

# Multilevel Modeling in Plain Language

Karen Robson & David Pevalin



Los Angeles | London | New Delhi  
Singapore | Washington DC

# ONE

## What Is Multilevel Modeling and Why Should I Use It?

### CHAPTER CONTENTS

<b>Mixing levels of analysis</b>	<b>4</b>
<b>Theoretical reasons for multilevel modeling</b>	<b>6</b>
<b>What are the advantages of using multilevel models?</b>	<b>7</b>
<b>Statistical reasons for multilevel modeling</b>	<b>7</b>
<b>Assumptions of OLS</b>	<b>8</b>
Dependence among observations	8
Group estimates	12
Varying effects across contexts	13
Degrees of freedom and statistical significance	16
<b>Software</b>	<b>17</b>
<b>How this book is organized</b>	<b>19</b>

You are probably reading this book because someone – a professor, a supervisor, a colleague, or even an anonymous reviewer – told you that you needed to use multilevel modeling. It sounds pretty impressive. It is perhaps even more impressive that multilevel modeling is known by several other names, including, but not limited to: hierarchical modeling, random coefficients models, mixed models, random effects models, nested models, variance component models, split-plot designs, hierarchical linear modeling, Bayesian hierarchical linear modeling, and random parameter models. It can seem confusing, but it doesn't need to be.

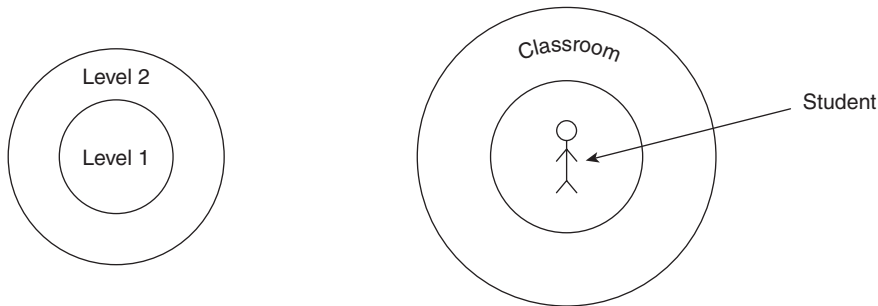
This book is for a special type of user who is far more common than experts tend to recognize, or at least acknowledge. This book is for people who want to learn about this technique but are not all that interested in learning all the statistical equations and strange notations that are typically associated with teaching materials in this area. That is not to say we are flagrantly trying to promote bad research, because we are not. We are trying to demystify these types of approaches for people who are intimidated by technical language and mathematical symbols.

Have you ever been in a lecture or course on a statistical topic and felt you understand everything quite well until the instructor starts putting equations on the slides, and talking through them as though everyone understands them? ‘As this clearly shows ... theta ... gamma ....’. Have you ever been in a course where several slides of equations are used to justify and demonstrate a procedure and you just didn’t understand? Perhaps you were thinking, ‘These equations must mean something in words. Why can’t they just use the words?’ You might also just want some practical examples that are fully explained in plain language that you may be able to apply to your own research questions. If this sounds like you, then this book is for you. If you are really fond of equations, then we’re afraid that our approach in this book won’t appeal to you.

Before we get started, we also want to emphasize that this book is not about ‘dumbing down’ complicated subject matter; it’s about making it accessible. We are not endorsing using modeling techniques without understanding them. This leads to sloppy and unscientific analyses that are painful to read. What this book does is unpack these sophisticated techniques and explain them in non-technical language. We assume that you understand the principles of hypothesis testing, sampling, research design, and statistical analysis up to and including ordinary least squares regression (OLS) with interaction terms. The techniques discussed here are just extensions of regression. Really! We do try to keep the jargon to a minimum but, as with all things new, there are some new terms and phrases to get to grips with.

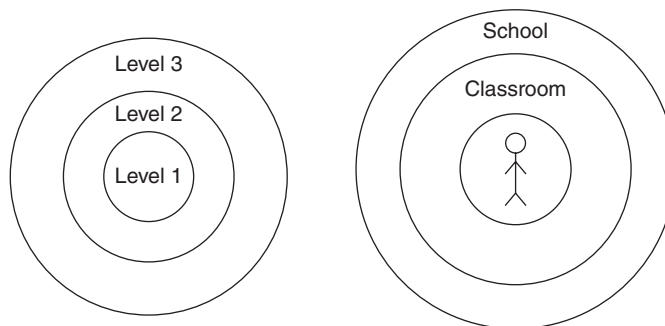
So, why might you have been told that you need to use multilevel modeling? The chances are that it is because your data have a ‘nested’, ‘clustered’, or ‘grouped’ structure and you are being guided to a (regression) technique that accounts for this type of data structure.

The idea of nesting, or clustering, is central to multilevel modeling. It simply means that smaller levels of analysis are contained within larger grouping units (Figure 1.1). The classic example is that individual students are nested within schools, but nesting can take other forms, such as individuals (Level 1) within cities (Level 2), patients within hospitals, siblings within families, or employees within firms.



