MULTILEVEL MODELLING NEWSLETTER

The Multilevel Models Project
Mathematical Sciences
Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, ENGLAND
E-mail: temsmya@ioe.ac.uk
Web site http://www.ioe.ac.uk/hgoldstn/home.html
Tel: +44 (0)171 612 6682 Fax: +44 (0)171 612 6686

Vol. 7 No. 3 **December, 1995**

Release of *MLn* Version 1.0a, November 1995

The latest version of *MLn* provides:

- new macros for analysis of multicategory response models, time series models and survival time models,
- improved macros for analysis of models with binary and count responses,
- new commands, mainly of interest to advanced users, to support these macros,
- a new command, BXSEarch, which is an improvement on XOMIt and XSEArch,
- fixes to bugs in Version 1.0.

Up-to-date versions of all macros and their documentation are held in files which can be downloaded from the Web site:

http:/www.ioe.ac.uk/hgoldstn/home.html.

Please refer to that page regularly for details of all new developments.

This upgrade is free to existing purchasers of *MLn* version 1.0.

To obtain your upgrade, together with details of new commands etc, send email quoting your name and *MLn* licence number to:
mln.order@ioe.ac.uk or FAX to *+44 (0)171 612 6032.*

Workshop in Norwich: 25-27 March 1996. Using *MLn,* this general introductory workshop will cover topics such as basic principles, setting up two and higher level models, repeated measures, logistic models, multivariate analysis and diagnostics. Course fees for academics and non-academics respectively are £350 and £600 inclusive. For further information please contact *Anne-Lise McDonald* at *Health Policy and Practice Unit, UEA, Norwich, NR4 7TJ. Tel +44 (0)1603 593631, email a.cox@uea.ac.uk.*

MLn Clinics in London 1996

Tuesday January 2
Tuesday February 6
Tuesday March 5
Tuesday April 2
at
Multilevel Models Project
11 Woburn Square, 2nd floor
London WC1A 0SN
Contact Min Yang for appointment
Tel: (0)171 612 6682
Email: temsmya@ioe.ac.uk

≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈≈

Also In This Issue

References & Conferences

Some Recent Publications to Multilevel Modelling:

Anderson, J. E., and Louis, T. A. (1995). Survival analysis using a scale change random effects model. *J. American Statistical Association*, **90**, 669-79.

Belin, T. B., and Rubin, D. B. (1995). The analysis of repeated measures data on schizophrenic reaction times using mixture models. *Statistics in Medicine*,**14**, 747-68.

Chinchilli, V. M., Eisenhart, J. D.,& Miller, W. G. (1995). Partial likelihood analysis of within unit variances in repeated measurement experiments. *Biometrics*, **51**, 205-16.

Follmann, D., and Wu, M. (1995). An approximate generalised linear model with random effects for informative missing data. *Biometrics*, **51**, 151-68.

Gibbons, R. D., Hedeker, D., Charles, S. C., & Frisch, P.(1994). A random effects probit model for predicting medical malpractice claims. J. *American Statistical Association*, **89**, 760-67.

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

Gray, J., Jesson, D., Goldstein, H., Hedger, J., & Rasbash, J. (1995). A multilevel analysis of school improvement: changes in schools' performance over time. *School effectiveness and school improvement*, **6**, 97-114.

Hedeker, D., and Gibbons, R. D. (1994). A random effects ordinal regression model for multilevel analysis. *Biometrics*, **50**, 933-44.

Kreft, G. G., DeLeeuw, J., & VanderLeeden, R. (1994). Review of five multilevel analysis programs. *The American Statistician*, **48**, 324-335.

Kreft, I. G. G., and Aschbacher, P. R. (1994). Measurement and evaluation issues in education: the value of multivariate techniques in evaluating an innovative high school reform program. *International Journal of Educational Research*, **21**, 181-96.

Lancelot, R., Lescourret, F. and faye, B. (1995). Multilevel modelling of pre-weaning kid mortality during the cold, dry season 1991-1992 in the outskirts of N'Djaména, Chad. *Preventive Veterinary Medicine*, **24**, 171-186

Lipsitz, S. R., Laird, N. M., & Harrington, D. P. (1994). Weighted least squares analysis of repeated categorical measurements with outcomes subject to nonresponse. *Biometrics*, **50**, 11-24.

McGilchrist, C. A. (1994). Estimation in gemeralised mixed models. *J. Royal Statist. Soc. B.*, **56**, 61-9.

Miller, M. E. (1995). Analysing categorical responses obtained from large clusters. *Applied Statistics*, **44**, 173-86.

Pickles, A., and Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, **14**, 1447-61.

Raudenbush S.W., Brennan, R.T. & Barnett, R. C. (1995). A multivariate hierarchical model for studying psychological change within married couples. *Journal of Family Psychology*. **9**, 2, 167-174.

Snijders, T., Spreen, M. and Zwaagstra, R. (1995). The Use of multilevel modeling for analysing personal networks: networks of cocaine users in an urban area. *Journal of Quantitative Anthropology*, **5**, 85-105.

Ten Have, T. R., Landis, J. R., & Weaver, S. L. (1995). Association models for periodontal disease progression: a comparison of methods for clustered binary data. *Statistics in Medicine*, **14**, 413-29.

