MULTILEVEL MODELLING NEWSLETTER

The Multilevel Models Project:

Dept. of Mathematics, Statistics & Computing
Institute of Education, University of London

20 Bedford Way, London WC1H 0AL, ENGLAND
e-mail: temsmya@ioe.ac.uk

Telephone: 071-612-6682 Fax: 071-612-6686

Vol. 5 No. 2 July, 1993

The new version of A Guide to ML3 for New Users: A new edition of the *Guide to ML3 for New Users*, also known as the 'Red Book', is now available. In addition to the introductory chapter, there are now three further chapters showing how to use ML3 to develop

- multivariate models
- logistic models for binary responses
- repeated measures models

Each chapter provides step-by-step instructions for you to follow on your own computer, and with interpretations.

The new edition, with accompanying data sets, is available from the Multilevel Models Project, price £20.

The new version of ML3-E: Version 2.3 of the ML3-E is now available. It can handle models with upto 200 explanatory variables and 200 random parameters. The respective limits in version 2.2 were 75 and 75. Version 2.3 also contains new commands for fitting randomly cross-classified models. If you would like to be sent a copy of the new version please contact the Multilevel Models Project, quoting your user registration number.

Workshops on Three-level Modelling: The next course on using the ML3 package on **2-5 November 1993** is full already. A further workshop is scheduled for **1-4 March 1994.** If you wish to book a place or to be put on the waiting list for further workshops, please contact *Min Yang* at the project address.

ML3 Clinics in London 1993

Free for users of *ML3/ML3-E*

Tuesday September 14
Tuesday October 12
Tuesday November 9
Tuesday December 7

10.30 am - 5.30 pm,
Multilevel Models Project
11 Woburn Square, Second floor
London WC1A 0SN
Call *Min Yang* for an appointment
Tel: 071 612 6682

Also In This Issue

The Multilevel Models Project is supported by a grant from the Economic and Social Research Council, U.K.

Multilevel models and generalised estimating equations

Harvey Goldstein

Liang and Zeger (1986) propose an estimation procedure, Generalised Estimating Equations (GEE), for generalised linear models with complex residual structures. The purpose of this note is to point out that, with certain differences, GEE parallels the iterative generalised least squares procedure for multilevel models introduced by Goldstein (1986) and more specifically the extension of this to nonlinear models, including generalised linear models (Goldstein, 1991). In the case of the identity link function and Normality the multilevel estimates are maximum likelihood or restricted maximum likelihood, as are those obtained with the EM algorithm (Bryk and Raudenbush, 1992) or the Fisher scoring algorith (Longford, 1987)

The GEE equations can be written as the solution of

$$\left(\frac{\partial\mu}{\partial\beta}\right)^T V^{-1}(Y-\mu)=0 \qquad (1)$$

where the linear predictor $E(Y)=\mu$ and V is a general weight matrix. Equation (1) defines a general quasilikelihood model (McCullagh and Nelder, 1989) and is essentially equivalent to Goldstein's (1991) procedure. The principal difference lies in the form of the estimators for V.

Even though GEE does not lay stress on modelling V explicitly in terms of random coefficients, V *is* assumed to depend on a vector of parameters α which is chosen to yield a structure for V specific to each application. The estimation of α is based upon current estimates of the (standardised) raw residuals, typically using straightforward regression or moment type estimators. This gives a new estimate of V and thus updated estimates for β via an application of (1) and the estimation proceeds iteratively, given suitable starting values.

It is shown by Liang and Zeger (1986) that the estimates so obtained are consistent and asymptotically Normal, even where V is misspecified. GEE is mainly concerned with the 'fixed' part of the model, that is the coefficient vector β, and it uses an empirically based robust estimator for $cov(\hat{\beta})$ based upon a consistent estimator of $cov(Y)$, namely $\tilde{Y}\tilde{Y}^T$ where $\tilde{Y}=Y-\hat{\mu}$.

This estimator is consistent but not fully efficient and while the GEE estimators of β are also consistent they are not fully efficient unless V is correctly specified. It is also worth noting that robust estimators for parameter covariance matrices similar to those used by GEE can be used for the standard multilevel estimators.

In contrast to GEE the multilevel approach explicitly models V as a function of coefficients of explanatory variables where the coefficients can be random at one or more levels of the data hierarchy and where the explanatory variables can be quite general, allowing for example autoregressive or cross-classified structures to be specified. It uses estimators which, under the assumed model, are efficient for V. More important perhaps than its theoretical improved efficiency, is that it views the structure of the random variation as substantively important, and provides very general procedures for modelling that structure.

One may regard the GEE procedure as providing simple approximate estimators for the fixed parameters of a multilevel model which could be useful for providing starting values for a full multilevel analysis, especially since typically it will be computationally faster. It may also be useful if the data analyst does not wish to make strong parametric assumptions about V.

References

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

Goldstein, H. (1986). Multilevel mixed model analysis using iterative generalised least squares. Biometrika, 73, 43-56

Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. Biometrika, 78, 45-51

Liang,K. and Zeger, S.L. (1986). Longitudinal data analysis using generalised linear models. Biometrika, 73, 13-22

Longford, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Biometrika, 74, 817-827.

