



New Statistical Methods for Analysing Social Structures: An Introduction to Multilevel Models

Lindsay Paterson; Harvey Goldstein

British Educational Research Journal, Vol. 17, No. 4. (1991), pp. 387-393.

Stable URL:

<http://links.jstor.org/sici?sici=0141-1926%281991%2917%3A4%3C387%3ANSMFAS%3E2.0.CO%3B2-9>

British Educational Research Journal is currently published by Taylor & Francis, Ltd..

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/taylorfrancis.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

New Statistical Methods for Analysing Social Structures: an introduction to multilevel models

LINDSAY PATERSON, *Centre for Educational Sociology, University of Edinburgh*

HARVEY GOLDSTEIN, *Institute of Education, University of London*

ABSTRACT *An introductory account is given of developments in multilevel modelling of educational and other social data. The technique is introduced with some simple examples and its importance is explained. Examples of applications in a number of areas are given, including repeated measures designs, school effectiveness studies, area-based studies and political opinion sample surveys. Almost all data collected in the social sciences have some form of inherent hierarchical structure, and this structure should be reflected in the statistical models that are used to analyse them. It is suggested that multilevel techniques and associated software packages have reached the stage when they can and should be applied routinely in the analysis of social data, and that failure to do so can result in potentially serious misinterpretations.*

1. Introduction

Recent developments in multilevel modelling have made available to social scientists powerful statistical techniques for analysing individuals as members of social groups. The techniques are also especially useful for repeated measures data. This note is an outline of the concepts of multilevel modelling. A fuller account and technical details can be found in, for example, Goldstein (1987a) and Raudenbush & Bryk (1986).

2. As a Generalisation of Ordinary Regression

Multilevel modelling is one extension of ordinary multiple regression. Consider a survey of 5000 pupils drawn from 100 schools; thus the groups here are the schools. Suppose we wish to investigate the relationship between a score based on examina-

tion attainment at age 16 (y_i for pupil i) and a measure of verbal-reasoning ability (x_i) on entry to secondary school. Ordinary regression would estimate a single equation by pooling all 5000 cases, expressing the examination attainment, say, as a linear function of ability:

$$y_i = b_0 + b_1 x_i + e_i, \quad (1)$$

where b_0 is the 'intercept' and b_1 is the slope coefficient, and both are parameters to be estimated. The term e_i is a random variable, often called a residual and usually assumed to be normally distributed with mean 0 and a constant variance.

Equation (1) simply says that the exam score (y_i) increases by a fixed amount (b_1) for a unit increase in verbal reasoning score (x_i). The difficulty with this model is that it does not allow for school differences. For two students with the same verbal reasoning score, the average or predicted exam score in general would be expected to depend on which school they attended, and the empirical evidence bears this out.

A multilevel model provides the appropriate generalisation of (1) to take account of such school differences. Changing the suffixes so that y_{ij} is the attainment of the i th pupil in the j th school, we can incorporate school membership into the model by writing:

$$y_{ij} = b_0 + b_1 x_{ij} + u_j + e_{ij}, \quad (2)$$

where the addition of a term specific to each school, u_j , indicates that over and above any given verbal reasoning score each school has its own contribution. Model (2) in essence is the one which has come to be used in school effectiveness research. This research seeks to identify apparently effective (large u_j) or apparently ineffective (small u_j) schools.

It would, in principle, be possible to estimate 100 values of u_j using the standard extension of (1) known as analysis of covariance. In practice, however, this would be cumbersome, and not feasible at all if the number of schools was very large. Moreover, we would typically be interested in making inferences about the variation between all schools, not merely the 100 which have been sampled. The key technical advance of multilevel modelling is to assume that the u_j vary randomly across schools (that is, at the school level).

