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. 6 No. 1 March, 1994

Multilevel Modelling Course: A two-week course on multilevel modelling in Social Science Data Analysis and Collection will be held at the Essex Summer School from 25th July to 5th August 1994. For further details contact Janet Brightmore, Essex Summer School, Department of Government, University of Essex, Colchester, CO4 3SQ, United Kingdom. E-mail: sum_sch@essex.ac.uk.

OS2 Version of ML3-E Now Available ML3-E is bound to Pharlap's DOS extender. We have just upgraded to Pharlap V6.0 which claims to be OS2 compatible. If you are running ML3-E under complex configurations (eg. a Network) and have experienced problems, the new version may well coexist more easily with your system. If you would like a copy of this version contact the Multilevel Models Project, and you will be sent a copy free of charge.

Also In This Issue

ML3 Clinics in London 1994

Free for users of *ML3/ML3-E*

Tuesday May 10
Tuesday June 7
Tuesday October 11
Tuesday November 15
Tuesday December 6

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

aLCD analysis of large and complex datasets

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Theory and Applications

Improved Estimation for Logit and Loglinear Multilevel Models

Harvey Goldstein

An increasing number of researchers are fitting logit linear multilevel models to data where the response is a proportion, and often binary (0,1). Examples are mortality data, examination passes, voting preferences, etc. Both VARCL and ML3 have had facilities for fitting such models for some time and give the same estimates.

Recently, however, it has become clear that for some kinds of data, notably where the response is binary, the estimation procedures used by VARCL and ML3 can seriously underestimate the parameters, especially the variances and covariances. *Breslow* and *Clayton* (J. American Statist. Assn., 1993, 9-25) carry a detailed discussion and show that it is possible to improve somewhat on these procedures. *Rodriguez* and *Goldman* (J. Royal Statis Soc., 1994, to appear) show that for a 3-level variance components model with binary responses and large amount of higher level variation the underestimation of the variances can be marked.

Using an extension of the proposal in *Goldstein* (Biometrika, 1991, 45-51) to add a second order term to the Taylor expansion in the nonlinear approximation used, and incorporating the PQL model of *Breslow* and *Clayton*, we have been able to obtain considerable improvements to existing estimates, although, there still appears to be some bias in extreme cases. To illustrate the improvement obtained, the following table shows the new estimates, compared to those obtained using the old procedures.

The following model is used

$logit(\pi_{ij}) = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j}x_{ij}$

$y_{ij} \sim Bin(\pi_{ij}, 1)$

$var(u_{0j}) = var(u_{1j}) = 0.5, cov(u_{0j}, u_{1j}) = 0.25$

$\beta_0 = 0.5, \beta_1 = 1.0$

There are 50 level 2 units with 20 level 1 units in each level 2 unit. The following results are based upon 100 simulations of the above model.

Param.	Old	New
$\sigma^2_{u_0}$	0.39(0.13)(0.16)	0.54(0.22)(0.21)
$\sigma^2_{u_1}$	0.26(0.23)(0.27)	0.51(0.39)(0.47)
$\sigma_{u_{01}}$	0.14(0.15)(0.15)	0.24(0.25)(0.22)
β_0	0.45(0.13)(0.13)	0.51(0.15)(0.14)
β_1	0.75(0.16)(0.16)	1.15(0.26)(0.21)

Mean values of 100 simulations. Empirical standard error in first bracket; mean of estimated standard errors in second bracket. Convergence was obtained in 99 cases for the old macros and in 86 cases within 15 iterations for the new.

Clearly, the existing estimation procedure has been considerably improved. Whereas using the old procedure there is consistent underestimation of both the random and fixed parameters, the new procedure produces estimates closer to their true values. The estimated standard errors also appear to be quite accurate.

A detailed description and a set of new ML3 macros to carry out the estimation for this model is available from the Multilevel Models Project. The macros can be altered to carry out the estimation for other data sets. The next version of the software, Mln, designed to fit very general multilevel models will carry a comprehensive set of commands for carrying out this estimation, together with a variety of other loglinear and general nonlinear models.

A Crossed Random Effects Model for Studying Social Context Effects on Individual Growth

Stephen W. Raudenbush

This note considers analysis of repeated measures data when the goal is to incorporate effects of social contexts on individual development over time. *Bryk* and *Raudenbush* (1988) employed a three-level hierarchical linear model in studying school effects on children's growth during the primary years. The level-1 observations were the time-series data for each child; the level-2 units were the children, and the level-3 units, the schools. Unfortunately, this three-level model applies only to those children who remain in a single school during the course of the investigation. To study children as they migrate across social contextual boundaries requires a crossed random effects model.

Just such migration occurs when we study the effects of teachers on children's cognitive growth during the primary years (Figure 1). The data constitute a sub-sample from the 'Immersion Study' (*Ramirez* et al.,1991), a national evaluation of alternative programs for children in the US having limited English proficiency. Each row in Figure 1 is a child and each column a teacher. The histories of children 5009 and 5010 illustrate the shifting social memberships that characterize many primary schools in the U.S.: These children shared membership during grade one in the classroom of teacher 4 but were split apart in grade 2, when child 5009 was assigned to teacher 14 and child 5010 to teacher 11. During grade 2, child 5009 became a classmate of child 5246 for the first time.

Raudenbush (1993) derived maximum likelihood estimates for the two-way random cross-classification and applied it to the data in Figure 1. This note summarizes the analysis without considering the estimation theory. The outcome measure is the California Test of Basic Skills, scaled to reflect growth over time in mathematics and administered each year from grade 1 to grade 4.

Figure 1: Organization of Data in a Sub-Sample from the Immersion Study

Each X indicates an Observation

Grade	Grade 1										Grade 2									Grade 3					Grade 4			
Teacher	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Totals
Student																												
5003	X										X																	2
5005		X										X																2
5006	X																											1
5007			X										X															2
5009				X										X							X					X		4
5010				X							X																	2
5011	X										X										X							3
5013			X																									1
5014														X								X						2
5015					X												X					X						3
.																												.
.																												.
.																												.
5232										X																		1
5234						X						X											X					3
5236					X												X						X					3
5237			X								X																	2
5238				X								X																2
5241										X																		1
5242								X												X								2
5243	X																		X									2
5245								X											X									2
5246					X								X											X		X		4
Totals	5	13	15	18	14	15	13	13	10	7	2	9	10	11	13	4	6	15	11	12	12	3	3	2	11	2	1	250

A two-level growth analysis

We first present an analysis that ignores the teacher effects. A two-level model is convenient in studying psychological change on independently responding persons.

Level-1 model. Using the 'HLM' program (*Bryk, Raudenbush, & Congdon*, 1994), we formulate a straight-line growth model for each child. Exploratory analysis showed that this linear growth model was reasonable.

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + e_{it}, \quad e_{it} \sim N(0, \sigma^2) \quad (1)$$

where

Y_{it} is the math outcome for child i at time t, t = 1,...,4; i = 1,...,123;

a_{it} = grade - 1, so that a_{it} = 0, 1, 2, 3, 4 at primary grades 1, 2, 3, 4, respectively;

π_{0i} is therefore the expected outcome at first grade for child i;

π_{1i} is the expected gain per year for child i; and e_{it} is a random error.

Level-2 model. We view the pair of π's as varying randomly over the population of children according to a bivariate normal distribution, i.e.,

$$\pi_{0i} = \beta_{00} + u_{0i}$$
$$\pi_{1i} = \beta_{10} + u_{1i} \quad (2)$$

where $$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \quad (3)$$

and

β_{00} is the population mean outcome at grade one;

β_{10} is the population mean rate of change;

u_{0i} is the 'child effect' on initial status, that is, the deviation of child i's true initial status from the population mean; and u_{1i} is the 'child effect' on the rate of growth, i.e., deviation of child i's true growth rate from the population mean growth rate.

