

Module 4: Multilevel structures and classifications

Jon Rasbash
Centre for Multilevel Modelling

Contents

Aims	1
Introduction	1
C4.1 Two-level hierarchical structures	3
C4.1.1 Students within schools	3
C4.1.2 Issues of sample size	5
C4.1.3 Variables and levels, fixed and random classifications	6
C4.1.4 Other examples of a two-level structure.....	8
C4.1.5 Repeated measurements within individuals, panel data.....	8
C4.1.6 Multivariate responses within individuals	10
C4.1.7 Two-stage sample survey design	12
C4.1.8 An experimental design in which the intervention is at the higher level	14
C4.2 Three-level structures	17
C4.2.1 Students within classes within schools	17
C4.2.2 A repeated cross-sectional design: students within cohorts within schools....	19
C4.3 Four-level structures	22
C4.3.1 Doubly nested repeated measures	22
C4.4 Non-hierarchical structures	25
C4.4.1 Cross-classifications: students cross-classified by school and neighbourhood .	25
C4.4.2 Repeated measures within a cross classification of patients by clinician.	27
C4.4.3 Multiple Membership Structures	28
C4.5 Combining structures: hierarchies, cross-classifications and multiple membership relationships	31
C4.6 Spatial structures	34
C4.7 Summary	36

Some 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 4: Multilevel Structures & Classifications**
- Click "[4.1 Two-level hierarchical structures](#)" to open Lesson 4.1
- Click [Q1](#) to open the first question

Aims

After completing this chapter you will be able to:

- Recognise a range of multilevel structures and classifications and how they correspond to real-world situations, research designs, and/or social-science research problems;
- Appreciate the different types of data frames associated with each structure and how subscripts are used to represent structure;
- Begin to appreciate ‘targets of inference’;
- Distinguish between levels and variables, and fixed and random classifications;
- Appreciate that multilevel structures are likely to generate dependent, correlated data that requires special modelling;
- Recognise the difference between long and wide forms of data structures;
- Begin to appreciate the advantages, both technical and substantive, of using a multilevel model, and the disadvantages of not doing so.

Introduction

Multilevel modelling is designed to explore and analyse data that come from populations which have a complex structure. In any complex structure we can identify *atomic* units. These are the units at the lowest level of the system. Often, but not always, these atomic units are individuals. Individuals are then grouped into *higher-level* units, for example schools. By convention we then say that students are at level 1 and schools are at level 2 in our structure.

This module aims to give a ‘pictionary’ of structures that underlie multilevel models. We give ‘pictures’ of common structures as *unit diagrams*, as *classification diagrams*, as data frames and in words. Note that the terms *classification* and *level* can be used somewhat interchangeably but level implies a nested hierarchical relationship of units (in which lower units nest in one, and only one, higher-level unit) whereas classification does not. The data frames, in addition to showing the structure, will also provide some example *explanatory* (predictor) variables and a *response* (y variable) as discussed in Module 2. We have chosen the following examples to show a range of population structures where multilevel modelling is useful, and often necessary. We have also tried to introduce what are often seen

Module 4 (Concepts): Multilevel structures and classifications
as demanding and difficult concepts in a straightforward manner (e.g. fixed and random classifications, missing at random). While we have given the basic structures in a schematic and rather abstract form, we always point to published examples where the structure has been used in research.

C4.1 Two-level hierarchical structures

Hierarchical structures arise when the lower-level unit nests in one and only one higher-level unit. Such a relatively simple structure can, as we shall see, accommodate a wide range of study designs and research questions.

C4.1.1 Students within schools

Figure 4.1 is a *unit diagram* which aims to show the underlying structure of a research problem in terms of individual units; the nodes on the diagram are specific population units. In this case the units are students and schools which form two levels (or classifications). The lower units form the student classification (St1, St2 etc.) and the higher units form the school classification (Sc1, ... , Sc4). This unit diagram is just a schema to convey the essential structure of students nested within schools. In a real data set we would have many more than four schools and 12 students. The hierarchical structure means that a student only attends one school and has not moved about. Such a structure may arise when we are interested in school performance and we make repeated measurements of this by assessing student performance for multiple students from each school. This structure is likely to give rise to correlated or non-independent data, in the sense that students in the same school will often have a tendency to be similar on such variables as exam performance. Even if the initial allocation to a group was at random, social processes usually act to create this dependence. Traditionally, statistical modelling has faced difficulties with such dependence, indeed it has largely assumed it does not exist, but with multilevel modelling such correlation is expected and explicitly modelled.

