Multilevel Models for analysing social data

by

Harvey Goldstein

Institute of Education

University of London

h.goldstein@ioe.ac.uk

1 Introduction

Almost all kinds of social data have a *hierarchical* or *clustered* structure. For example, studies of inheritance deal with a natural hierarchy where offspring are grouped within families. Offspring from the same parents tend to be more alike in their physical and mental characteristics than individuals chosen at random from the population at large. For instance, children from the same family may all tend to be small, perhaps because their parents are small or because of a common impoverished environment or both.

Many designed experiments also create data hierarchies, for example evaluation studies carried out in several randomly chosen centres or institutions. In formulating models that take account of such hierarchies we are concerned only with the *fact* of such hierarchies not their provenance. While the focus here is on applications in the social sciences, the techniques are applicable more generally.

I refer to a hierarchy as consisting of *units* grouped at different *levels*. Thus offspring may be the level 1 units in a 2-level structure where the level 2 units are the families: students may be the level 1 units clustered within schools that are the level 2 units.

The existence of such data hierarchies is neither accidental nor ignorable. Individual people differ as do institutions. This is mirrored in all kinds of social activity where differences among institutions are often a direct result of differences among people, for example when students with similar motivations or aptitudes are grouped in highly selective schools or colleges. In other cases, the groupings may arise for reasons less strongly associated with the characteristics of individuals, such as the allocation of young children to elementary schools, or the allocation of subjects to different experimental groups. Nevertheless, once groupings are established, even if their establishment is random, they will tend to become differentiated, and this implies that the group and its members can both influence and be influenced by the composition of the group. To ignore this relationship risks overlooking the importance of group effects, and may also render invalid many of the traditional statistical analysis techniques used for studying data relationships.

A simple example will illustrate the importance of statistical validity. A well known and influential study of primary (elementary) school children carried out in the 1970's (Bennett, 1976) claimed that children exposed to so called 'formal' styles of teaching reading exhibited more progress than those who were not. The data were analysed using traditional multiple regression techniques which recognised only the individual children as the units of analysis and ignored their groupings within teachers and into classes. The results were statistically significant. Subsequently, Aitkin et al, (1981) demonstrated that when the analysis accounted properly for the grouping of children into classes, the significant differences disappeared and the progress of 'formally' taught children could not be shown to differ from the others.

This reanalysis is the first important example of a *multilevel* analysis of social science data, using information from both 'higher' level units (teachers) and 'lower' level units (children). In essence what was occurring here was that the children within any one classroom, because they were taught together, tended to be similar in their performance. As a result they provided rather less information than would have been the case if the same number of students had been taught separately by different teachers. In other words, the basic unit for purposes of comparison should have been the teacher not the student. The function of the students can be seen as providing, for each teacher, an estimate of that teacher's effectiveness. Increasing the number of students per teacher would increase the precision of those estimates but not change the number of teachers being compared. Beyond a certain point, simply increasing the numbers of students in this way hardly improves things at all. On the other hand, increasing the number of teachers to be compared, with the same or somewhat smaller number of students per teacher, considerably improves the precision of the comparisons.

To motivate and describe the multilevel approach we shall take our first example from education. This will be followed by a discussion of other areas of application which will also serve to introduce some of the more recent extensions to multilevel modelling and to illustrate their potential for understanding social processes.

2 Educational data

Schooling systems present an obvious example of a hierarchical structure, with pupils grouped or nested within schools, which themselves may be clustered within education authorities or boards.

Consider the common example where test or examination results at the end of a period of schooling are collected for each school for a randomly chosen sample of schools. The researcher wants to know whether a particular kind of subject streaming practice in some schools is associated with improved examination performance. She also has good measures of the pupils' achievements when they started the period of schooling so that she can control for this in the analysis. The traditional approach to the analysis of these data would be to carry out a regression analysis, using performance score as response, to study the relationship with streaming practice, adjusting for the initial achievements.