The covariance components include the population variances in initial status (τ_{00}) and growth rate (τ_{11}), and the covariance between initial status and growth rate (τ_{01}).

Results (Table 1, Model 1) indicate an estimated population mean at grade one of 256.54 and an estimated mean rate of growth of 47.28 per year. More important, there is evidence of substantial between-subject variance both in initial status and rate of growth.

Table 1(a) Immersion Data - Fixed Effects

Effects	Models*	Coeff.	SE	t-ratio
Expected 1st grade status	1	265.54	2.51	---
	2	257.51	1.71	---
	3	257.51	1.86	---
Expected linear growth rate	1	47.28	2.25	21.00
	2	45.44	1.48	30.74
	3	44.90	1.65	27.20
Effect of masters' degree	1			
	2			
	3	5.37	3.62	1.48

* Model 1: teacher variance Ignored.
Model 2: teacher variance estimated.
Model 3: teacher variance predicted.

Table 1(b) Immersion Data - variance components

Parameter	Model 1	Model 2	Model 3
Intital status: $Var(\pi_{0i})$	484.22	440.43	435.25
Growth rate: $Var(\pi_{1i})$	133.58	91.50	101.37
Covariance: $Cov(\pi_{0i}, \pi_{1i})$	28.02	97.47	88.25
Teacher effect: $Var(v) = \delta^2$		79.58	68.63
Residual error: $Var(e_{it})$	319.60	274.37	281.36
Model fit deviance	2277.84	2266.18	2265.08

Crossed random effects analysis

The two-level analysis treats children as independent. However, we know that the children were nested within teachers each year. A teacher might have a beneficial or deleterious effect on her children during a given year, effectively deflecting the growth curves of those children in a positive or negative direction. Such a teacher effect, if it exists, is interesting in its own right and creates dependence among the test scores of children sharing membership in the same classroom, invalidating the independence assumption of the two-level analysis.

A simple way to incorporate a teacher effect is to assign each teacher a positive or negative number, α_q, that represents the teacher q's effect on the children she encounters. Then, if child i gains $\pi_\chi + \alpha_q$ during the year that child encounters teacher q. This conceptualization can be implemented by incorporating 27 dummy variables as time-varying covariates in the level-1 model:

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + \sum_{q=1}^{Q} \alpha_q \delta_{qit} + e_{it} \quad e_{it} \sim N(0, \sigma^2) \qquad (4)$$

where π_{0i} and π_{1i} again represent the initial status and rate of growth for child i but now α_q represents the effect associated with teacher indicator δ_{qit} $q = 1, \ldots, 27$ ($\delta_{qit} = 1$ if child i encounters teacher q at time t and 0 otherwise).

Variation between children is modelled exactly as in Equations 2 and 3. In modeling the teacher effects, Equation 4 is overfitted, with 27 teacher coefficients. However, if we have no a priori information on the background, training, or skill of the teachers, we can view their effects as 'exchangeable' (*Lindley* and *Smith*, 1972), functionally equivalent to assuming they are sampled randomly from a distribution of teacher effects. We model the teacher effects as normally distributed, i.e.,

$$\alpha_q \sim N(0, \sigma_\alpha^2) \qquad (5)$$

This model is now much more parsimonious: rather than estimating 27 fixed teacher parameters, we estimate a single variance, σ_α^2, the variance of the teacher effects.

Combining the models. When we substitute Equations 2 and 4 into Equation 1, we have the cross-classified random effects model

$$Y_{it} = \beta_{00} + \beta_{10}\alpha_{it} + u_{0i} + u_{1i}a_{it} + \sum_{q=1}^{Q} \alpha_q \delta_{qit} + e_{it} \qquad (6)$$

with fixed effects (the $\beta's$), random effects of children (the $u's$), and random effects of teachers (the $\alpha's$).

Comparing results. A comparison of results between the two-level model that ignores classrooms (Table 1, Model 1) and the crossed random effects model (Table 1 Model 2) reveals important differences. Estimation of the teacher effect variance of 79.58 in Model 2 is accompanied by reductions in the estimates of variance among children in intercepts (440.43 for Model 2 versus 484.22 for Model 1) and in slopes (91.50 versus 133.58). Thus, part of the variability that had been attributed to individual differences is now attributed to teachers. Also, the variance within children is reduced (274.37 versus 319.60), indicating that part of the variability attributable to temporal instability in Model 1 is accounted for by teachers Model 2.

How important are classroom effects? The estimated teacher variance of 79.58 is equivalent to a standard deviation of $\sqrt{79.58} = 8.92$. Now the average growth rate is 45.44, so that the expected learning gain in a given year for a teacher having an effect one SD above average would be 45.44 + 8.92 = 54.36 while the expected gain for a teacher with an effect one SD below average would be 36.52. This is a non-trivial difference, especially given the possibility that it could be compounded by a 'run of good or bad luck,' that is, a series of good or bad teachers.

Comparing the fit of Models 1 and 2 yields a difference between deviances of 2277.84 - 2266.18 = 11.66, which, when compared against the critical value of chi-square with df = 1, gives quite strong evidence in favour of the model that incorporates the teacher variance component, $p < .001$.

The effect of teacher education. The assumption of Equation 6 that no prior knowledge exists about the relative magnitude of teacher effects. In fact, we know that teacher education levels vary, so we incorporate this information into the model:

$$\alpha_q = \gamma(Masters)_q + V_q, \quad V_q \sim N(0, \sigma^2_{\alpha|masters}) \qquad (7)$$

where $(Master)_q$ takes on a value of one if teacher q has received a masters or doctoral degree, zero if not. The addition of the teacher education indicator to the model (Table 1, Model 3) resulted in a small reduction in the estimated teacher-level variance component from 79.58 to 68.63. The effect of higher education was in the expected direction, $\hat{\gamma} = 5.37$, but failed to achieve conventional significance levels as judged either by the ratio of the coefficient to its standard error, $t = 1.48$, or by a comparison of fit between Model 2 and Model 3, which yields a difference between deviances of 2266.18 - 2265.08 = 1.10, df = 1.

Conclusions

The identification of social context effects invites research into the structure of that variation. What kinds of teachers and teaching methods account for these effects? Are some teachers improving? Are some teachers more effective for certain kinds of children than for others? These questions can be posed by collecting the relevant data and elaborating the model presented here. We can also examine posterior 'teacher effect' estimates for each teacher, computed automatically as a by-product of the EM algorithm used for estimation (see Appendix to *Raudenbush*, 1993).

However, extreme care is necessary in interpreting the social context effects, especially when persons are changing contexts during the course of the study. In the illustrative data example we need to know whether children are assigned to teachers on the basis of their progress during the prior year. If so, the teachers would be endogenous to achievement, and the results of our analysis would be misleading. Fortunately, there is no evidence of such selective assignment to teachers in these data.

The formulation of models and estimation procedures for nested and crossed random effects may be viewed as part of a long-term effort to develop a family of analytic tools that correspond to the classical experimental designs but that are far more flexible than the classical analysis of variance. Such flexibility includes the capacity to handle unbalanced data, covariance components, covariates measured either discretely or continuously at each level of analysis, and discrete or continuous outcomes.

References

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101(1), 147-158.

Bryk, A.S., & Raudenbush, S.W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. American Journal of Education, 97(1), 65-108.