McCullagh,P. and Nelder, J.A. (1989). Generalized Linear Models. London, Chapman and Hall.

Pattern probabilities in paired binary data

N. T. Longford

Introduction

Logistic regression with random coefficients is a natural adaptation of the normal random coefficient models for binary data. In the normal case we consider for individuals $i = 1, \ldots, n_j$ within clusters $j = 1, \ldots, N_2$ the model

$$y_{ij} = X_{ij}\beta + Z_{ij}\delta_j + \varepsilon_{ij} \qquad (1)$$

where $\{y_{ij}\}$ are the outcomes, $\{X_{ij}\}$ and $\{Z_{ij}\}$ are known vectors, β is a vector of unknown (regression) parameters, and $\{\delta_j\}$ and $\{\varepsilon_{ij}\}$ are two mutually independent random samples from $N(0,\Sigma)$ and $N(0,\sigma_W^2)$, respectively. Conventionally, the first components of X and Z correspond to the intercept (are identically equal to unity). For binary outcomes $\{y_{ij}\}$ we consider the model

$$P(y_{ij} = 1 \mid \delta_j) = logit^{-1}(X_{ij}\beta + Z_{ij}\delta_j),$$

with the same assumptions for X_{ij}, Z_{ij} and δ_j as in (1). The variables in X may be defined for individuals or clusters, those in Z are defined for individuals. Since binary data contain very sparse information about the between-cluster variation, it is often meaningful to consider only models with

$$Z_{ij} \equiv 1,$$
$$P(y_{ij} = 1 \mid \delta_j) = logit^{-1}(X_{ij}\beta + \delta_j).$$

The between-cluster variance, the sole element of Σ, is denoted by σ_B^2 ($var(\delta_j)$). See Longford (1993) for background and a review of alternative approaches.

In normal random coefficient models the between-cluster variance σ_B^2 has a natural interpretation; it is the covariance of two observations in the same cluster, adjusted for the explanatory variables. Also, the decomposition of the total variance $\sigma_W^2 + \sigma_B^2$ into its within- and between-cluster components in helpful; hence the term *variance components.*

Of course, this interpretation does not carry over to the binary case. Although the covariance, or correlation, of a pair of binary observations in a cluster can be obtained, e.g., by application of the delta method, it does not convey the association of the binary outcomes within a cluster as clearly as the probabilities of the patterns (0,0), (0/1), and (1,1) for a pair of binary outcomes in a cluster. These probabilities, as functions of the between-cluster variance σ_B^2 and the common linear predictor $\mu = X\beta$, cannot be expressed in a closed form, but can be obtained by simulation. For instance, the probability of the pattern (0,1) is

$$P\{(0,1)\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp\frac{(\mu+\sigma_B\delta)}{\{1+\exp(\mu+\sigma_B\delta)\}^2}\exp\left(-\frac{\delta^2}{2}\right)d\delta$$

$$= \frac{1}{\sqrt{2\pi\sigma_B^2}}\int_{-\infty}^{+\infty} \delta\exp\frac{(-\delta^2/2)}{\{1+\exp(\mu+\sigma_B\delta)\}}d\delta \qquad (2)$$

(the identity is obtained by integration by parts). Either expression in (2) can be evaluated by Gaussian quadrature for an array values of μ and σ_B^2. Owing to symmetry it suffices to consider $\mu \geq 0$. the probability in (2) can be evaluated efficiently by generating a random sample of δ's, calculating the corresponding conditional probabilities $P\{(0,1)\} = P(0 \mid \delta) \times P(1 \mid \delta)$, and drawing binary outcome with these probabilities. The procedure for the other patterns is analogous.

The unexpected finding is that the marginal probabilities of these patterns are approximately linear functions of the standard deviation σ_B for a wide range of values of μ and σ_B. This may promote understanding of the between-cluster variation in binary data.

The top left-hand panel of Figure 1 displays the plot of probabilities of agreement (pattern (0,0) and (1,1)) and disagreement ((0,1) or (1,0)) for a pair of binary outcomes from the same cluster, as a function of the cluster-level variance σ_B^2, in the range $0 < \sigma_B < 3$. The solid line is the function, and the dotted line is its ordinary least squares approximation, given in the first row of Table 1. Although the approximation is far from perfect, it is sufficient for the practitioner' orientation: for unconditional probabilities around 0.5 the probability of agreement increases by almost 0.1 for every unit of σ_B on the logit scale.

Note that $0 \leq \sigma_B \leq 3$ is a very wide range; $\sigma_B = 3$ corresponds to extreme between-cluster variation. For instance, if the linear predictor μ is equal to zero (probability 0.5 in an 'average' cluster), in a sample of 100 clusters there are likely to be several with conditional probabilities $logit^{-1}(\mu + \sigma)$ greater than 0.9 or smaller than 0.1.

