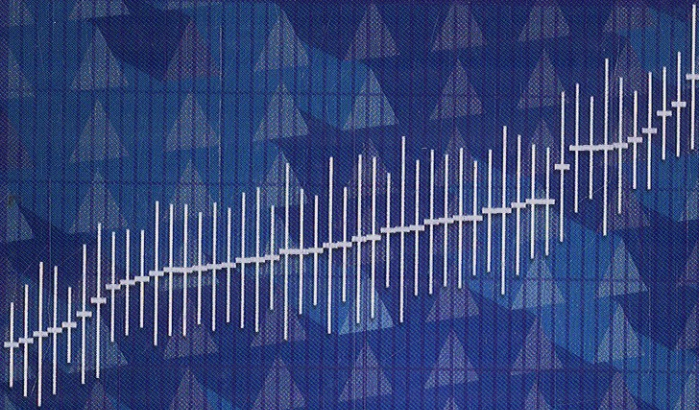


 WILEY

Multilevel Statistical Models

4th Edition



Harvey Goldstein

WILEY SERIES IN PROBABILITY AND STATISTICS

Contents

Preface	xv
Acknowledgements	xvii
Notation	xix
A general classification notation and diagram	xx
Glossary	xxiii
1 An introduction to multilevel models	1
1.1 Hierarchically structured data	1
1.2 School effectiveness	3
1.3 Sample survey methods	5
1.4 Repeated measures data	5
1.5 Event history and survival models	6
1.6 Discrete response data	7
1.7 Multivariate models	7
1.8 Nonlinear models	8
1.9 Measurement errors	9
1.10 Cross classifications and multiple membership structures	9
1.11 Factor analysis and structural equation models	10
1.12 Levels of aggregation and ecological fallacies	10
1.13 Causality	11
1.14 The latent normal transformation and missing data	13
1.15 Other texts	14
1.16 A caveat	14
2 The 2-level model	15
2.1 Introduction	15
2.2 The 2-level model	17
2.3 Parameter estimation	19
2.3.1 The variance components model	19
2.3.2 The general 2-level model with random coefficients	21
2.4 Maximum likelihood estimation using iterative generalised least squares (IGLS)	22
2.5 Marginal models and generalised estimating equations (GEE)	25

2.6	Residuals	25
2.7	The adequacy of ordinary least squares estimates	27
2.8	A 2-level example using longitudinal educational achievement data	28
	2.8.1 Checking for outlying units	30
	2.8.2 Model checking using estimated residuals	31
2.9	General model diagnostics	32
2.10	Higher level explanatory variables and compositional effects	34
2.11	Transforming to normality	36
2.12	Hypothesis testing and confidence intervals	39
	2.12.1 Fixed parameters	39
	2.12.2 Random parameters	41
	2.12.3 Hypothesis testing for non-nested models	42
	2.12.4 Inferences for residual estimates	43
2.13	Bayesian estimation using Markov Chain Monte Carlo (MCMC)	45
	2.13.1 Gibbs sampling	47
	2.13.2 Metropolis-Hastings (MH) sampling	48
	2.13.3 Convergence of MCMC chains	48
	2.13.4 Making inferences	49
	2.13.5 An example	50
2.14	Data augmentation	55
Appendix 2.1	The general structure and maximum likelihood estimation for a multilevel model	57
Appendix 2.2	Multilevel residuals estimation	60
	2.2.1 Shrunken estimates	60
	2.2.2 Delta method estimators for the covariance matrix of residuals	61
Appendix 2.3	Estimation using profile and extended likelihood	63
Appendix 2.4	The EM algorithm	65
Appendix 2.5	MCMC sampling	67
	2.5.1 Gibbs sampling	67
	2.5.2 Metropolis-Hastings (MH) sampling	70
	2.5.3 Hierarchical centring	71
	2.5.4 Orthogonalisation of the explanatory variables and parameter expansion	72
3	3-level models and more complex hierarchical structures	73
3.1	Complex variance structures	73
	3.1.1 Partitioning the variance and intra-unit correlation	79
	3.1.2 Variances for subgroups defined at level 1	80
	3.1.3 Variance as a function of predicted value	83
	3.1.4 Variances for subgroups defined at higher levels	85
3.2	A 3-level complex variation model example	85
3.3	Parameter constraints	88
3.4	Weighting units	90