Thus u_j is a random variable, assumed to have a zero mean and a constant variance. One of the aims of the data analysis will be to see whether the addition of further variables to the model (for example social status) reduces or 'explains' any of the between-school variation. Extensions to further levels of nesting (for example Local Education Authorities (LEAs)) are straightforward, as are extensions which assume that coefficients such as b_1 in (2) are random variables, varying across schools.

Thus multilevel modelling can respect the heterogeneity of social experience. Populations need not be treated as homogeneous, and the methods yield statistical tests of whether the groups that have been represented in the model really are different from each other, and confidence intervals for the extent of any differences. In the example, this would involve testing whether, for instance, there really was a non-zero variance for u_j , and also estimating how large that variance was.

Since all data analysis in the social sciences deals with populations where there are inbuilt hierarchies, the scope for application of multilevel models is very wide. In economics, groups in the hierarchy could be markets or firms (Raffe & Willms, 1989). In political science the groups could be constituencies, and in geography they could be neighbourhoods (Garner & Raudenbush, 1990). A particularly interesting

application is to repeated measurements on individuals, where the lowest-level units (level 1) are 'occasions' and the higher-level units (level 2) are individual subjects (see below). Simple versions of these techniques have in fact been used for many years in agriculture, genetics and medicine, often under the name of variance components. The significant advance recently has been to develop very general models and to provide efficient software capable of handling complex models and large amounts of data.

3. What is Lost by not Using Multilevel Techniques?

3.1. *Aggregate Data are Unreliable*

A common approach to analysing hierarchical data has been to aggregate to group level and then use only group means. Thus instead of using information about the 5000 pupils in the above example, only the 100 school means would be used: that is, for each school there would be available (perhaps from administrative records) a mean examination attainment and a mean entry-score. Ordinary regression would be used to relate the mean attainment to the mean score.

There are two drawbacks to this approach, one technical and the other substantive. The technical one is that the statistical estimates can be very unreliable in the sense that slight perturbations to the data or to the model can produce markedly different results. (This shows up as large standard errors for estimates, and as high collinearity among predictors.) An empirical demonstration of such unreliability is given by Woodhouse & Goldstein (1988), for a published analysis of examination results in English LEAs.

One reason why lack of statistical robustness is a problem is that aggregate data are remote from the social and educational processes that are of interest. This is the second drawback of using aggregate data: it is so removed from causal processes that it provides no help to explanation. For example, learning is done by children, not schools or LEAs, and we cannot offer explanations based upon aggregate level analyses that would make explanatory sense. As Woodhouse & Goldstein (1988) show, the results from an analysis of aggregated data bear no necessary relationship to those results which would be derived from a proper multilevel analysis.

3.2. *Disaggregated Data are More Revealing*

A more positive way of putting that last point is that only if we let regression relationships vary among groups can we see where effects are occurring, and understand how they are occurring. For example, Aitkin & Longford (1986) showed that the relationship between attainment in the General Certificate of Education (GCE) and a verbal reasoning test score (VRQ) on entry to secondary school varied among schools in a Local Education Authority. For one school which they studied, pupils of above-average VRQ did exceptionally well in GCE examinations, whereas those below average did exceptionally badly. The conclusions of this statistical analysis thus pointed to questions that could be of interest to educational practitioners: what was happening in that school to produce this result? Letting the regression relationships vary therefore allowed statistical modelling to move closer to the interests of what are usually known as qualitative investigations. Multilevel

analysis lets the uniqueness of each institution be respected, and at the same time retains the capacity to generalise which all statistical analysis offers.

3.3. Clustering Affects Standard Errors

Allowing individual relationships to vary among groups is necessary for technical reasons as well, because the grouping might have an effect on standard errors. This is a well-known issue in design-based inference from clustered data, and various devices for dealing with it approximately have been developed over the years (such as the design effect). It is a problem also for model-based inference, because in practice no model can incorporate all group-level influences on individual behaviour, and so such influences become part of the random error.