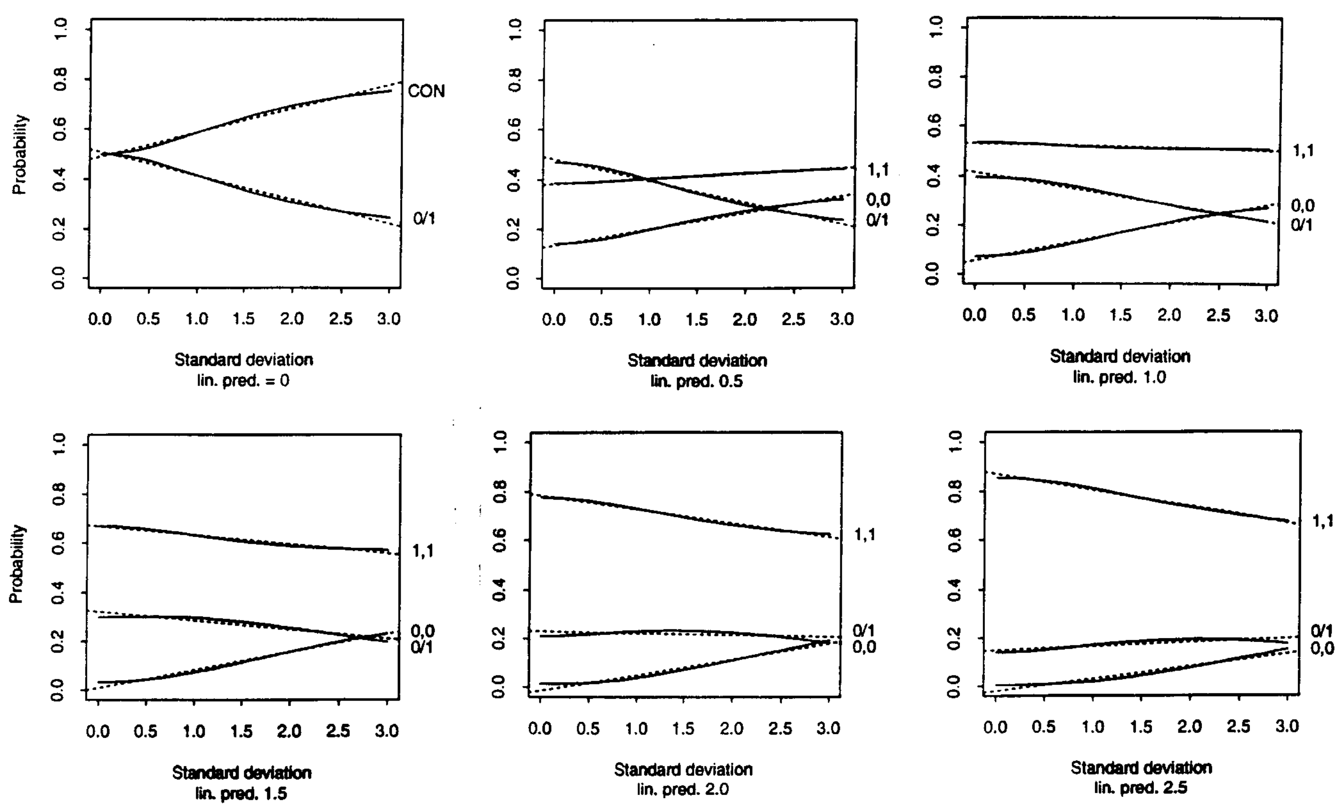
Table 1: Linear approximations to the pattern probabilities for a pair of binary outcomes as functions of the (logistic) between-cluster variance σ_B^2 and the (logistic) linear predictor μ.

μ	(1,1)	(0/1)	(0,0)
0	$0.246 + 0.0955\sigma_B$	$0.508 - 0.0955\sigma_B$	$0.246 + 0.0955\sigma_B$
0.5	$0.384 + 0.0205\sigma_B$	$0.482 - 0.0874\sigma_B$	$0.134 + 0.0669\sigma_B$
1.0	$0.531 - 0.0097\sigma_B$	$0.413 - 0.0657\sigma_B$	$0.056 + 0.0754\sigma_B$
1.5	$0.669 - 0.0372\sigma_B$	$0.322 - 0.0364\sigma_B$	$0.010 + 0.0736\sigma_B$
2.0	$0.784 - 0.0573\sigma_B$	$0.229 - 0.0070\sigma_B$	$-.013 + 0.0634\sigma_B$
2.5	$0.870 - 0.0675\sigma_B$	$0.149 + 0.0161\sigma_B$	$-.020 + 0.0514\sigma_B$

The other panels of Figure 1 display plots of the probabilities of the three patterns as functions of $0 \le \sigma_B \le 3$ for the values of the linear predictor $\mu = 0.5, 1, 1.5, 2.0$ and 2.5. The dotted lines are the linear approximations given in Table 1. The linear approximations are very good for $\mu \le 1.5$ and even for larger μ for the pattern (1,1). The linear predictor $\mu = 1.5$ corresponds to the unconditional probability $p = 0.82$.

More complex approximations, such as by bilinear functions are not very useful, and would not be as informative. A practitioner with experience in Splus or another statistical package can easily compute these probabilities.

Figure 1: Probabilities of patterns of paired binary outcomes as functions of the linear predictor μ and the between-cluster standard deviation σ_B; logistic scale. The functions are drawn by solid and the linear approximations by dotted lines. Table 1 contains the equations for the linear approximations. For $\mu = 0$ the probabilities of agreement, patterns (0,0) or (1,1), and disagreement, (0,1) or (1,0) are plotted, denoted respectively by 'CON' and '0/1'. For the other values of μ the probabilities of the three patterns are plotted.



