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. 2 September, 1994

Multilevel Modelling Course: Another three-day course on multilevel modelling using ML3 will be held at the Institute of Education, University of London from 22 to 24 November 1994. For further information contact the Multilevel Models Project at the address above on the top of this page.

A New Project on Multilevel Modelling: A new project supported by Medical Research Council (MRC) has been started at the Institute of Education, University of London. This project will focus on the development of methodology and software for the efficient construction and use of population norms. The project will explore and develop new procedures for efficient longitudinal norms by utilizing existing methodology for multilevel modelling. The research officer on the project is Huiqui Pan who is working with Jon Rasbash and Harvey Goldstein.

Also In This Issue

ML3 Clinics in London 1994/1995

Free for users of *ML3/ML3-E*

Tuesday October 11
Tuesday November 15
Tuesday December 6
Tuesday January 3
Tuesday February 7
Tuesday March 7

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

Project News

Advanced Multilevel Modelling Workshop for Political Scientists, Institute of Education: 4-5 July 1994

There were eighteen participants from the U.K., Ireland, Belgium and the Netherlands. The workshop was supported by a grant of £1,000 from the ESRC under the Analysis of the Large and Complex Datasets (ALCD) programme and was jointly organised by *Anthony Heath* (Nuffield College), *Min Yang* and *Harvey Goldstein* (Institute of Education), assisted by Professor *Kelvyn Jones* (University of Portsmouth).

The first day was devoted to an introduction to the British Election Study data (BES) which formed the basis of the workshop's data analysis. A subset of these data for the years 1964, 1970, 1974, 1983, 1987, 1992 was used. The principal analyses used as a response whether or not the respondent voted Conservative, and a logistic-multilevel model was developed for this . Predictor variables included social class and region, The participants were introduced to the ML3 software package and shown how to fit this class of models. They were then encouraged to explore the data set with models of increasing complexity and to ask substantively interesting questions.

On the second day the workshop concentrated on the analysis of some of the BES panel data (1983, 1986, 1987) and participants were shown how to fit a 3-level repeated measures model for a binary response. Participants continued with their own data analyses and the afternoon was taken up with presentations from participants as follows:

1. *Paul Nieuwbeerta* on trends in class voting using time series data from 20 industrialised countries after 1945.

2. *Paula Surridge* on the effects of constituency contextual variables, such as average employment status, on the probability of voting Conservative.

3. *Kelvyn Jones* on the relationship between the probability of voting Labour and constituency level variables such as the percent of mining activity.

The discussion of the these presentations raised a number of interesting substantive issues, including the possibility of modelling vector outcomes with choice among several parties, and the possibility of random cross classifications. As a result a number of further cooperations were set up.

The Analysis of Large and Complex datasets (ALCD) Commissioning of Phase 2

The ALCD programme aims to help ensure that the ESRC's investment in the collection and storage of large and complex datasets is better exploited. Phase 1 of the research programme started in September 1993, and 15 mainly methodological projects are being funded. Topics covered include:

Duration and event history analysis of longitudinal data;
multilevel-modelling;
selection, attrition and other sources of error;
advances in computer analysis;
merging different datasets.

Phase 2 will concentrate on the methodology of Data Resource Management under the theme of Data Exploitation. It is scheduled to start in September 1995. The ALCD working group are currently discussing the form of the Phase 2 programme. In broad terms they would like this phase to:

- include research projects which exploit many datasets of different types and which promise long term contributions to applied research using complex data
- emphasise deliverables, including software, documentation and training courses, which will promote access to complex data and improve data resource management
- promote dissemination of deliverables drawn from Phase I projects.

It is hoped that the Fellowship scheme introduced in Phase 1 will continue in Phase 2. (The next round of Fellowships will be advertised shortly).

In order to flesh out these aims and to identify priorities the ALCD Co-ordinator, Professor *Fred Smith* of the university of Southampton convened a workshop, which included both potential users and researchers, held in Loughborough in May. The results of the workshop will help inform the members of the ALCD working group. All those with an interest in ALCD research are invited to send their views about Phase II to *Fred.* These views will then be transmitted to the members of the work group.