Dempster, A.P., Rubin, D.B., & Tsutakawa, R.K. (1981). Estimation in covariance components models. Journal of the American Statistical Association, 76, 341-353.

Goldstein, H. (1987b). Multilevel covariance components models. Biometrika, 74, 430-431.

Hartley, H.O., & Rao, J.N.K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. Biometrika, 74, 430-431.

Lindley, D.V., & Smith, A.F.M. (1972). Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B, 34, 1-41.

Patterson, H.D., & Thompson, R. (1971). Maximum likelihood estimation of variances. Biometrika, 58, 545-554.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. Journal of Educational Statistics, 18,4,321-349.

Multilevel Logistic Regression Models in Educational Research: GCSE Estimated Grades

Martin R Delap

Introduction

Multilevel modelling techniques are increasingly being used to analyse the results of educational research. For example *Tymms* and *Fitzgibbon* (1992) have analysed the influence of homework upon A Level results, whilst *Sammons, Nuttall* and *Cuttance* (1993) used multilevel models to investigate differential school effectiveness. In this paper I shall show how a multilevel logistic regression model may be used to assist in the analysis of teachers' estimates of the performance of their pupils in public examinations; in particular to answer the following questions:

i. Are estimates of some grades more likely to be correct than for other grades?

ii. Is the proportion of correct estimates dependent upon the gender of the candidate?

iii. Does the proportion of correct estimates vary between teacher groups? Or between schools?

Data structure

When formulating a strategy for the collection of data for a programme of research, careful thought must be given to the experimental design. Where multilevel modelling techniques are to be applied, particular attention must be paid to the hierarchical structure of the data. In the field of educational research there is a natural tendency to assume that the most appropriate structure is that of candidates nested within classes which are in turn nested within schools. However, this structure may not necessarily be the most appropriate one to use. This study is a case in point. Teachers provide estimates of how they expect the pupils will perform in an examination. I shall therefore regard the individual teachers who make the estimate as the second level in the data structure rather than the classes.

In more complex models it would be possible to cross classify teachers with their classes, however this is an unnecessary complication which is ignored here. One can of course envisage situations where the structure I have adopted is not appropriate, for example two classes within a school may be taught by two different pairs from a 'pool' of three teachers.

Sample

The schools participating in the study were asked to distinguish between estimates made by different teachers within the school. In cases where pupils were taught by more than one teacher, the schools were asked to ensure that a single estimated grade was arrived at by agreement between the teachers.

Teachers were asked to supply their estimates of the single letter grade which they expected each candidate to obtain. The results of GCSE examinations are given in terms of a single letter estimated grade in the range A to G. Candidates who do not meet the minimum requirement for Grade G are ungraded and are denoted by the letter 'U'. A total of 68 schools participated in the study. There were 223 identifiable teacher groups covering 5,036 candidates.

The model

The analysis considered the accuracy of the estimated grade as the response; candidates were assigned a value 1 if the estimated grade matched the actual grade, 0 otherwise. The explanatory variables consisted of a set of dummy variables. These were generated for each estimated grade. A further set of dummy variables were used to explore whether there were any effects due to the gender of the candidates. The gender variable was assigned the value 0 for males, 1 for females. Since the effect of gender may be different for each estimated grade, the interaction terms between gender and the estimated grade were included as fixed explanatory variables.

The fixed terms of the model were given as a set of contrasts with estimates of Grade C. Thus the fixed terms were given by

$$f_{ijk} = \beta_{Cj}(CONS)_{ijk} + \beta_{Aj}(EST\_A)_{ijk} + \beta_{Bj}(EST\_B)_{ijk} + \ldots$$
$$+\beta_{GC}(GENDER)_{ijk} + \beta_{GA}(GEN\_A)_{ijk} \ldots \beta_{GU}(GEN\_U)_{ijk}$$

The coefficients of each of the explanatory variables of estimated grade were allowed to vary at level 2 (teacher group). This enabled calculation of the variance of the coefficients at level 2. Binomial variation was assumed at level 1. The model was run using the ML3E program with macros for binary response variables available from the Multilevel Models Project (1993).

Results

Fixed Parameters: For the fixed parameters the model was constructed so that the 'CONSTANT' term referred to estimates of Grade C. The remaining coefficients are the additional terms for each estimated grade and gender. The results are given in Table 1. It can be seen that the coefficient for estimates of Grade A is positive and significant (more than twice its standard error). This indicates that a significantly greater proportion of estimates of Grade A are accurate than estimates of Grade C. Conversely, a significantly lower proportion of estimates of Grade E are accurate than estimates of Grade C. The fixed parameters of the model can be used to determine the modelled proportion of accurate estimates of each estimated grade for males and females. For example,

Estimated Grade C boys = antilogit (-0.25) ≈ 44%
girls = antilogit (-0.25 + 0.06) ≈ 45%

Estimated Grade A boys = antilogit (-0.25 + 1.565) ≈ 79%
girls = antilogit (-0.25 + 1.565 + 0.06 - 0.16) ≈ 77%

The proportions obtained from the fixed parameters are, as we would expect, very similar to those obtained from the 'raw' data for each subject. Thus we have modelled the proportions and are in a position to answer our first two questions. The proportion of accurate estimates varies according to the grade, furthermore the proportions of accurate estimates of Grades A and E are significantly different from those of Grade C. The gender coefficients are, with the exception of estimates of Grade U, not significant. This indicates that the modelled proportion of accurate estimates is approximately the same for both girls and boys.

Random parameters: The main advantage of the use of the logistic regression models over a simple analysis of the raw data is that it is possible to estimate the variance of the parameters at the higher levels. Hence it is possible to determine the extent to which the proportion of accurate estimates varies between teacher groups, or between centres.

At level 3 (school level) there is virtually no unexplained variance, suggesting that there is no significant variation attributable to schools in the proportion of accurate estimated grades. At level 2, because of the way in which the model is specified, the coefficient Constant/Constant refers to the variance of the estimates of Grade C. The variance of the log odds ratio at level 2 for each of the estimated grades may be readily calculated from the data in Table 1.

Table 1 Results of the multilevel logistic regression model

Fixed parameters	Estimate	Standard Error
Constant	-0.250	0.022
Estimate A	1.565	0.190
Estimate B	0.023	0.125
Estimate D	-0.217	0.132
Estimate E	-0.506	0.163
Estimate F	-0.143	0.204
Estimate G	-0.093	0.274
Estimate U	-0.833	1.035
Gender	0.060	0.113
Gender*Est A	-0.160	0.220
Gender*Est B	0.045	0.173
Gender*Est D	-0.059	0.178
Gender*Est E	0.254	0.215
Gender*Est F	-0.468	0.270
Gender*Est G	-0.321	0.341
Gender*Est U	3.314	1.528

Random parameters		
level 3		
CONS/CONS	0.001	0.001
level 2		
CONS/CONS	0.017	0.004
Est A/CONS	-0.055	0.022
Est A/Est A	0.731	0.290
Est B/CONS	-0.031	0.014
Est B/Est B	0.098	0.113
Est D/CONS	-0.022	0.014
Est D/Est D	0.053	0.114
Est E/CONS	0.0	0.0
Est E/Est E	0.0	0.0
Est F/CONS	0.0	0.0
Est F/Est F	0.0	0.0
Est G/CONS	-0.087	0.030
Est G/Est G	0.577	0.374
Est U/CONS	-0.162	0.134
Est U/Est U	6.194	4.702

Table 2 provides the results of these computations. It can be seen that there is much greater variation for estimates of Grades A, G and U than for the other estimates. Using these variances we may determine appropriate confidence intervals within which we expect the corresponding proportion of candidates to lie. Thus for example we find that the 95% confidence interval for the proportion of accurate estimates of Grade C (P_c) within a teacher group is approximately,

$$antilogit(-0.25 - 2\sqrt{0.017}) \le P_c \le antilogit(-0.25 + 2\sqrt{0.017})$$
$$\rightarrow 38\% \le P_c \le 50\%$$

The use of a multilevel model also allows computation of the covariances of the level 2 random terms. However, these are not directly interpretable and have not been calculated here.

