Module 5: Introduction to Multilevel Modelling Concepts

Fiona Steele
Centre for Multilevel Modelling

Pre-requisites

- Modules 1-4

Contents

All of the sections within this module have online quizzes for you to test your understanding. To find the quizzes:

EXAMPLE
From within the LEMMA learning environment

- Go down to the section for **Module 5: Introduction to Multilevel Modelling**
- Click " 5.1 Comparing Groups Using Multilevel Modelling" to open Lesson 5.1
- Click Q 1 to open the first question

All of the sections within this module have practicals so you can learn how to perform this kind of analysis in MLwiN or other software packages. To find the practicals:

From within the LEMMA learning environment

- Go down to the section for Module 5: Introduction to Multilevel Modelling, then

Either

- Click "5.1 Comparing Groups Using Multilevel Modelling" to open Lesson 5.1
- Click MLwiN practical

Or

- Click Print all Module 5 MLwiN Practicals

What is multilevel modelling?

In the social, medical and biological sciences multilevel or hierarchical structures are the norm. Examples include individuals nested within geographical areas or institutions (e.g. schools or employers), and repeated observations over time on individuals. Other examples of hierarchical and non-hierarchical structures were given in Module 4. When individuals form groups or clusters, we might expect that two randomly selected individuals from the same group will tend to be more alike than two individuals selected from different groups. For example, children learn in classes and features of their class, such as teacher characteristics and the ability of other children in the class, are likely to influence a child’s educational attainment. Because of these class effects, we would expect test scores for children in the same class to be more alike than scores for children from different classes. By a similar argument, measurements taken on the same individual at different occasions, e.g. physical attributes or social attitudes, will tend to more highly correlated than two measurements from different individuals. Such dependencies can therefore be expected to arise and we need multilevel models – also known as **hierarchical linear models**, **mixed models**, **random effects models** and **variance components models** - to analyse data with a hierarchical structure. Throughout this module we refer to the lowest level of observation in the hierarchy (e.g. student or measurement on a given occasion) as level 1, and the group or cluster (e.g. school or individual) as level 2.

One assumption of the single-level multiple regression model is that the measured units are independent (see Module 3). Specifically, we assume that the residuals e_i are uncorrelated with one another. If data are grouped and we have not taken account of group effects in our regression model, the independence assumption will not hold. One way to allow for grouping is to include a set of dummy variables for groups as explanatory variables in the model. For example, in C3.2.2 we allowed for between-country differences in levels of hedonism by including two dummy variables for Germany and France (treating the UK as the reference group). A model with dummy variables for groups is called a fixed effects model but, for reasons summarised in C3.2.3, there are problems with adopting this approach when the number of groups is large. An alternative strategy to allow for group effects is to include in the model explanatory variables that measure group characteristics that are believed to influence individual outcomes. We might, for example, collect data on teachers' experience and their teaching methods. In practice, however, the processes which lead to clustering are complex and important sources of group effects are likely to be unmeasured and therefore not fully accounted for by including group-level variables. So, on its own, this approach is not enough to allow for clustering.

What are the implications of ignoring clustering? Suppose we are interested in the predictors of children's educational attainment and, in particular, whether there are inequalities by gender and ethnicity. We are not concerned with differences among schools and therefore fit a multiple regression model with gender, ethnic group and some family background measures as explanatory variables. If attainment is clustered by school, however, and this is not taken into account in the analysis, the standard errors of the regression coefficients will generally be underestimated (see C5.2.4 for a non-technical explanation for this). Consequently confidence intervals will be too narrow and p-values will be too small, which may in turn lead us to infer that a predictor has a 'real' effect on the outcome when in fact the effect could be ascribed to chance. Underestimation of standard errors is particularly severe for coefficients of predictors that are measured at the group level, e.g. an indicator of whether a school is mixed or single sex. Correct standard errors will be estimated only if variation among groups is allowed for in the analysis, and multilevel modelling provides an efficient means of doing this.

