Contents

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Chapter 9

Chapter 10

Chapter 11

References

Preface

In the mid 1980's a number of researchers began to see how to introduce systematic approaches to the statistical modelling and analysis of hierarchically structured data. The early work of Aitkin et al (1981) on the teaching styles' data and Aitkin's subsequent classic work with Longford (1986) initiated a series of developments that, by the early 1990's had resulted in a core set of established techniques, experience and software packages that could be applied routinely. These methods and further extensions of them are described in this book and are coming to be applied widely in areas such as education, epidemiology, geography, child growth, household surveys and many others.

In addition to the first edition of the present text (Goldstein, 1987b), two expository volumes appeared in the early 1990's. That by Bryk and Raudenbush (1992) discusses 2 and 3-level linear multilevel models with applications especially to educational data and also to repeated measures designs. Longford (1993) gives a more theoretically oriented account and includes additionally discussion of a multilevel factor analysis model, models with categorical responses and multivariate models. The present volume aims to integrate existing methodological developments within a consistent terminology and notation, provide examples and explain a number of new developments, especially in the areas of discrete response data, time series models, random cross classifications, errors of measurement, missing data and nonlinear models. In many cases these developments are the subject of continuing research, so that we can expect further elaborations of the procedures described.

The main text seeks to avoid undue statistical complexity, with methodological derivations occurring in appendices. Examples and diagrams are used to illustrate the application of the techniques and references given to other work. The book is intended to be suitable for graduate level courses and as a general reference.

Harvey Goldstein

August, 1994

Preface to the first Internet edition

It is now nearly 5 years since the second edition was completed. Since then there have been many developments; in methodology, in applications and in computation. A new edition of Multilevel Statistical Models is now being planned and it will incorporate these developments. In the meantime the second edition has been corrected and one or two topics amplified, with some additional references. This edition does not contain a subject index; readers can search the text electronically for topics. Information about current issues in multilevel modelling can be obtained from the folowing web site which has further useful links; www.ioe.ac.uk/multilevel/ .

Harvey Goldstein

h.goldstein@ioe.ac.uk

April 1999

Acknowledgements

This book would not have been possible without the support and dedication of all those who have worked on or been closely associated with the Multilevel Models Project at the Institute of Education: Jon Rasbash, Min Yang, Bob Prosser, Geoff Woodhouse, Pan Huiqui, Michael Healy, Ian Plewis and Bill Browne. I am also most grateful to the Economic and Social Research Council for their continuing funding support since 1986, most recently under the auspices of the programme for the Analysis of Large and Complex Datasets.

In addition, the following most generously have allowed me access to datasets, commented critically on chapter drafts or pointed out errors: Bob Carpenter, Arto Demirjian, David Draper, Peter Egger, Anthony Heath, Kelvyn Jones, Ita Kreft, Iuri da Costa leite, Toby Lewis, Mac MacDonald, Rod McDonald, Colm O'Muircheartaigh, Lindsay Paterson, Steve Raudenbush, German Rodriguez, Pamela Sammons, Richard Tanguay, and Sally Thomas.

Notation

The following definitions refer to a 2-level model. The extension to three and higher level models is usually straightforward. Where this is not clear, a three level definition is included.

Definition	Symbol
Response variable vector	Y
Explanatory variable design matrix	X
Fixed part explanatory variable design matrix for a single unit	X_{ij} for a level 1 unit X_j for a level 2 unit
Total residuals at each level for a 3-level model	$v_k = \sum_{h=0}^{q_3} v_{hk} z_{hk}^{(3)}$ $u_{jk} = \sum_{h=0}^{q_2} u_{hjk} z_{hjk}^{(2)}$ $e_{ijk} = \sum_{h=0}^{q_1} e_{hijk} z_{hijk}^{(1)}$
Explanatory variable design matrix for level 2 and level 1 random coefficients	$Z^{(2)}$, $Z^{(1)}$
Predicted value from fixed part of model	$\bar{y}_{ij} = X_{ij} b = (Xb)_{ij}$
Raw or total residual for level 1 unit	$\tilde{y}_{ij} = y_{ij} - \bar{y}_{ij}$
Mean raw residual for level 2 unit	$\tilde{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \tilde{y}_{ij}$
Estimated residual or posterior residual estimate	$\bar{u}_j$, $\bar{e}_{ij}$
Covariance matrix of random coefficients at level i	Ω_i, $\Omega = \{\Omega_i\}$
Parentheses denoting vector or matrix of elements	{}
Covariance matrix of response vector for k-level model	V_k or just V
Contribution to covariance matrix of response vector from level *i* for k-level model	$V_{k(i)}$, or just $V_{(i)}$
Direct sum of matrices $A_1, ..., A_k$	$\bigoplus_{i=1}^{k} A_i$
Kronecker product of conformable matrices A_1, A_2	$A_1 \otimes A_2$

vec operator on matrix A	$vec(A)$

Glossary

Cluster	A grouping containing 'lower level' elements. For example in a sample survey the set of households in a neighbourhood.
Design matrix	In the fixed part of the model, the matrix of values of the explanatory variables X. In the random part the matrix of explanatory variables Z.
Explanatory variable	Also known as an 'independent' variable. In the fixed part of the model usually denoted by x and in the random part by z.
Fixed part	That part of a model represented by Xb, that is the average relationship.
Level	A component of a data hierarchy. Level 1 is the lowest level, for example students within schools or repeated measurement occasions within individual subjects.
Level $\boldsymbol{n}$ variation	The variation of level $\boldsymbol{n}$ unit measurements about the fixed part of a model.
Nesting	The clustering of units into a hierarchy
Random part	That part of a model represented by Zu, that is the contribution of the random variables. at each level.
Response variable	Also known as a 'dependent' variable. Denoted by y.
Unit	An entity defined at a level of a data hierarchy. For example an individual student will be a level 1 unit within a level 2 unit which is a school.