Many examples have been published where the clusters or the groups are schools. But to see the scope for wider application, consider a model of electoral behaviour. Here clustering could arise at the constituency level, for wards within constituencies, and for voters within wards. Correct standard errors would be estimated only if the levels of wards and constituencies were respected in the analysis: multilevel modelling provides an efficient way of achieving that (Goldstein & Silver, 1989). Multilevel modelling would provide the bonus, moreover, of being able to model the influence of ward characteristics or of constituency characteristics as well as individual-level predictors such as social class.

Multilevel analysis allows characteristics of the group to be incorporated into models of individual behaviour, while also producing correct estimates of standard errors so that valid tests and intervals can be constructed. Current developments can handle the analysis of either continuous or discrete response variables, or indeed mixtures of these: see below.

4. Time

As mentioned above, repeated measures on individuals can be represented as grouping time-points or occasions within people (who in turn can be arranged within further hierarchies). Goldstein (1989) has explained how this representation can be used to gain insights into the physical growth of children. The two levels of random variation—among individuals, and among occasions within individuals—take account of the fact that growth characteristics of individual children, such as their average growth rate, vary around a population mean, and also that each child's observed measurements vary around his or her own growth trajectory. Again we find that statistical analysis is better able to represent qualitative experience: the idiosyncratic characteristics of each child's growth can be respected in a model which also retains the capacity for generalisation.

Furthermore, as before, characteristics at higher levels—here that of the child—can be used to explain processes at lower levels. Raudenbush (1988), for example, reports a study which found greater variation among children in reading growth than in arithmetic growth. The reading growth itself could be predicted by characteristics of the child's home, such as the educational level of the mother.

Time can be incorporated even if multiple measurements are not available on individuals. For example, regular cohorts of school leavers are sampled by the Scottish Young People's Survey from secondary schools in Scotland. The multilevel structure is thus pupils within time points within school, and not all schools need to

be represented at each time point. The advantage of this structure is that time-related characteristics can be incorporated without ignoring also the issue of whether schools remain stable institutionally over time. Of course, as before, the individual characteristics of pupils are the strongest predictors of measures such as examination attainment, but school-level and time-level variables are relevant too (Paterson, 1991). Willms & Raudenbush (1989) have used such a structure to study the stability of school effects, a question of considerable policy interest for developing reliable performance indicators (Coopers & Lybrand, 1988). Using data from schools in the Inner London Education Authority, Nuttall *et al.* (1989) have analysed examination results at age 16+, and have shown how schools differ in terms of the achievement of students from different ethnic minorities, and how some schools have the effect of compensating for initial differences and other schools have the effect of magnifying them.

5. Non-linear Models and Random Cross Classifications

The questions introduced in Section 2 made the tacit assumption that the response variable (y_{ij}) was continuous, but many social scientists need methods for proportions, or simply variables scored as 0 or 1, and there are other cases where a non-linear model relationship might be appropriate. Multilevel analogues of generalised linear models are being developed at the Institute of Education in London University. In particular, multilevel log-linear modelling can be carried out using the program ML3 (Goldstein, 1991).

Another important development is to represent structures that are not strictly hierarchical. An example would be where pupils are recognised as belonging to schools as well as residential neighbourhoods. Schools and neighbourhoods form a cross classification of units at level 2 and we are interested in studying both the variation between schools and the variation between neighbourhoods. Further work on such models is currently being carried out at the Institute of Education (Goldstein, 1987b).

6. Conclusions

Multilevel modelling offers both a statistical and a conceptual advance. It allows statistical analysis to be more flexible in that it can respect the multiple groupings of society—incorporating both explanatory processes and random variation at several levels. Because it opens up this flexibility, it should also stimulate new thinking about the scope for quantitative analysis in the social sciences, allowing statistics to be used where researchers have understandably felt that conventional methods did violence to the richness of their data and the complexity of the world. Because multilevel modelling is a generalisation of single-level techniques, and because software is now available, we can recommend that it be used routinely.