Summary

The purpose of this paper is to bring attention of researchers using logistic regression with random effects to the fact that the between-cluster standard deviation is approximately linearly related to the probabilities of patterns for paired outcomes for a wide range of values of the linear predictor. This may be helpful in understanding the substantive importance of the between-cluster variance of a given size.

Reference

Longford, N.T. (1993) Logistic regression with random coefficients. Computational Statistics and Data Analysis. To appear.

...*...

ESRC (U.K.) Funds for Developing and Using Multilevel Models in the Analysis of Large and Complex Datasets

The Economic and Social Research Council (U.K.) has recently funded a group of projects for developing the analysis of large and complex datasets in the social sciences, aiming to advance and improve statistical analysis and modelling in the context of substantive research. Several of the projects aim to apply and develop multilevel models substantively for tackling the issues of complexity in large datasets in their relevant areas during a 2 - 3 years' period. Full details can be obtained from ***Paul Rouse, ESRC, Polaris House, North Star Avenue, Swindon, SN2 1UJ.***

The Multilevel Models Project team has been funded for three years under this initiative. The project will develop existing methodology for modelling multilevel data directly to address the problems associated with large and complex datasets. It will collaborate with social researchers jointly to tackle methodological and substantive issues and to explore the potential of the methodology for providing new insights into complex data structures. The project will disseminate knowledge and expertise in this area by means of workshops, a newsletter, talks and academic publications. It will extend existing software to exploit new hardware and software environments.

CONTRIBUTORS

We are grateful to the people who provided articles for this issue.

Van der Beek, M.C.J.
ACTA, Department of Orthodontics
Louwesweg 1
1066 EA, Amsterdam.

Goldstein, H.
Rasbash, J.
Woodhouse, J.
Department of Mathematics, Statistics and Computing
Institute of Education, University of London
20 Bedford Way, London WC1A 0AL
U.K.

Hoeksma, J.B.
Free University,
Department of Developmental Psychology
Van der Boechorstraat 1
1081 BT, Amsterdam

Longford, N.T.
Education Testing Service
Princeton, NJ 22T
USA

Jones, K.
Department of Geography, University of Portsmouth
Buckingham Building, Lion Terrace, PO1 3HE
U.K.

Tanguay, R
Faculte de medecine dentaire, Universite de Montreal
C.P. 6128, Succ. A, Montreal, Quebec
Canada H3C 3J7

APPLICATIONS

Analyzing age dependent correlations in the study of growth and development

(Jan B. Hoeksma and Mirjam C.J. Van der Beek)

The field of growth and development offers prolific opportunities for applying the longitudinal multilevel model. Goldstein (1989) for instance showed how the longitudinal multilevel model can be used to predict adult height. Hoeksma and Koomen (1992) demonstrated how the multilevel model relates to the five rationales of longitudinal developmental psychological research, as they were formulated by Baltes and Nesselroade (1979). And to name just another example, Bryk and Raudenbush (1992) explained how the longitudinal model can be used to model educational growth.

The longitudinal multilevel model draws its value from the fact that it enables researchers to study growth and development at an individual level without resorting to analyses of individual cases! This is an important advantage, because substantive research questions with respect to growth and development generally refer to the individual, but are nevertheless often analyzed at a mean level. In other words the longitudinal multilevel model narrows the gap between research question and analysis. From this point of view the most valuable information gained from a longitudinal multilevel analysis comes from the random part of the model.

In this article we further substantiate the value of the longitudinal multilevel model by using it to estimate the correlation between two dynamic variables as a function of time or age. The data are taken from the field of physical growth, more specifically from the field of orthodontics. The application is meant as an example. As such it can be transferred easily to other research areas within the large field of growth and development.

Aim and Data

One of the main objectives of orthodontic treatment is to guide cranio-facial growth. To attain this goal, orthodontists must be able to gauge the child's growth status. This is generally based on the child's standing height relative to its age. Orthodontists seem to assume tacitly that the relationship between physical and facial growth does *not* depend on age. In addition they seem to assume that changes in one dimension are parallelled by changes in the other. The aim of our analysis is to find out whether these implicit assumptions are correct.

The sample consisted of 71 girls from the Nijmegen growth study. The number of measurement occasions, in the age range from 7.9 to 14.3 years, varied form seven to ten per girl. The two variables of interest were facial height and standing height. Facial height was measured from cephalograms. It roughly corresponds with the distance from the midpoint between the eyes to the most prominent point on the chin. Orthodontists call this distance "Nasion-Gnathion" and use it to distinguish short and long faces. The number of measurements for physical height amounted to 503 and for Nasion-Gnathion to 506.

Model

Is the correlation between facial and standing height constant across age? If not, how does it change with age? And are changes in physical height parallelled by changes in facial height and vice versa. Does this assumed parallelism change with age? To answer these questions we use the bivariate polynomial model.