Table 2 Variance of the log odds ratio at level 2 for each estimated grade

Estimated Grade	Variance
Estimate A	0.637
Estimate B	0.053
Estimate C	0.017
Estimate D	0.027
Estimate E	0.017
Estimate F	0.017
Estimate G	0.420
Estimate U	5.887

Conclusions

The first two questions posed at the start of this paper could have been answered from the 'raw' estimated grades data. However, multilevel logistic regression modelling has been used to show that there is a significant variation between teachers in the proportion of estimates which were accurate. Furthermore the results show that this variation is substantially greater for estimates of Grades A, G and U than for the other grades. We have also seen that there is little variation between schools which has not been accounted for by the variation between teacher groups.

References

Multilevel Models Project (1993) A Guide to ML3 Macros: Multilevel Binary Response Models. London: Institute of Education.

Sammons P, Nuttall D and *Cuttance P* (1993) Differential School Effectiveness: Results from a re-analysis of the Inner London Education Authority's Junior Schools Project Data. British Educational Research Journal 19 pp 381-405.

Tymms P B and *Fitzgibbon C T* (1992) The Relationship of Homework to A Level Results. Educational Research 34 pp 3-10.

Using Multilevel Multinomial Regression to Analyse Line-up Data

Daniel B. Wright & Anne T. Sparks

There is a great deal of interest in the validity of witness identifications in line-ups (Wells, 1993); court cases are sometimes decided on line-up identifications alone. In a line-up there are three possibilities for the witness. S/he can pick the suspect (not necessarily the culprit), pick a foil (someone other than the suspect) or not make any identification. This means a multinomial model is necessary to analyse the data. The response category, Line-up data is, in general, multilevel. In our data as many as 21 witnesses viewed the same suspect and these are nested within suspect.

The substantive focus of this research was whether two specialist line-up suites produced different results from those at police stations in the Greater London area. These suites call 'regulars' to take part in line-ups. At ordinary stations, the foils are members of the public picked 'off the street' for that single line-up. Where and how line-ups are conducted are what *Wells* (1978) refers to as system variables because theoretically they are under the control of the judicial system. Several other variables were available for these analyses; the gender of the witness, the severity of the crime, the race of the suspect, etc. *Wells* (1978) calls these estimator variables; the judicial system has no control over them.

Table 1. Response patterns for the different locations (Percentage in bracket)

	Suite 1	Suite 2	Stations	Total
Picked Suspect	256 (42.7)	168 (33.7)	191 (40.5)	 (39.2)
Picked Foil	127 (21.2)	116 (23.3)	68 (14.4)	 (19.8)
No identification	216 (36.1)	214 (43.0)	213 (45.1)	 (41.0)
Total N	599	498	472	1569

Table 1 gives the response patterns for the two suites and the ordinary police stations. A higher percentages of witnesses at the two suites picked foils than at the ordinary stations. It is important to note that because these are real line-ups, we do not know whether the suspect is in fact the culprit.

The line-ups all took place in 1992 in the Greater London area. In total, 1,569 witnesses made judgements in which 623 suspects were involved. The number of witnesses per suspects varied from one to 21. There were eleven variables in the initial file.

System variables:	LOC1 (1 if at suite 1, 0 otherwise) LOC2 (1 if at suite 2, 0 otherwise) LOCOTHER (1 if at an ordinary police station, 0 otherwise)
Estimator variables:	WITMALE (1 if the witness is male, 0 if female) SUSWHITE (1 if the suspect is white, 0 otherwise)
Response variable:	VIOLENT (1 if the crime is violent, 0 otherwise) PICKSUS (1 if the suspect is picked, 0 otherwise)
Identifiers:	PICKFOIL (1 if a foil is picked, 0 otherwise) NO_IDENT (1 if no one is picked, 0 otherwise) SUSPECT (Level 3 - suspect code) WITNESS (Level 2 - witness code)

We followed the *Multilevel Models Project*'s guide to multilevel multinomial models (1993) for this analysis (see *Goldstein* (1991, 1992) for aspects of the theory). The final model included all the predictor variables listed above. The reference category for the location was LOCOTHER and for the response was NO_IDENT. Let π_{sij} be the probability that the i^{th} witness of the j^{th} suspect chooses the s^{th} alternative, either picking the suspect $s = 1$ or a foil ($s = 2$), and let π_{tij} be the corresponding probability for suspects not making any identification. If we let $u_j^{(s)}$ represent the suspect level residuals then the full model equation is

$$\ln\left(\frac{\pi_{sij}}{\pi_{tij}}\right) = \beta_0^{(s)} + \beta_1^{(s)} LOC1_{ij}^{(s)} + \beta_2^{(s)} LOC2_{ij}^{(s)} + \beta_3^{(s)} SUSWH_{ij}^{(s)} + u_j^{(s)}$$

$s = 1, 2$

There are two parameters estimated for each variable; one for choosing the suspect and one for choosing a foil ($s = 1$ and 2, respectively) being contrasts with no identification. When testing for extra-multinomial variation (sometimes called overdispersion), using only suspects who were viewed by more than one witness (as was done in *Goldstein* (1991) for the binomial case), we found $\hat{\sigma}_e^2$ (the variance of the witness level residuals) was 0.81 (SE=0.03) rather than the theoretical value of 1. Extra-multinomial variation exists, indicating that additional explanatory variables may be useful, or that there may be lack of independence among witnesses.

Table 2 shows the estimates for this model both assuming multinomial variation and allowing for extra-multinomial variation. The fixed part of these models are nearly identical. Each shows an increased number of people choosing a foil in the two specialist suites and a suggestion (not statistically significant) that white suspects are less likely to be chosen.

In concluding it is worth briefly discussing the central limitation of using data from real line-ups: the inability to make causal conclusions. There are several possible reasons why more foils were chosen in the two suites. One explanation is that the staff at the suites may not ensure that witnesses know they should only make an identification if they are positive that person is the culprit. However, other explanations exist. It may be that because the staff at the suites have lists of volunteer foils they are able to choose foils who look similar to the culprit, thereby producing fairer line-ups. Equally, because the cases chosen for the specialist suites in many cases were deemed inappropriate for the police stations, it may be that other characteristics (those corresponding to the estimator variables) created the difference. These issues are discussed in greater detail in *Sparks* (1993).