Yashchin, E. (1995). Likelihood ratio methods for monitoring parameters of a nested random effect model. *J. American Statistical Association*, **90**, 729-38

Statistics Education Discussion

The Summer 1995 issue of the *Journal of Educational and Behavioral Statistics* (volume 20, No. 2) is a special issue on Multilevel Models: Problems and Prospects. Guest editor for the issue is Ita G.G. Kreft. Contributors are David Draper, David Rogosa and Hilary Saner, Jan de Leeuw and Ita Kreft and Carl Morris.

Comments on the main papers by Harvey Goldstein, Nicholas Longford, Stephen Raudenbush, William Mason. Replies by Draper, Rogosa and Saner, De Leeuw and Kreft.

The issue has 140 pages. Single copies are available from AERA Publications Sales, 1230 17th Street NW, Washington, DC20036-3078.

4th International Social Science Methodology Conference, ESSEX' 96 1-5 July, 1996

-------------*Call for papers* -------------

Essex' 96, to be held at the Institute for the Social Sciences, Colchester, UK, is the fourth in a series of conferences which have become the major international forum on Social Science Methodology, sponsored by the Research Committee on Logic and Methodology of the International Sociological Association.

One of the themes of the conference is Multilevel Modelling. One stream within this theme is (computational) statistics. This would bring together a number of statistical developments and present them together, to further understanding of this growing field. A second stream focusses on theory and applications. The goal is to broaden the range of applications and to bridge the gap between theory and research in this area.

The organizers are: *Joop Hox* (Department of Education, University of Amsterdam, Wibautstraat 2-4, NL-1091 WB, Amsterdam, the Netherlands: hox@educ.uva.nl), *Tom Snijders* (University of Groningen: t.a.b.snijders@ppsw.rug.nl), *Uwe Engel* (University of Potsdam, *DE:* engel@rz.uni-potsdam.de). Coordination: Joop Hox.

If you want to present a paper in the multilevel theme, please submit an abstract directly to the conference organizers, and send a copy to Joop Hox at the address given above. The deadline for the abstract is 25 ***January, 1996***.

The conference organisation has a brochure with an abstract form. To obtain this, send email to: *David Rose, ESSEX '96, University of Essex. Colchester CO4 3SQ, UK at CONF96@ESSEX.AC.UK.*

Annual conference of British Society for Population Studies

The BSPS was held at the University of Brighton during 30the August - 1st September 1995. It focussed on the current population issues, several studies using multilevel modelling approach as were presented. Here are some references.

1. A multilevel modelling approach to the determinations of urban and rural fertility rates in Bangladesh (*Khan, H.T.A.*, Department of Mathematics, Napier University, Sighthill Court, Edinburgh EH11 4BN. Email:*a.khan@central.napier.ac.uk*)

2. Immunisation in rural Bangladesh: a multilevel analysis (*Steele, F., Diamond, I.* and *Amin, S.*, Department of Social Statistics, University of Southampton, Southampton SO17 1BJ. Email: *fas@alcd.soton.ac.uk*)

3. Role of government level family planning workers and health centres as determinants of contraceptive use in Bangladesh (*Kamal, N.*, Centre for Pupolation Studies, London School Hygiene & Tropical Medicine, 99 Gower Street, London WC1E 7AZ)

4. Demographic and Socio-economic influences on patterns of leaving home in the post-war period in Britain (*Murphy, M.* and *Wang, D.L.*, Population Studies, London School of Economics, Houghton Street, London WC2A 2AE. Email: *d.wang@lse.ac.uk*)

Theory & Applications

Multilevel unit specific and population average generalised linear models

Harvey Goldstein, Institute of Education, University of London, UK

Zeger et al (1988) make a distinction between two kinds of models for hierarchically structured data where there is a non-identity link function such as the logit or log. We illustrate the distinction using the following 2-level logit model where the (0,1) response y_{ij} is binary and there is just an intercept and a covariate in the linear predictor, namely

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \beta_0 + \beta_1 x_{ij} + u_j \qquad (1)\\ \text{var}(y_{ij}) &= \pi_{ij}(1-\pi_{ij}) \\ \text{var}(u_j) &= \sigma_u^2 \end{aligned}$$

assuming binomial variation at level 1.

This model is referred to a a 'subject specific' model which derives from their consideration of a repeated measures model where 'subject' is level 2. A more general description is 'unit specific' (US) which we shall adopt. It assumes a specific covariance structure for the responses. An alternative specification is to write what is termed a 'population average' (PA) or 'marginal' (*Diggle* et al., 1994) model as

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \beta_0^* + \beta_1^* x_{ij} \qquad (2)\\ \text{var}(y_{ij}) &= \alpha.\pi_{ij}(1-\pi_{ij}) \\ \text{cov}(Y_j) &= V \end{aligned}$$

where V can assume particular or general structures, for example an equicorrelation structure. Specifically, it is not an explicit function of the covariance matrix of the random coefficients, although its form is sometimes derived from considering a particular US model and integrating over the random coefficients to obtain the marginal distribution (see for example *Bock* and *Aitkin*, 1981).

The two models in general will differ in their covariance structures and hence will provide differing estimates of the fixed coefficients for the same data. The PA model provides no specific information about higher level variation and is therefore useful only for making inferences about average population effects. Thus, (2) allows us directly to estimate the change in response probablity corresponding to a unit change in x_{ij} whereas in (1) a unit change in x_{ij} allows us to estimate a change in the response probability *for any given level 2 unit*. Since the link function is nonlinear, this change will depend on u_j.