One important issue which arises is the sample size necessary to carry out a satisfactory analysis. It is difficult to provide general guidelines here, but a key factor is the number of higher-level units, for example schools. Experience suggests that studies with fewer than 25 schools are unable to provide very precise model estimates. Studies with more than 100 schools, on the other hand, have been able to provide useful insights. Likewise, for the estimation of the intercept or slope parameters for a single school, at least about 25 students are required.

There are current limitations due to the incomplete development of underlying methods and computational algorithms. Among these are models for random cross classifications (see section 5) and certain kinds of non-normally distributed data. Much remains to be learned both theoretically and practically, but the more multilevel models are applied in a wide variety of contexts the better we shall come to understand their strengths and their limitations.

7. Software and Further Information

The three main programs for carrying out multilevel analysis are available as follows; each comes with a manual containing worked examples of how to use it, and describing some of the theory of the techniques:

ML3: The Multilevel Models Project,
Department of Mathematics, Statistics and Computing,
Institute of Education,
University of London,
20 Bedford Way,
London WC1H 0AL,
United Kingdom.

HLM: S. W. Raudenbush,
Department of Counselling, Educational Psychology and Special Education,
460 Erikson Hall,
Michigan State University,
East Lansing,
MI 48824, USA.

VARCL: N. T. Longford,
Educational Testing Service,
Princeton, NJ,
USA.

The topic of multilevel modelling is developing rapidly, and papers appear regularly in the statistical and social science literature. Up-to-date news and discussion is available in the Multilevel Modelling Newsletter, available from the Multilevel Models Project, which also runs introductory workshops and more advanced data analysis 'clinics'.

Acknowledgements

This work was partly supported by the Economic and Social Research Council (UK) through its funding of the Centre for Educational Sociology as a Designated Research Centre (grant number C000280004) and by its research grant (R000232157) to the Multilevel Models Project at the Institute of Education.

REFERENCES

- AITKIN, M. & LONGFORD, N. (1986) Statistical modelling issues in school effectiveness, *Journal of the Royal Statistical Society A*, 149, pp. 1–42.
- COOPERS & LYBRAND (1988) *Local Management of Schools: a report to the Department of Education and Science* (London, HMSO).
- GARNER, C. L. & RAUDENBUSH, S. W. (1991) Neighbourhood effects on educational attainment, *Sociology of Educational*.
- GOLDSTEIN, H. (1987a) *Multilevel Models in Educational and Social Research* (London, Griffin).
- GOLDSTEIN, H. (1987b) Multilevel covariance components, *Biometrika*, 74, pp. 430–431.
- GOLDSTEIN, H. (1989) Models for multilevel response variables with an application to growth curves, in: R. D. BROCK (Ed.) *Multilevel Analysis of Educational Data* (New York, Academic Press).
- GOLDSTEIN, H. (1991) Nonlinear multilevel models, *Biometrika*, 78, pp. 45–52.
- GOLDSTEIN, H. & SILVER, R. (1989) Multilevel and multivariate models in survey analysis, in: C. J. SKINNER, D. HOLT & T. M. F. SMITH (Eds) *Analysis of Complex Surveys* (Chichester, Wiley).
- NUTTALL, D. L., GOLDSTEIN, H., PROSSER, R. & RASBASH, J. (1989) Differential school effectiveness, *International Journal of Educational Research*, 13, pp. 769–776.
- PATERSON, L. (1991) Trends in attainment in Scottish secondary schools, in: S. W. RAUDENBUSH & J. D. WILLMS (Eds) *Schools, Classrooms and Pupils: international studies of schooling from a multilevel perspective*, pp. 85–100 (New York, Academic Press).
- RAFFE, D. & WILLMS, J. D. (1989) Schooling the discouraged worker: local-labour-market effects on educational participation, *Sociology*, 23, pp. 559–581.
- RAUDENBUSH, S. W. (1988) *The Analysis of Longitudinal, Multilevel Data* (available from the Centre for Educational Sociology, University of Edinburgh).
- RAUDENBUSH, S. W. & BRYK, A. S. (1986) A hierarchical model for studying school effects, *Sociology of Education*, 59, pp. 1–17.
- WILLMS, J. D. & RAUDENBUSH, S. W. (1989) A longitudinal hierarchical linear model for estimating school effects and their stability, *Journal of Educational Measurement*, 26, pp. 1–24.
- WOODHOUSE, G. & GOLDSTEIN, H. (1988) Educational performance indicators and LEA league tables, *Oxford Review of Education*, 14, pp. 301–320.