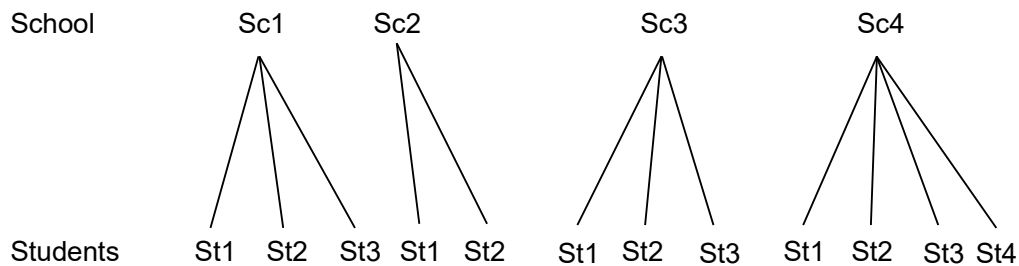


Figure 4.1. Unit diagram of a two-level nested structure; students in schools

This two-level nested structure can also be represented by a *classification diagram* (Figure 4.2). Classification diagrams have one node per classification (or level). Nodes joined by a single arrow indicate a nested (strict hierarchical) relationship between the classifications.

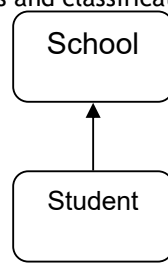


Figure 4.2. Classification diagram of a two-level nested structure; students in schools

Classification diagrams are more abstract than unit diagrams and are particularly useful, as we shall see, when the population being studied has a complex structure with many classifications.

Table 4.1 shows a data frame for the structure shown in Figure 4.1. We have also included a response (exam score in the current year), one school-level explanatory variable (school type), and two student-level explanatory variables (gender and previous exam score, say two years earlier). You will notice that the response is measured on the atomic unit, that is, level 1 (students); and that school 1 has three students, while school 4 has four students. That is, the data are not *balanced*; multilevel models do not require that there are the same number of lower level units in each and every higher level unit. In this example (and by common convention) the subscript i is used to index (represent) the lower level unit of the Student, while the subscript j indexes Schools.

With such a data frame we could ask a very rich set of questions by using a two-level multilevel model in which a student's current attainment is related to prior attainment (a previous test score) and there are data available on the gender of the student and the public/private nature of the school; these include

- i) Do males make greater progress than females?
- ii) Does the gender gap vary across schools?
- iii) Are males more or less variable in their progress than females?
- iv) What is the between-school variation in students' progress?
- v) Is School X (that is, a specific school) different from other schools in the sample in its effect?
- vi) Is there more variability in progress between schools for students with low prior attainment?
- vii) Do students make more progress in private than public schools?
- viii) Are students in public schools less variable in their progress?
- ix) Do girls make greater progress in state schools¹

¹ A classic study of school effects with an extended discussion of the issues involved is given by Aitkin, M. and Longford, N.T. (1986) Statistical modelling issues in school effectiveness studies (with Discussion). *J. Roy. Statist. Soc. A* 149, 1-43. Other examples include Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., et al. (1993). A multilevel analysis of school examination results. *Oxford Review of Education* 19: 425-433, Thomas, S (2001) Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions. *School Effectiveness and School Improvement* 12(3), 285 - 322

Questions ii, iii, iv, vi, and viii can be addressed by modelling variability as functions of explanatory variables, whereas questions i, v, vii, and ix are about modelling the mean as a function of explanatory variables. The defining strength of multilevel modelling is that it can do both, that is, model the mean and the variance simultaneously (traditional techniques can only model the mean). This idea may seem a little confusing at the moment but it is a theme we will be returning to throughout these training materials.

Table 4.1. Data frame representation of Figure 4.1 and 4.2: a two-level study for examining school effects on student progress

Classifications or levels		Response	Explanatory variables		
<i>Student i</i>	<i>School j</i>	<i>Student exam score_{ij}</i>	<i>Student previous examination score_{ij}</i>	<i>Student gender_{ij}</i>	<i>School type_j</i>
1	1	75	56	M	State
2	1	71	45	M	State
3	1	91	72	F	State
1	2	68	49	F	Private
2	2	37	36	M	Private
3	2	67	56	M	Private
1	3	82	76	F	State
2	3	85	50	F	State
1	4	54	39	M	Private
2	4	91	71	M	Private
3	4	43	41	M	Private
4	4	66	55	F	Private

C4.1.2 Issues of sample size

A question that often comes up at this point is how many units are needed at each level. It is difficult to give specific advice but there are some general principles that are worth stating now. The key one is the *target of inference*: in other words, are the units in your dataset special ones that you are interested in in their own right, or are you regarding them as representatives of a larger population which you wish to use them to draw conclusions about? If the target of inference in an educational study is a particular school then you would need a lot of students in that school to get a precise effect. If the target of inference is between-school differences in general, then you would need a lot of schools to get a reliable estimate. That is, you could not sensibly use a multilevel model with only two schools even if you had a sample of 1000 students in each of them. In the educational literature it has been suggested that, given the size of effects that are commonly found for between-school differences, a minimum of 25 schools is needed to provide a precise estimate

This document is only the first few pages of the full version.

To see the complete document please go to learning materials and register:

<http://www.cmm.bris.ac.uk/lemma>

The course is completely free. We ask for a few details about yourself for our research purposes only. We will not give any details to any other organisation unless it is with your express permission.