**Figure 1.1** Two-level nesting

In these types of two-level models, the lower level consisting of the smaller units (often individuals) is called Level 1, and the next level is referred to as Level 2. You may also have a third (or higher) level within which Level 2 units are nested (see Figure 1.2) such as students (Level 1) within classrooms (Level 2) within schools (Level 3), patients within hospitals within districts, siblings within families within neighbourhoods, or employees within firms within nations.

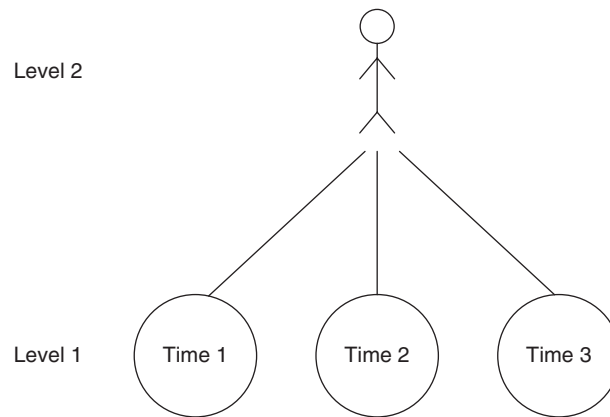


**Figure 1.2** Three-level nesting

Multilevel modeling can deal with three or more levels of nesting. However, in this book we will focus on two-level models. As you gain more expertise in multilevel modeling, you may want to explore more complex structures, but for the scope of this introductory book, analysis with two levels will serve as the foundation upon which illustrative examples are created.

This book will also only consider cross-sectional nested data. There is another variation of nesting with longitudinal data. Longitudinal data are obtained when information is collected from respondents at more than one point in time. For example, people are interviewed annually in a longitudinal survey such as the

British Household Panel Survey. The way nesting is conceptualized with longitudinal data is a bit different than with cross-sectional data. Data collection events (Time 1, Time 2, Time 3, etc.) are nested within individual respondents. Therefore, the time is Level 1 and the respondent is Level 2 (Figure 1.3). If your data look like this then you may still start with this book as an introduction to cross-sectional multilevel models before branching out into longitudinal data. On the positive side, there are a number of other potential techniques for analysing this type of longitudinal or panel data. Halaby (2004), for example, offers a sociologically based primer for examining issues of causality in longitudinal data. Fitzmaurice et al. (2012) have written a more detailed textbook on the topic of applied longitudinal data analysis that is targeted at a multidisciplinary audience.

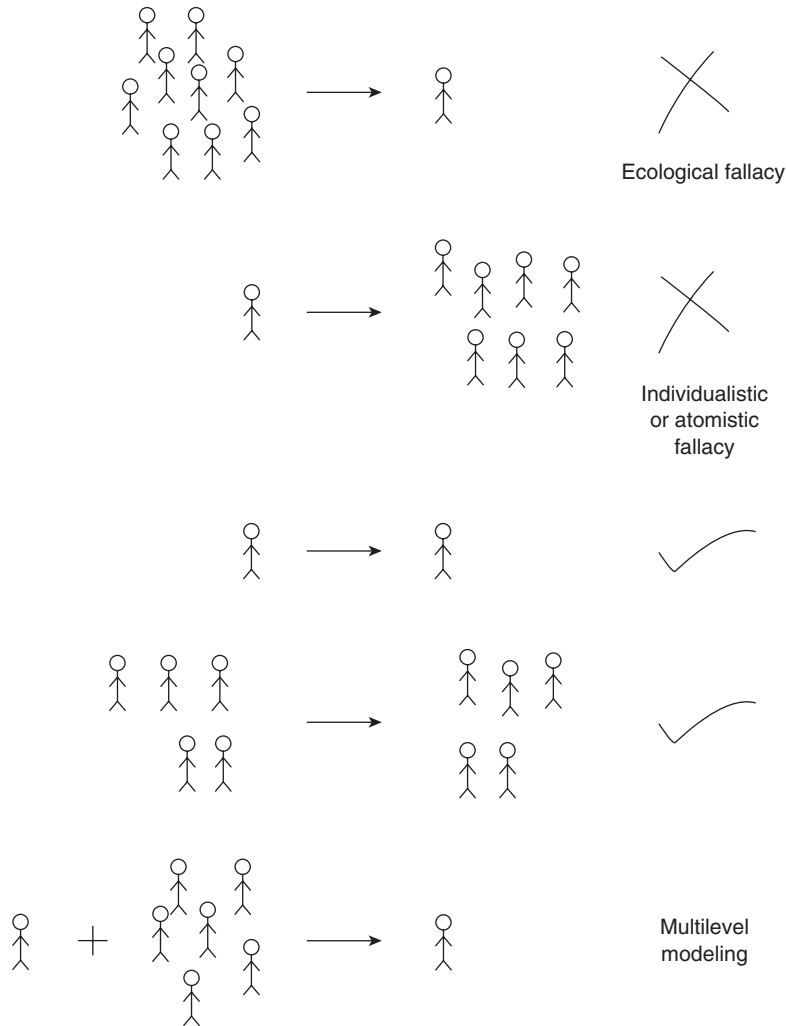


**Figure 1.3** Longitudinal nesting

## MIXING LEVELS OF ANALYSIS

There are two errors in causal reasoning that have to do with mixing different levels of analyses, which are illustrated in Figure 1.4. The first is known as the *ecological fallacy* and has to do with generalizing group characteristics to individuals. If we analyse the effect that average neighbourhood income has on the crime rates of that neighbourhood, we are comparing group characteristics with group characteristics. To extend this argument to *particular individuals* in the neighbourhood can be misleading. It is not appropriate to apply group-level characteristics to individual-level inferences. We may well find that as average income in a neighbourhood decreases, the crime rate increases – but we cannot say that if an individual's income decreases, he or she is more likely to commit crime!