An analysis that explicitly models the manner in which students are grouped within schools has several advantages. First, it enables the data analyst to obtain statistically efficient estimates of regression coefficients. Secondly, by using the clustering information it provides correct standard errors, confidence intervals and significance tests. These generally will be more 'conservative' than the traditional ones which are obtained simply by ignoring the presence of clustering - just as Bennett's (1976) previously statistically significant results became non-significant on reanalysis. Thirdly, it makes it possible to explore the complexities of any variation among schools. For example, we can study the extent to which schools differ for different kinds of students, for example to see whether the variation between schools is greater for initially high scoring students than for initially low scoring students (Goldstein et al, 1993). Fourthly, by allowing the use of covariates or predictors measured at any of the levels of a hierarchy, it enables the researcher to explore the extent to which such variation between schools is accountable for by factors such as the organisational practice of the school or the social backgrounds of the students. Finally, there is often an interest in the relative ranking of individual schools, based upon the performances of their students after adjusting for intake characteristics and achievements. This can be done most efficiently using a multilevel modelling approach. In practice, however, such rankings are of limited use. They may identify institutions which have extremely high or low rankings for further study, but they have too much imprecision (large confidence intervals) for fine comparisons (see Goldstein and Spiegelhalter, 1996).

To fix the basic notion of a level and a unit, consider figures 1 and 2 based on hypothetical relationships.

Figure 1 shows the exam score and intake achievement scores for five students in a school, together with a simple regression line fitted to the data points. The residual variation in the exam scores about this line, is the *level 1 residual variation*, since it relates to level 1 units (students) within a sample level 2 unit (school). In figure 2 the three lines are the simple regression lines for three schools, with the individual student data points removed. These vary in both their slopes and their intercepts (the point at which they would cross the exam axis), and this variation is *level 2 variation*. It is an example of multiple or complex level 2 variation since both the intercept and slope parameters vary.
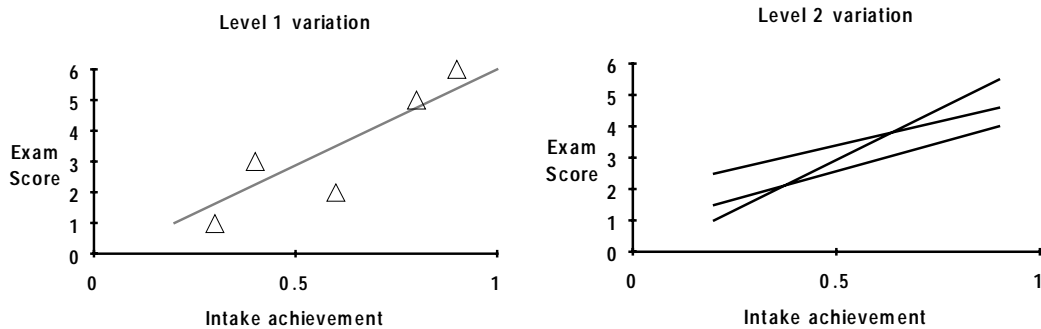
Figure 1   **Figure 2**

One procedure for incorporating school effects into a model is to fit a different regression line for each school. In some circumstances, for example where we have very few schools and moderately large numbers of students in each, this may be efficient. It may also be appropriate if we are interested in making inferences about just those schools. If, however, we regard these schools as a (random) sample from a population of schools and we wish to make inferences about the variation between schools in general, then a full multilevel approach is called for. In other words, just as we regard students as having been sampled from a population of students and where the sample is used to make inferences about that population rather than about each individual student, so we regard the schools as instruments for making inferences about the relevant population of schools. Moreover, if some of our schools have very few students, fitting a separate model for each of these will not yield reliable estimates: we can obtain more precision by regarding the schools as a sample from a population and using the information available from the whole sample to make 'smoothed' estimates for any one school.

A simple multilevel model

Consider a simple regression type model where an outcome or response, say a test score, is related to an input measure, say a pretest score for a random sample of students from a set of schools. Write

$$y_{ij} = b_0 + b_1 x_{ij} + u_j + e_{ij} \tag{1}$$

where the subscript i refers to the i-th student in the j-th school and u_j is the 'effect' for the j-th school, that is the additional contribution (positive or negative) that this school makes to the prediction of the response score (y_{ij}), given the pretest score (x_{ij}). The 'fixed' coefficients b_0, b_1 have the usual regression interpretations. The student or 'level 1' residual term e_{ij}, as in the ordinary regression case, typically is assumed to have a Normal distribution with a zero mean and a variance (σ_e^2) which needs to be estimated along with b_0, b_1.

