level 2. Since the researcher almost always employs predictor variables and is not simply interested in the null model, most mixed models are conditional. The central point of LMM often is to assess the difference between the researcher's conditional model and the null model without predictors. The likelihood ratio test can be used to assess this difference.

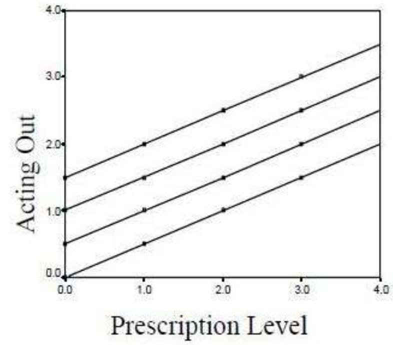
4.1. Random Vs Fixed Coefficients

The following are different scenarios from an experiment on the acting time as influenced by prescription levels of a particular drug.

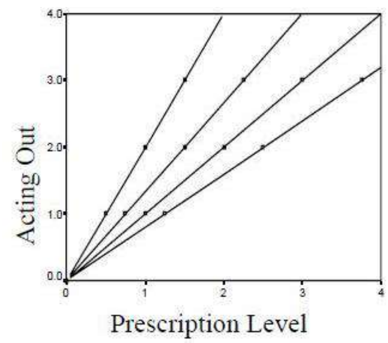
A model with Fixed Intercepts and Slopes



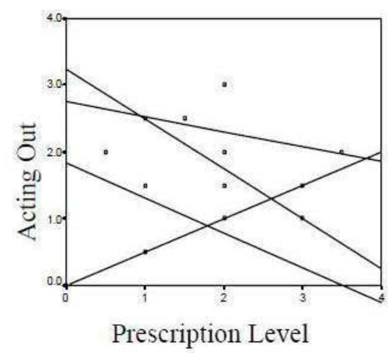
Random intercepts and fixed slopes



Fixed Intercepts and Random Slopes



Random Intercepts and Slopes



EXERCISE 8. When should we use HLM?

EXERCISE 9. What are the assumptions of the Linear regression approaches?

EXERCISE 10. What are the problems with hierarchical data as compared to data used in linear regression?

In many research studies, we start by drawing a sample of individuals and randomly assign them to either treatment or control. Here, each individual represents an independent observation and traditional data analytic techniques are appropriate. However, we are not always able to separate people from their contexts. For instance, students learn in schools, children grow up in neighborhoods and patients are treated in hospitals. When the cluster is a necessary part of a research design, the resultant data will be nested, or hierarchically structured.

However, individuals in the same group are not independent, for example people in a group tend to be more like others in their group than they are like others in a different group due to selection, shared history, shared experiences, common geography etc.

5. Generalized Linear Mixed Models

Generalized linear mixed models serve similar purposes to the models already discussed except that the "generalized" label means that new algorithms have been added to support a variety of link functions. Link functions, of course, are transforms of the dependent variable similar to that found, for instance, in binary logistic regression, where what is predicted is not the dependent variable itself (using the identity link function of OLS regression) but instead is the logit (the natural log of the odds that the dependent equals 1) of the dependent variable. Although the predictor side of the equation must be linearly related to the link function of the dependent, the original values of the predictor variables may be nonlinearly related to the original values of the dependent variable. A large number of link functions are possible, only some of which are currently supported by statistical packages for hierarchical linear modeling. Of fundamental importance is that generalized linear mixed modeling supports dependent variables that are not continuous and not normally distributed, as is required by ordinary regression and other general linear model procedures.
Analysis with generalized linear mixed models involves selecting the type of data distribution and link function that corresponds to the nature of the researcher's dependent variable. The commonly used distributions and their corresponding link functions are:

Normal (continuous). This alternative assumes a normal distribution of the dependent variable with an identity link function. The outcome variable at level 1 may have any value on a continuous scale (ex., employee-level performance scores). This option creates the same models as for ordinary linear mixed modeling.

Bernoulli. This alternative assumes a Bernoulli distribution, which is a special case of the binomial distribution, employing a logit link function. In a Bernoulli model, the outcome variable at level 1 (ex., employee-level retirement status) has only two outcomes (ex., not retired = 0, retired = 1).

Binomial (number of trials). This alternative assumes a dependent variable with a binomial distribution and a logit link function, corresponding to binary logistic regression.