The ecological fallacy can be demonstrated in a number of ways. Another common misinterpretation of group characteristics is to look at the average income in two very different communities that are about the same size. In Wealthyville,



**Figure 1.4** Units of analysis and making generalizations

the average household income is \$500,000 per year. In Poorville, the average household income is only \$15,000 per year. If 10,000 households live in each community, we would say that the average household income across both communities is \$257,500 per year. This would give a completely inaccurate representation of the communities, however, because it doesn't represent the household income of *anyone*. It is far too little to represent Wealthyville (just about half the actual household income) and too high to represent Poorville (over 17 times the actual household income). By taking group characteristics and trying to generalize to individual households, we have committed the ecological fallacy.

Keeping your units of analysis comparable also applies to arguments made in the opposite direction – generalizing individual processes to group processes.

This problem is known as the *atomistic fallacy* (or individualist fallacy). People make this mistake when they take results from individual-level data and apply them to groups, where the context may be very different. We may find, for example, that being an immigrant is associated with an increased risk of mental health problems. A policy solution, however, of creating mental health programmes for all immigrants may be misguided, if contextual variables (at the group level) are not taken into account. It may be that immigrants in large cities have better mental health than immigrants in small communities (where they may be isolated) (Courgeau, 2003). If we simply take individual-level characteristics and apply them to groups, failing to take contexts into consideration, we may come to conclusions based upon flawed logic.

Both the ecological and atomistic fallacies are errors that researchers make when they take data at one level and try to make generalizations to another level. As social scientists, we know that individual characteristics (e.g. age, gender, race) and contextual-level variables (e.g. school, neighbourhood, region) are important determinants for many different outcomes of interest. In multilevel modeling, we use both individual and group characteristics and our outcomes can be modeled in ways that illustrate how individual and group characteristics both affect outcomes of interest, and how group characteristics may influence how individual characteristics affect the outcome of interest, given certain contexts.

## THEORETICAL REASONS FOR MULTILEVEL MODELING

Your models should always be theory-driven, and the best model choice is one that corresponds to a sound theoretical rationale. One that is often overlooked is the general theoretical arguments around how the social world is portrayed. Education researchers, such as Bronfenbrenner (1977, 2001), have argued that the outcomes of individuals, particularly children, cannot be understood without taking different contexts into perspective. His ecological systems approach identifies a number of different contexts to be taken into consideration in terms of how they work independently and together to create the environments in which children live. By looking at data collected from individuals, we are focusing on the micro-level (i.e. individual) effects of specific characteristics on outcomes of interest, but it is more likely the case that these micro-level effects vary significantly across larger units at the meso (school or community) and macro (municipal or national) levels. The micro and the macro (and the micro and meso) interact. This theoretical perspective is most readily tested with statistical techniques that recognize these important distinctions.

Although discussions in this vein invariably resort to examples from education research, the applications and theoretical motivations apply across a range

of disciplines, including health, political science, criminology, sociology, and management research. Scholars from all these disciplines have noted the importance of linking the individual (the micro) and the contexts in which he or she lives (the macro). The popularity of theories that focus exclusively on the individual or solely on higher levels (groups, firms, nations) is being overshadowed by approaches that try to mix the two and presumably give a more accurate depiction of the complexity of the social world.

## **WHAT ARE THE ADVANTAGES OF USING MULTILEVEL MODELS?**

Well, as the name implies, multilevel models are equipped to analyse multiple levels of data. The information about individual and group characteristics is retained and separate estimates are produced for both. Adjustments are made for correlated error terms and for the different degrees of freedom (i.e. Level 1 degrees of freedom and Level 2 degrees of freedom). Perhaps most importantly, we are also able to do cross-level interactions so that we can explain how Level 1 predictors affect our dependent variable according to different contexts (Level 2 predictors). Additionally, we can look at how Level 1 predictors interact with each other and how Level 2 predictors interact with each other. And because multilevel modeling is just an extension of OLS, much of the knowledge you already have about OLS will come in handy for model-building and the interpretation of the results.

We now show that using regular single-level regression techniques to address multilevel issues is fraught with problems. The main problem is that your results are likely to be filled with errors that originate from various violations of the regression assumptions. In many cases, we get poorly estimated results that will be statistically significant, leading us to reject our null hypotheses when we really shouldn't be rejecting them. In other words, finding associations where none exist, erroneous conclusions, and possibly leading to ineffective policies.

