Multilevel Modeling

Goals

- To introduce the format and structure of multilevel modeling.
- To introduce estimation and inference for multilevel modeling.
- To describe mixed-effects models.

Introduction to Multilevel Data

Recall that one of the assumptions of the standard linear model is that of **independence**. The observations are sampled independently: Any pair of errors ϵ_i and $\epsilon_{i'}$ (or, equivalently, of conditional response-variable values, Y_i and $Y_{i'}$) are independent for $i \neq i'$. The assumption of independence needs to be justified by the procedures of data collection.

The standard linear model and OLS regression are generally inappropriate for dependent observations.

Dependent (or clustered) data arise in many contexts, the two most common of which are hierarchical data and longitudinal data.

Hierarchical data are collected when sampling takes place at two or more levels, one nested within the other. Some examples:

- Students within schools (two levels).

- Students within classrooms within schools (three levels).

- Patients within physicians (two levels).

- Patients within physicians within hospitals (three levels).

There can also be non-nested multi-level data — for example, high school students who each have multiple teachers, but we will not cover this situation.

Longitudinal data are collected when individuals (or other units of observation) are followed over time. Some examples:

- Biannual data on weight-preoccupation and exercise among adolescent girls.

- Data collected at irregular intervals on recovery of IQ among coma patients.

Another example of dependent data is individuals in families, for example, if you sampled mothers' and their daughters or if you sampled couples.

In all of these cases, it is not generally reasonable to assume that observations within the same higher-level unit, or longitudinal observations within the same individual, are independent of one another.

Specifically, there may be unobserved group characteristics that affect the outcomes of the individuals in each group.

Suppose Y_{ij} is the value of the response variable for the ith observation in the jth group or cluster. There are $i = 1, \ldots, n_j$ observations in group j and $j = 1, \ldots, J$ groups or clusters.

Note that it is not required that all the groups have the same number of observations in them, which is why we let the number of observations vary by group j (i.e., n_j).

One way to deal with the dependence is to introduce a group-specific parameter representing the effect of unobserved group characteristics,

$$Y_{ij} = \alpha_j + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \epsilon_{ij}$$

where α_j is a group-specific parameter representing the effect of unobserved group characteristics and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ and are independent of one another. Basically, this is like introducing a dummy variable for each group.

There are some disadvantages of this approach.

1. Depending on the number of groups you have, you could have a large number of dummy variables and therefore, a large number of paramaters.

2. We cannot include group-level covariates among the predictors. This means that we can control for group characteristics, but we cannot estimate their effects.

3. The ordinary least squares estimator is consistent as the number of individuals in each group approaches infinity but is not consistent if the number of groups approaches infinity but the number of individuals per group does not. This may or may not be plausible in any particular study.

Mixed-effect models make it possible to take account of dependencies in hierarchical, longitudinal, and other dependent data and to examine the effects of group-level characteristics.

Unlike the standard linear model, mixed-effect models include more than one source of random variation — i.e., more than one random effect.

Mixed-effects models have been developed in a variety of disciplines, with varying names and terminology: random-effects models, variance component models, hierarchical linear models, multi-level models, random-coefficient models, repeated-measures models.

Let's begin with the simplest random-effect model:

$$Y_{ij} = \alpha_j + \epsilon_{ij}$$

where again ϵ_{ij} are independent error terms and α_j is a *random* variable representing a group-specific effect. This differs from the previous model in that you can think of α_j as another error term. Thus, α_j is an error term at the group level and ϵ_{ij} is an error term at the individual level (i.e., for observation i in group j). Because α_j is random, it has a distribution (just like ϵ_{ij} does). Thus, $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$ and as before, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. Also, α_j and ϵ_{ij} are assumed to be independent of one another. Thus, the errors must be independent across groups but not within a group. In fact, we can compute the correlation between any two observations within the same group as:

$$\rho = \text{cor}(Y_{ij}, Y_{i'j}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$$

This is usually called the intraclass correlation coefficient.

We have two variance components - σ_α^2 is the variation across groups (or between-groups) and σ_ϵ^2 is the variation within groups.

There are two further advantages of the random-effect model over the fixed effect model.

1. Random-effect models can be extended to more than two-levels (unlike fixed effect models).

2. In the random-effect model, we can include level 1 and level 2 predictors and let any of the level 1 coefficients vary across groups.

Longitudinal Data

In most respects, modeling longitudinal data — where there are multiple observations on individuals who are followed over time — is similar to modeling hierarchical data.

We can think of individuals as analogous to level-2 units, and measurement occasions as analogous to level-1 units.

Just as it is generally unreasonable to suppose in hierarchical data that observations for individuals in the same level-2 unit are independent, so is it generally unreasonable to suppose that observations taken on different occasions for the same individual are independent.

An additional complication in longitudinal data is that it may no longer be reasonable to assume that the "level-1" errors ϵ_{ij} are independent, since observations taken close in time on the same individual may be more similar than observations farther apart in time.

When this happens, we say that the errors are autocorrelated. The linear mixed model makes provision for autocorrelated errors.

Estimation and Inference

Estimation

There are primarily two methods to estimate mixed-effects models:

Full maximum-likelihood (ML) estimation of the model maximizes the likelihood with respect to all of the parameters of the model simultaneously (i.e., both the fixed-effects parameters and the variance components).

Restricted (or residual) maximum-likelihood (REML) estimation integrates the fixed effects out of the likelihood and estimates the variance components; given the estimates of the variance components, estimates of the fixed effects are recovered.

A disadvantage of ML estimates of variance components is that they are biased downwards in finite samples (much as the ML estimate of the error variance in the standard linear model is biased downwards). The REML estimates, in contrast, correct for loss of degrees of freedom due to estimating the fixed effects.

The difference between the ML and REML estimates can be important when the number of "clusters" (i.e., level-2 units) in the data is small.

A common source of estimation difficulties in mixed models is the specification of overly complex random effects. Interest usually centers on the fixed effects, and it often pays to try to simplify the random-effect part of the model.

Statistical Inference

The usual manner of obtaining standard errors for variance and covariance components tend not to be accurate so confidence intervals and p-values based on them are also inaccurate.

We can, however, test variance and covariance components by a likelihood-ratio test, contrasting the (restricted) log-likelihood for the fitted model with that for a model removing the random effects in question.

Cautionary Remarks:

Because REML estimates are calculated integrating out the fixed effects, one cannot legitimately perform likelihood-ratio tests across models with different fixed effects when the models are estimated by REML.

Likelihood-ratio tests for variance-covariance components across nested models with identical fixed effects are perfectly fine.

The null hypothesis for the likelihood-ratio test of a variance sets the variance to 0, which is on the boundary of the parameter space; the p-value should be adjusted for this constraint.

There are two options for obtaining p-values for the fixed effects: the Satterthwaite approximation and the Kenward-Roger approximation.