The funding for Phase 2 has been set at £1.2 mn over about three years. This is your opportunity to influence the form of this important research programme, and anybody wishing to express views about the content of Phase 2 of ALCD should contact *Fred Smith* at *Department of Mathematics, University of Southampton, Highfield, Southampton SO17 1BJ. Tel. 0703 593655, Fax. 0703 593939, Email. tmfs@uk.ac.soton.maths.*

Further information about the ALCD Programme is available from ESRC's Research Resources Division, Tel: 0793 413103

* ... * ... * ... * ... * ... * ... * ... * ... * ... * ... * ... * ... *

Some New References to Multilevel Modelling

Baxter-Jones, A., Goldstein, H., and Helms, P. (1993). The Development of aerobic power in young athletes. *Journal of Applied Physiology*, **75**, 1160-1167.

Belin, T. R., Diffendal, G. J., Mack, S., Rubin, D. B., Schaffer, J. L., & Zaslavsky, A. M. (1993). Hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation. *J.American Statistical Association*, **88**, 1149-1159.

Carriere, K. C. (1994). Incomplete repeated measures data analysis in the presence of treatment effects. *J. American Statistical Association*, **89**, 680-86.

Drum, M. L., & McCullagh, P. (1993). REML estimation with exact covariance in the logistic mixed model. *Biometrics*, **49**, 677-690.

DuToit, S. H. C. (1993). Analysis of multilevel models. Pretoria: *Human Sciences Research Council.*

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., & Thomas, S. (1993). A multilevel analysis of school examination results. *Oxford Review of Education*, **19**, 425-433.

Goldstein, H., Healy, M.J.R., and Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, **13**: 1643-1656.

Hox, J. J. (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological methods & research*, **22**, 300-318.

Jones, K. (1993). Everywhere is Nowhere: multilevel perspectives on the importance of place. Portsmouth: University of Portsmouth. (Unpub)

Kang, S. J., & Raudenbush, S. W. (1994). A covariance components model with two-way crossed random effects and ML estimation via the EM algorithm. (Unpub)

Kristiansen, I.S., & Holtedahl, K. (1993). Effect of the remuneration system on the general practitioner's choice between surgery consultations and home visits. *Journal of Epidemiology and Community Health*, **47**: 481-484.

Langford, I. H., Bateman, I. J., & Langford, H. D. (1994). A multilevel approach to triple-bounded dichotomous choice contingent valuation. *CSERGE Working Paper GEC 94-04.*

Lipsitz, S. R., & Fitzmaurice, G. M. (1994). Sample size for repeated measures studies with binary responses. *Statistics in Medicine*, **13**, 1233-1240.

Lipsitz, S. R., Kim, K., & Zhao, L. (1994b). Analysis of repeated categorical data using generalised estimating equations. *Statistics in Medicine*, **13**, 1149-64.

Longford, N. T. (1993). Inference about variation in clustered binary data. (Unpub)

Longford, N. T. (1993). Regression analysis of multilevel data with measurement (Unpub)

Mark, S. D., & Gail, M. H. (1994). A comparison of likelihood based and marginal estimating equation methods for analysing repeated ordered categorical responses with missing data. *Statistics in Medicine*, **13**, 479-93.

McCulloch, C. E. (1994). Maximum likelihood variance components estimation for binary data. *J. American Statistical Association*, **89**, 330-35.

McDonald, B. W. (1994). Two random effects models for multivariate binary data. *Biometrics*, **50**, 164-72.

McDonald, R. P. (1993). A general model for two-level data with responses missing at random. *Psychometrika*, **58**, 575-585.

Miller, M. E., Davis, C. S., & Landis, J. R. (1993). The analysis of longitudinal polytomous data: generalised estimating equations and connections with weighted least squares. *Biometrics*, **49**, 1033-1044.

Rutter, C. M., & Elashoff, R. M. (1994). Analysis of longitudinal data: random coefficient regression modelling. *Statistics in Medicine*, **13**, 1211-32.

Snijders, T. A. B., & Bosker, R. J. (1994). Modelled variance in two-level models. *Sociological methods & research*, **22**, 342-363.

Thomas, N., & Longford, N. T. (1994). Empirical Bayes methods for estimating hospital specific mortality rates. *Statistics in Medicine*, **13**, 889-903.

Verbyla, A. P. (1993). Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *J. Royal Statist.Soc.B.*, **55**, 493-508.