Obtaining correct standard errors is just one reason for using multilevel modelling. If you are interested only in controlling for clustering, rather than exploring it, there are other methods that can be used. For example, survey methodologists have long recognised the consequences of ignoring clustering in the analysis of data from multistage designs and have developed methods to adjust standard

errors for *design effects*. Another approach is to model the dependency between observations in the same group explicitly using a *marginal model*. Both methods yield correct standard errors, but treat clustering as a nuisance rather than something of substantive interest in its own right. Multilevel modelling enables researchers to investigate the nature of between-group variability, and the effects of group-level characteristics on individual outcomes. Some examples of research questions that can be explored using multilevel models are given below:

- Is there between-school variability in students' academic progress? Does the strength of the relationship between prior attainment and subsequent performance vary across schools? The first question is concerned with overall differences in school effectiveness while the second asks whether some schools are more effective for certain types of students, e.g. those with low or high ability. A school in which the mean attainment at age 16 depends little on a student's intake score may be said to show greater equity because it has decreased differences in outcomes across its intake spectrum.

- Do health outcomes vary across areas? Are between-area variations in health explained by differences in access to health services? Is the amount of variation between areas different for rural and urban areas?

- Does the rate of physical growth vary across children? Does variability in the growth rate differ for boys and girls?

Note that the data structures in the above examples are all hierarchical, that is each level 1 unit belongs to a single level 2 unit. More generally, structures can be non-hierarchical. Module 4 gave examples of cross-classified and multiple membership structures. In this module, we consider only models for hierarchical structures. We also restrict the discussion to models for continuous (normal) responses. Multilevel models for non-hierarchical structures and non-normal responses will be described in subsequent modules.

Table 5.1 summarises the alternative approaches that might be considered when analysing a dataset with hierarchical structure in which we anticipate some dependency.

Table 5.1. Alternative analysis strategies for hierarchical data

Strategy	Consequences
Fit a single-level model and ignore structure	Substantively you would not measure the importance of context. Technically, your standard errors would be too small, leading to incorrect inferences (concluding that effects that might be ascribed to chance are 'real', i.e. a high risk of Type I error).
Include a set of dummy variables for groups (a fixed effects model)	Group is treated as a fixed classification, so the target of inference is restricted to the groups represented in the sample. If the number of groups is large, there will be a large number of additional parameters to estimate. The effects of group-level predictors cannot be estimated simultaneously with group residuals.
Fit a single-level model with group-level predictors	High risk of Type I errors because standard errors of coefficients of group-level predictors may be severely underestimated. No estimate of the between-group variance that remains unaccounted for by the included group-level predictors.
Correcting standard errors for design effects, or fitting a marginal model in which the dependency is modelled directly	The standard errors will be correct (properly adjusted for clustering), but unable to assess the degree of between-group variation.
Multilevel modelling (random effects)	Correct standard errors and an estimate of between-group variance.

Introduction to the example dataset

The ideas of multilevel modelling will be introduced using data from the 2002 European Social Surveys (ESS). Measures of ten human values have been constructed for 20 countries in the European Union. According to value theory, values are defined as desirable, trans-situational goals that serve as guiding principles in people's lives. Further details on value theory and how it is operationalised in the ESS can be found on the ESS education net (http://essedunet.nsd.uib.no/cms/topics/1/).

We will study one of the ten values, *hedonism*, defined as the 'pleasure and sensuous gratification for oneself'. The measure we use is based on the extent to which respondents identify themselves with a person with the following descriptions:

- *He (sic) seeks every chance he can to have fun. It is important to him to do things that give him pleasure.*

- *Having a good time is important to him. He likes to "spoil" himself.*

Higher scores on the hedonism variable indicate more hedonistic beliefs.

Data for three countries – France, Germany and the UK – were analysed in Module 3 to illustrate multiple regression. Here, we analyse data from all 20 countries in the study. The combined sample size for these countries is 36,527. The data have a two-level hierarchical structure with individual respondents at level 1 and countries at level 2. We will treat country as a random classification.

In the following analyses, we investigate between-country variation in hedonism using different types of two-level model. We consider four explanatory variables:

- *Respondent's age in years*
- *Respondent's gender*
- *Respondent's monthly household income in bands* (less than 150 Euros, 150-300, 300-500, 500-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-5000, 5000-7500, 7500-10000, 10000+)
- *Country-level income* (the mean income band in a country); this is a level 2 variable.