Poisson (constant exposure). This alternative assumes a dependent variable reflecting count data (hence non-negative integer values) with a log link function. The "constant exposure" term, also called "equal exposure," means each level 1 subject had the same chance to accumulate the count (ex., the same time interval).

Poisson (variable exposure). An example of this type would be a count of people displaying some trait in multiple cities of differing populations. The "exposure" varies since, all other things equal, larger cities might be expected to have a larger count. Like Poisson-constant exposure models, this alternative also assumes Poisson distribution of count data with a log link function, but the Poisson variance is weighted by the exposure variable.

Multinomial. This alternative assumes a multinomial distribution of the dependent variable, with a generalized logit link function. Multinomial data are categorical, such as "career choice" with values 1 = administrative, 2 = clerical, 3 = other. The coding values are arbitrary. A multinomial model is an extension of the Bernouilli model for dependents with more than two categories.

Ordinal. This alternative also assumes the dependent variable has a categorical distribution, but the categories are ordered—for example, ordered from "strongly agree" to "strongly disagree." The link function is cumulative logit.

Chapter 5

Multilevel Modelling

1. Fixed and Random Effects

Fixed Effects: A variable whose domain represent all possible levels that are of theoretical interest. For instance, Treatment conditions, Gender, Race/Ethnicity etc.

Random Effects: A variable whose domain represents a sample from a larger population of potential values, eg. Individuals (always!!), Schools, Clinics etc.

2. Multilevel Model Building

In ordinary linear regression model, with one predictor variable (X):

$$Y_i = \beta_0 + \beta_1 X_1 + e_i$$

Y_i is composed of two parts, that is, a fixed effects component $(\beta_0 + \beta_1 X_1)$ and a random effects component (e_i).

In an HLM regression model, we can dis-aggregate the fixed effects and show that each of these is in turn composed of fixed and random components.

In two-level hierarchical models, separate level 1 models are developed for each of the J level 2 units. Consider the case of a continuous outcome or dependent variable, Y (for example, patient satisfaction), and a single, continuous level 1 predictor or covariate X (for example, patient's age). The level 1 models are of the form:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \tag{5.1}$$

where Y_{ij} is the dependent variable measured on the ith level 1 unit (for example, patient) nested within the jth level 2 unit (for example, physician), β_{0j} is the intercept for the jth level 2 unit (physician), X_{ij} is the level 1 predictor or covariate (for example, patient age), β_{1j} is the regression coefficient associated with level 1 predictor X for the jth level 2 unit (physician) and e_{ij} is the random error associated with the ith level 1 unit nested within the jth level 2 unit. Interpretation of estimated model parameters (particularly the intercept terms, $\hat{\beta}_{0j}$

) depends upon the way in which the level 1 covariates or predictors are modelled. If the level 1 predictor, X_{ij}, is considered in its original metric, then the intercept, β_{0j}, is the expected value of the dependent variable when X_{ij} is equal to zero. If the level 1 predictor, X_{ij} , is centered about its overall (or grand) mean ($\bar{X}..$), then the intercept, β_{0j}, is interpreted as the mean of the jth level 2 unit adjusted for X (for example, the mean outcome score for physician j adjusted for patient age). If the level 1 predictor is centered at the respective level 2 unit mean, $\bar{X}._j$, called centering at the group mean, then the intercept, β_{0j}, is interpreted as the unadjusted mean of the dependent variable. Finally, if the level 1 predictor is centered at some other meaningful value, then the intercept, β_{0j}, is interpreted as the expected value of the dependent variable when the predictor is equal to that value.

In the case of a dichotomous dependent or outcome variable, Y, and a single, continuous level 1 predictor or covariate, X, the level 1 models are of the form:

$$p_{ij} = \frac{1}{1 + \exp\left\{-\left(\beta_{0j} + \beta_{1j}X_{ij} + e_{ij}\right)\right\}}$$

where $p_{ij} = E\left(Y_{ij}\right)$. However, in this current section we focus on applications concerning continuous dependent variables, Y, which are assumed to be approximately normally distributed at each value of a single, continuous covariate, X, as described in Equation 5.1. Other types of models (those assuming other distributions other than the Normal distribution) are discuss in the section on Generalized Hierarchical Linear Models.