Waclawiw, M. A., & Liang, K.-Y. (1994). Empirical Bayes estimation and inference for the random effects model with binary response. *Statistics in Medicine*, **13**, 541-55.

Ware, J. H. (1993). Analysis of longitudinal data: choosing and interpreting regression models. *European Respiratory Journal*, **6**, 325-327.

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

Book Review

Rondom Coefficient Models by N T Longford. Oxford University Press 1993. Pp xiv + 270, £30.00. ISBN 0 19 852264 9

Goldstein (1987) and *Bryk* and *Raudenbush* (1992) have written books on multilevel modelling, and now *Nick Longford*, another of the leaders in this field of research, has done the same. The book covers very much the same ground as the other two. After a not particularly helpful introductory chapter (the reader of a book like this is unlikely to need to be told the definition of a symmetric matrix or the density function of the Normal distribution), the main subject is introduced in a chapter entitled *'Analysis of covariance with random effects'* which deals with variance component models. This chapter introduces five specimen datasets which are used as examples throughout the book. Next comes a chapter called *'Random regression coefficients'* describing two-level models. Later chapters introduce multiple levels of nesting, generalized linear models with random coefficients and factor-analysis models including measuring errors. Most chapter include extensive bibliographical notes and there are some 200 references.

What marks the book out as contrasting with its predecessors? The mathematical level is fairly sophisticated - most of the mathematical results are formally proved and the reader must not be put off by some formula-rich pages. Where computation is concerned, the author describes briefly the EM and interactively reweighted least squares methods, but strongly recommends the Newton-Raphson or Fisher scoring method. His approach can be characterised by the remark

> 'Using the new generation of statistical software (Splus, GAUSS or Metlab) gives the qualified analyst a diatinct advantage over the modules in general purpose statistical software (eg in BMDP or SAS) or specialized software for multilevel analysis (HLM, ML3 or VARCL).'

It is perhaps the qualified analyst referred to who will find this book the most useful.

References

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

Goldstein, H (1987) Multilevel Models in Educational and Social Research. Griffin.

M J R Healy

Some Software from *Pro*GAMMA

VARCL: VARCL (VARiance Component analysis by maximum Likelihood) by N.T. Longford employes the Fisher scoring algorithm for multilevel analysis. The main program allows three levels of nesting, but the package also contains a program version that allows up to nine levels (restricted to simple specifications of the random effects). VARCL can constrain regression coefficients and variance parameters to arbitrary values, and fix covariance parameters to zero. Users can choose between Normal and non-normal distributions (Poisson, Binomial and Gamma) for the dependent variables. It costs at $250.

HLM/2 & 3-Level: HLM (Hierarchical Linear Modeling) developed by A.S. Bryk and S. Raudenbush combines a two-level and a three-level version. It conveniently takes the user through the specification of the model at the respective levels. Estimates are three kinds of parameters: empirical Bayes estimates of randomly varying lowest level coefficients; GLS estimates of the level-2 coefficients or ML estimates of the level-3 coefficients and of the variance-covariance components. It provides options for multivariate hypothesis tests for the fixed effects and variance-covariance components. The program closely coordinates with the book HLM, *Applications and Data Analysis Methods*. The book describes use of both two- and three-level models in organizational research, studies of individual development, and meta-analysis applications. The software costs $415 and the book costs $40.

ML3-E v2.3: ML3-E, developed by Harvey Goldstein and Jon Rasbash, allows three levels of nesting with DOS extender. It uses Iterative Generalized Least Squares as the estimation algorithm. Data are stored in a 'worksheet', with extensive options for data manipulation and high resolution graphing. ML3-E allows linear constraints on both the fixed and random effects. Heterogeneous residual variances at different levels can be modelled. Multivariate models and cross-classified models can be done routinely. It can also fit nonlinear models by using special macros. Some simulation tools are available. Three books are coordinated with the program priced together at $460 for academic users.

For detailed information contact

iec *Pro*GAMMA
P.O.Box 841
9700 AV Groningen
The Netherlands
Tel: 31 50 636900
Fax: 31 50 636687
email: gamma.post@gamma.rug.nl

Contributors

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

Blien, Uwe, Institute for Employment, Research IAB, Postfach, D-90327 Nurnberg, Germany.