If we wish to use (1) to estimate the average population change in probability for a unit change in x_{ij} we can either use an approximation based on the Normality assumption (*Zeger* et al., 1988) or simulate from the fitted model. In the latter case we would generate a sample of N u_j's assuming Normality, and apply the antilogit transformation to each one for each relevant value of x_{ij}. These transformed values on the probability scale are then averaged to give an estimate of the population mean for the given x_{ij}. By increasing the value of N we can approximate the population mean as accurately as desired (*Goldstein*, 1995, Chapter 5).

One of the suggested advantages of PA models is the direct estimation of population effects on the probability scale. In view of the fact that these effects are readily estimated from US models this advantage seems neglible. On the other hand, the disadvantage of not being able to provide estimates for higher level structure variation seems in general to be a major disadvantage of PA models. If there really is a hierarchical structure it seems natural to incorporate it into the model directly. In this sense PA models are not multilevel models at all since there is no explicit hierarchical structure specified.

References

Bock, R.D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46, 443-59.

Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data.* Oxford, Clarendon Press.

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

Zeger, S.L., Liang, K. And Albert, P.S. (1988). Models for longitudinal data: a generalised estimating equation approach. *Biometrics*, 44, 1049-60

The Effects of Centering in Multilevel Analysis:Is the Public school the loser or the winner?A new analysis of an old question

Ita Kreft, Educational Foundations, UCLA, USA

Introduction

The analysis model in this paper is a replication of the model used by *Raudenbush* and *Bryk* (1986) (*R&B*). The decades old question regarding private versus public school sector effects in student outcomes was analyzed by *R&B* for the first time using a multilevel method. But controversy still exists as to how the micro level explanatory variables should be scaled for such analyses. The default option in ***HLM*** (*Bryk, Raudenbush*, *Selzer* and *Congdon*, 1988), which seems to be preferred by the authors of this software (see e.g. *Bryk* and *Raudenbush*, 1992, p.25 etc.) is to center explanatory variables around the context mean, when coefficients for these explanatory variables are allowed to be random. Centering is usually recommended in the literature for technical reasons, but as illustrated here with examples, has conceptual implications as well. These examples show that effects of centering are not the same as in single level analysis. Context centering (not the traditional grand mean centering) changes the variance of the response variable between and within schools. This change effects mostly the parameter estimates at the second level, the fixed as well as the random parameter estimates.

In the examples presented here, those changes are clearest in the estimation of a public sector effect on math achievement of students. For a mathematical explanation and technical discussion of context centering compared to raw scores (or grand mean centering) see *Kreft*, *De Leeuw* and *Aiken* (1995).

Analyses

The public-private sector controversy started with *Coleman* et al's (1982) assertion that Catholic schools are more effective than private schools. The effect was most pronounced for lower social economic status (SES) students, a sector-by-SES cross-level interaction. The results of *R&B*'s (1986) analyses of the High School and Beyond data also show that the Catholic sector is slightly more egalitarian, and "Lower-SES students fare better in Catholic schools, and higher SES students fare better in public schools." (l.c., p.13) The analyses with NELS-88 in Table 1 contain the same micro-level explanatory variables, Homework and SES, as in *R&B*, predicting math achievement. The difference over models is in the treatment of the student level explanatory variable homework and SES as either CWC or as RS. CWC scores are indicated in the table by either Homework_X.j or SES-X.j. The models have one macro level explanatory variable, Sector, coded private=0, public=1.

Table 1. Effects on Sector of different treatment of data (standard errors)

	1. RS	2. CWC no Means	3. CWC with Means
Intercept	50.16	55.06	47.53
Homework	1.24 (.05)		
Homework_X.j		1.18 (.05)	1.20 (.05)
SES	4.35 (.09)		
SES_X.j		3.84 (.10)	3.85 (.10)
Sector	-2.06 (.29)	-5.42 (.34)	0.62 (.28)
MeanSES	not fitted	not fitted	8.14 (.25)
MeanHomework	not fitted	not fitted	1.65 (.20)
Var. slope	0.49 (.33)	0.40 (.09)	0.40 (.09)
	153333	153968	153004

Note that RS = Raw score model; CWC = Homework and SES scores are centered within context; CWC with means = the context means are reintroduced in the model and two extra parameter are estimated.

Table 1 shows that the results for the sector effect are different, depending on the treatment of the data. Conclusions range from a Public Sector effect that is moderately negative -2.06, (model 1), large negative, -5.42 (model 2), to positive +0.62 (model 3). Model 3 is the *R&B* model. These results illustrate that the raw score model is behaving in the same direction for Sector as the CWC model without means for Homework and SES. But the negative effect for the Public sector becomes twice as large in the CWC model without means. Reintroducing the means omitted in Model 2, changes the effect in an opposite direction in Model 3, where sector has a positive effect. The difference in deviances of the models in Table 1 shows that they are fitting the data not equally well.

Another important difference among models is that the variance for the slope of SES is not significantly random in the RS model (1), while it is significant in both CWC models (models 2 and 3). In Model 1 the variance is 0.49, with a large standard error of 0.33, while the same variance is estimated in the CWC models is 0.40, but with a small standard error of 0.09. The stability of these estimates, given the number of second level observations (1,001 schools) is sufficient and cannot serve as the explanation for these differences. It shows again that centering has an effect, this time on the variance components of the model.