We have J models of the form shown in Equation (5.1), each model potentially having different intercept and slope coefficients (β_{0j} , β_{1j}). In the level 2 models, we consider these regression coefficients (β_{0j} and β_{1j}) as dependent variables and relate each to appropriate level 2 covariates. In the case of a single, continuous level 2 predictor or covariate, M (for example, physician's years in medical practice), the level 2 models are of the form:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}M_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}M_j + u_{1j} \end{aligned} \tag{5.2}$$

where β_{0j} and β_{1j} are the intercept and slope for the jth level 2 unit (physician), γ_{00} and γ_{10} are the overall mean intercept and slope adjusted for M, respectively, M_j is the level 2 predictor or covariate (for example, physician's years in medical practice), γ_{01} and γ_{11} are the regression coefficients associated with the level 2 predictor M relative to the level 2 intercepts and slopes, respectively and u_{0j} and u_{1j} are the random effects of the jth level 2 unit (physician) on the intercept and slope, respectively, adjusted for M. The level 2 predictor M can be modelled in its original metric (as in Equation (5.2)) or centered about its grand mean. Substituting (5.2) into (5.1) yields the combined model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}M_j + \gamma_{10}X_{ij} + \gamma_{11}M_jX_{ij} + u_{1j}X_{ij} + u_{0j} + e_{ij} \tag{5.3}$$

Notice that the combined model involves level 1 and level 2 covariates (X_{ij} and M_j, respectively), a cross-level term (M_jX_{ij}) and a complicated error term, $u_{1j}X_{ij} + u_{0j} + e_{ij}$. Model (5.3) is not of a form in which ordinary least squares (OLS) can be used to estimate parameters, since OLS assumes errors are independent with mean zero and common variance σ^2. In model (5.3), the errors are not independent across level 1 units (patients), instead there is dependency among level 1 units nested within each level 2 unit (physician practice) in the terms u_{0j} and u_{1j}. In addition, the variances of the errors may no longer be homogeneous if u_{0j} and u_{1j} take on different values within each level 2 unit (for example, physician practice). The estimation and hypothesis testing techniques which handle this types of models will be discussed later.

2.1. Assumptions of the model

For models concerning continuous dependent variables (Y_{ij} of (5.1)), we assume that the errors in the level 1 (for example patient-level) models are normal random variables with mean zero and common variance σ^2:

$$E\left(e_{ij}\right) = 0 \text{ and } \operatorname{var}\left(e_{ij}\right) = \sigma^2$$

In the level 2 (for example, physician-level) models we assume that the parameters β_{0j} and β_{1j} are distributed as multivariate normal with means γ_{00} and γ_{10}, respectively, and variances τ_{00} and τ_{11}, respectively. The covariance of β_{0j} and β_{1j} is denoted τ_{01}. For simplicity, we consider the situation in which the errors are homogeneous at both levels 1 and 2, although more complicated error structures are allowed. Finally, level 1 and level 2 errors are uncorrelated. These assumptions are summarized as below:

$$E\left(u_{0j}\right) = 0 \text{ and } E\left(u_{1j}\right) = 0$$

$$E\left(\beta_{0j}\right) = \gamma_{00} \text{ and } E\left(\beta_{1j}\right) = \gamma_{10}$$

$$\operatorname{var}\left(\beta_{0j}\right) = \operatorname{var}\left(u_{0j}\right) = \tau_{00} \text{ and } \operatorname{var}\left(\beta_{1j}\right) = \operatorname{var}\left(u_{1j}\right) = \tau_{11}$$

$$\operatorname{cov}\left(\beta_{01}, \beta_{1j}\right) = \operatorname{cov}\left(u_{0j}, u_{1j}\right) = \tau_{01}$$

$$\operatorname{cov}\left(u_{0j}, e_{ij}\right) = \operatorname{cov}\left(u_{1j}, e_{ij}\right) = 0$$

2.2. To Center or Not to Center

In regression, the intercept is interpreted as the expected value of the outcome variable, when all explanatory variables have the value zero. However, zero may not even be an option in our data (e.g., Gender). This will be especially important when looking at cross-level interactions. General rule of thumb: If you are estimating cross-level interactions, you should grand mean center the explanatory variables. Therefore, the centered version of Models 5.1 and 5.3 will be 5.4 and 5.5 respectively;

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + e_{ij} \tag{5.4}$$

$$Y_{ij} = \gamma_{00} + \gamma_{01} M_j + \gamma_{10}(X_{ij} - \bar{X}..) + \gamma_{11} M_j (X_{ij} - \bar{X}..) + u_{1j}(X_{ij} - \bar{X}..) + u_{0j} + e_{ij} \tag{5.5}$$

This two level model can be extended to three levels as follows:
The first level (individual) model looks a lot like the ordinary regression model:

