MULTILEVEL MODELLING NEWSLETTER

Produced through the Multilevel Models Project:

c/o Mathematics, Statistics & Computing Dep't
Institute of Education, University of London,
20 Bedford Way, London WC1H 0AL U.K.
e-mail: rprosser%uk.ac.lon.educ.isis@ac.uk
Telephone: 01-636-1500 ext 473

Vol. 1 No. 1 January 1989

EDITORIAL

During the past several years, considerable interest has developed in the application of multilevel models in many fields–education (school effects research), orthodontics (facial growth), and demography (fertility studies) to name just three. A number of groups in Europe, North America, and Australia are engaged in theoretical development of the models to inform data analytic practice. In addition, several recent conferences and workshops in Britain, the U.S., and Holland have focussed on applications, particularly in the field of education. The time seems right to develop a new means of sharing the ideas and experiences of multilevel modellers....hence, the **Multilevel Modelling Newsletter**.

The Multilevel Models Project based at the University of London Institute of Education (described on page 6) is coordinating the production and distribution of this publication. Assisting us with this is Dr. Tony Bryk of the University of Chicago who recently spent six weeks with the project.

Our aims in producing the newsletter are:

- to help applied researchers and statisticians communicate about the most recent developments in this field;
- to provide a forum for discussion of common problems;
- to stimulate new ideas for theory and analytic practice; and
- to serve as a guide to resources, particularly new literature and software.

The target audience includes anyone with an interest in the field–statisticians and researchers in the social sciences, medicine, biology, business, and other disciplines. We plan to distribute three issues each year, and we propose the following as regular features: summaries of new literature; software developments (new releases, upgrades, bug fixes, etc.); information on conferences and workshops; descriptions of applications in different fields; reports on theoretical advances; and a question and answer column. (In the next issue, for example, we hope to provide information on new developments in the *HLM* software, a recent multilevel conference in Nijmegen (The Netherlands), and a soon-to-be-published book on multilevel modelling in education edited by Darrell Bock.)

The success of this venture will depend to a large extent on the participation of readers in the communication process. We welcome your suggestions and ideas, and we request reports on your work, other news, and questions regarding analysis issues. A brief questionnaire is included with this issue. Please fill it in and return it to Bob Prosser, Research Officer, Department of Mathematics, Statistics, & Computing, University of London Institute of Education, 20 Bedford Way, London WC1H 0AL. If you know of interested colleagues who would like to receive future editions, please note their names and addresses on the questionnaire.

The **Multilevel Modelling Newsletter** is distributed free of charge, but reproduction and distribution are costly, and the Project budget is not unlimited. We would ask that if you are able, to please send a contribution towards these expenses (say, £3 or US$5). Thank you.

SOFTWARE

ML2 IS NOW BEING DISTRIBUTED

ML2, the latest PC version of the Multilevel Models Project's two-level software is now being shipped to interested researchers. Along with estimates of the fixed and random parameters and the associated standard errors, the following statistics can be obtained using Version 1.0B:

- residuals from each level;
- two types of covariance matrices for the residuals;
- predicted response variable values for individuals; and
- estimates of contrasts of the fixed parameters and associated χ^2 values for hypothesis tests.

An extensive graphics component enables the user to conduct diagnostic analyses with residuals. (A graphics card is required.)

ML2 provides a comprehensive interactive analysis environment through integration with the kernel of a general-purpose statistical package called Nanostat (Healy, 1987). Simple commands are available for

- creating, recoding, and transforming variables;
- sorting and selecting cases;
- deleting cases (listwise) with missing values;
- conducting exploratory plotting; and
- computing descriptive statistics and performing simple analyses, e.g., OLS regression.

The data manipulations indicated above and the modelling commands operate on reusable columns in a worksheet, facilitating efficient use of available storage. Worksheets and command sequences can easily be saved for future analyses.

Like earlier versions of the Project's software, *ML2* is extremely versatile in terms of model choice. The coefficients of any explanatory variable may be random at either level. Further, at each level, the random coefficients may have any pattern of variances and covariances. Thanks to some fancy matrix algebra (Goldstein & Rasbash, to appear) the program now runs, on average, considerably faster than its forerunners. New conveniences and features in *ML2* include:

- a simple model specification procedure;
- a choice of two estimation algorithms–iterative generalized least squares and restricted iterative generalized least squares.
- an aggregation utility–group means and standard deviations can easily be obtained from individuals' scores.
- optional "freezing" of estimates of individual parameters for part of the estimation process to speed the approach to convergence;

An online help facility is available as a reminder about the structure of the commands for implementing these and the other features of the program.