Eeden, Pieter van den, Department of Social Research Methodology, Vrye Universiteit, Knoingslaan 22-24, Amsterdam 1075 AD, Netherlands.

Engel Uwe and *Hurrelmann Klaus*, Research Center Prevention & Intervention Childhood & Adolescence, University of Bielefeld.

Heath, Anthony, Nuffield College, Oxford OX1 1NF, England.

Healy, M J R, Institute of Education, University of London, 20 Bedford Way, London WC1 0AL

Hox, Joop, Department of Education, University of Amsterdam, Ijsbaanpad 9, 1076 CV Amsterdam, Netherlands, *Jaap Dronkers* and *Hubert Schijf.*

Langford, Ian, Health Policy and Practice Unit, University of East Anglia, Norwich NR4 7TJ, England.

Maas. C.J.M., Department of Methodology, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, Netherlands.

Nieubeerta, P., *Graaf, N.D. de* and *Ultee, W.*, Department of Sociology, catholic University, Postbus 9108, 6500 HK Nijmegen, Netherlands.

Pan, Huiqi, Department of MSC, Institute of Education, University of London, 20 Bedford Way, London WC1A 0AL, England.

Smith, Fred, Department of Mathematics, University of Southampton, Highfield, Southampton SO17 1BJ.

Wiggins, Richard D. and *Wale, Christopher J.* Social Statistics Research Unit, City University, London EC1V 0HB.

Yang, Min and *Goldstein, Harvey*, Department of MSC, Institute of Education, University of London, 20 Bedford Way, London WC1A 0AL, England.

The Multilevel Models Project at the Institute of Education is joined by John Howarth from September 1994. He will work closely with Jon Rasbash on programming of ML3.

Theory and Applications

Applications of Multilevel Models in Sociology

The XIII International Sociology Congress was held in 18 - 23 July 1994 in Bielefeld, Germany. Following are some abstracts of papers using multilevel models presented to the conference.

Multilevel analysis of the changing relationship between class and party in Britain, 1964-1992, by *Anthony Heath* (Nuffield College, Oxford OX1 1NF, England), *Min Yang* and *Harvey Goldstein* (Department of MSC, Institute of Education, University of London, 20 Bedford Way, London WC1A 0AL, England). Previous analyses of the changing relationship between class and vote in Britain have assumed that the British Election Surveys constitute simple random samples. In fact, they are all clustered samples, and the number of sampling points has varied substantially over time. The paper uses the statistical technique of multi-level modelling to investigate the effects of this clustering and compares the results with those obtained with single-level logistic models. In general, the multilevel and single-level models lead to similar conclusions about the changing relation between class and vote: they both show evidence of a change in the class/vote relationship over time. However, the multilevel models also show that, while the clustering does not affect conclusions about the class dealignment debate, there are other important substantive findings which emerge from the multilevel approach. First, there is clear evidence of substantial constituency differences in the intercepts; that is individuals from the same social class had very different propensities to vote Conservative in different constituencies. Second, there were also significant constituency differences in the slopes for the working classes although not for other classes; that is, constituencies seemed to vary in their level of class polarization.

Analysing regional wage structures with multilevel models, by *Uwe Blien*, (Institute for Employment, Research IAB, Postfach, D-90327 Nurnberg, Germany). In the paper a study is presented that uses a multilevel linear model with random coefficients to estimate a wage function. Multilevel models allow the combination of variables recorded at the level of aggregates.

The analysis shows that the regional unemployment rate affects the regional wage: every additional percent of the rate is associated with a reduction of daily income by 1.7 percent. Thus, evidence corroborating the so called wage curve was found.

There are additional wage differences. A general direct regional effect is relatively small. Another effect causes regional variation in the gap between women's and men's wages. This result indicates that women are a 'mobile reserve army' in the labour market. Therefore, their wage reacts flexible to different labour market conditions in different regions.

The various effects can be attributed to the existence of regional segmented labour markets.