## **STATISTICAL REASONS FOR MULTILEVEL MODELING**

There are many statistical reasons to choose multilevel modeling techniques over standard linear OLS regression. You may have accepted that you just need to learn this and don't really care about all the technical reasons why, but we would argue that you should at least grasp the basic reasons why OLS is deficient in estimating models with nested data. You may have already tried some of the 'workarounds',



which we discuss below, in OLS to model nested data. These ‘workarounds’ have been, and continue to be, used by many researchers and it is not difficult to find examples of them in the literature. They are still technically flawed, however, and we explain below why it is problematic to choose OLS, despite these ‘workarounds’, when trying to deal with nested data structures.

## ASSUMPTIONS OF OLS

Multilevel modeling is an extension of OLS. The thing that makes multilevel modeling special is that it addresses specific violations of the ‘assumptions’ of OLS. Remember the assumptions? These are the conditions under which OLS regression works properly. All statistics have a set of assumptions under which they perform as they were intended.

One assumption is that the relationship between the independent and dependent variable is linear. Another is that the sample used in the analysis is representative of its population. Yet another is that the independent variables are measured without error. We more or less follow these assumptions in our day-to-day usage of OLS – and we should check that we are meeting some of these assumptions by doing regression diagnostics.

### Dependence among observations

There are assumptions that relate to ‘independence of observations’ that we might think about less often. But it is this particular violation of the assumptions that multilevel modeling techniques are best suited to address. The assumption of independence means that cases in our data should be independent of one another, but if we have people clustered into groups, their group membership will likely make them similar to each other. Once people (Level 1) start having similar characteristics based on a group membership (Level 2), then the assumption of independence is violated. If you violate it, *you get incorrect estimations of the standard errors*. This isn’t just a niggly pedantic point. If you violate the assumptions, you are more likely to wrongly achieve statistical significance and make conclusions that are simply incorrect.

Perhaps an overlooked common-sense fact is that even if you don’t really care about group-level factors in your analysis (i.e. they aren’t part of your hypotheses), ignoring them does not make the problem go away. This is easy to demonstrate.

Suppose we are interested in how gender and parental occupational status influence academic achievement. Table 1.1 presents results from an OLS regression of reading scores on gender and parental occupational status. Reading scores are standardized within the sample to a mean of 0 and a standard deviation of 1.

Gender is a dummy variable with 1 denoting males. Parental occupational status is a 64-category ordinal scale (the lowest value presenting low occupational status) that is treated as an interval-level variable here. The data come from the Australian sample from the 2006 Programme for International Student Assessment (PISA) organized by the Organisation for Economic Co-operation and Development (OECD, 2009). Data were obtained when all children were 15 and 16 years of age from schools in all eight states and territories of Australia.

**Table 1.1** OLS regression of standardized reading scores on gender and parental occupational status ( $N = 13,646$ )

Variables	<i>b</i>	s.e.
Male	-0.381***	(0.016)
Parental occupational status	0.020***	(0.001)
Constant	-0.859***	(0.028)
$R^2$	0.141	

*b* – unstandardized regression coefficients; s.e. – standard errors

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

Table 1.1 indicates that males (compared to females) have lower reading scores by 0.381 and that each unit increase in parental occupational status is associated with increases in reading scores of 0.020. These results are both statistically significant and have small standard errors. Our model has no group-level indicators, such as class or school. Just because we haven't included group-level indicators, it does not mean that our problems of dependence among observations and thus correlated errors do not exist.

First, we need to predict our residuals from the regression equation whose coefficients we have just identified. Remember that the residuals are the difference between our predicted reading score based on these characteristics and the actual reading score we see in the data. Next, we can test if the assumption is violated if we run an analysis of variance (ANOVA) of these residuals by a grouping factor. Our grouping factor here is the region of Australia – the state or territory. If the residuals are independent of the regions, that is great – that means our errors are randomly distributed. This is not, however, the case in our data as the ANOVA returns a result of  $F = 55.8$ ,  $df = 7$ ,  $p < 0.001$ .

It might be helpful to think of it this way: we have several thousand students within the eight different regions in Australia. The ANOVA tells us that our individual-level results violate that assumption of uncorrelated errors because we find that the ANOVA gives statistically significant results. Table 1.2 shows the mean reading scores by region.

**Table 1.2** Means and standard deviations of standardized reading scores by region

<i>Region</i>	<i>Mean</i>	<i>Std dev.</i>
Australian Capital Territory	0.269	0.946
New South Wales	0.090	0.980
Victoria	-0.610	1.280
Queensland	-0.039	0.997
South Australia	0.070	0.915
Western Australia	-0.148	1.005
Tasmania	-0.075	0.937
Northern Territory	0.140	0.954

We could assume that a 'fix' to this would be to add dummy variables to the model that represent the different regions. We add dummy variables to a model so that we can include nominal-level variables in our estimation. As the regions are nominal, we can then add them as a set of dummies with an omitted reference category. Table 1.3 shows how many students are in each region in Australia, while Table 1.4 displays the regression results for the model including the region dummy variables.