Let y_{hij} refer to the measurement of variable h $(h = 1,2)$ at occasion i for child j. More specifically, let y_{1ij} refer to standing height and y_{2ij} to facial height. If child j's age at occasion i is designated by x_{ij} the model can be written as:

$$y_{hij} = \Sigma_t \beta_{ht} x_{ij}^t + \Sigma_t u_{hjt} x_{ij}^t + e_{hij} \qquad (1)$$
$$(t = 0, 1, \ldots, s)$$

where β_{ht} refer to the average growth curves for physical height $(h = 1)$ and facial height $(h = 2)$. The u_{hjt} are random variables referring to individual differences with: $E(u_{hjt}) = 0$, $Var(u_{hjt}) = \sigma^2_{ht}$ and $Cov(u_{hjt}, u_{h'jt'}) = \sigma_{ht,h't'}$. The level 1 (between occasion) variance is designated by $Var(e_{hij}) = \sigma^2_h$.

The random part of the model contains the information central to our research questions. The correlation between standing height and facial height at age x is found directly by computing the covariance between the two measures, $Cov(\Sigma_t u_{1t}x_{ij}^t, \Sigma_{t'} u_{2t'}x_{ij}^{t'}) = \Sigma_t\Sigma_{t'} x_{ij}^t x_{ij}^{t'} \sigma_{1t,2t'}$ and their variances, $Var(\Sigma_t u_{hjt}x_{ij}^t) = \Sigma_t x_{ij}^{2t}\sigma^2_{ht} + 2\Sigma\Sigma_t x_{ij}^t x_{ij}^{t'} \sigma_{ht,ht'}$ from the fitted model random parameters. In this calculation we might choose to omit the level 1 variation if this was considered largely to the measurement error.

The next question is whether changes in the two variables are related to each other. To what extent is the growth velocity of facial height related to the growth velocity of physical height? Velocity curves are obtained by taking the first derivative of the growth curve. Then the variances and covariance of the velocity are based on $\Sigma_t t x_{ij}^{t-1} u_{hjt}$, and the correlation between the growth velocity of standing height and facial height is readily estimated.

Results

The parameters of the bivariate model with standing height and facial height as dependent variables were estimated using ML3 (Prosser, Rasbash and Goldstein, 1991). Before the analyses the age variable was centred around 11 years of age. Physical height is expressed in centimetres facial height in millimetres.

The analysis took several steps. At each step a variance and associated covariance terms were added. The number of iterations amounted to approximately 380. In the final model the average growth curve of both physical and facial height followed a fourth degree polynomial. The parameter estimates are displayed in Table 1.

Table 1. Estimates of fixed parameters, including standard errors.

	est.	s.e.
β_{10}	145.6	0.73
β_{11}	6.37	0.16
β_{12}	0.28	0.082
β_{13}	-0.055	0.014
β_{14}	-0.036	0.006
β_{20}	109.2	0.649
β_{21}	2.28	0.119
β_{22}	0.10	0.073
β_{23}	-0.008	0.012
β_{24}	-0.017	0.007

Table 2: Variance covariance matrix of random coefficients, standard errors in brackets

	β_{10}	β_{11}	β_{12}	β_{20}	β_{21}	β_{22}
β_{10}	37.95 (6.4)					
β_{11}	3.73 (1.1)	1.66 (0.31)				
β_{12}	-1.86 (0.46)	-0.40 (0.10)	0.25 (0.05)			
β_{20}	14.64 (4.36)	1.40 (0.90)	-0.55 (0.36)	29.49 (5.00)		
β_{21}	1.86 (0.66)	0.79 (0.17)	-0.26 (0.06)	1.82 (0.60)	0.56 (0.13)	
β_{22}	-0.79 (0.24)	-0.20 (0.06)	0.10 (0.02)	-0.01 (0.20)	-0.08 (0.03)	0.052 (0.02)

For both dimensions the random coefficients could be estimated reliably up to the second degree (Table 2).

The estimated residual (level 1) variance for physical height was $\sigma_1^2 = 0.48$ (s.e.=0.038) and for facial height $\sigma_2^2 = 1.12$ (s.e.=0.086). Note that physical height is measured in centimetres, whereas facial height is measured in millimetres. For both dimensions the residual variance is rather small. The level 1 residuals were assumed adequate.

Next we computed the correlation between standing height and facial height conditional on age, in order to find out whether the relationship between physical and facial height changes with age. The level 1 variances were not used in the calculation.

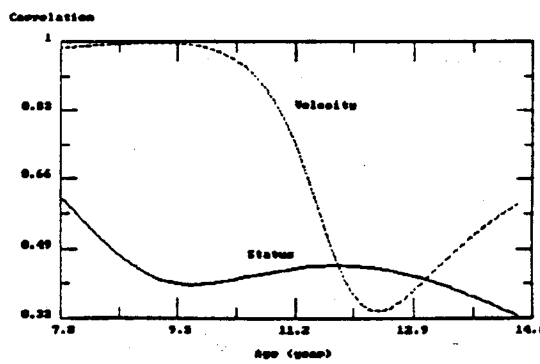


Figure 1. Estimated correlations of growth status and velocity as functions of age

The result is displayed in figure 1. In accordance with observations from clinical practice, facial and physical height are related to each other. The relationship is medium at early age, then declining afterwards. There is a rising wave of the correlation during the pubertal growth period. By about 15 years old the relationship becomes rather weak.