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} X_{1jk} + e_{ijk}$$

We find that members are nested in groups which are nested in conditions, that is,

We have m persons $(i = 1, 2, ..., m)$
We have g groups $(j = 1, 2, ..., g)$
We have c conditions $(k = 1, 2, ..., c)$

We can specify a second level model that accounts for the clustering of students (i) within schools (j):

$$\beta_{0jk} = \gamma_{00k} + u_{0jk}$$

This model implies that each j-unit has a unique mean that is drawn from an underlying distribution that can be described by a mean (γ_{00k}) and variance (u_{0jk}). We can combine the two levels through simple substitution into a MIXED MODEL:

$$Y_{ijk} = \underbrace{[\gamma_{00k} + \beta_{1jk}X_{1jk}]}_{Fixed\ Effects} + \underbrace{[u_{0jk} + e_{ijk}]}_{Random\ Effects}$$

EXERCISE 11. Write the assumptions in the three level model framework.

2.3. Repeated Measures Model

Time represents repeated observations nested in the same person, that is, the lowest level of observation in this model:

$$Level\,1\,Model\ \ \{Y_{tijk} = \beta_{0ijk} + \beta_{1ijk}X_{1ijk} + e_{tijk}$$

$$Level\,2\ Model\ \{\beta_{0ijk} = \gamma_{00jk} + u_{0ijk}$$

$$Level\,3\ Model\ \{\gamma_{00jk} = \lambda_{000k} + u_{00jk}$$

$$\underbrace{Y_{tijk} = \gamma_{00jk} + \lambda_{000k} + \beta_{1ijk}X_{1ijk} + [u_{00jk} + u_{0ijk} + e_{tijk}]}_{The\ combined\ Model}$$

2.4. Random Slopes and Intercepts

$$Level\ 1\ Model\ \ \{Y_{ijk} = \beta_{0jk} + \beta_{1jk}X_{1jk} + e_{ijk}$$

$$Level\ 2\ Model\ \begin{cases} \beta_{0jk} = \gamma_{00k} + u_{0jk} \\ \\ \beta_{1jk} = \gamma_{10k} + u_{1jk} \end{cases}$$

$$Y_{ijk} = \gamma_{00k} + \gamma_{10k}X_{1jk} + [u_{0jk} + u_{1jk}X_{1jk} + e_{ijk}]$$

Example. Illustration

In the archetypical cross-sectional example, a researcher is interested in predicting test performance as a function of student-level and school-level characteristics. Using the model-building notation, an empty (i.e. lacking predictors) student-level model is specified first:

$$Y_{ij} = \beta_{0j} + r_{ij} \tag{5.6}$$

The outcome variable Y for individual i nested in school j is equal to the average outcome in unit j plus an individual-level error r_{ij} . Because there may also be an effect that is common to all students within the same school, it is necessary to add a school-level error term. This is done by specifying a separate equation for the intercept:

$$\beta_{oj} = \gamma_{00} + u_{oj} \tag{5.7}$$

where γ_{00} is the average outcome for the population and u_{0j} is a school-specific effect. Combining equations 5.6 and 5.7 yields:

$$Y_{ij} = \gamma_{00} + u_{oj} + r_{ij} \tag{5.8}$$

Denoting the variance of r_{ij} as σ^2 and the variance of u_{0j} as τ_{00} , the percentage of observed variation in the dependent variable attributable to school-level characteristics is found by dividing τ_{00} by the total variance:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \tag{5.9}$$

In this case, ρ is referred to as the **intraclass correlation coefficient**. The percentage of variance attributable to student-level traits is easily found according to $1 - \rho$. A researcher who has found a significant variance component for τ_{00} may wish to incorporate macro level variables in an attempt to account for some of this variation.

For example, the average socioeconomic status of students in a district may impact the expected test performance of a school, or average test performance may differ between private and public institutions. These possibilities can be modeled by adding the school-level variables to the intercept equation,

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\left(MEANSES_j\right) + \gamma_{02}\left(SECTOR_j\right) + u_{0j} \tag{5.10}$$

and substituting 5.10 into equation 5.6. MEANSES stands for the average socioeconomic status while SECTOR is the school sector (private or public).