**Table 1.3** Australian students by region

<i>Region</i>	<i>N</i>	<i>%</i>
Australian Capital Territory	954	7.0
New South Wales	3,270	24.0
Victoria	702	5.1
Queensland	2,322	17.0
South Australia	1,548	11.3
Western Australia	1,221	9.0
Tasmania	2,183	16.0
Northern Territory	1,446	10.6
Total	13,646	100

As you can see in Table 1.3, students from this sample are dispersed among the eight regions of Australia. From reviewing the literature, we may have reason to believe that regional effects are determinants of academic achievement in Australia, as they are in many other countries around the world. For example, educational policies and resources are governed at the state level in Australia, and those regions containing the largest cities may offer the best resources for students (Australian Government, no date).

In Table 1.4 the gender and parental occupational status variables are the same as in Table 1.1. The region variable is entered as seven dummy variables, with the Australian Capital Territory as the omitted category.

**Table 1.4** OLS regression of standardized reading scores on region, gender and parental occupational status ( $N = 13,646$ )

<i>Variables</i>	<i>b</i>	<i>s.e.</i>
Region <sup>a</sup>		
<i>New South Wales</i>	-0.075*	(0.034)
<i>Victoria</i>	-0.747***	(0.046)
<i>Queensland</i>	-0.168***	(0.035)
<i>South Australia</i>	-0.089*	(0.038)
<i>Western Australia</i>	-0.222***	(0.040)
<i>Tasmania</i>	-0.189***	(0.036)
<i>Northern Territory</i>	-0.026	(0.038)
Male	-0.377***	(0.016)
Parental occupational status	0.019***	(0.000)
Constant	-0.687***	(0.042)
$R^2$	0.165	

<sup>a</sup>Reference category is Australian Capital Territory

*b* – unstandardized regression coefficients; *s.e.* – standard errors

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Now is a good time to review the interpretation of coefficients as this will be important for understanding multilevel model outputs as well. The unstandardized coefficients (all in their own units of measurement) in Table 1.4 would be interpreted as:

- Compared to being in Australian Capital Territory, being in New South Wales is associated with a decrease in standardized reading scores by 0.075, controlling for the other variables in the model.
- Compared to being in Australian Capital Territory, being in Victoria is associated with a decrease in reading scores by 0.747, controlling for the other variables in the model.
- Compared to being in Australian Capital Territory, being in Queensland is associated with a 0.168 decrease in reading scores, controlling for the other variables in the model.
- Compared to being in Australian Capital Territory, being in South Australia is associated with a 0.089 decrease in reading scores, controlling for the other variables in the model.
- Compared to being in Australian Capital Territory, being in Western Australia is associated with a 0.222 decrease in reading scores, controlling for the other variables in the model.
- Compared to being in Australian Capital Territory, being in Tasmania is associated with a 0.189 decrease in reading scores, controlling for the other variables in the model.

- The coefficient for the Northern Territory is not statistically significant.
- Compared to being a female, being a male is associated with a 0.377 decrease in reading scores, controlling for the other variables in the model.
- Each unit increase in the parental occupational status scale is associated with a 0.019 increase in reading scores, controlling for all the other variables in the model.
- The constant is the reading score of  $-0.687$ , which is the value when all the independent variables have a value of zero. In this case it would be a female living in Australian Capital Territory whose parents have an occupational status score of zero – which isn't possible with this particular occupational status measure.

Clearly there are 'region' effects here – the Australian Capital Territory (ACT) seems to be the best place for reading scores. The problem with this type of analysis is that children are nested within the regions. Furthermore, they are nested within schools, and even classrooms. While our analysis is at the individual level, the observations aren't completely independent in the sense that there are eight regions into which pupils are divided.

An OLS regression assumes that the coefficients presented in the table are 'independent' of the effects of the other variables in the model, but, in this type of model, this assumption is false. If we believe that regions – an overarching structure – affect students differentially, the effect that regions have on reading scores is not independent of the effects of the other variables in the model. The structures themselves share similarities that we cannot observe but, nevertheless, influence our results. It is probably the case that parents' occupational status and the gender of the student are also differentially associated with reading scores, depending on the region in which the student attends school. The OLS assumption of independent residuals is probably violated as well. It is likely that the reading scores within each region may not be independent, and this could lead to residuals that are not independent within regions. Thus, the residuals are correlated with our variables that define structure. We need statistical techniques that can handle this kind of data structure. OLS is not designed to do this.

### Group estimates

Maybe at this point you are wondering, 'If groups are so important, maybe I should just focus on group-level effects.' You may think that a possible solution to these problems is just to conduct analyses at the group level. In other words, to avoid the problem of giving group characteristics to individuals, just aggregate the data set so that we focus on groups, rather than individuals. In addition to having far less detailed models and violating some theoretical arguments (i.e. supposing your hypotheses are actually about individual- and group-level processes), a common

problem with this approach is usually the sample size. If we have data on 13,000 students in eight different regions, aggregating the data to the group level would leave us with just eight cases (i.e. one row of data representing one region). To conduct any sort of meaningful multivariate analysis, a much larger sample size is required. As we mentioned earlier in this chapter, focusing on group estimates may lead to an error in logic known as the ecological fallacy where group characteristics are used to generalize to individuals. Thus, there are several reasons not to rely on group estimates.