The school effect terms u_j can be treated in one of two ways. We can set out to estimate each one separately, so that if there are, say 100 schools, we will have to estimate a further 99 parameters. If we regard the schools as a random, however, then we treat these terms in a similar fashion to the student level residuals and assume that the u_j come from a Normal distribution with a zero mean and a variance to be estimated, say σ_u^2. Once we have made such an assumption we have a model in which there are two random variables, one at each level of the data structure, and it is this feature which makes it a *multilevel* statistical model. It also means that standard statistical software packages, which typically assume only a single random variable, cannot be used to fit these models. Special purpose software is available (see appendix)

and some of the standard packages such as SAS and STATA have begun to introduce some multilevel modelling features.

In the remaining sections I will describe some of the applications where multilevel models have found a useful application and also introduce some of the extensions of the basic model (1). A more thorough technical treatment can be found in Goldstein (1995). For a general reference and information on current developments see the World Wide Web multilevel models site at http://www.ioe.ac.uk/multilevel/.

3 Sample survey methods

In a household survey, the first stage sampling unit will often be a well-defined geographical unit. From those which are randomly chosen, further stages of random selection are carried out until the final households are selected. Because of the geographical clustering exhibited by measures such as political attitudes, special procedures have been developed to produce valid statistical inferences, for example when comparing mean values or fitting regression models (Skinner et al, 1989).

While such procedures usually have been regarded as necessary they have not generally merited serious substantive interest. In other words, the population structure, insofar as it is mirrored in the sampling design, is seen as a 'nuisance factor'. By contrast, the multilevel modelling approach views the population structure as of potential interest in itself, so that a sample designed to reflect that structure is not merely a matter of saving costs as in traditional survey design, but can be used to collect and analyse data about the higher level units in the population. The subsequent modelling can then incorporate this information and obviate the need to carry out special adjustment procedures, which are built into the analysis model directly.

4   Repeated measures data

A different example of hierarchically structured data occurs when the same individuals or units are measured on more than one occasion. A common example occurs in studies of human growth or, for example, in longitudinal studies of attitude change. Here the occasions are clustered within individuals that represent the level 2 units with measurement occasions the level 1 units. Such structures are typically strong hierarchies because there is much more variation between individuals in general than between occasions within individuals.

In a multilevel framework this involves, in the simplest case, each individual having their own trend line with the intercept *and* slope coefficients varying between individuals (level 2). Thus, equation (1) will be extended by allowing the coefficient of x_{ij}, which now refers to time or age, to vary across individuals who are now the level 2 units. Thus at level 2 we have two random variables, the intercept and slope, with a corresponding correlation (see Figure 2). This addition of a *random coefficient* can be extended so that any of the predictor variables can have coefficients varying at any level of the hierarchy.

5   Discrete response data

Until now I have assumed implicitly that our response or dependent variable is continuously distributed, for example an exam score. Many kinds of statistical modelling, however, deal with categorised responses, in the simplest case with proportions. Thus, we might be interested in whether or not a respondent votes for a particular political party, or in an examination pass rate and how these vary from area to area or school to school.

Such models, part of the class known as generalised linear models, have been available for some time for single level data (McCullagh and Nelder, 1989), with associated software. By allowing for the modelling of higher level effects we obtain an analogous set of multilevel generalised linear models to those for continuous responses.

6   Random cross classifications and multiple unit membership

Whilst the title of this article refers to multilevel, that is hierarchical models, there are many examples where units are cross-classified as well as clustered. In geographical research, the definition of an individual's geographical area is contingent upon the context being considered. Thus, the relevant location unit for purposes of leisure may not be the same as that surrounding the environment of work or schooling. We can conceive formally of individuals belonging simultaneously to both types of unit each of which may have an influence on a person's life.

In most schooling systems, students move from elementary to secondary or high school. We might expect that both the elementary and secondary schools attended will influence a student's achievements, behaviour and attitudes. Thus the level 2 units are of two types, elementary school and secondary school, where each 'cell' of their cross classification contains some, or possibly no students. In this example, a third way of classification could be the area or neighbourhood where the student lives.

A related structure occurs where for a single level 2 classification, level 1 units may belong to more than one level 2 unit. An example from sociology concerns children's' and adults' friendship patterns where an individual may belong to several groups simultaneously. The characteristics of the members of each group will influence such an individual, in relation to the individual's exposure to the group. Spatial data is another example where an individual will be influenced by the characteristics of the area in which they live and also by neighbouring areas. Thus the individual can be considered to 'belong' simultaneously to several units with the contributions of each unit being weighted in relation to its 'distance' from the individual.