The following countries were included in the study: Austria, Belgium, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. The target of inference could be a wider population of countries from which those in the study can be considered a random sample. However, it is not clear which countries such a population would contain. In this case, it is more natural to think of the sample data as if they were a set of realisations from some underlying process that could extend through time and possibly space[1]. This process has driven the observations, but the statistics we compute from the observed data refer to a particular point in time and are subject to random fluctuations. We are interested in the underlying process that has generated the data we observe, and use the 'sample' data to make inferences about this process.

[1] In survey sampling this abstract notion of a target population is called a *superpopulation*. A superpopulation is infinite, while a population consisting of a fixed number of countries (e.g. all European countries) is finite.

C5.1 Comparing Groups using Multilevel Modelling

C5.1.1 A multilevel model for group effects

A single-level regression model for the mean

Before introducing multilevel models, let's consider the simplest possible regression model: a model for the mean of the dependent variable y with no explanatory variables. Such a null or empty model may be written

$$y_i = \beta_0 + e_i \qquad (5.1)$$

where y_i is the value of y for the ith individual ($i = 1,\ldots,n$), β_0 is the mean of y in the population, and e_i is the 'residual' for the ith individual, i.e. the difference between an individual's y value and the population mean. *Figure 5.1* shows the residuals for four observations ($n = 4$). We usually assume that the residuals follow a normal distribution with mean zero and variance σ^2, i.e. $e_i \sim N(0,\sigma^2)$. The variance summarises the variability around the mean; if this is zero all the points would have the same y-value and would therefore lie on the $y = \beta_0$ line. The larger the variance, the greater the departures about the mean.

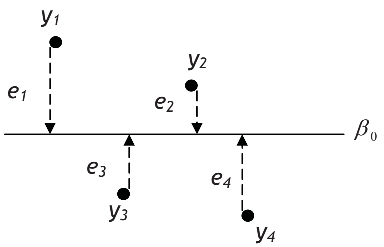

Figure 5.1. Residuals for four data points in a single-level model for the mean

A multilevel model for group means

Now let's move to the simplest form of a multilevel model, which allows for group differences in the mean of y. We now view the data as having a two-level structure with individuals at level 1, nested within groups at level 2. To indicate the group that individual i belongs to, we add a second subscript j so that y_{ij} is the value of y for the ith individual in the jth group. Suppose there are a total

of J groups with n_j individuals in the jth group, and that the total sample size is $n = n_1 + n_2 + \ldots + n_J$.

In a two-level model we split the residual into two components, corresponding to the two levels in the data structure. We denote the group-level residuals, also called **group random effects**, by u_j and the individual residuals by e_{ij}. The two-level extension of (5.1) which allows for group effects is given by

$$y_{ij} = \beta_0 + u_j + e_{ij} \qquad (5.2)$$

where β_0 is the overall mean of y (across all groups).

The mean of y for group j is $\beta_0 + u_j$, and so the group-level residual u_j is the difference between group j's mean and the overall mean. The individual-level residual e_{ij} is the difference between the y-value for the ith individual and that individual's group mean, i.e. $e_{ij} = y_{ij} - (\beta_0 + u_j)$. *Figure 5.2* shows y-values for eight individuals in two groups, with individuals in group 2 denoted by black circles and those in group 1 denoted by grey squares. The overall mean is represented by the solid line and the means for groups 1 and 2 are shown as dashed lines. Also shown are the group residuals and the individual residual for the 4th individual in the 2nd group (e_{42}). Group 1 has a below-average mean (negative u_j), while group 2 is above average (positive u_j).

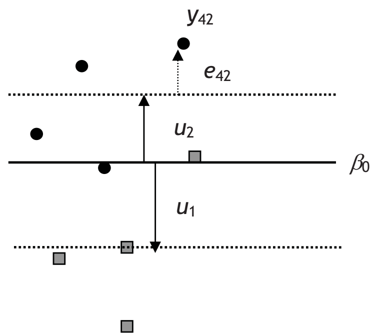

Figure 5.2. Individual and group residuals in a two-level model for the mean

Residuals at both levels are assumed to follow normal distributions with zero means: $u_j \sim N(0,\sigma_u^2)$ and $e_{ij} \sim N(0,\sigma_e^2)$. The total variance is therefore partitioned into two components: the between-group variance σ_u^2, based on departures of group means from the overall mean, and the within-group between-individual