Table 2. Parameter Estimates for Line-Up Data (Standard errors in bracket)

Parameters		Multinomial variation assumed	Multinomial variation not assumed
Fixed part			
Intercept	$\beta_0^{(1)}$	0.18 (0.21)	0.21 (0.20)
	$\beta_0^{(2)}$	-1.12 (0.24)	-1.13 (0.23)
LOC1	$\beta_1^{(1)}$	-0.18 (0.18)	-0.19 (0.17)
	$\beta_1^{(2)}$	0.46 (0.20)	0.41 (0.20)
LOC2	$\beta_2^{(1)}$	0.18 (0.18)	0.19 (0.18)
	$\beta_2^{(2)}$	0.57 (0.20)	0.57 (0.20)
SUSWHITE	$\beta_3^{(1)}$	-0.26 (0.15)	-0.26 (0.15)
	$\beta_3^{(2)}$	-0.01 (0.16)	0.00 (0.16)
VIOLENT	$\beta_4^{(1)}$	-0.14 (0.15)	-0.14 (0.15)
	$\beta_4^{(2)}$	0.02 (0.17)	0.00 (0.17)
WITMALE	$\beta_5^{(1)}$	0.11 (0.12)	0.09 (0.11)
	$\beta_5^{(2)}$	0.08 (0.14)	0.08 (0.13)
Random part			
level 3	$\sigma_u^{2(1)}$	0.89 (0.15)	1.17 (0.16)
	$\sigma_u^{2(2)}$	0.52 (0.16)	0.97 (0.18)
	$\sigma_u^{(12)}$	-0.42 (0.12)	-0.58 (0.13)
level 2	σ_e^2	1.00 (---)	0.81 (0.03)

The superscripts for the parameters indicate whether it refers to the estimate for choosing the suspect (superscript (1)) or a foil (superscript (2)).

References

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

Goldstein, H. (1992) Multilevel Multinomial Response Models. Working paper of the Multilevel Models Project, Institute of Education, University of London.

Multilevel Models Project (1993) A Guide to ML3 Macros: Multilevel Multinomial Response Logistic Models. Multilevel Models Project, Institute of Education, University of London.

Sparks, A.T. (1993) Identification parades - New styles versus old. How do they compare? Unpublished thesis. University of London.

Wells, G.L. (1978) Applied eyewitness testimony research: System variables and estimator variables. Journal of Personality and Social Psychology, 36, 1546-1557.

Wells, G.L. (1993) What do we know about eyewitness identification? American Psychologist, 48, 553-571.

The cooperation of the Metropolitan Police Service in allowing Anne Sparks to conduct this research is gratefully acknowledged. The statistical analysis was conducted with support from a British Academy Fellowship to Daniel Wright. A longer version of this paper is currently in preparation and is available from *Daniel Wright.*

Multilevel Modelling in the Western Australian Research Institute for Child Health: modelling the accuracy and precision of portable peak exploratory flow meters

Paul Burton: Much of the clinical research in our Institute is based upon experimental or observational designs which involve repeated measures and a response variable that is either normally distributed or can easily be transformed to near normality. In this setting multilevel modelling is proving to be an invaluable analytical tool. In addition to numerous one-off analyses, we have recently completed a comprehensive analysis of the accuracy and precision of portable peak exploratory flow (PEF) meters when they are used by asthmatic children in the non-laboratory setting.

Introduction

It has been recommended that children with asthma should monitor their airway function with twice daily measurements of peak exploratory flow (PEF) using a portable PEF meter. In the laboratory setting, using pulsed airflows with known wave-forms, it has been demonstrated that, apart from a non-linear bias that can easily be corrected by modifying the measurement scale, portable PEF meters fulfil internationally agreed criteria for clinical accuracy and precision; as an example, the American Thoracic Society guide-lines are that a meter should read within ±5% of the true value or within ±12 Litres/min, whichever is the larger. Unfortunately, there are grounds for believing that the accuracy and precision of portable PEF meters may be considerably less good when they are used by real children out in the community. We therefore decided to investigate this possibility in a group of asthmatic children in Western Australia.

The data set

The primary data set consisted of 629 measurement of peak exploratory flow (PEF) obtained using a portable PEF meter, each observation being paired with a second estimate of PEF obtained using a spirometer. In every case the spirometer measurement was taken immediately before or after the estimate obtained with the portable meter. In keeping with standard clinical practice the spirometer estimate was viewed as being the true one. The paired observations were taken over a 3 month period in 12 boarding school boys aged between 11 and 17 years. The number of paired observations in individual boys ranged from 28 to 101. All measurements were supervised by a trained school nurse.

The parameters of interest

In order for portable PEF monitoring to be of clinical value, it is essential that a clinically relevant drop in true peak flow (as measured by the spirometer) is accompanied by a fall in meter PEF of a magnitude that is sufficiently large to be discriminated from random within-child variation. The principal parameters of interest were therefore the gradient of the relationship between meter-based PEF (MPEF) and true (spriometer-based) PEF (SPEF) and the functional relationship between ***total within child variability*** and SPEF.

Preparation for modelling

The MPEF measurements were firstly adjusted to take account of the non-linear bias noted above, using an algorithm derived from the published work of *Miller and Pederson* [1993]. MPEF and SPEF were then centralised (by subtracting 300 L/min) prior to analysis.

Multilevel modelling

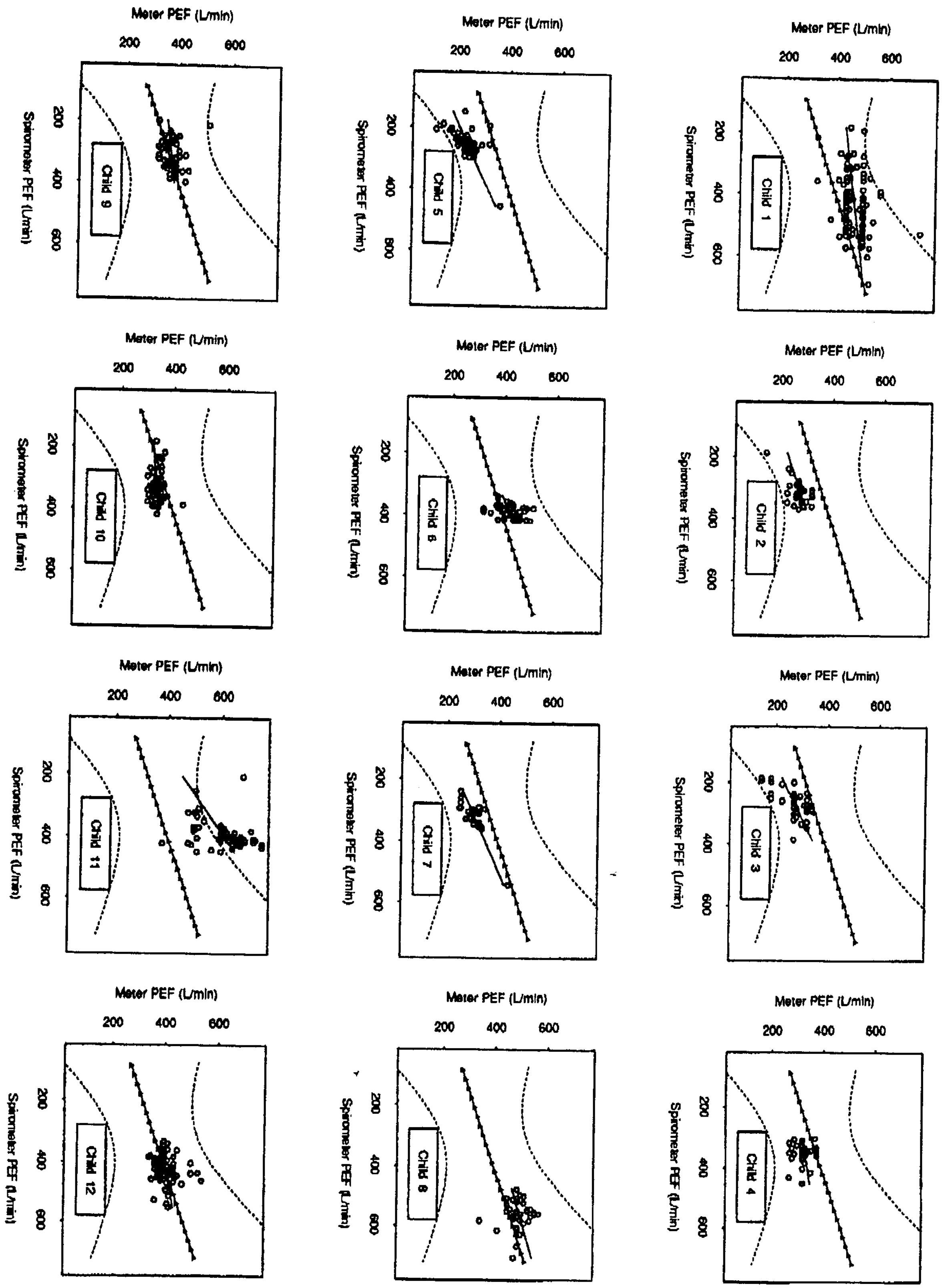
Where appropriate statistical significance testing was based upon the likelihood ratio test. The ML3 package was used.