3.4.1	Maximum likelihood estimation with weights	91
3.4.2	Weighted MCMC estimation	92
3.5	Robust (sandwich) estimators and jackknifing	93
3.6	The bootstrap	95
3.6.1	The fully nonparametric bootstrap	95
3.6.2	The fully parametric bootstrap	95
3.6.3	The iterated parametric bootstrap and bias correction	96
3.6.4	The residuals bootstrap	98
3.7	Aggregate level analyses	101
3.7.1	Inferences about residuals from aggregate level analyses	103
3.8	Meta analysis	104
3.8.1	Aggregate and mixed level analysis	106
3.8.2	Defining origin and scale	107
3.8.3	An example: meta analysis of class size data	107
3.8.4	Practical issues in meta analysis	108
3.9	Design issues	109
4	Multilevel models for discrete response data	111
4.1	Generalised linear models	111
4.2	Proportions as responses	112
4.3	Examples	115
4.3.1	A study of contraceptive use	115
4.3.2	Modelling school segregation	117
4.4	Models for multiple response categories	119
4.5	Models for counts	122
4.6	Ordered responses	123
4.7	Mixed discrete-continuous response models	124
4.8	A latent normal model for binary responses	126
4.9	Partitioning variation in discrete response models	127
4.9.1	Model linearisation (Method A)	128
4.9.2	Simulation (Method B)	128
4.9.3	A binary linear model (Method C)	129
4.9.4	A latent variable approach (Method D)	129
4.9.5	An example of VPC calculations	130
Appendix 4.1	Multilevel generalised linear model estimation	132
4.1.1	Approximate quasilielihood estimates	132
4.1.2	Differentials for some discrete response models	134
Appendix 4.2	Maximum likelihood estimation for multilevel generalised linear models	135
4.2.1	Simulated maximum likelihood estimation	135
4.2.2	Residuals	138
4.2.3	Cross classifications and multiple membership models	139
4.2.4	Computing issues	139
4.2.5	Maximum likelihood estimation via quadrature	140

Appendix 4.3	MCMC estimation for generalised linear models	142
4.3.1	Metropolis-Hastings (MH) sampling	142
4.3.2	Latent variable models for binary data	142
4.3.3	Proportions as responses	144
Appendix 4.4	Bootstrap estimation for multilevel generalised linear models	145
4.4.1	The iterated bootstrap	145
5	Models for repeated measures data	147
5.1	Repeated measures data	147
5.2	A 2-level repeated measures model	148
5.3	A polynomial model example for adolescent growth and the prediction of adult height	149
5.4	Modelling an autocorrelation structure at level 1	153
5.5	A growth model with autocorrelated residuals	154
5.6	Multivariate repeated measures models	156
5.7	Scaling across time	156
5.8	Cross-over designs	157
5.9	Missing data	157
5.10	Longitudinal discrete response data	159
6	Multivariate multilevel data	161
6.1	Introduction	161
6.2	The basic 2-level multivariate model	162
6.3	Rotation designs	164
6.4	A rotation design example using Science Survey test scores	164
6.5	Informative response selection: subject choice in examinations	167
6.6	Multivariate structures at higher levels and future predictions	168
6.7	Multivariate responses at several levels	170
6.7.1	Fitting responses at several levels using random data augmentation	171
6.8	Principal components analysis	172
6.9	Multiple discriminant analysis	173
Appendix 6.1	MCMC algorithm for a multivariate normal response model with constraints	175
6.1.1	Constraints among parameters	176
7	Latent normal models for multivariate data	179
7.1	The normal multilevel multivariate model	179
7.2	Sampling binary responses	180
7.3	Sampling ordered categorical responses	180
7.4	Sampling unordered categorical responses	182
7.5	Sampling count data	182
7.6	Sampling continuous non-normal data	183

7.7	Sampling the level 1 and level 2 covariance matrices	183
7.8	Model fit	184
7.9	Partially ordered data	185
7.10	Hybrid normal/ordered variables	185
	7.10.1 Ordered data with known thresholds	186
7.11	Discussion	187
8	Multilevel factor analysis, structural equation and mixture models	189
8.1	A 2-stage 2-level factor model	189
8.2	A general multilevel factor model	191
8.3	MCMC estimation for the factor model	192
	8.3.1 A 2-level factor example	193
8.4	Structural equation models	195
8.5	Discrete response multilevel structural equation models	197
8.6	More complex hierarchical latent variable models	198
8.7	Multilevel mixture models	198
9	Nonlinear multilevel models	201
9.1	Introduction	201
9.2	Nonlinear functions of linear components	201
9.3	Estimating population means	202
9.4	Nonlinear functions for variances and covariances	203
9.5	Examples of nonlinear growth and nonlinear level 1 variance	204
Appendix 9.1	Nonlinear model estimation	207
	9.1.1 Modelling variances and covariances as nonlinear functions	208
10	Multilevel modelling in sample surveys	211
10.1	Sample survey structures	211
10.2	Population structures	212
	10.2.1 Superpopulations	212
	10.2.2 Finite population inference	213
10.3	Small area estimation	214
	10.3.1 Information at domain level only	215
	10.3.2 Longitudinal data	216
	10.3.3 Multivariate responses	216
11	Multilevel event history and survival models	217
11.1	Introduction	217
11.2	Censoring	218
11.3	Hazard and survival functions	218
11.4	Parametric proportional hazard models	219
11.5	The semiparametric Cox model	220
11.6	Tied observations	221