The program is designed for AT-type computers with at least 560K of *available* RAM and a hard disk. A co-processor version is available. Enquiries about ordering *ML2* may be sent to Bob Prosser at the Project address on the front page.

References

Goldstein, H., & Rasbash, J. (submitted for publication). Efficient computational procedures for the estimation of parameters in multilevel models, based on iterative generalized least squares.

Healy, M. (1987). *Nanostat users' guide*. Unpublished documentation.

MODELLING WITH VARCL

Nick Longford

VARCL is an interactive program for analysis of hierarchically structured data. It consists of two independent FORTRAN 77 subprograms: (a) *VARL3* for data with up to three levels of nesting and (b) *VARL9* for data with up to nine levels. The random term covariance structure which can be fitted using the former is quite general, whereas the latter is for modelling spherically symmetric covariance structures only.

The general formulation of the models that can be fitted with *VARL3* is

(1) $$Y_{ijk} = X_{ijk}\beta_{jk} + \epsilon_{ijk}$$

where X_{ijk} is the $r \times l$ vector of regressors for subject i in group j in area k. The vector of regression coefficients β_{jk} has a decomposition

(2) $$\beta_{jk} = B + b_{jk} + b_k$$

DEVELOPMENTS

and $\{b_{jk}\}$ and $\{b_k\}$ form two mutually independent random samples from $N(0, \Sigma_2)$ and $N(0, \Sigma_3)$, respectively. The elementary level random terms ϵ_{ijk} are assumed i.i.d. and $N(0, \sigma_e^2)$, and independent from the other two random samples. The user can specify which variances in Σ_2 and Σ_3 are constrained to 0, or to a positive constant, and which are to be estimated. If the variance is constrained to zero then all the covariances in the same row and column of the variance matrix are inflexibly constrained to zero. If two variances in the same variance matrix are positive, their covariance can be constrained to zero. Similarly, the regression parameters, i.e., the elements of B, can be constrained to a value provided by the user, or left free to be estimated.

The intercept, formally a variable with the value of 1 for every subject, is included in the list of regression variables by default. In order to maintain invariance with respect to shift of origin of the variables, an intercept-by-slope covariance cannot be constrained to zero unless one of the constituent variances is also set to zero.

Two-level models correspond to having one level three unit in equation 1 and $\Sigma_3 = 0$. If every group contains exactly one subject then (1) becomes a variance heterogeneity model, in which the variance of an observation is a quadratic function of the regression variables. The confounding of σ_e^2 and Σ_2 can be resolved be constraining the former to be a suitably chosen small positive number. Model (1) involves variance heterogeneity in the general case also, but usually the interpretation in terms of variable relationships is preferred.

Using VARCL

Before beginning an interactive analysis session with the program, the user prepares a (flat fixed-format) data file and a *basic information file* containing the job title, number of levels, number of units and variables on each level, maximum number of iterations in model fitting, required precision, labels for variables, format for the data, and nesting structure for the data. The user inputs these files and specifies two kinds of model to the program: a *maximal* model and the *actual* model(s) to be fitted. The maximal model should represent the envelope of all the models which could be fitted during a session.

The first specification tasks are identification of the response/ outcome variable and description of any a priori case weighting. Next the regression variables (the fixed part of the model) are selected from the variable list declared in the basic information file. Additionally, in *VARL3* the variables associated with variation on levels 2 and/ or 3, i.e., represented in Σ_2 and/ or Σ_3 by variances to be estimated, have to be sub-selected from the list of regression variables. The latter list is known as the random part.

Models to be fitted are described to the program by constraining particular parameters (regression coefficients, variances and covariances) associated with the maximal model to zero. The program defaults to fitting a variance components model with no slope variances and all regression parameters included. This model can be altered by, for example, exclusion of an explanatory variable and/ or removal of the restriction that a particular slope variance be zero. The selection and alteration routines have a menu format in which incorrect declarations can be easily ammended.

After a model is fitted the user can generate a *dump* file and end the analysis session. The dump file can be restored in a future session and further submodels of the maximal model can be fitted.

Program Availability

VARCL was originally developed for use on VAX computers running VMS, but a version for AT-type computers is currently being tested. Experimentally, a quasi-likelihood adaptation (Longford, 1988) has been installed for fitting variance components models when normality assumptions are violated.

For further information, write N. T. Longford at Educational Testing Service, 21-T Rosedale Road, Princeton, NJ 08540 or telephone (609) 734-5151. Electronic mail on BITNET can be sent to NTL6600@rosedale.