The final model that was adopted had 2 levels (level 1 = observations, level 2 = children) and invoked MPEF as its response variable. CONS (the regression constant) and SPEF were fixed explanatory variables. The level 1 variance structure included all three variance/covariance terms for CONS and SPEF, while level 2 contained the two variance terms only. There was found to be no requirement for a covariance term between SPEF and CONS at level 2 ($\chi^2_{(1)} = 0.01$) and in practice the addition of such a term made no material difference to the remaining parameters in the model. The table details the parameter estimates of the final model. The level 1 variance structure indicated that the relationship between level 1 variability and SPEF was quadratic with a minimum value when spirometer PEF was 317 L/min; this was consistent with clinical impressions. The estimates overall gradient for MPEF:SPEF was 0.385 (95% confidence interval 0.212 to 0.557) and, presuming that the spirometer readings are not subject to serious random measurement error (an assumption that clinicians believe to be completely appropriate), this indicates that the overall response of the portable meter is markedly attenuated; a 100 L/min fall in true PEF is associated with an estimated 38.5 L/min fall in meter based PEF.

Level 2 residuals for SPEF were combined with the fixed effect for SPEF in order to obtain gradient estimates in each individual child, and variances for these estimates were obtained by combining the ***comparative variances*** for the residuals with the ***squared standard error*** for the fixed effect of SPEF. The estimated gradient in individual children ranged from +0.10 (95% CI -0.17 to +0.38), to +0.82 (+0.55 to +1.09). The point estimate of the gradient was lower than +0.25 in five of the twelve children.

The level 1 variance structure was used to estimate the total within-child variance (VAR_i) that would be expected at the mean SPEF value for the j^{th} child and the equivalent standard deviation (SD_i) was derived as $(VAR_i)^{0.5}$.

Let us make the somewhat arbitrary - but if anything conservative - assumption that a fall in true SPEF is unlikely to be detected as abnormal if it leads to a reduction in MPEF to a level that lies within 1.28 standard deviations (based on random within child variability) of an individual child's baseline mean. In order for a true fall in SPEF to be identified as abnormal, it would therefore be necessary for the true fall to be in excess of $1.28SD_i + G_i$, where G_i is the estimated gradient in the j^{th} child. On the basis, it would appear that in six of the twelve children, it would be quite possible for true PEF to fall to an extent that would be clinically very important and yet for the resultant fall in meter based PEF to be too small to be discriminated from random within child variation.


Meter PEF (L/min)
200 400 600
Spirometer PEF (L/min)
200 400 600 800
Child 9
Child 5
Child 1
Child 10
Child 6
Child 2
Child 11
Child 7
Child 3
Child 12
Child 8
Child 4

Figure

Table The final multilevel model (S.E. in bracket)

	Parameter	Estimate
Fixed	CONS	53.8 (22.49)
	[95% CI]	[9.7, 97.9]
	SPEF	0.385 (0.0878)
	[95% CI]	[0.212, 0.557]
Random		
Level 2	CONS	5791 (2461)
	SPEF	0.067 (0.037)
Level 1	CONS	935.4 (94.36)
	CONS/SPEF	-2.14 (1.12)
	SPEF	0.126 (0.019)

Twelve plots were constructed to display the child-specific models and the raw data simultaneously. The j^{th} plot detailed the distribution of the raw data in the j^{th} child and the estimated MPEF:SPEF relationship for that child predicted from the fixed effects and level 2 residuals. The overall model (based upon the fixed effects alone), and the approximate 95% confidence intervals for a single observation (based upon the fixed effects and the variance structure at both levels 1 and 2) were superimposed upon the plot for each individual child.

A residual analysis based upon the level 1 residuals for both CONS and SPEF (standardised using their diagnostic variances) provided no evidence of a poor fit. Observations with a potentially large regression leverage were identified (by eye) from the 12 child-specific plots and the analysis repeated having deleted these observations. Principal inferences were unchanged.

Conclusions

On the basis of this analysis, we have concluded that, despite recent recommendations, it is inappropriate for all asthmatic children to monitor their PEF using a portable meter. In a sizeable proportion of children, the fall in meter-based PEF in response to a clinically significant fall in true (spirometer) PEF would appear to be so small as to be completely obscured by random within child variation. Monitoring PEF with an instrument that produces equivocal results is at best worthless and at worst potentially dangerous.

Random-Effects Regression Models a) with Autocorrelated Errors and b) for Ordinal Outcomes

Donald Hedeker & Robert D. Gibbons

The 2-level random-effects regression model for the $n_i \times 1$ response vector y can be written as:

$$y_i = W_i\alpha + X_i\beta_i + e_i, \quad i = 1,\ldots N \; level-2units \qquad (1)$$

where W_i is a known $n_i \times p$ design matrix for the fixed effects, α is the $p \times 1$ vector of unknown fixed regression parameters, X_i is a known $n_i \times r$ design matrix for the random effects, β_i is the $r \times 1$ vector of unknown individual effects, and e_i is the $n_i \times 1$ error vector. Typically, the distribution of the random effects is assumed to be multivariate normal with mean vector **0** and covariance matrix Σ, and the errors are assumed to be independently distributed as multivariate normal with mean vector **0** and covariance matrix $\sigma_\epsilon^2 I_{n_i}$. Recent work has extended this model to allow for autocorrelated errors, especially errors which follow an autoregressive process of order 1, (*Chi & Reinsel*, 1989; *Hedeker*, 1989). Here one assumes that the residual covariance matrix is of the form $\sigma_\epsilon^2\Omega_i$ and though Ω_i carries the i subscript, it depends on i only through its dimension n_i, that is, the number of parameters in Ω_i does not depend on i, but depends only on the type of autocorrelation structure that is being assumed. For this, one can define ω to be the $s \times 1$ vector of autocorreletion terms that Ω_i depends on. Thus, for the first-order AR and MA processes $s = 1$, for an ARMA(1,1) process $s = 2$, and for the general autocorrelation structure s would be set to the number of autocorrelated lags to be estimated. This is the model and forms of autocorrelated error structures that is estimated by the MIXREG program (*Hedeker*, 1993b). Thus far, MIXREG has been produced for DOS and can be run either directly in batch mode or by utilizing a graphical menu-driven user interface. A manual accompanies the program which describes the estimation procedure and provides a few examples of the program's use. Estimation is accomplished using maximum marginal likelihood methods, utilizing both the EM algorithm and a Fisher scoring solution. Empirical Bayes estimates of the random coefficients can also be obtained.