The slope is not significantly random in Model 1, while highly significant and worth exploring in Models 2 and 3. *R&B*, using a CWC model, did explore this significant variance of the slope for SES by adding cross-level interactions. The *R&B* model is given in Model 3 of Table 2.

Table 2 Cross level interactions in RS and CWC (standard errors)

	1. RS	2. CWC no Means	3. CWC with Means
Intercept	54.00	60.19	49.57
Homework	0.86 (.10)		
Homework_X.j		0.48 (.10)	0.82 (.10)
SES	3.84 (.10)		
SES_X.j		2.94 (.26)	2.97 (.26)
Sector	-3.35 (.37)	-7.12 (.40)	-0.60 (.38)
Ratio	-0.16 (.02)	-0.29 (.03)	-0.06 (.02)
MeanSES	not fitted	not fitted	7.97 (.26)
MeanHomework	not fitted	not fitted	1.30 (.21)
PublicHomework	0.48 (.11)	0.88 (.11)	0.48 (.11)
PublicXSES	not fitted	1.03 (.28)	1.03 (.28)
	153272	153808	152964

The models in Table 2 have again SES and Homework as explanatory micro-level variables, either RS or CWC. Two school level explanatory variables are Sector (or public) and Ratio (teacher-student ratio). In the RS model no cross-level interaction with SES and Public is fitted. This decision is based on one tradition in multilevel modeling, that slopes which are not significantly random, are not examined further. Because the SES slope is significantly random in the CWC models (Models 2 and 3), a logical next

step is to fit an interaction with Sector and SES.
The slope for Homework is significant over all models and the cross-level interaction with the student level explanatory variable Homework and the school level explanatory variable Sector is fitted over all three models.

Again different conclusions for the effect of school level characteristics are reached when comparing the results of the RS model (Model 1) to the CWC models (Model 2 and 3) in Table 2. Sector and ratio have a coefficient of higher magnitude in the CWC model without the means reintroduced, which drops considerably in the CWC model where these means are present (model 3). The effect of the Public Sector drops close to zero with the reintroduction of SES and Homework school means. The correlation between Sector and Mean SES is r = -0.30, and between Sector and Mean Homework r = -0.54 (negative = high means go with the private sector, coded zero). Leaving the means out changes the Sector effect and the conclusions based on such a model; a phenomenon worth exploring further.

Theory driven choice of model

If theory development is the goal of the analysis, centering or not centering becomes an interesting issue (see earlier discussions regarding centering in the Newsletter by *Raudenbush* 1989a and b, *Longford*, 1989, and *Plewis*, 1989). Using centered scores (without context means reintroduced at the macro-level) yields a less well fitting model, because important context variation is eliminated. Researchers choosing this type of centering are either implicitly or explicitly assuming that this school level variable is not meaningful for their theory.

When including means in the model, the choice between a CWC model and a raw score model still needs theoretical rather than technical considerations. Technical considerations are, for instance, computational ease or interpretation. But centering a variable may change the meaning of the variable, because it splits the variable into a within and a between part. For instance, centering the variable gender makes the score of boys in a dominant girls class different from scores of boys in a dominant boys class. Boys are no longer boys. Same sex has no longer same score, but instead a deviation score from the class percentage of gender. If gender is centered around the class mean it follows logically that this class mean is introduced back in the model for the explanation of the rest of the gender differences in the response variable.

The relation between math achievement and homework across sectors provides another good example. The average homework of a school may reflect to some extent parental and school climate values, as may the average SES. The relationship of homework to math achievement in CWC models reflects the importance of the effort of each student within schools, not a general effect. A centered score for homework changes the meaning to the relative number of hours dedicated to homework by a student in comparison to the average number of hours of homework in that school. CWC models change the interpretation of parameter estimates and also the predictions based on such models compared to RS models. If homework is put in as a deviation from the school mean, homework can be seen as a relative effect, partly determined by school factors. Mean-Homework can be put back in the model as a school characteristic based on theory. For instance, more quality in teaching leads on average to less homework.

I believe that centering in multilevel analysis is an important issue that needs serious attention. The examples have illustrated that the new multilevel methodology has added some choices to the researcher, choices that make a difference. As a tool for theory development these choices create new opportunities. It offers more and different ways to analyze the same data, while it forces researchers to conceptually rethink their models. For policy research, I think, multilevel models can only be used if a strong theory directs the way either to center or not to center.

References

Bryk , A.S. and S.W. Raudenbush (1992). *Hierarchical Linear Models. Applications and Data Analysis Methods*. Newbury Park: Sage.

Bryk, Anthony S, Raudenbush, Stephen W, Seltzer Michael and Congdon Richard T (1988) *An Introduction to HLM: Computer Program and Users' Guide*. Coleman, J.S.

Hoffer, Th, and Kilgore, S. (1982). Cognitive Outcomes in Public and Private Schools. *Sociology of Education*, 55, 162-182.

Kreft, Ita G.G. and Jan De Leeuw and Leona S Aiken (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research ,* 30, 1, 1-21

Longford, Nicholas T. (1989) To Center or not to Center. *Multilevel Modelling Newsletter*, Vol **1**, 3 7,8 and 11

Plewis, Ian. (1989). Comment on Centering Predictors. *Multilevel Modelling Newsletter*, Vol.**1**, 3,.6 and 11

Raudenbush, Stephen W. and Anthony S. Bryk (1986). A Hierarchical Model for Studying School Effects. *Sociology of Education*, **59**, 1-17.