Intergenerational transmission in longitudinal birth cohort data using multilevel modelling, by *Wiggins, Richard D.* and *Wale, Christopher J.* (Social Statistics Research Unit, City University, London EC1V 0HB). The study of relationship within families across generations is of great social importance. During the fifth sweep of the 1958 National Child Development Study (NCDS) study when cohort members were 33 years old a subsample of one third of cohort members children were investigated (2,617 children from 1761 families). Analysis to explore intergenerational transmission from parent to child gives rise to a very complex data structure requiring a multi-level approach, which recognizes the nesting of generational units. These nesting levels will be classified as follows: the highest level is that of the cohort member with their own personal characteristics. The next level contains two types of units. The first is the set of repeated measurements on a cohort member (eg., educational attainment). The other type of unit is that formed from the relationships with the partner(s) producing different offspring. By viewing the offspring as 'nested' within the cohort member, it becomes possible to interpret the variation and covariation of offspring characteristics, partner characteristics and the characteristics of the cohort member, at the level of the cohort member.

The paper demonstrates two principal advantages of adopting a multilevel perspective: (a) the ease with which unbalanced data can be accommodated (eg, different numbers of offspring per each cohort member; (b) study of variation at different levels of the hierarchy. Finally, the approach opens up new areas of secondary analysis of existing data which are expected to have an impact on research designs concerned with the collection of intergenerational data. The work has been funded under the ESRC's research programme into the Analysis of Large and Complex Datasets (ALCD) and is carried out in collaboration with the Multilevel Models Project at the Institute of Education.

An application of a multilevel structural equation model, by *Joop Hox* (Department of Education, University of Amsterdam, Ijsbaanpad 9, 1076 CV Amsterdam, Netherlands), *Jaap Dronkers* and *Hubert Schijf*. Social science data often have a hierarchical structure, with variables defined at both a group and an individual level. This article describes a recently developed method to analyse multilevel data with structural models and latent variables, and presents an example of a multilevel path analysis on educational data in the Netherlands. Although specialized software would have its advantages, the model described here can be estimated with conventional structural modeling software. Similarities and dissimilarities with other multilevel approaches are also discussed.

Primate of multilevel Analysis with Respect to Hierarchical Nested data, by *C.J.M. Maas* (Department of Methodology, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, Netherlands). The last few years nested data are more and more often analyzed with multilevel analysis programs which are specially developed for this purpose. The advantage of analyzing with such programs, with respect to the techniques used up until now, are stated in this article both in theory as with an empirical example. The following techniques are discussed: aggregation, disaggregation, two-stage regression analysis and covariance analysis. The presentation of the random coefficient model shows how to handle the problems of these four techniques. In an empirical example the different techniques are compared and the usefulness of the multilevel analysis is shown.

Contextual Effects of Social Mobility on Political Party Preference in 15 OECD Countries: 1964-1991, by *P. Nieubeerta, N.D. de Graaf* and *W. Ultee* (Department of Sociology, catholic University, Postbus 9108, 6500 HK Nijmegen, Netherlands). In this paper we test several individual and contextual hypotheses about the impact of intergenerational class mobility on political party preferences. We test these hypotheses by employing multilevel models. We run these models on 150 cross-sectional data sets representing 15 OECD countries (Australia, Austria, Canada, Denmark, Britain, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Sweden, Switzerland, and USA) over the period 1958-1989. The contextual analysis shows that: (a) a class with a high level of inflow mobility has a bigger impact on newcomers than does a class with a lower level of inflow; (b) that the higher the level of outflow mobility in the manual class the more right-wing the political preference of this class is, and the higher the level of outflow in a non-manual class is, and the higher the level of outflow in a non-manual class the more left-wing the political preferences of that class is; and (c) that the higher level of 'deviant political inflow' into a class, the more political preference of the immobile will change in the direction of the newcomers' preferences. The relevance of these findings is discussed.

The effects of interviewer and respondent characteristics on answer behaviour in survey research: A multilevel approach, by *Pieter van den Eeden* (Department of Social Research Methodology, Vrye Universiteit, Knoingslaan 22-24, Amsterdam 1075 AD, Netherlands), *Johannes H. Smit, Dorly J.H. Deeg* and *Aartjan T.F. Beekman*. Until recently, the study of interviewer effects has focused on establishing direct effects of interviewer characteristics on respondent response. Recently, an alternative approach has ben developed which emphasizes the conditioning influence of the interviewer characteristics on the respondent's answering process. The objective of this paper is to illustrate this alternative approach with empirical evidence, using the random coefficient hierarchical regression model.