### Varying effects across contexts

In OLS, there is an assumption that the effects of independent variables are the same across contexts. For example, the effect of gender on school achievement is the same for everyone regardless of the region in which they go to school.

**Table 1.5** OLS regressions of standardized reading scores on gender and parental occupational status, by region

	<i>ACT</i>	<i>NSW</i>	<i>VIC</i>	<i>QLD</i>
Male	-0.246*** (0.056)	-0.446*** (0.031)	-0.408*** (0.093)	-0.395*** (0.039)
POS	0.024*** (0.002)	0.019*** (0.001)	0.020*** (0.003)	0.019*** (0.001)
Constant	-1.047*** (0.122)	-0.720*** (0.056)	-1.436*** (0.179)	-0.840*** (0.067)
<i>N</i>	954	3,270	702	2,322
	<i>SA</i>	<i>WA</i>	<i>TAS</i>	<i>NT</i>
Male	-0.410*** (0.043)	-0.386*** (0.052)	-0.336*** (0.038)	-0.279*** (0.047)
POS	0.017*** (0.001)	0.024*** (0.002)	0.017*** (0.001)	0.018*** (0.001)
Constant	-0.653*** (0.076)	-1.132*** (0.084)	-0.801*** (0.065)	-0.666*** (0.086)
<i>N</i>	1,548	1,221	2,183	1,446

Unstandardized regression coefficients. Standard errors in parentheses. POS – parental occupational status; ACT – Australian Capital Territory; NSW – New South Wales; VIC – Victoria; QLD – Queensland; SA – South Australia; WA – Western Australia; TAS – Tasmania; NT – Northern Territory.

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

We have many reasons to suspect that the effects of individual characteristics vary across contexts – that their impacts are not the same for everyone, regardless of group membership. We may find that the effects of gender are more pronounced in particular regions, for example. The context may influence the impact of gender on school achievement – for example, some regions may have an official policy around raising the science achievement of girls or the reading achievement of boys (as they are typical problems in the school achievement literature). Multilevel models allow regression effects (coefficients) to vary across different contexts (in this case, region) while OLS does not.

We might now think that one possible solution would be to run separate individual-level OLS regressions for each group. Table 1.5 displays the results of such an exercise.

This may seem to solve the problem of examining how group differences affect the impact of independent variables on the outcome of interest. You can see, for example, that the effect of being male ranges from being  $-0.446$  in New South Wales (NSW) to  $-0.246$  in the Australian Capital Territory (ACT). Likewise with parental occupational status (POS), the coefficients range from  $0.017$  in South Australia (SA) to  $0.024$  in the ACT and Western Australia (WA). These results do not tell us if the values are statistically significantly different from each other, without further calculations, and they also do not tell us anything about group properties which may influence or interact with individual-level outcomes. In addition to being a poor specification, this technique can get unwieldy if you have a large number of groups. Here we have only eight and the presentation of results is already rather difficult.

Another possible solution in OLS to effects varying across contexts might be to run interaction terms. You probably learned in your statistics training about interaction effects or moderating effects. If we thought that an independent variable affects an outcome differentially based upon the value of another independent variable, we could test this by using interaction terms. Based on the criticism above of running separate regressions for each group, a reasonable solution may seem to be to create interaction terms between the region and the other independent variables. We were making a similar argument earlier when we suggested that gender might impact on student achievement depending on region. We create the interaction terms by multiplying gender by region and parental occupational status by region (gender \* regions; parental occupational status \* regions) and we add them to the OLS regression as a set of new independent variables. If the interaction terms are statistically significant, it means that there is evidence that the effect of gender on reading and/or parental occupational status is contingent on the regions in which students go to school. The results of this estimation are presented in Table 1.6.

**Table 1.6** OLS regression of standardized reading scores on gender, region, parental occupational status and interactions ( $N = 13,646$ )

<i>Variables</i>	<i>b</i>	<i>s.e.</i>
Male	-0.246***	(0.059)
Region <sup>a</sup>		
<i>New South Wales</i>	0.326*	(0.139)
<i>Victoria</i>	-0.389*	(0.184)
<i>Queensland</i>	0.207	(0.143)
<i>South Australia</i>	0.394**	(0.151)
<i>Western Australia</i>	-0.085	(0.153)
<i>Tasmania</i>	0.246	(0.144)
<i>Northern Territory</i>	0.380*	(0.154)
Male * New South Wales	-0.200**	(0.067)
Male * Victoria	-0.162	(0.091)
Male * Queensland	-0.149*	(0.070)
Male * South Australia	-0.164*	(0.075)
Male * Western Australia	-0.140	(0.079)
Male * Tasmania	-0.090	(0.071)
Male * Northern Territory	-0.033	(0.076)
Parental occupational status (POS)	0.024***	(0.002)
POS * New South Wales	-0.005*	(0.002)
POS * Victoria	-0.005	(0.003)
POS * Queensland	-0.005*	(0.002)
POS * South Australia	-0.007**	(0.002)
POS * Western Australia	0.000	(0.003)
POS * Tasmania	-0.007**	(0.002)
POS * Northern Territory	-0.007**	(0.003)
Constant	-1.047	(0.127)
$R^2$	0.167	

<sup>a</sup>Australian Capital Territory is the reference category

b – unstandardized regression coefficients; s.e. – standard errors

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

The results in Table 1.6 suggest that region impacts on how individual characteristics affect the dependent variable. There are many statistically significant interactions. To really unravel what they mean, we have to look at them along with the main effects of the variables in the model. We can graph the main effects with the interaction effects and demonstrate the overall effect. We're not going to do that here, but we do address graphing interaction effects later. The main point from these results is that there are significant interactions. Perhaps we have just solved the problem of group effects?