Raudenbush, Stephen W. (1989a) RCenteringS predictors in multilevel analysis: choices and consequences. *Multilevel Modelling Newsletter*, Vol.**1**, 2 , 10-12

Raudenbush, Stephen W. (1989b). A Response to Longford and Plewis. *Multilevel Modelling Newsletter*, Vol.**1**, 3, 8-10.

(*The complete article can be obtained as a Technical Report #30. from National Institute of Statistical Sciences, North Carolina, or from the author Kreft@math.ucla.edu)*

The Use of Multilevel Models for Screening Data Accumulated from a Number of Studies

Vanessa Simonite, Oxford Brookes University, UK.

In a recent issue of the Multilevel Modelling Newsletter, Langford and Lewis (1995) discussed in general terms the complexities of outlier detection in multilevel analyses and emphasised the importance of developing a multilevel approach to searching for outliers. Results are presented here to show how a multilevel approach to screening data was used to screen a database consisting of information collected within a number of studies.

Predicting Basal Metabolic Rate

Basal metabolic rate (BMR) is a key factor in estimating the food requirements of individuals or populations, but is costly and difficult to measure. For several decades, researchers have used predictive equations to estimate BMR from weight, which is easily obtainable under almost all circumstances. The predictive equations in use are those published by the FAO/WHO/UNU and are based on analyses of data collected world-wide in a large number of studies (Schofield, 1985). A new database consisting of data collected from approximately 200 studies provides the opportunity to derive new equations based on a larger number of individuals and to investigate variation between studies. To maintain the quality of the database, strict inclusion criteria defining appropriate methods and conditions, in particular for the measurement of BMR, were used to decide which studies would be used. The database consists of measurements for approximately 1,400 individuals within 200 separate studies; these individuals are divided into groups on the basis of age and sex and separate equations derived for each group. As the predictive equations must be capable of being applied world-wide and in a variety of situations, the objective is to use a simple linear regression of BMR on weight rather than to obtain a model with the best fit, perhaps by introducing additional independent variables.

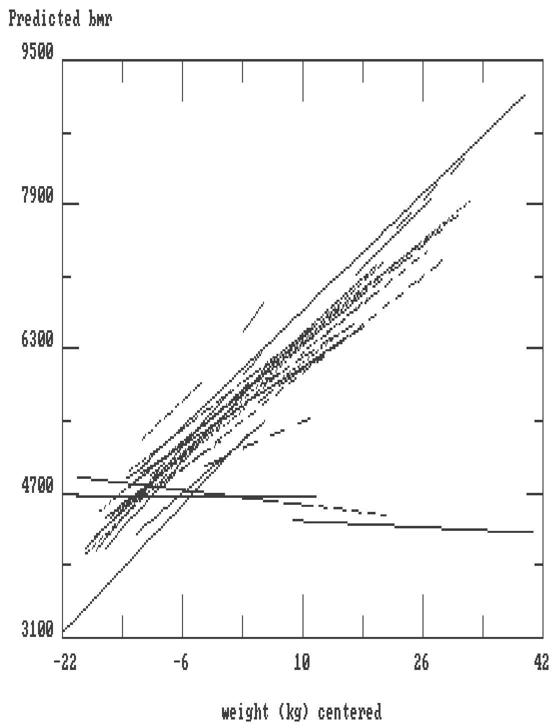
Results below are based on analyses of the data for male subjects aged between 11 and 18 years, using data from 28 studies. Screening was carried out at two levels ; first to identify outlying individuals and second to identify outlying or unusual studies. The results of screening at the study level are described below.

Screening for Outliers

Screening of individual cases identified a number of outlying individuals for whom there was strong evidence that BMR or other anthropometric measurements had been incorrectly recorded. These individuals were checked as far as possible and either corrected or excluded from all further analyses.

The next step was to fit a two-level, random slopes model for the regression of BMR on weight to the data ; for this, individuals were the level 1 units and studies the level 2 units. Inspection of the results showed substantial level 2 variation in both the intercepts and slopes; estimated variation in intercepts being $\sigma_u^2 = 1.65 \times 10^5 (se = 5.05 \times 10^4)$ and the estimated variation in slopes, $\sigma_v^2 = 8.64 \times 10^3 (se = 2.85 \times 10^2)$. The estimate of the correlation between the correlation was fairly high $(r = 0.73)$. Thus the slope, or mean BMR per additional unit of weight is correlated with the average BMR for individuals within a study. This latter result is not supported by the findings of past studies of BMR, in which graphs of BMR against weight have shown that where deviations from linearity occur, the points tend to follow a curve which becomes flatter, rather than steeper, with increasing BMR (Schofield, *op cit*.).

Estimates of the random coefficients for individual studies were examined and revealed three studies with particularly unusual values for both intercept and slope. These are illustrated by the following figure, which shows the fitted regression lines for all of the studies in the analysis. The most striking feature is the set of three regression lines which are flat, in contrast to the other regression lines with similar slopes.



The relationship between BMR and weight, being a physiological one, should vary little from study to study although minor variations due to a number of factors, such as climate or ethnicity may be expected. However, major differences between studies would occur mainly as the result of technical problems in the measurement of basal metabolic rate. For example, a consistent bias in the recording of BMR would generate regression lines with unusual intercepts, while studies with measurement errors correlated with the level of BMR would be expected to produce regression lines with unusual slopes.