Reference

Longford, N. T. (1988). A quasilikelihood adaptation for variance component analysis. Paper presented at the Annual Meeting of the American Statistical Association, New Orleans, LA, August, 1988.

APPLICATIONS

MODELLING FACIAL GROWTH

Richard Tanguay

Polynomial regression models are well suited for studying craniofacial growth. Polynomials are easy to compute, can be fitted over short age spans, and do not impose a predetermined structure on the growth curve. Covariates are readily incorporated. As traditionally applied, these models typically assume equally spaced target ages and complete serial data for individuals. As Strenio et al. (1983) first pointed out, one can view longitudinal data as a two-level structure with subjects at the higher level and measurement occasions at the lower level. The fixed coefficients estimate the mean growth curve; the random parameters model variance within and between subjects.

We have conducted exploratory analyses applying a two-level model to craniofacial data. Our objective is to show the potential of multilevel models for evaluating craniofacial growth.

The data are derived from 572 lateral cephalograms of 105 girls, followed serially between 10 and 15 years of age by the Human Growth Research Center at the University of Montreal. The sample is representative of the French-Canadian population in Quebec. All of the cephalograms were traced and digitized by one technician. The analyses pertain to the cephalometric landmark gnathion. To facilitate the distinction between the amount and direction of craniofacial growth, the digitized coordinates were transformed from rectangular into polar form.

Only the increase in gnathion length (corresponding to the distance between the gnathion and the sella) is described here. The fixed part of the model relates gnathion length to a cubic polynomial in age centered at 12.5 years. The cubic model indicates that growth velocity increases until a maximum is reached and then decreases. The within-subject variance describes the dispersion about the individuals' growth curves. At level 1 we fitted the variance as a linear function of age, but the coefficient was very small and not significant. At level 2, the intercept, linear and quadratic terms were allowed to vary randomly between subjects; the cubic term was fixed. It is of interest that, between individuals, the intercept and linear coefficients show a low correlation (0.10), while the intercept and quadratic coefficients are moderately correlated (-0.61). The linear and quadratic coefficients show low negative correlations (-0.12). The between-person variance, as calculated from the level 2 random parameters, is a quartic function of age. Of practical importance, these estimates can be used to generate percentiles for the age of maximum velocity for clinical evaluations. Using simulations as described by Goldstein (1987), the mean age for the maximum velocity is estimated at 11.98 years. The 5th and 95th percentiles are 10.7 and 13.3 years respectively. The mean growth rate (velocity) at the age of maximum velocity is 0.24 cm per year and the 5th and 95th percentiles are respectively 0.18 and 0.31 cm per year.

In conclusion, multilevel procedures offer an effective means of estimating mean growth, patterns of variation, and the timing and association of developmental events. The correlations between the random parameters within a given level are important because they provide a basis for estimating an individual's growth curve when only limited data are available. We should soon be able to evaluate and treat patients relative to their own curves rather than the population's mean curve. The efficiency of the growth predictions might be further improved by modelling several variables simultaneously. This involves incorporating a third level into the analysis. Finally, the procedures provide a means of directly testing hypotheses. They can be used to evaluate growth differences that may be explained by factors such as sexual dimorphism, malocclusion, or clinical treatments.

Parameter Estimates for a Cubic Model

Fixed Parameters

Explanatory Variable	*Estimate* ($\widehat{SE}$)
Intercept	10.63 (0.041)
Age	0.227 (0.0067)
Age^2	-0.0077 (0.0019)
Age^3	-0.0050 (0.0011)

continued on p. 7

WORKSHOPS & CONFERENCES

SUMMER HLM WORKSHOP A SUCCESS

Peter Cuttance

A workshop attended by about 30 researchers from various parts of the globe was held at Edinburgh University in August 1988. The aim was to introduce multilevel modelling methods to researchers from a range of disciplines, although the majority of the participants were engaged in some aspect of educational research. Steve Raudenbush and I were the principal tutors with assistance being provided by Cathy Garner and Alan Alexander.

Participants were provided with hands-on computing experience for about sixty percent of the time during the five days of the workshop. The first two days consisted of alternate sessions of lectures and computing. After the second day, tutorial discussions of the examples that were set for the computing sessions replaced the lectures. By the fourth day most of the participants had commenced analyses of their own data.

We considered it important that participants should work with their own data once the basics of the methodology had been introduced. A particular benefit of this strategy was the range of examples that it provided for the discussion sessions. Several participants prepared presentations for these sessions based on their analyses during the last two days of the workshop. *HLM* was the main software employed for the workshop, but sessions were allocated to demonstrations of Nick Longford's updated *VARCL* program and the Multilevel Models Project's *ML2*.