This model's structure is basically as follows. First, the answering process is described at the level of the respondent. Subsequently, respondent specific parameters are related to interviewer specific variables. This structure allows inclusion of the coefficient resulting from the intra-interviewer regression in the regression equation at the interviewer level (inter-interviewer regression model). Thus, the variance to be explained is split up in a respondent part (level 1) and an interviewer part (level 2). This two-level model is applied to data collected in the Longitudinal Aging Study Amsterdam (LASA; 2838 respondents within 43 interviewers).

The dependent variable is a scale indicating well-being (Center for Epidemiological Studies Depression Scale); background variables on respondent level are age, sex and self-perceived health. Interviewer variables are age, education, personality traits and social skills.

Causal analysis of panel data, by *Engel Uwe* and *Hurrelmann Klaus* (Research Center Prevention & Intervention Childhood & Adolescence, University of Bielefeld). Though it seems reasonable to regard longitudinal data as highly useful or even necessary to study causal relationships empirically, such data alone can not provide a sufficient base for that purpose. Panel data designs are as useful as they are subject to complicating methodological factors. They let causal inferences draw on the way putative cause & effect measures develop over time, but to take advantage of this diachtonic perspective, special care is necessary to handle well-known problems in repeated measurement designs, including panel attrition, measurement error, and the possibility of complex temporal effect pattern. Focussing on this last problem, hierarchical growth curve modeling is discussed as a tool for causal analysis, using panel data from the Bielefeld (Germany) Study on Risk behaviour in Youth.

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

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

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

A cross-classified multi-level model of district mortality rates

Ian H. Langford

Mortality rates in England and Wales display a persistent regional pattern indicating generally poorer health in the north and west. Some of this is simply a reflection of regional differences in the extent of social deprivation which is known to exert a profound influence on health. Part of the pattern may also be the result of regional differences in urbanisation which also affect mortality rates. However, there may be important regional differences over and above these compositional effects. This study uses a cross-classified multi-level model in an attempt to establish the magnitude of such independent regional differences in mortality rates.

Data

The data consisted of age standardised mortality ratios (SMRs) for males and females under 65 for 1989-91 in 401 local authority districts of England and Wales (excluding the City of London and Isles of Scilly). Nine ACORN families described by *Craig* (1986) were used as a means of classifying the districts on a demographic and socio-economic basis (outer and inner London were aggregated), whilst the nine standard regions of England and Wales gave a purely geographical classification. The data on SMRs were then cross-classified using a multi-level model to produce an analysis of variance which takes into account both ACORN classification and regional location within the same model. The SMRs, male and female, for each district are at level 1, then level 2 may be ACORN family or region. The main point of interest is that a district in any one region may be in any of the ACORN families, and vice-versa. Two covariates were included in the analysis. The Townsend Index (x_1) was used as a measure of social deprivation measured on a continuous scale, centred around zero (*Townsend* et al 1988). Sex was included as a dummy variable x_2 to allow males and females to be included in the same model.

Heterogeneous population sizes

The natural logarithm of the standardised mortality ratios (LSMR) was used as the dependent variable. This does not represent individual data, but an aggregate measure of mortality risk in districts with heterogeneous denominator populations. Hence, we have a 'hidden' level which consists of the individual deaths within each district. One method of dealing with the situation is to model the joint distribution of cases within districts and SMRs between districts. For example, the cases within districts may be assumed to follow a Poisson distribution, and the rates between districts a gamma distribution (*Clayton* and *Kaldor* 1987; *Langford* 1994). In Bayesian terms, the likelihood of cases occurring within districts is conditional upon prior information on the distribution of SMRs between districts. The posterior distribution can then be expressed in terms of empirical Bayes estimates of risk, or as a negative binomial distribution of individual cases. However, here we are dealing with large numbers of deaths and the random effects generated by heterogeneous population size can be modelled more directly by a weighting factor proportional to $(n_{i(jk)})^{-0.5}$, for the *i'th* district in the *j'th* ACORN category, and *k'th* region, where $n_{i(jk)}$ is the denominator population of the *i'th* district. Hence, a simple variance components model can be written as:

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

$$e_{i(jk)} \sim N(0, \sigma_e^2)$$

$$f_{i(jk)} \sim N(0, \sigma_f^2)$$

where u_j and u_k are random effects due to ACORN family and region respectively around an overall mean a, and

$$f_{i(jk)} = \frac{n_{i(jk)}^{-0.5}}{N^{-1}\Sigma_i n_{i(jk)}^{-0.5}}$$

where N is the number of districts. Hence, $f_{i(jk)}$, the terms associated with heterogeneous population size, have been scaled to have a mean of unity to allow for direct comparison of the magnitude of their effect compared to $e_{i(jk)}$, the unconditional estimates of variance at level 1. The level 1 variance is equivalent to the random effects model of *Breslow* (1984) except that the dispersion parameter used to account for extra-Poisson variation is dependent on population size.