Moreover, the researcher may wish to include student-level covariates. A student's personal socioeconomic status may affect his or her test performance independent of the school's average socioeconomic (SES) score. Thus equation 5.6 would become:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\left(SES_{ij}\right) + r_{ij} \tag{5.11}$$

If the researcher wishes to treat student SES as a random effect (that is, the researcher feels the effect of a student's SES status varies between schools), he can do so by specifying an equation for the slope in the same manner as was previously done with the intercept equation:

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{5.12}$$

Finally, it is possible that the effect of a level-1 variable changes across scores on a level-2 variable. The effect of a student's SES status may be less important in a private rather than a public school, or a student's individual SES status may be more important in schools with higher average SES scores. To test these possibilities, one can add the MEANSES and SECTOR variables to equation 5.12.

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\left(MEANSES_j\right) + \gamma_{12}\left(SECTOR_j\right) + u_{1j} \tag{5.13}$$

A random-intercept and random-slope model including level-2 covariates and cross-level interactions is obtained by substituting equations 5.10 and 5.13 into 5.11:

$$\begin{aligned} Y_{ij} &= \gamma_{00} + \gamma_{01}\left(MEANSES_j\right) + \gamma_{02}\left(SECTOR_j\right) + \gamma_{10}\left(SES_{ij}\right) \\ &\quad + \gamma_{11}\left(MEANSES_j * SES_{ij}\right) + \gamma_{12}\left(SECTOR_j * SES_{ij}\right) \\ &\quad + u_{0j} + u_{1j}\left(SES_{ij}\right) + r_{ij} \end{aligned} \tag{5.14}$$

This approach of building a multilevel model through the specification and combination of different level-1 and level-2 models makes clear the nested structure of the data. However, it is long and messy, and what is more, it is inconsistent with the notation used in much of the documentation for general statistical packages. Instead of the step-by-step approach taken above, the pithier, and more general, matrix notation is often used:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon \tag{5.15}$$

Here $\mathbf{y}$ is an $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times p$ matrix containing the fixed effects regressors, is a $p \times 1$ vector of fixed-effects parameters, $\mathbf{Z}$ is an $n \times q$ matrix of random effects regressors, $\mathbf{u}$ is a $q \times 1$ vector of random effects, and ε is an $n \times 1$ vector of errors. The relationship between equations 5.14 and 5.15 is clearest when, in the step-by-step approach, the fixed effects are grouped together in the first part of the right-hand-side of the equation and the random effects are grouped together in the second part.

$$
\begin{aligned}
Y_{ij} \; = \; & \underbrace{\gamma_{00} + \gamma_{01}\,(MEANSES_j) + \gamma_{02}\,(SECTOR_j) + \gamma_{10}\,(SES_{ij})}_{Fixed\ Effects} \\
& + \underbrace{\gamma_{11}\,(MEANSES_j * SES_{ij}) + \gamma_{12}\,(SECTOR_j * SES_{ij})}_{Fixed\ Effects} \\
& + \underbrace{u_{0j} + u_{1j}\,(SES_{ij})}_{Random\ Effects} + r_{ij} \qquad (5.16)
\end{aligned}
$$

Note that it is possible for a variable to appear as both a fixed effect and a random effect (appearing in both **X** and **Z** from 5.15). In this example, estimating 5.16 would yield both fixed effect and random effect estimates for the student-level SES variable. The fixed effect would refer to the overall expected effect of a student's socioeconomic status on test scores; the random effect gives information on whether or not this effect differs between schools.

3. Impact on the Analysis

The dependence will create a positive intraclass correlation (ICC) that reflects an extra component of variance attributable to the group:

$$ICC_{m:g:c} = corr\,(y_{i:j:k} : y_{i':j:k})$$

The positive ICC reduces the variation among the members of the same group so the within-group variance is:

$$\sigma_e^2 = \sigma_y^2\,(1 - ICC_{m:g:c})$$

And the between group component of variance is:

$$\sigma_{g:c}^2 = \sigma_y^2\,(ICC_{m:g:c})$$

The intraclass correlation is the fraction of the total variation in the data that is attributable to the unit of assignment:

$$ICC_{m:g:c} = \frac{\sigma^2_{g:c}}{\sigma^2_e + \sigma^2_{g:c}}$$

The total variance is the sum of the two components:

$$\sigma^2_y = \sigma^2_e + \sigma^2_{g:c}$$

The Variance Inflation Factor (VIF) is $[1 + (m - 1)ICC]$. VIF is an index of the relative increase in variation found in a HLM analysis as compared to the standard OLS. The variance of the condition mean based on hierarchical data is difference from the variance of the condition mean expected in an trial with individual randomization. When the ICC > 0, VIF increases both as m and as ICC increase.