One possibility is that within a particular study, influential values might be responsible for an unusual slope, but none of the three deviant lines could be attributed to influential values. Further, all three studies were large, with more than 100 individuals in each. Each of the three studies provided data which was unexceptional at an individual level, but in each case the study as a whole portrayed an aberrant model of the relationship between BMR and weight which contradicted the well established physiological relationship between weight and BMR. The conclusion of substantive discussions of these regression lines was that the measurement of one of the variables must have been at fault. The most likely explanation is that the recording of BMR was faulty, perhaps because one of the many requirements for measuring it accurately had not been met by the equipment or practice of the laboratory. As a result, the decision was that the measurement of BMR could not be relied on and therefore the studies should be excluded.

Analysis of Screened Data

Repeating the analysis after removing the three level 2 outliers produced both a dramatic decrease in level 2 variation and changes in the estimates of the fixed parameters. Variation in the intercepts fell by 69%, and variation in slopes fell by 84%. The correlation between random coefficients is changed from a high positive value to a moderate negative value of -0.4 .

The estimates of the fixed parameters also changed, to give a new equation for estimating basal metabolic rate. Removing the outliers reduced both the standard errors for these fixed parameter estimates and changed the estimate of the slope by a significant amount. For an individual with average weight, the effect of removing the outlying studies on predicted BMR is relatively small, however, the change in estimated slope means, for example, that for individuals with weights one standard deviation above the mean, the estimated BMR is 5% higher than estimated by the equation based on data from all studies. In practice, the underestimation of calorie requirements could have serious consequences.

Conclusion

Three studies, or level 2 units were identified as outliers. In each case it appears that failure to measure basal metabolic rate adequately may have produced large sets of data portraying a distorted relationship between BMR and weight. It is interesting to observe that as a result of detecting outliers at the second level , individuals from these studies were excluded, and these included individuals or level one units resembling individuals in other studies, who were not excluded.

For the analysis reported here, all the data from the outlying studies were rejected for substantive reasons: however in similar analyses for other age groups, some unusual values for the random coefficients were found to be the result of influential, individual data points within otherwise unexceptional studies.

Where data are merged from a number of sources, the value of the results rests on assumptions about the quality and conduct of contributing studies or centres. A multilevel approach to detecting outliers provides a valuable statistical tool for screening in this situation and is especially useful when, as in the example presented here, the researcher has no direct control or influence over the conduct of contributing studies.

References

Langford, I.H. and Lewis, T.L. (1995) Detecting Outliers in Multilevel Models: an overview. *Multilevel Modelling Newsletter*, 7(2), 6-7.

Schofield, W.N. (1985) Predicting Basal Metabolic Rate, New Standards and Review of Previous Work. *Human Nutrition: Clinical Nutrition*, 39C (Supplement1) 5-41.

MLA: Software for Two-Level Analysis with Resampling Options

Frank M.T.A. Busing, Rien van der Leeden & Erik Meijer, The Netherlands

Introduction

The MLA program is designed to analyse data with a two-level hierarchical structure. MLA has been especially designed for our own research purposes, in particular to study several resampling approaches. Nevertheless, the program contains some specific features that could make it a useful addition to the major packages for multilevel analysis available today: MLn, VARCL and HLM.

MLA can be characterized by four major properties:

- A user-friendly interface.
- One-step and two-step OLS estimation that may serve as a kind of diagnostic compared to the results of the usual maximum likelihood methods (a non-iterative weighted least squares method will also be implemented).
- A fast algorithm that uses the Broyden-Fletcher-Goldfarb-Shanno optimisation method to obtain maximum likelihood estimates of all model parameters simultaneously.
- Extensive options for simulation: four approaches to bootstrapping multilevel models and two jackknife methods.

The MLA program is able to fit a general class of two-level models, but with a simple Level-1 covariance structure. It has a limited option for imposing parameter constraints.

Using MLA

MLA uses ASCII text files as input and output. The input file consists of statements starting with a keyword. Models are specified by simply formulating the model equations. For example, maximum likelihood and bootstrap estimates for a random coefficient model are obtained by specifying:

/TITLE

MLA: random coefficient model

/DATA

file = subset.dat %data taken from NELS (see below)
vars = 17 %a total of 17 variables
id2 = 1 %Level-2 identification by first variable

/MODEL

v9 = b1 + b2*v5 + e % Level-1 equation
b1 = g1 + g2*v17 + u1 % Level-2: random intercepts, dependent on v17
b2 = g3 + g4*v17 + u2 % Level-2: random slopes, dependent on v17

/SIMULATION

kind = bootstrap % use simulation option bootstrap
method = error % resample error terms
type = shrunken % use shrunken residuals for error resampling
replications = 500 % repeat bootstrap simulation 500 times