7   Further topics

I have outlined only some of the more common models available in this article. Other developments include the following topics.

- Multivariate multilevel models where, for example, several measures are made simultaneously on each level 1 unit and these then vary and covary at higher levels. A special case is where some of the responses are made at level 1 and some at higher levels, for example measures of student *and* teacher attitudes. The analysis can be carried out when some measurements are missing (randomly or by design) and this leads to procedures for the efficient design and modelling of rotation or matrix samples.

- Most measurements made in the social sciences contain some error component. This may be due to observer error as when measuring the weight of an animal, or an inherent result of being able to measure only a small sample of behaviour as in educational testing. It is well known that when variables in statistical models contain relatively large components of such error the resulting statistical inferences can be very misleading unless careful adjustments are made. The same is true in multilevel models where errors can occur at several levels and a discussion and illustration is given by Woodhouse et al (1996).

- In structural equation modelling, were individuals are grouped within hierarchies, for all the same reasons discussed above, it is important to carry out such analyses in a multilevel framework. For example, we may be interested in underlying individual attitudes based upon a number of indicators. Data on such indicators may be available over time and we can postulate a model whereby the underlying attitude varies from individual to individual (level 2) and also varies randomly over time within individuals (level 1). The model can then be elaborated further by studying whether there is any systematic change over time and whether this varies across individuals. (Muthen, 1994)

- Another extension, particularly important for data such as attitude and other scales as well as educational data, is to allow for complex variation at level 1. For example, when analysing educational achievement, it is known that boys tend to have a greater variance than girls and in general the level 1 variance may depend on the value of any explanatory variable.

- Modelling time spent in various states or situations is important in a number of areas. In industry the 'time to failure' of components is a key factor in quality control. In medicine the survival time is a fundamental measurement in studying certain diseases. In economics the duration of employment periods is of great interest. In education, researchers often study the time students spend on different tasks or activities. Such 'event history' or 'survival' models need to be embedded within a multilevel framework, for example where individuals (level 2 units) repeatedly pass through various employment or unemployment periods or where survival times of patients are measured in a number of different clinics (level 2 units).

8 A caveat

The application of multilevel modelling has already begun to yield new and important insights in a number of areas and, as the software becomes more widely available, the application of these techniques should become relatively straightforward, even routine.

All this is welcome, yet despite their usefulness, models for multilevel analysis cannot be a panacea. In some circumstances, where there is little structural complexity, they may be hardly necessary, and traditional single level models may suffice, both for analysis and presentation. On the other hand multilevel analyses can bring extra precision to attempts to understand causality, for example by making efficient use of student achievement data in attempts to understand differences between schools.

These models are not, however, substitutes for well grounded substantive theories, nor do they replace the need for careful thought about the purpose of any statistical modelling. Furthermore, by introducing more complexity they can extend but not necessarily simplify interpretations. Multilevel models are tools to be used with care and understanding.

Appendix

In addition to some of the general purpose statistical packages, there are two major software packages for multilevel modelling: HLM (Bryk and Raudenbush, 1992) and MLn (Rasbash and Woodhouse, 1995). A review of these and other packages has been carried out by Kreft et al (1994).

References

Aitkin,M., Anderson,D. and Hinde,J. (1981). Statistical modelling of data on teaching styles (with discussion). *Journal of the Royal Statistical Society,* A., **144**, 148-61

Bennett,N.(1976). Teaching Styles and Pupil Progress. London, Open Books.

Bryk, A. S. and S. W. Raudenbush (1992). *Hierarchical Linear Models.* Newbury Park, California, Sage

Goldstein, H. (1995). *Multilevel Statistical Models.* London, Edward Arnold. New York, Halsted Press.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society,* A. **159**: 385-443.

Kreft, G. G., J. De Leeuw and R. Van der Leeden (1994). Review of five multilevel analysis programs. *The American Statistician,* **48**: 324-335

Longford, N. T. (1993). *Random Coefficient Models.* Oxford, Clarendon Press

Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological methods & research* **22**: 376-398.

Rasbash, J. and Woodhouse, G. (1995). *MLn Command Reference*. London, Institute of Education

Skinner,C.J., Holt,D. and Smith,T.M.F (1989). *Analysis of complex surveys*, Chichester:Wiley.

Woodhouse, G., Yang, M., Goldstein, H. and Rasbash, J. (1996). Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society, A*. **159**: 201-12.