The observed relationship between the two variables may be inflated somewhat because the level 1 variance is not taken into account. This variance is however very small. As a consequence the correlations with and without error are hardly different.

Finally we computed the relationship between the growth velocities of the two dimensions, in order to see wether the changes in the two dimensions parallel each other. The result is also displayed in figure 1. The correlation between the growth statuses and between the growth velocities follow a different pattern. The positive correlation of the latter is very strong before the pubertal growth spurt starts, followed by a dramatically decrease and reaches the lowest point at around age 12. After that the correlation goes up again.

Discussion

Our major goal was to explain how the longitudinal multilevel model can be applied to describe the correlation between two dynamic variables as a function of age. The application described can be easily transferred to other domains, for instance developmental psychology in order to answer questions about synchronous and co-development.

An important question is whether the results presented are valid. A complete answer to this question is beyond the scope of the present article. Our provisional answer is yes. The dramatic changes in the correlations around age ten, both with respect to status and velocity are the result of pubertal growth, which mainly affects physical height. As a result the original close relationship between physical and facial height is lost for some time. After puberty the relationship for growth velocity is reestablished again.

Multilevel Models Project Gets Its Own "ANONYMOUS FTP SITE SERVER"

A new Anonymous FTP (File Transfer Protocol) site is now available for users of ML3 around the world. You can reach this site via Internet at the following address:

amalgame.medent.umontreal.ca
or at IP address
132.204.154.20

The home directory for the ML3 PROJECT is */pup/ML3*. At this early stage, only the */pub/ML3/bin_resp/* directory is of interest. It contains all the macros developed by the Multilevel Models Project for analyzing Multilevel binary response logistic models. Included in this directory a postscript file of the "Guide to ML3 Macros" in both British and American paper size.

Each sub-directory contains the following two files : README and INDEX. The README file gives a brief description of the purpose of the files included in the sub-directory and special instructions if needed. The INDEX file contains a list all the files in the sub-directory with a short description. The README and INDEX files may be downloaded in ASCII mode.

For those who are not familiar with the ftp procedure consult your local system ftp help. Basically you connect to the anonymous server as follows:

ftp amalgame.medent.umontreal.ca
Name: anonymous
Password: type in your e-mail address. (e.g. tangr@medent.umontreal.ca)
ftp> cd /pub/ML3

You can now issue different commands to set the transfer mode in ASCII (***ascii***) or in BINARY (***bin***) mode, to get a file (***get*** "filename"), to list the directory (***dir***), to change directory (***cd***) and to exit from the server (***quit***). It is however good practice to first get the README and INDEX files in order to know which files you have to transfer for your own needs.

Feel free to report bugs and tips in an ascii or postscript file that you put in the ***/pub/incoming*** directory by issuing the ftp command : ***put*** "filename" and make sure you give your e-mail address in your document. Note however that the bugs and tips you will put in the ***/pub/incoming*** directory will be submitted to the Multilevel Team and thereafter put in the ***/pub/ML3/BUGS_TIPS*** directory .

Have a nice FTP session!

Richard Tanguay
tangr@medent.umontreal.ca

Modelling Random Cross Classifications in ML3

Jon Rasbash

The motivation for doing multilevel modelling is that most social processes we wish to model take place in the context of an hierarchical social structure. If we ignore the hierarchy in our models we are prevented from asking substantive questions about it. Also the correlations induced by the hierarchy will be ignored thereby biasing the standard errors estimated in our model. All this is well known amongst practitioners of multilevel modelling.

The assumption that social structures are purely hierarchical is often rather simplistic. People will often belong to more than one grouping at any given level of a hierarchy and each grouping can be a source of random variation. For example, in an educational model both the neighbourhood a child comes from and the school a child goes to may have important effects. One school will contain children from many neighbourhoods and different children from any neighbourhood will attend several different schools. Therefore school is not nested within neighbourhood and neighbourhood is not nested within school. We now have a cross classified model. The consequences of ignoring an important cross classification are similar to those from ignoring an important hierarchical classification.

We can represent this model as

$$y_{i(jk)} = \alpha + u_j + u_k + e_{i(jk)} \qquad (1)$$

This models the achievement score of the i'th child from the jk'th school/neighbourhood combination ($y_{i(jk)}$) as being composed of an overall mean (α), a random departure due to school j (u_j), a random departure due to neighbourhood k (u_k) and an individual level random departure ($e_{i(jk)}$).

The model can be elaborated by adding individual level explanatory variables, which may also be allowed to vary across schools or neighbourhoods. Also, school or neighbourhood level variables can be added to attempt to explain variation across schools or neighbourhoods.

There are many other types of problem that fit into a cross-classified structure. For example, each pupil's exam paper may be assessed by a set of raters. Provided we have a different set of raters in each school we then have a pupil/rater cross classification at level 1 nested within schools at level 2. This model can be written as

$$y_{(ij)k} = \alpha + u_k + e_{ik} + e_{jk}$$

Where the rater and pupil effects are modelled by the level 1 random variables e_{ik} and e_{jk}.