For analysis of non-continuous outcomes, *Hedeker & Gibbons* (in press) extend the above model for ordinal (and dichotomous) outcome variables. This extension builds on the work of *Gibbons & Bock* (1987) who proposed a random-effects probit regression model for longitudinal dichotomous outcomes. The *Hedeker & Gibbons* model is developed for both the probit and logistic response functions and model either clustered or longitudinal 2-level data. The threshold concept is used in which it is assumed that the observed ordered category is determined by the value of a latent unobservable continuous response that follows a linear regression model incorporating random effects, that is, the model given above in (1). Thus, unlike other models for ordinal responses which allow for only 1 random effect (*Jansen*, 1990; *Ezzet & Whitehead*, 1991), this model allows for multiple random-effects. For estimation, maximum marginal likelihood (MML) methods are used with Gauss-Hermite quadrature to numerically integrate over the distribution of random-effects. At present, the model does not allow for autocorrelated errors. The MIXOR program(Hedeker, 1993a) has been developed to estimate this random-effects ordinal regression model. Like MIXREG, MIXOR runs under DOS either directly in batch mode or through a graphical menu-driven user interface. An accompanying manual describes the estimation procedure (a Fisher scoring solution) and provides examples of the program's use.

We have also worked on some application-orientated articles which describe and illustrate the use of multilevel models in a more applied manner. *Gibbons* et al.,(1993) describe traditional and random-effects models applied to longitudinal data for an audience of psychiatric researchers. This paper presents no equations, but instead, tries to describe in words the ideas and the actions associated with the use of various statistical models, and in particular, random-effects models for longitudinal data. At a slightly more advanced level, *Hedeker, Gibbons & Flay* (in press) describe a random-intercepts model for clustered data at an introductory level for an audience of applied psychological researchers. While this article presents nothing new from a statistical point of view, it attempts to introduce multilevel models and, in particular, the maximum marginal likelihood solution in an accessible manner. An SPSS program accompanies this paper which can be used to estimate a 2-level random-intercepts model for continuous outcome variables. A similar paper in terms of level of presentation, (*Gibbons & Hedeker*, in press) describes the use of the random-effects probit regression model for longitudinal dichotomous data to an audience of applied psychological researchers. Finally, *Gibbons, Hedeker, Charles & Frisch* (in press) present and illustrate an application of the random-effects probit regression model to an applied statistical audience. This article focuses on modeling vulnerability to a medical malpractice alarm in a sample of physicians who were repeatedly observed across 10 years in terms of the presence or absence of a malpractice claim.

The MIXREG and MIXOR programs can be obtained from*Ann Hohmann*, Ph.D., M.P.H., NIMH Services Research Brach, 5800 Fishers Lane, Room #10C-06, Rockville MD, 20857. Reprints of the articles can be obtained from *Don Hedeker*, Division of Biostatistics/Epidemiology (M/C 922), School of Public Health, 2121 West Taylor Street, Room 510, Chicago, Illinois 60612-7260, U.S.A; or from Robert D. Gibbons, Biometric Laboratory, University of Illinois at Chicago, 912 S. Wood, Chicago, Illinois 60612, U.S.A. EMAIL notes can be sent to *Don Hedeker* through BITNET at U41098@UICVM.

References

Chi, E.M. & Reinsel, G.C. (1989) Models for longitudinal data with random effects and AR(1) errors. Journal of the American Statistical Association, 84: 452-459.

Ezzet, F., & Whitehead, J. (1991) A random effects model for ordinal responses from a crossover trial. Statistics in Medicine, 10, 901-907.

Gibbons, R.D., & Bock, R.D. (1987) Trend in correlated proportions. Psychometrika, 52, 113-124.

Gibbons, R.D., & Hedeker, D. (in press) Application of random-effects probit regression models. Journal of Consulting and Clinical Psychology.

Gibbons, R.D., Hedeker, D., Charles, S.C., & Frisch, P.R.(in press) A random-effects probit model for predicting medical malpractice claims. Journal of the American Statistical Association.

Gibbons, R.D., Hedeker, D., Elkin, I., Waternaux, C.,Kraemer, H. C., Greenhouse, J.B., Shea, M.T., Imber, S.D., Sotsky, S.M., & Watkins, J.T.(1993) Some conceptual and statistical issues in analysis of longitudinal psychiatric data. Archives of General Psychiatry, 50, 739-750.

Hedeker, D. (1989) Random Regression Models with Autocorrelated Errors: Investigating Drug plasma Levels and Clinical Response. Doctoral dissertation. Chicago: The University of Chicago.

Hedeker, D.(1993a) MIXOR: A Fortran program for mixed-effects ordinal probit and logistic regression. Technical report, Prevention Research Center, School of Public Health, University of Illinois at Chicago.

Hedeker, D.(1993b) MIXREG: A Fortran program for mixed-effects linear regression with autocorrelated errors. Technical report, Prevention Research Center, School of Public Health, University of Illinois at Chicago.

Hedeker, D. & Gibbons, R.D. (in press) A random-effects ordinal model for multilevel analysis. Biometrics.

Hedeker, D., Gibbons, R.D. & Flay, B.R.(In press) Random regression models for clustered data: With an example from smoking prevention research. Journal of Consulting and Clinical Psychology.

Jansen, J. (1990) On the statistical analysis of ordinal data when extravariation is present. Applied Statistics, 39, 75-84.

Editor's note: A paper by *Goldstein, Healy and Rasbash* will shortly appear in 'Statistics in Medicine' which describes the multilevel modelling of autocorrelated data in continuous as well as discrete time.

Explained Variance in Two-level Models

Tom A.B. Snijders & Roel J. Bosker

One way to approach the concept of 'explained proportion of variance' is to transfer its customary treatment straightforwardly to the hierarchical random effects model: treat proportional reductions in the estimated variance components as analogous to R^2 values. This approach leads to several R^2 value, one for each variance component. It can happen that adding explanatory variables increases rather than decreases some of the variance components. Even negative values for R^2 are possible. For this reason another approach is proposed.

For variance component models a measure of modelled (or explained) variation can be based on the principle of proportional reduction of prediction error. This can be defined for a single level model as

$$R^2 = \frac{var(Y_i) - var(Y_1 - X_i\beta)}{var(Y_i)} = 1 - \frac{var(Y_i - X_i\beta)}{var(Y_i)}$$

This formula expresses one of the equivalent ways to define R^2, and it is based on the reduction in the difference between the observed and the predicted outcomes.

The same principle can be used to define 'explained proportion of variance' in the hierarchical linear model. The level-1 explained proportion of variance R_1^2 is defined as the proportional reduction in mean squared prediction error:

$$R_1^2 = 1 - \frac{var(Y_{ij} - X_{ij}\beta)}{var(Y_{ij})}$$

The most straightforward way to estimate R_1^2 is to consider $\hat{\sigma}^2 + \hat{\tau}^2$ (the estimated level 1 plus level 2 variance) for the empty model (without predictor variables), and compute 1 minus the ratio of these values. In other words, R_1^2 is just the proportional reduction in the value of $\hat{\sigma}^2 + \hat{\tau}^2$ due to including the X-variables (the predictor variables) in the model.

Now we turn to the level-2 explained proportion of variance. It is natural to define this as the proportional reduction in mean squared prediction error for the prediction of $\overline{Y}_{.j}$. The level-2 explained proportion of variance is now defined as the proportional reduction in mean squared prediction error for $\overline{Y}_{.j}$ (the reduction in the difference between the observed and predicted mean level-2 outcomes):

$$R_2^2 = 1 - \frac{var(\overline{Y}_{.j} - \overline{X}_{.j}\beta)}{var(\overline{Y}_{.j})}$$

To estimate the level-2 explained proportion of variance, we follow a similar approach as for estimating R_1^2, for data with a constant group size n, we estimate R_2^2 as the proportional reduction in the value of $\hat{\sigma}^2/n + \hat{\tau}^2$. For differing group sizes one could use as the value for n either a value deemed a priori to be 'representative', or the harmonic mean.