Unfortunately not. There are still problems with this model. While there is no shortage of examples of this type of analysis in published work, one major problem with this approach is the nature of the cross-level interaction term. Cross-level interaction terms refer to interaction terms which have variables at different levels of aggregation. In this case, we have interacted individual characteristics (Level 1) with group characteristics (Level 2). This approach is fraught with problems. Treating group-level variables as though they are properties of individuals may result in flawed parameter estimations and downwardly biased standard errors (Hox and Kreft, 1994), and so we are more likely to find significant results. This also results in problems in the calculation of degrees of freedom, which leads to flawed estimates of the standard errors and faulty results. The problems associated with degrees of freedom are explained in more detail below.

### Degrees of freedom and statistical significance

Degrees of freedom are a problem when using OLS to model multilevel relationships. When we use OLS and simply add group-level variables, such as region in the example above, we create a model that assumes individual-level degrees of freedom. At this point you may well be wondering, 'What are degrees of freedom?' – fair enough. As the name implies, degrees of freedom are related to how many of the values in a formula are able to vary when the statistic is being calculated. Our example data contains 13,646 students in eight different regions. These students then have 13,646 individual pieces of data. We use this information to estimate statistical relationships. In general, each statistic that we need to estimate requires one degree of freedom – because it is no longer allowed to vary. Many equations contain the mean, for example. Once we calculate a group mean, it is no longer able to vary. Again, once we calculate a standard deviation, we lose another degree of freedom. In our examples above, degrees of freedom are determined from individual data, but if we have group characteristics in this individual-level data set, OLS calculates the degrees of freedom as though they are simply related to characteristics of individuals. In terms of group characteristics, the degrees of freedom should be based on the number of regions (8) rather than the number of pupils (13,646). The numbers – 13,646 versus 8 – are obviously very different. Degrees of freedom are integral in calculating tests of statistical significance. The resulting error from using the wrong degrees of freedom in OLS calculations is that it increases our likelihood of rejecting the null hypothesis when we should not. In other words, we are more likely to get statistically significant results – when we shouldn't – if we use the individual level degrees of freedom instead of the group level degrees of freedom.

**Table 1.7** Summary of OLS ‘workarounds’ and their associated problems

<i>Proposed solution</i>	<i>Unresolved problem</i>
Ignore group-level characteristics, focusing solely on individual attributes.	Misspecification of model if group-level variables are indeed important predictors of outcome of interest. Also, if outcomes for individuals are correlated by group membership, even if groups are not considered in the model, there is still a problem with independence of observations (i.e. ignoring it doesn't make it go away). Possibility of committing atomistic fallacy.
Focus only on group variation and aggregate data to group level.	Ignores role of individual attributes in explaining outcome. May have problems if there are small numbers of groups. Possibility of committing ecological fallacy.
Separate regression estimations for each group.	Fails to address how the group properties may influence or interact with individual-level outcomes.  There may be a problem if there are many groups and if there are small groups.
Create dummy variables for groups and create interaction terms between group dummies and individual-level characteristics.	Same problems as fitting regression estimates for each group separately. Degrees of freedom for cross-level interaction terms problematic. As well, the groups are treated as though they are unrelated although they may have things in common if they are drawn from a larger population of groups.

Adapted from Diez-Roux (2000: 173)

Table 1.7 summarizes the OLS ‘workarounds’ discussed above and their associated problems. The overarching problem is that when you use OLS models on data better suited to multilevel techniques you are very likely to *underestimate standard errors* and therefore increase the likelihood of results being statistically significant, possibly rejecting a null hypothesis when you should not. In other words, you are more likely to make a Type I error. If you correctly model your multilevel data then your results will be more accurately specified, more elegant, and more convincing, and your statistical techniques will match your conceptual model.

Multilevel modeling, in general and specific aspects of it, has also come in for some criticism. As with many debates of this kind, there is unlikely to be a firm and final conclusion, but we do advocate that users of any technique are aware of the criticisms and current debates. So we suggest that you start with this series of papers: Gorard (2003a, 2003b, 2007) and Plewis and Fielding (2003).

## SOFTWARE

In the main text of this book we use Stata 13 software. We assume that the reader is familiar with the Stata software program as we do not see this book as an introduction to Stata – see Pevalin and Robson (2009) for such a treatment for an earlier version of Stata. At the time of writing, Stata 13 was the current version, with some changes for the main commands used in this book from version 12 – namely, the

change from the `xtmixed` to the `mixed` command. While this changed little of the output and only a few options that are now defaults, it did impact on some of the other commands written by others for use after `xtmixed`. So, at times we use the Stata 12 command `xtmixed`, which works perfectly well in Stata 13 but is not officially supported. Stata 14 was released while this book was in production and we have checked that the do-files run in version 14.