The cross classification need not be balanced, some pupils' papers may not be assessed by all the raters. If the same set of raters are used in different schools then raters are cross classified with schools and this model is described by equation 1. Schools could also be crossed by neighbourhoods. In this case pupils are nested within a three-way rater /school /neighbourhood classification. Equation 1 can be extended to represent a three way classification by adding a term for the rater classification (u_l) producing :

$$y_{i(jkl)} = \alpha + u_j + u_k + u_l + e_{i(jkl)}$$

Similar applications are found in survey analysis for example where interviewers are crossed with areas.

Further elaboration of these models are possible, for example by adding random interaction terms. A detailed paper setting out such models has been submitted for publication.

Consider an example where we are analysing children's' overall exam attainment at age sixteen. The secondary school a child attends is cross classified by the primary school the child attended. The model is of the form given in equation 1. The u_j is a random departure due to secondary school and u_k is a random departure due to primary school. The data are on 3,435 children who attended 148 primary schools and 19 secondary schools in Fife. The result of the analysis is as follows:

parameter	estimate	se
α	5.50	0.18
$\sigma^2_{u_j}$	0.35	0.16
$\sigma^2_{u_k}$	1.12	0.20
$\sigma^2_{ei(jk)}$	8.11	0.20

This analysis shows that the variation in achievement at age sixteen attributable to primary school is three times greater than the variation attributable to secondary school. This type of finding is an intriguing one for educational researchers and raises many issues for further study.

Macros for running cross classified models are now available from the Multilevel Models Project.

Acknowledgment

I would like to thank Fife Regional Council's Department of Education, the Centre for Educational Sociology(CES), University of Edinburgh and the ESRC who support the research centre at CES for the use of the data set.

Reference

Rasbash, J. and Goldstein, H. (1993) Efficient analysis of mixed hierarchical and cross-classfied random structures using a multilevel model.

GRAPHING with ML3

ML3 macros for plotting random terms for set of categorical predictors

Kelvyn Jones

One of the great strengths of ML3 is its ability to readily plot the results of fitted models. Especially useful is the PLGL command which allows the plotting, and subsequent identification and labelling, of lines representing higher-level 'groups'. An example is given in Figure 9 of ***A guide to ML3 for new users***. Unfortunately, while this command is ideal for models when a continuous predictor is allowed to be random at a higher level; it cannot be used as it stands when the predictors are a set of categories. The aim of this note is to provide a macro procedure for predictor variables with between 1 and 4 categories when those categories, as usual, are defined as a 'constant' and a set of 'dummy contrasts'. The macros make no claim to be elegant or efficient, but I have found them effective for interpretation and presentation of results (*Jones et al, 1992; Jones, 1993*).

In essence the macros work as follows:

1. the model with categorical predictors allowed to be random at a higher level is estimated to convergence;
2. a number of columns and boxes are used as 'pointers' which identify the number, level and location of the categorical variables that are random;
3. a macro generates the 'predictions' for each category for each higher-level group; these are the values for the vertical axis; a crude 'skipping' procedure is used to deal with differing numbers of random terms;
4. the horizontal axis is produced as a combination of the category codes;
5. the PLGL command is used to produce the desired plots;
6. the high-resolution graphic menus are used to label axes and identify unusual or extreme groups.

Stage 4 is the key to the plot, and to appreciate how this achieved, I will consider the procedure for a single group, in this case one postcode area (the example is taken from the ML3 Manual, section 9.4). There are a set of four categories (a constant and three dummy contrasts) with the following predictions (log-odds of employment), variables and definitions:

Pred	Variable		
0.5	constant	male-unqualified	a set of 1's
0.6	unfem	female-unqualified	a dummy contrast
0.45	qualmal	male-qualified	a dummy contrast
0.7	qualfem	female-qualified	a dummy contrast.

To produce a 'bar-graph' for each category as shown in Figure 1, we require the predictions to be duplicated, and the horizontal codes to be generated as follows:

0.5	0.5	0.6	0.6	0.45	0.45	0.7	0.7
1	2	2	3	3	4	4	5

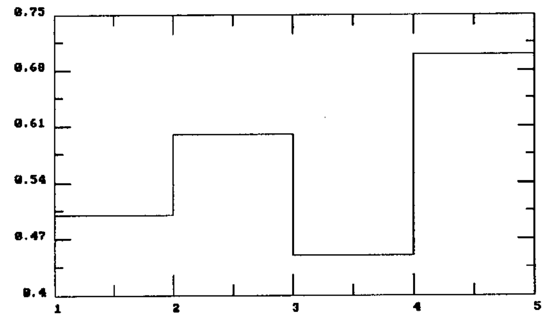


Figure 1 Plotting a single group

where 1-2 represents the base category of unqualified-males, 2-3 stands for unqualified-females and so on.