What happens to R_1^2 and R_2^2 when predictor variables are added to the multilevel model? Population values of R_1^2 and R_2^2 in correctly specified models become smaller when predictor variables are deleted. For estimates of R_1^2 and R_2^2, however, the situation is different: these estimates sometimes do increase when predictor variables are deleted. When it is observed that an estimated value for R_1^2 or R_2^2 becomes smaller by the addition of a predictor variable, or larger by the deletion of a predictor variable, There are two possibilities: either this is a chance fluctuation, or the larger model is misspecified. In this sense changes R_1^2 or R_2^2 in the 'wrong' direction serve as a diagnostic for possible misspecification.

Proofs, examples and extension of the formulae to models with random slopes can be found in: *Snijders, T.A.B. and R.J. Bosker* (1994), Modelled variance in two-level models, *Sociological Methods and Research*, To be published.

Information in the February 1994, special issue of Sociological Methods & Research on Multilevel Analysis Methods.

This information is provided by *Joop Hox*. Back issues may be ordered from Sage. Inquiries to: Sage, 6 Bonhill Street, London EC2A 4PU, United Kingdom (Europe, Middle East) or Sage, 2455 Teller Road, Newbury Park, CA 91320, USA (USA, rest of world).

Multilevel Analysis Methods by *Joop J. Hox* and *Ita G.G. Kreft*. This special issue of SMR is about the analysis of data collected at different levels of observation and about the methodological problems that are present when natural experimentation and observations nested within existing social groups are the object of study. The methodological problems are summarized in the term multilevel problems. This article discusses some traditional approaches to the analysis of multilevel data and their statistical shortcomings. The random coefficient linear model is presented, and the currently available software is discussed. Next, some more general developments in multilevel modeling are discussed. The authors end with an overview of this special issue.

Hierarchical Regression Models for Interviewer and Respondent Effects by *Joop J. Hox*. It is generally recognized that interviewers may have an important effect on the quality of the data collected in survey research. This article presents an application of the hierarchical regression model in the analysis of interviewer effects. The hierarchical regression model offers an elegant way of analyzing the simultaneous effects of specific interviewer and respondent characteristics. It is especially attractive if the research design does not provide for a random assignment of respondents to interviewers, because it allows the researcher to use statistical rather than experimental control by modeling the interviewer effects conditional on the respondent effects.

The Gender Gap in Earnings A Two-Way Nested Multiple Regression Analysis With Random Effects by *Ita G.G.Kreft* and *JAN de LEEUW*. The gender income gap is a much debated subject both at an analytical and economic level. This article considers both, but emphasizes the different ways the data can be analyzed. The authors show that a hierarchical linear model is the best way to evaluate male-female wage differentials. Both interindustry and intraindustry wage disparities between men and women are measured by using a technique that assumes that observations within the same industry have correlated error terms. By simultaneously testing human capital factors and environmental factors, the analysis model serves as a link between theory and empirical analysis. The results show that the wage differences are larger in some industries than in others, so that it can be assumed that a gender income gap is not only a function of individual differences in qualification, but also differences between industries. The between-industry differences in gender income gaps contradict the hypothesis that gender income differential is largely due to female work preferences and the resulting segregation.

Modelled Variance in Two-Level Methods by *TOM A. B. SNIJDERS* and *ROEL J. BOSKER*. The concept of explained proportion of variance or modelled proportion of variance is reviewed in the situation of the random effects hierarchical two-level model. It is argued that the proportion reduction in (estimated) variance components is not an attractive parameter to represent the joint importance of the explanatory (independent) variables for modeling the dependent variable. It is preferable instead to work with the proportional reduction in mean squared prediction error for predicting individual values (for the modelled variance at level I) and the proportional reduction in mean squared prediction error for predicting group averages (for the modelled variance at level 2). (See an article on page 15).

Multilevel Cross-Classified Models by *Harvey Goldstein*. Random cross-classifications of units can arise at any level of a data hierarchy. For example, school students may be classified both by the schools they attend and their neighbourhoods of residence. This article explores the issues of efficiently modeling such data and gives an example from a study of parental choice of schools.

Multilevel Covariance Structure Analysis by *BENGT O. MUTHEN*. This article gives an introduction to some new techniques for multilevel covariance structure modeling with latent variables. Although these techniques only incorporate a subset of models that are relevant to multilevel data, the techniques do provide a large set of new analysis possibilities and have the advantage that they only require conventional structural equation modeling software. The presentation draws on methodology presented in earlier works by the author.

The Bilevel Reticular Action Model for Path Analysis With Latent Variables by *RODERICK P. McDONALD*. A two-level (hierarchical) model for path analysis with latent variables is described, together with some properties of a computer program written to implement the model. A simple illustrative example is given.

Some New References to Multilevel Modelling

Draper, D. (1993) Inference and Hierarchical Modeling in the Social Sciences (unPub).

Duncan, C., Jones, K. and Moon, G. (1993) Do places matter? A multilevel analysis of regional variations in health-related behaviour in Britain, Social science and medicine, 37(6): 725-733.

Jones, K. (1993) Using multilevel models for survey analysis, Journal of the Market Research Society, 35(3):249-265.

Jones, K. & Bullen, N. (1993) A multilevel analysis of the variations in domestic property prices: Southern England 1980-87, Urban Studies, 30(8):1409-1426.

Jones, K., Johnston, R.J. & 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.

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

Raudenbush S.W. (1993) A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research, Journal of Educational Statistics, 18(4): 321-349.

Please send us any multilevel modelling publications for inclusion in this section in future issues.

Correction: The contact address for the ALCD project, Latent Variable Models in Social Science, which appeared in our last issue of the newsletter (Vol.5 no. 3) on page 3 should be to Dr. ***Lilian de Menezes***, Methodology Institute, London School of Economics & Political Sciences, Houghton Street, London WC2A 2AE.

Contributors

We are grateful to the following who provided articles or news for this issue.

Roel J. Bosker: Dept of Education, University of Twente, The Netherlands (R.J.Bosker@edte.utwente.nl)

Paul Burton: The Western Australian Research Institute for Child Health, Pertth W.A.6001, Australia (paulb@weema.chi.uwa.edu.au)

Martin R Delap: Research and Statistics Group, Associated Examining Board, Stag Hill House , Guildford, Surrey, GU2 5XJ, UK.

David Draper: University of California, Los Angeles (USA) and University of Bath (UK) (d.draper@maths.bath.ac.uk)

Craig Duncan & Kelvyn Jones: Department of Geography, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth PO1 3HE, UK (jonesk@csovax.portsmouth.ac.uk)

Donald Hedeker: The University of Illinois Chicago, Prevention Research Center, (U41098@UICVM)

Joop Hox: Faculty for educational Sciences, University of Amsterdam, Amsterdam, The Netherlands (a716hox@hasara11)

Steve Raudenbush: College of Education, Michigan State University. East Lansing, MI 48824, USA (19112hlm@edu.msu)

Tom A.B. Snijders: Dept of Statistics & Measurement Theory, Univ. of Groningen, The Netherlands (t.b.snijders@ppsw.rug.nl)

Daniel Wright and Anne T. Sparks: Memory and Cognition Group, Psychology Department, City University, Northampton Square, London, EC1V 0HB, UK.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

A change in procedure
To order ML3 please now contact:
Mrs. Milene Adaken
Finance Department, Institute of Education
20 Bedford way, London WC1H 0AL
Tel: 071 612 6024, Fax: 071-612-6032
Credit cards accepted

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~