/END

For this analysis (data from the 1988 National Education Longitudinal Study, National Center for Education Statistics, U.S. Department of Education, see also *Kreft* and *Van der Leeden*, 1994), the output includes:

```
FULL INFORMATION MAXIMUM LIKELIHOOD
ESTIMATES

  FIXED PARAMETERS

LABEL    ESTIMATE    SE     T     PROB(T)

G1    59.098244   6.547975   9.03    0.0000
G2   -15.827270   6.925261  -2.29    0.0223
G3     1.108726   4.648499   0.24    0.8115
```

```
G4    0.922200   4.916968   0.19   0.8512

  RANDOM PARAMETERS

LABEL    ESTIMATE     SE     T    PROB(T)

U1*U1  39.862886  20.314502  1.96  0.0497
U2*U1 -28.697576  14.153063 -2.03  0.0426
U2*U2  21.390295  10.257549  2.09  0.0370

E      42.782546   3.902927 10.96  0.0000

  (CONDITIONAL) INTRA-CLASS CORRELATION =
39.8629/(42.7825+39.8629) = 0.4823

  BOOTSTRAP ESTIMATES

    # REPLICATIONS         = 500
    # CORRECT REPLICATIONS = 500

  FIXED PARAMETERS

  LABEL     ESTIMATE            SE

    G1      59.393917       5.921939
    G2     -16.048766       6.311083
    G3       0.959815       4.268724
    G4       1.050352       4.525572

  RANDOM PARAMETERS

  LABEL     ESTIMATE            SE

  U1*U1    47.116121      12.025256
  U2*U1   -33.646603       8.159356
  U2*U2    25.036045       5.995707

    E      42.869828       4.190016
```

Additional output consists of summary statistics, one- and two-step OLS estimates, technical information, residuals, posterior means, separate output of simulation results and simple diagnostics. Preliminary results from an evaluation study of the different bootstrap methods are reported in *Meijer*, *Van der Leeden* and *Busing* (1995) and *Busing*, *Meijer* and *Van der Leeden* (1995).

Distribution information

MLA runs as a stand-alone batch program on any IBM-PC/AT, PS/2 or compatible under DOS or OS/2. A minimum of 256Kb of free RAM is necessary. MLA also runs in a DOS environment under WINDOWS'95 or OS/2. The program can handle up to 16 equations with 32 terms each. Limitations are 16000 Level-2 units, 8000 Level-1 units per Level-2 unit, 16000 variables, 16000 simulation replications and 64 constraints.

The MLA program has been, and is being further developed by *Frank Busing*, *Erik Meijer* and *Rien van der Leeden*, and can be obtained from the authors (for contact address see below). Although MLA is not intended to be a commercial product, we ask a fee of U.S. $20 to cover some of the expenses concerning administration and distribution. The distribution disk includes executables, examples and a postscript file containing a copy of the comprehensive user's guide for Version 1.0b (Busing, Meijer and Van der Leeden, 1994). If you have interest in MLA, please contact:

Rien van der Leeden
Department of Psychometrics and Research Methodology, University of Leiden
P.O. Box 9555, 2300 RB Leiden
phone +31 71 527 3763 / 3761
fax +31 71 5273619
email vanderleeden@rulfsw.leidenuniv.nl / busing@rulfsw.leidenuniv.nl

(Editor's note: It is planned to make this software available from the Multilevel Models Project Web site shortly)

References

Busing, F.M.T.A., Meijer, E. & *Van der Leeden, R.* (1994). MLA. Software for multilevel analysis of data with two levels. PRM 94-01. Leiden, The Netherlands: Leiden University, Department of Psychometrics and Research Methodology.

Busing, F.M.T.A., Meijer, E. & *Van der Leeden, R.* (1995). The MLA program for two-level analysis with resampling options. In: T.A.B. Snijders et al (Eds.). Toeval zit overal [Randomness is everywhere]. Symposium Statistische Software 1995 (pp. 37-57). Groningen, The Netherlands: iec ProGAMMA.

Kreft, I.C.G. & *Van der Leeden, R.* (1994). Random coefficient linear regression models. PRM 94-03. Leiden, The Netherlands: Leiden University, Department of Psychometrics and Research Methodology.

Meijer, E., Van der Leeden, R. & *Busing, F.M.T.A.* (1995). Implementing the bootstrap for multilevel models. Multilevel Modelling Newsletter, 7(2), 7-11.

Multilevel Analysis Discussion List

John Roberts, University of Manchester, UK

This electronic mail discussion list was set up in August 1995. It provides a forum for discussion, and a means of disseminating information easily, for people using multilevel analysis and any associated software. There are now over 700 members, with one or two messages per day on average.

Any messages relevant to multilevel analysis are welcome. Information sent to the list can be, for example, about seminars, courses, conferences, job vacancies, and new software. Discussion can be about methods of analysis, interpretation of results, and so on. Examples on discussion during the last few months are as such as: <u>extra-binomial variation</u>, <u>proportion variance explained</u>, <u>path model with dichotomous dependent variable</u>, <u>sample size for level-2 variable</u>, <u>missing data problems</u>, <u>incomplete hierarchies and MLn</u>, <u>interpretation of HLM coefficients in growth study</u>, <u>intraclass correlation</u> and *et al*.

Hints, comments, and problems with software can also be sent to the list. The software can be one of the specialist multilevel analysis packages such as MLn, HLM, VARCL, GENMOD, or more general packages with some multilevel capabilities (e.g. SAS).

Files containing previous messages can be retrieved if required. It is intended that list will also have associated with it other files. These will include reports and other documents of long-term value, and also software programs (e.g. MLn macros), deposited by list members for the use of others.

The list uses the Mailbase system at the University of Newcastle, and was set up with encouragement from the Multilevel Models Project (Institute of Education, London), and the Census Microdata Unit (University of Manchester). The list owner is *John Roberts*, MIDAS Service, University of Manchester, who can be contacted by emailing *multilevel-request@mailbase.ac.uk*

Anyone with an interest in multilevel analysis can join. The list is UK-based, but list members can be from any country.