LINKED CITATIONS

- Page 1 of 2 -



You have printed the following article:

New Statistical Methods for Analysing Social Structures: An Introduction to Multilevel Models

Lindsay Paterson; Harvey Goldstein

British Educational Research Journal, Vol. 17, No. 4. (1991), pp. 387-393.

Stable URL:

<http://links.jstor.org/sici?sici=0141-1926%281991%2917%3A4%3C387%3ANSMFAS%3E2.0.CO%3B2-9>

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

References

Statistical Modelling Issues in School Effectiveness Studies

M. Aitkin; N. Longford

Journal of the Royal Statistical Society. Series A (General), Vol. 149, No. 1. (1986), pp. 1-43.

Stable URL:

<http://links.jstor.org/sici?sici=0035-9238%281986%29149%3A1%3C1%3ASMIISE%3E2.0.CO%3B2-6>

Neighborhood Effects on Educational Attainment: A Multilevel Analysis

Catherine L. Garner; Stephen W. Raudenbush

Sociology of Education, Vol. 64, No. 4. (Oct., 1991), pp. 251-262.

Stable URL:

<http://links.jstor.org/sici?sici=0038-0407%28199110%2964%3A4%3C251%3ANE0EAA%3E2.0.CO%3B2-V>

Multilevel Covariance Component Models

H. Goldstein

Biometrika, Vol. 74, No. 2. (Jun., 1987), pp. 430-431.

Stable URL:

<http://links.jstor.org/sici?sici=0006-3444%28198706%2974%3A2%3C430%3AMCCM%3E2.0.CO%3B2-R>

Nonlinear Multilevel Models, with an Application to Discrete Response Data

Harvey Goldstein

Biometrika, Vol. 78, No. 1. (Mar., 1991), pp. 45-51.

Stable URL:

<http://links.jstor.org/sici?sici=0006-3444%28199103%2978%3A1%3C45%3ANMMWAA%3E2.0.CO%3B2-J>

LINKED CITATIONS

- Page 2 of 2 -



A Hierarchical Model for Studying School Effects

Stephen Raudenbush; Anthony S. Bryk

Sociology of Education, Vol. 59, No. 1. (Jan., 1986), pp. 1-17.

Stable URL:

<http://links.jstor.org/sici?sici=0038-0407%28198601%2959%3A1%3C1%3AAHMFSS%3E2.0.CO%3B2-1>

A Longitudinal Hierarchical Linear Model for Estimating School Effects and Their Stability

J. Douglas Willms; Stephen W. Raudenbush

Journal of Educational Measurement, Vol. 26, No. 3. (Autumn, 1989), pp. 209-232.

Stable URL:

<http://links.jstor.org/sici?sici=0022-0655%28198923%2926%3A3%3C209%3AALHLMF%3E2.0.CO%3B2-8>

Educational Performance Indicators and LEA League Tables

Geoffrey Woodhouse; Harvey Goldstein

Oxford Review of Education, Vol. 14, No. 3. (1988), pp. 301-320.

Stable URL:

<http://links.jstor.org/sici?sici=0305-4985%281988%2914%3A3%3C301%3AEPIALL%3E2.0.CO%3B2-U>