Results and discussion

Table 1 shows the results for the variance components model (with the addition of the Townsend Index in the fixed part to control for the overall effect of social deprivation), i.e.

$$LSMR_{i(jk)} = \alpha + \beta_1 x_{1i(jk)} + u_j + u_k + e_{i(jk)} + f_{i(jk)}$$

Table 1 Variance components for cross-classified model (controlling for TINDEX in fixed part of model)

Parameters	Estimate	S.E.
Random		
Level 2 (region)		
σ_u^2	0.004331	0.002116
Level 2 (ACORN)		
σ_u^2	0.001112	0.000594
level 1		
σ_e^2	0.003907	0.000459
σ_f^2	0.000320	0.000396
Fixed		
α	4.582	0.0248
β_1	0.0496	0.0009

-2 ln likelihood = -2041.92, d.f. = 795 (males and females included in the same response variables).

As can be seen, the effect of heterogeneous populations σ_f^2 is relatively small in this particular case. The interesting fact is that around four times as much variation is occurring between regions as between ACORN categories, suggesting that regional location is of great importance. Table 2 shows the results of a more complex model, with variance and co-variance terms involving SEX x_2 and TINDEX being included in the random part of the model. SEX and was constrained to be zero in the fixed part of the model, and not included at regional level in the full model to reduce time between iterations, but had very little effect when modelled on its own at this level.

Some of the parameter estimates in the random part at the higher levels are estimated as zero. This is partly due to having a low number of regions and ACORN categories (nine of each), but random coefficients models with ordinary leastsquares also estimated very small differences between regions or ACORN categories for these variables.

Table 2 Full cross-classified model

Parameters	Estimate	S.E.
Random		
Level 2 (region)		
σ_u^2	0.004331	0.002116
σ_{u,x_1}	0.0	0.0
$\sigma_{x_1}^2$	0.0	0.0
Level 2 (ACORN)		
σ_u^2	0.001358	0.000732
σ_{u,x_1}	0.0	0.0
$\sigma_{x_1}^2$	0.0	0.0
σ_{u,x_2}	-0.000366	0.000345
σ_{x_1,x_2}	0.0	0.0
$\sigma_{x_2}^2$	0.000309	0.000249
Level 1		
σ_e^2	0.003907	0.000459
σ_{u,x_1}	-0.000007	0.000131
$\sigma_{x_1}^2$	0.000024	0.000014
σ_{u,x_2}	0.001091	0.000865
σ_{x_1,x_2}	0.000089	0.000071
σ_{f,x_1}	-0.000070	0.000131
σ_{f,x_2}	-0.000609	0.000851
σ_f^2	0.000598	0.000507
Fixed		
α	4.583	0.0243
β_1	0.02534	0.00097

-2 ln likelihood = -2056.53, d.f. = 782 (males and females included in the same response variables).

One of the main aims of the study was to examine in the residual variation left over from the model in Table 2, and so variables were not omitted on grounds of significance. The diagnostic, or level 1 residuals were calculated for each district, representing the proportional excesses or deficits in mortality at district level when social deprivation and the crossed effects of ACORN categorisation and regional location were accounted for. Figures 1 and 2 show the distribution of diagnostic residuals for males and females, and perhaps surprisingly, a very strong geographical patterning is evident with a rough line between the Severn and the Wash distinguishing excess mortality in the North and West from deficits in the South and East. This suggests that some factor operating at a very large spatial scale is influencing mortality patterns beyond the variation which can be accounted for by social deprivation, regional and socio-economic classification. The possible explanations for this North-South divide could include differences in:

(a) climate - health has been related to a number of climatic variables (*Langford* and *Bentham* 1993);

(b) geology - differences in factors such as hardness of water, and mineral intake may influence health;

(c) cultural practices - do Northerners really subsist on beer, meat pies and cigarettes?