The macros as presented below are for use with the logistic\binomial macros and are self-documenting. To see how they can be used, I will analyze the model of proportion employed for four categories, with three of them allowed to vary at level 2, and extra-binomial variation at level 1. The edited log-file is as follows:

```
names
```

	name	n	missing	min	max
1	postcade	401	0	2.8e+005	3.1e+005
2	cell	401	0	1.0000	401.00
3	proporti	401	0	0.0000	1.0000
4	constant	401	0	0.0000	1.0000
5	unfem	401	0	0.0000	1.0000
6	qualmal	401	0	0.0000	1.0000
7	qualfem	401	0	0.0000	1.0000
8	runmal	401	0	0.0000	1.0000
9	runfem	401	0	0.0000	1.0000
10	rqualm	401	0	0.0000	1.0000
11	rqualf	401	0	0.0000	1.0000

GRAPHING with ML3

```
sett
```

```
EXPLanatory variables in        CONSTANT UNFEM    QUALMAL  QUALFEM  RUNMAL
FPARameters                     CONSTANT UNFEM    QUALMAL  QUALFEM
FMEAns
RMEAns
RESPonse variable in            RATE
IDENtifying codes for level 1: CELL     level 2: POSTCODE level 3:
RESEtting covariances level 1: ON       level 2: ON       level 3: ON
MAXIterations   5    TOLErance   2      METHod is IGLS    BATCh is OFF


LEVEL 3 RANDOM PARAMETER MATRIX unspecified
LEVEL 2 RANDOM PARAMETER MATRIX
        CONSTANT UNFEM    QUALFEM
CONSTANT 1
UNFEM    1        1
QUALFEM  1        1        1
LEVEL 1 RANDOM PARAMETER MATRIX
        RUNMAL   RUNFEM   RQUALM   RQUALF
RUNMAL   1
RUNFEM   0        1
RQUALM   0        0        1
RQUALF   0        0        0        1
```

After convergence the following results were obtained.

```
fixe
```

```
->fixe

 PARAMETER            ESTIMATE    S. ERROR     PREV. ESTIMATE
CONSTANT                0.515       0.125           0.515
UNFEM                   0.172       0.153           0.172
QUALMAL                 1.004       0.153           1.004
QUALFEM                 1.157       0.162           1.157
```

```
rand
```

```
->rand

                        LEVEL 3
 PARAMETER            ESTIMATE    S. ERROR     PREV. ESTIMATE

                        LEVEL 2

 PARAMETER            ESTIMATE    S. ERROR     PREV. ESTIMATE
CONSTANT /CONSTANT      0.185      0.0872           0.185
UNFEM    /CONSTANT     0.0761      0.0937          0.0761
UNFEM    /UNFEM        0.0638       0.229          0.0638
QUALFEM  /CONSTANT          0           0               0
QUALFEM  /UNFEM             0           0               0
QUALFEM  /QUALFEM           0           0               0

                        LEVEL 1

 PARAMETER            ESTIMATE    S. ERROR     PREV. ESTIMATE
RUNMAL   /RUNMAL        1.231       0.204           1.231
RUNFEM   /RUNFEM        0.822       0.229           0.822
RQUALM   /RQUALM        0.904       0.142           0.904
RQUALF   /RQUALF        1.108       0.166           1.108
```

The 'pointers' are now set as follows:

Note: set level at which variables are random
Calc b78=2
Note: set number of variables allowed to be random
calc b79=3
Note: input column numbers of variables that are random
Input c78
4 5 7
Note: input order of random variables in fixed part
Input c79
1 2 4

That is the third variable in the fixed part is skipped over as QUALMALE is not allowed to vary at level 2.

Note: toggle off echo
Echo
Note: call macro to produce plot
obey c:\ml3\bincat.mac

The resultant plot is shown in Figure 2; the cursor, scale, title and labelling commands on the menu bar for the high-resolution graphics are then used to produce Figure 3. The greater variation for unqualified females is clearly seen; the labels refer to the postcode areas with the highest and lowest predictions for this category.

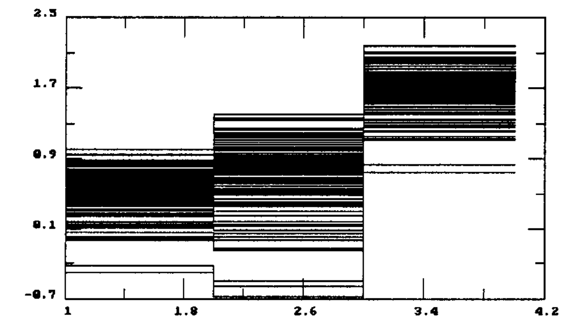


Figure 2 Output from the macro

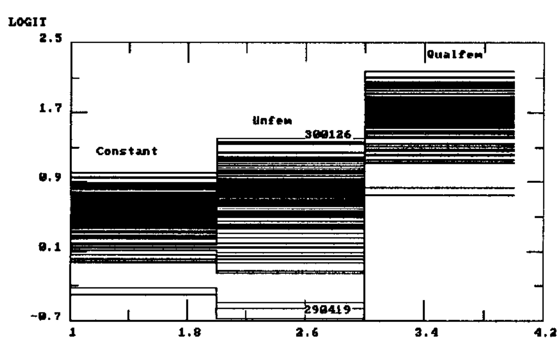


Figure 3 Output after modification

The macros are available from the author.

References

Jones, K, Johnston, R J and *Pattie, C J* (1992) 'People, places, regions: exploring the use of multilevel modelling in the analysis of electoral data', British Journal of Political Science, 22,343-380.

Jones, K (1993) Modelling ecologies: a multilevel perspective on 'ecological correlations and behaviour of individuals'