CONFERENCE ADVANCE NOTICE

The *International Conference on Applications of Multilevel Methods in Educational Research* to be held in Edinburgh on August 12th and 13th, 1989, will bring together some of the key researchers from the U.K., North America, and Europe who have applied multilevel methods of analysis to educational problems. In recent years there has been a rapid increase in the use of multilevel methods for studying the effects of schools, schooling, and educational reform. The conference will share results from these studies and, from the vantage point of this experience, it will examine the potential uses of multilevel methods and the conceptual issues involved in the design and interpretation of multilevel analyses. The conference will concentrate on non-technical aspects of multilevel methods; participants will not need to be familiar with computing or statistical aspects of multilevel techniques.

The conference is being organized jointly by the Centre for Educational Sociology (CES) and by Stephen Raudenbush (Michigan State University) and Douglas Willms (CES and University of British Columbia). For further information, please contact Eileen McCormack, the conference administrator, at the Centre for Educational Sociology, Edinburgh University, 7 Buccleuch Place, Edinburgh EH8 9LW Scotland. The Fax number is 031 668 3263, the telephone number is 031 667 1011 ext 6803, and the e-mail address is Eileen McCormack@uk.ac.edinburgh

LONDON WORKSHOP SERIES BEGINS

The Multilevel Models Project has initiated a series of one and two day workshops to be held at the Institute of Education of the University of London. The first, *An Introduction to Multilevel Modelling*, designed primarily for social scientists, was held December 12th and 13th, 1988.

This training session featured discussion of theory as well as hands-on work with the *ML2* program. Topics included:

- "Why multilevel models?"
- analysis of two-level data
- use of *ML2*
- residuals and parameter contrasts
- modelling of longitudinal/ growth data

The emphasis was on data analytic practice rather than mathematical statistics. Everyone worked on example analyses using the Junior School Project data, and several of the 17 participants brought their own data for analysis. An optional evening session was arranged for this.

A second introductory workshop will be held on February 2nd and 3rd. The Project plan for 1989 includes additional offerings of this training, possibly in May and October. Subjects such as multilevel analysis of binary data/ proportions, multiple matrix sampling, three (and higher) level models, and cross classifications will be covered in one day

continued on p. 7

PROJECT NEWS

THE MULTILEVEL MODELS PROJECT

The (Development and Dissemination of) Multilevel Models Project directed by Professor Harvey Goldstein of the University of London Institute of Education has three primary aims:

(a) to extend existing methodology;
(b) to study the practical application of the models to real data sets; and
(c) to disseminate information about the theory and practice of multilevel analysis.

This research and development program which extends the work of an earlier two year project is supported by Britain's Economic and Social Research Council.

The project team includes Jon Rasbash whose responsibilities include data analysis and software development and Bob Prosser who is responsible for organizing workshops, conducting simulation research and analyzing data.

The work of the Project has benefitted from the visits of Rod McDonald of Macquarie University (Australia) and Tony Bryk of the University of Chicago during the summer and autumn of 1988. During Professor McDonald's visit, a paper dealing with the application of multilevel models to structural equation analysis was completed (McDonald & Goldstein, submitted for publication). An important outcome of Dr. Bryk's visit was a comparison of the EM algorithm involved in hierarchical linear modelling (HLM) and IGLS, the Project's method of estimation. A new estimation procedure–Restricted Iterative Generalized Least Squares–was developed (Goldstein, to appear) and is implemented in the Project's new *ML2* software.

Methodology for *constrained estimation* has recently been developed whereby random parameters in a model are allowed to be functions of the fixed coefficients. This extension has important applications in the analysis of discrete data and in models for educational achievement when test scores are subject to a ceiling effect. Some progress has been made on developing algorithms for the analysis of random cross-classifications, but this needs further work before an efficient implementation is possible. A simulation study is being conducted to compare several methods for treating missing data in multilevel analysis.

Future plans include:

(a) production of software for 3-level and higher level models;
(b) extension to multilevel time series modelling;
(c) development of software to handle errors in explanatory variables; and
(d) further theoretical and practical study of multilevel models for discrete data.

The Project team will be active in disseminating information about multilevel analysis through further workshops, seminars, and talks.

References

Goldstein, H. (to appear). Restricted (unbiassed) iterative generalized least squares estimation. *Biometrika*

McDonald, R. P. & Goldstein, H. (submitted for publication). Balanced versus unbalanced designs for linear structural relations in 2-level data.