Further research will examine age disaggregated mortality rates, as well as examining self-reported morbidity as measured in the 1991 Census.

References

Breslow, N.E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, **33**, 38-44.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671-681.

Craig, J.A. 1981 socio-economic classification of local and health authorities of Great Britain. *OPCS Studies on Medical and Population Subjects* No. **48**. London: HMSO 1985.

Langford, I.H. (1994). Using empirical Bayes estimates in the geographical analysis of disease risk. *Area*, (summer issue - in press).

Langford, I.H. and Bentham, G. (1993). The potential impacts of climate change on winter mortality in England and Wales. *CSERGE Working Paper GEC 93-25*, University of East Anglia, U.K.

Townsend, P., Phillimore, P. and Beattie, A. (1988). Health and deprivation: inequality and the north. *Croon Helm, London.*

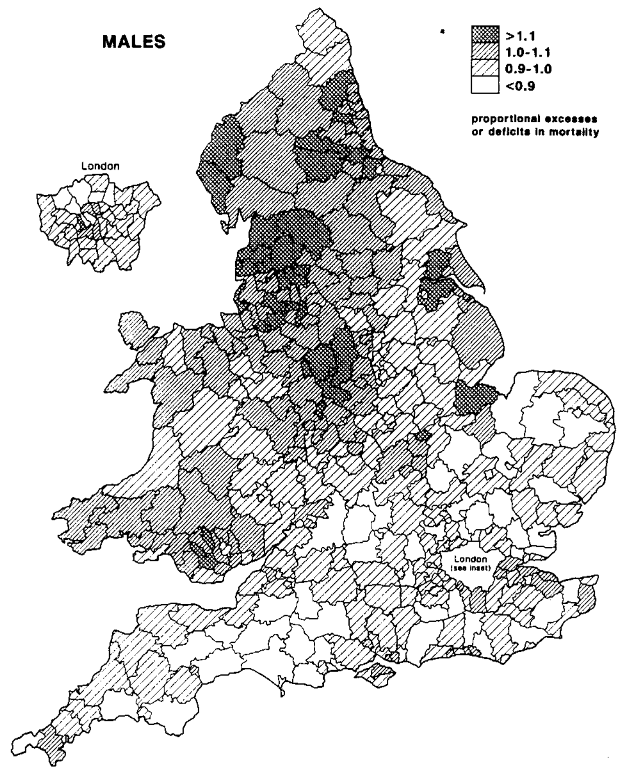

Figure 1 Residual variation in SMRs at district level for males

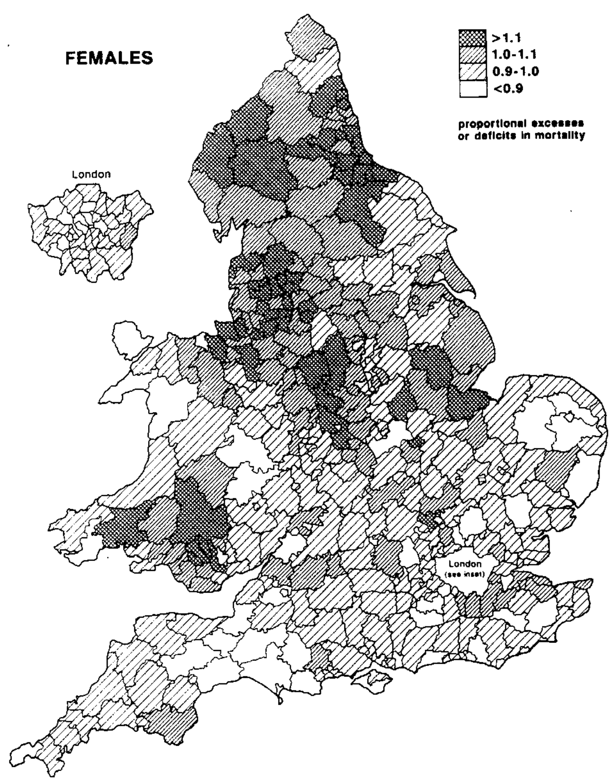

Figure 2 Residual variation in SMRs at district level for females