As you may have gathered from the previous paragraph we use **this font** for Stata commands in the text. We italicize *variable names* in the text when we use them on their own, not part of a Stata command, but we also use italics to emphasize some points – we hope that the difference is clear.

In the text we use the `///` symbol to break a Stata command over more than one line. For example:

```
mixed z_read cen_pos || schoolid: cen_pos, ///  
      stddev covariance(unstructured) nolog
```

This is only done to keep the commands on the book page. In the do-files (available on the accompanying webpage at the URL given below), the command can run on much further to the right. We have tried to keep what you see on these pages and what you see in the do-file the same, but if you come across a command without the `///` in the do-file then don't worry about it.

At a number of points we include the Stata commands that we used to perform certain tasks, including data manipulation and variable creation. As with all software, there are a number of ways to get what you want, some elegant and some a bit cumbersome. Our rationale is to start by using commands that are easy to follow and then move on to some of the more integrated features in Stata. In doing so, we hope that what we are doing is more transparent. If your programming skills are more advanced than those we demonstrate then we're sure you'll be able to think of more elegant ways of coding in no time at all. In the do-files to accompany this book we sometimes include alternate ways of programming to illustrate the versatility of the software.

There is a webpage to accompany this book at <https://study.sagepub.com/robson>. On this webpage you will find do-files and data files for each of the chapters so you can run through them in Stata and amend them for your own use. You will also find the chapter commands in R, with some explanation how to use R for these multilevel models. The webpage will be very much a 'live' document with links to helpful sites and other resources. In the list of references we have noted which papers are 'open access'. Links to these papers are also on the webpage.

We have chosen to use Stata in this book for two reasons. First, we wanted to use a general-purpose statistical software package rather than a specialized multilevel

package. If readers already have experience of Stata then we avoid the challenges of getting to grips with new software. Even so, we could have then chosen from Stata, SPSS, SAS, and R. Which brings us to the second reason: Stata is our preferred software. It's what we use. It's what we teach with. And it's what we have previously written about. We chose to produce matching R code because R is a general-purpose software package and it is freely available. We both have previous experience with SPSS, and so SPSS code for the examples may well materialize on the webpage. For those who can't wait, Albright and Marinova (2010) have produced a succinct primer for 'translating' multilevel models across SPSS, Stata, SAS, and R which is available at: <http://www.indiana.edu/~statmath/stat/all/hlm/hlm.pdf>

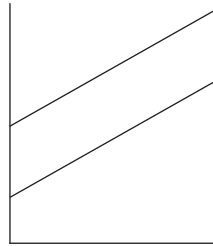
If your preference is for a specialized multilevel package then we strongly recommend that you start with the resources available from the Centre for Multilevel Modelling at the University of Bristol and their MLWiN software (free to academics) at: <http://www.bristol.ac.uk/cmm/>

## HOW THIS BOOK IS ORGANIZED

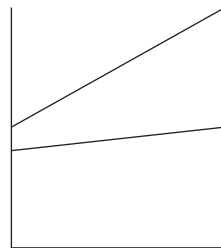
This book is organized into four chapters. This introductory chapter is followed by a chapter on random intercept models. This starts with a very quick review of OLS single-level regression to get us all on the same page and then gradually introduces random intercept models, how these differ from single-level OLS regression, and how to explore the nesting structures in your data. There are a few points where we touch on ongoing debates in the multilevel modeling community. Not everything is clear-cut, but we try to steer you through these without getting stuck in the technicalities (and probably a never-ending debate!). We continue with developing random intercept models, building up slowly with plenty of illustrations and example commands and output which is explained. We try not to throw too many equations at you, but when we do, we explain them in words. We look at adding independent variables at different levels, interaction effects, model fit statistics, and model diagnostics. Throughout, we try to follow the same example, but at times we have to deviate from that to demonstrate some points.

Chapter 3 is about random coefficient models, sometimes called random slope models. These are more complicated models than random intercepts, and we recommend that you read through Chapter 2 before starting out on this chapter. We pretty much follow the same topics as with random intercept models.

Of course there will be more detail in the chapters, but to conclude this introductory chapter we will simply remark that random intercept models look like this, with parallel regression lines for each group:



And random coefficient models look like this, with group regression lines which are not parallel to each other:



In Chapter 4 we turn to how to present results from these models and how to use Stata to produce publication-quality tables.

---

## CHAPTER 1 TAKEAWAY POINTS

---

- If your data contain nested levels, you should probably use multilevel modeling.
- Ignoring the levels in your data and simply using ordinary regression techniques requires you to violate important assumptions in regression theory.
- Multilevel modeling techniques allow you to better specify your model and achieve more accurate results.
- This book only considers cross-sectional nested data.
- Your models should be theory-driven and not motivated by the perceived need to add arbitrary complexity to your estimations.