THE MULTILEVEL CAUSAL AND FACTOR MODELS PROJECT

R. P. McDonald

Computer programs for structural models for multivariate observations are now routinely applied in the social sciences either for causal analysis (path analysis) or factor analysis or a combination of the two. Given a number of measures (test scores for example) on a large number of sampling units (usually human subjects), we use the programs to test complex hypotheses of causal relationship between the measured variables, or to determine the number and nature of the common factors (latent variables). These are abstract attributes of the examinees that the tests measure in common. (See McDonald, 1985.) More generally we may seek to examine the causal relations between the factors measured by the tests, doing what is known as "path analysis with latent variables."

Existing computer programs for the structural analysis of multi-variate data require the sampling of units to follow a simple, single-level sampling scheme. The object of the present work is to develop theory and computer programs for path analysis with latent variables, including, of course, pure factor analysis and causal models for observable variables, for data sets with sampling at several levels.

Suppose we have a random sample of students from a random sample of classes from a random sample of schools. In a typical case of pure multilevel factor analysis, we would give a large number of cognitive tests to the students, and fit a model in which the factors have between-students-within-classes correlations, between-classes-within-schools correlations, and between-schools correlations. The interesting question to be tested would be whether the relations beween tests and factors ("factor loadings") remain the same at all three levels. In a typical example of causal analysis, we would have measures of educational outcomes from the students, measures of class quality, and measures of school quality. We would set up and test causal models of the type: "school quality affects educational outcomes only through the intermediation of class quality", vs.: "school quality directly affects outcomes along with class quality".

Goldstein & McDonald (in press) contains the foundations of the current project. Theory developed so far shows that the mathematics and computer programming is much simpler for balanced designs (with equal numbers of students in each class etc.) than for unbalanced designs, and that cases with missing data present very much worse computational problems. A series of programs is to be written, working up from the simpler to the more complex cases.

References

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum Associates.

Goldstein, H. & McDonald, R. P. (in press). A general model for the analysis of multilevel data. *Psychometrika*

Workshops....continued from p. 5

advanced workshops. (Suggestions for additional advanced topics are welcome.) The first advanced workshop is scheduled for March 15, 1989.

No fee is charged for participation in these workshops, and each participant receives a copy of the current version of the Project's software. Readers interested in attending should obtain an application form from Bob Prosser at the Institute of Education.

Facial Growth....continued from p. 4

Estimates for a Cubic Model cont'd

Level 2 Variances and Covariances

	Estimate
Intercept	0.1746
	(0.0245)
Age	0.0013
	(0.0002)
Age^2	0.00014
	(0.00005)
Int. $\times$ Age	0.0015
	(0.0017)
Int. $\times$ Age^2	-0.0030
	(0.0008)
Age $\times$ Age^2	-0.00005
	(0.00007)
Level 1 Variance	
Intercept	0.0068
	(0.00058)

References

Goldstein, H. (1987). *Multilevel models in educational and social research.* London: Griffin.

Strenio, J., Weisberg, H., & Bryk, A. (1983). Empirical Bayes estimation of growth curve parameters and their relationship to covariates. *Biometrics, 39,* 71-86.

CONTRIBUTORS

Thanks very much to the people who provided news for the inaugural issue!

Peter Cuttance & Cathy Garner
Centre for Educational Sociology
7 Buccleuch Place
Edinburgh EH8 9LN SCOTLAND

Rod McDonald
School of Education
Macquarie University
North Ryde
New South Wales 2113 AUSTRALIA

Richard Tanguay
Departemente de Sante Buccale
Universite de Montreal
Casse Postale 6121, Succursale A
Montreal, Quebec H3C 3JF CANADA

Nick Longford
Educational Testing Service
21-T Rosedale Road
Princeton, NJ 08540 U.S.A.

Questionnaire

We hope you will take a few minutes to complete this short questionnaire which will help us in designing future issues of the newsletter. Thank you.

Name Telephone

Address e-mail

Are you currently doing research on theoretical aspects of multilevel modelling? If so, please describe.

Please suggest theoretical issues you would like to see discussed in future issues.

Are you using multilevel modelling in an applied study? If so, please describe the topic briefly.

Would you be willing to write a short article on your work for a future issue?

Please give us your suggestions about (a) general topics and features that would be helpful to you and (b) format/ style.

Do you think this issue was too technical in relation to your needs and interests? or, about right? or, not technical enough?

Please note names of colleagues who would be interested in receiving future issues.

Please return your questionnaire to:

Bob Prosser, Research Officer
Multilevel Models Project
Department of Mathematics, Statistics, & Computing
Institute of Education
20 Bedford Way
London WC1H 0AL U.K.