To join the list, send a message to

mailbase@mailbase.ac.uk

The message should contain a single line, with a command of the form

join multilevel firstname(s) lastname

Do not copy this exactly, but substitute your name! Lastnames must be hyphenated, or use underscores to preserve lower-case parts of the name. Valid examples are:

join multilevel John Smith

join multilevel Professor John Arthur Smith

join multilevel John Blashford-Smith

join multilevel Johannes _van_der-Waals

Further instructions about the use of the list will be sent after joining.

Letters & Responses

Dear Editor,

Although articles in the Newsletter are primarily concerned with the methodology of multilevel modelling other statistical considerations should not be ignored.

I refer to *Lambert* & *Abrams* (1995) which gives a nice example of multilevel modelling with given level 1 variances but unfortunately regresses the log odds ratio on the baseline risk without giving a health warning!

It is well known *Oldham* (1968), *Cox* & *Snell* (1981), that for matched pair data, u, v the measurement errors of $(u-v)$ and of v are negatively correlated so that obtaining a negative regression coefficient does not imply a structural relationship unless the measurement error (level 1) of v is small compared to the variation between studies (level 2).

Figure 1, Page 19 of *Lambert* and *Abrams* (1995), shows a clear negative association between the log odds ratio and the level of baseline risk - but is this spurious, arising out of measurement error in the baseline risk?

Not according to the source article, *Davey Smith, Song* & *Sheldon* (1993) pages 1369-70, who state that more valid regression on the overall risk gives a virtually identical result to that on the baseline risk. It is a pity that Lambert and Abrams did not refer to this.

Also they do note make it clear whether they are analysing total mortality or mortality from chronic heart disease.

Finally including such very different studies (*Davey Smith*, *Song* & *Sheldon*, 1993, Table 1) in a single meta-analysis casts doubt on the validity of the random effects (multilevel) model. maybe some residual analyses (following *Langford* and *Lewis*, 1995) could be used to examine this!

(Mrs J.I. Galbraith)

Dear Mrs Galbraith,

Thank you for your interest in the above article.
The purpose of the article was to demonstrate on a previously analysed data-set (*Davey Smith*, *Song* & *Sheldon*, 1993), that similar results could be obtained using *ML3* when compared to the standard analysis (*Dersimion* & *Laird*, 1986).

You correctly state that the negative association observed between the log-odds ratio and the level of baseline risk could be biased. As you point out, this issue was addressed in the original article where a regression analysis was performed using mortality from coronary heart disease averaged over both the treatment and control groups. This yielded very similar results to the original analysis. Also in the original article it was argued that the large range of coronary heart disease death rates observed makes it unlikely that using the baseline risk would be a major source of bias.

You also state that the inclusion of such different studies in a single meta analysis casts doubt over the validity of the random effects model. There has been much debate on the various approaches to meta-analysis (*Jones*, 1995) and the purpose of our article was not to address this issue but to demonstrate that the mixed effects models can be performed in *ML3*. However, as we discuss, the differences in the trials could be modelled. For example, one could model level 1 if the trial concerned drug or non-drug treatments.

(*Paul Lambert, Keith Abrams*)

References

DR Cox & EJ Snell (1981), *Applied Statistics*, Chapman & Hall, London, pages 73, 74.

Davey Smith G., Song F. and Sheldon T. A. (1993), Cholesterol lowering and mortality: the importance of considering initial level of risk, *British medical Journal* **306**: 1367-1373.

DerSimonian R.D. and Laird N. (1986), Meta-analysis in clinical trials, *Controlled Clinical Trials*, **7**: 177-188.

PC Lambert & KR Abrams (1995), Meta-Analysis Using Multilevel Models, *Multilevel Modelling Newsletter*, Vol. **7** No. **2**, 17-19.

IH Langford & T Lewis (1995), Detecting outliers in multilevel Models: an overview, *Multilevel Modelling Newsletter*, Vol. **7** No. **2** , 6-7.

PD Oldham (1968), *Measurement in Medicine*, English Universities Press, London, 148-151, 198-200.

Jones D.R. (1995), Meta-analysis: weighting the evidence, *Statistics in Medicine*, **14**: 137-149.

Contributors

We are most grateful to the following people who contributed articles and information to this issue.

Joop Hox, Department of Education, University of Amsterdam, IJsbaanpad 9, NL-1076 CV Amsterdam, the Netherlands. Tel +.31.20.5703530, Fax 5703500. Email: hox@educ.uva.nl.

Harvey Goldstein, Mathematical Sciences, Institute of Education, University of London, 20 Bedford Way, London WC1 0AL, England. Tel: 44 (0)171 612 6652, Fax: 44 (0)171 612 6686. Email:hgoldstn@ioe.ac.uk.

Rien van der Leeden, Dept Psychometrics & Research Methodology, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands. Email: vanderleeden@rulfsw.leidenuniv.nl.

Ita Kreft, California State University. Email: kreft@smath.ucla.edu.

John Roberts, MIDAS Manchester Computing (London Office), 20 Guildford Street, London WC1N 1DZ, UK. Email: r.j.roberts@mcc.ac.uk.

Vanessa Simonite, School of Computing & Mathematical Sciences, Oxford Broodkes University, Headington Oxford, OX3 0BP, England. Tel: 44 (0)1865 483652, fax: 44 (0) 1865 483666. Email: vsimonite@brookes.ac.uk