11.7	Repeated events proportional hazard models	222
11.8	Example using birth interval data	222
11.9	Log duration models	223
	11.9.1 Censored data	225
	11.9.2 Infinite durations	226
11.10	Examples with birth interval data and children's activity episodes	227
11.11	The grouped discrete time hazards model	229
	11.11.1 A 2-level discrete time event history model for repeated events	232
	11.11.2 Partnership data example	234
	11.11.3 General discrete time event history models	234
11.12	Discrete time latent normal event history models	237
	11.12.1 Censored data	238
	11.12.2 Missing data	238
	11.12.3 Information about the timing of events in an interval	238
	11.12.4 Modelling the threshold parameters and time varying covariates	239
	11.12.5 An example using partnership durations	240
12	Cross-classified data structures	243
12.1	Random cross classifications	243
12.2	A basic cross-classified model	245
12.3	Examination results for a cross classification of schools	247
12.4	Interactions in cross classifications	248
12.5	Cross classifications with one unit per cell	248
12.6	Multivariate cross-classified models	249
12.7	A general notation for cross classification	249
12.8	MCMC estimation in cross-classified models	250
Appendix 12.1	IGLS estimation for cross-classified data	252
	12.1.1 An efficient IGLS algorithm	252
	12.1.2 Computational considerations	253
13	Multiple membership models	255
13.1	Multiple membership structures	255
13.2	Notation and classifications for multiple membership structures	256
13.3	An example of salmonella infection	257
13.4	A repeated measures multiple membership model	258
13.5	Individuals as higher level units	259
	13.5.1 Example of research grant awards	261
13.6	Spatial models	261
13.7	Missing identification models	263
Appendix 13.1	MCMC estimation for multiple membership models	265
14	Measurement errors in multilevel models	267
14.1	A basic measurement error model	267

14.2	Moment-based estimators	268
14.2.1	Measurement errors in level 1 variables	268
14.2.2	Measurement errors in higher level variables	269
14.3	A 2-level example with measurement error at both levels	271
14.4	Multivariate responses	273
14.5	Nonlinear models	274
14.6	Measurement errors for discrete explanatory variables	274
14.7	MCMC estimation for measurement error models	275
Appendix 14.1	Measurement error estimation	277
14.1.1	Moment based estimators for a basic 2-level model	277
14.1.2	Parameter estimation	278
14.1.3	Random coefficients for explanatory variables measured with error	279
14.1.4	Nonlinear models	279
14.1.5	MCMC estimation for measurement error models: continuous variables	280
14.1.6	MCMC estimation for measurement error models: discrete variables	281
14.1.7	A latent normal model for discrete and continuous variables with measurement errors	284
15	Smoothing models for multilevel data	285
15.1	Introduction	285
15.2	Smoothing estimators	285
15.2.1	Regression splines	285
15.3	Smoothing splines	286
15.4	Semiparametric smoothing models	287
15.5	Multilevel smoothing models	290
15.6	General multilevel semiparametric smoothing models	292
15.7	Generalised linear models	293
15.8	An example	293
15.9	Conclusions	298
16	Missing data, partially observed data and multiple imputation	301
16.1	Introduction	301
16.2	Creating a completed dataset	302
16.3	Joint modelling for missing data	304
16.4	A 2-level model with responses of different types at both levels	305
16.4.1	Sampling level 1 non-normal responses	305
16.4.2	Sampling level 2 non-normal responses	306
16.5	Multiple imputation	306
16.6	A simulation example of multiple imputation for missing data	307
16.7	Longitudinal data with attrition	308

xiv	CONTENTS	
16.8	Partially known data values	309
	16.8.1 An application to record linkage	310
	16.8.2 Estimating a probability distribution using record linkage weights	311
16.9	Conclusions	312
17	Multilevel models with correlated random effects	315
17.1	Introduction	315
17.2	Non-independence of level 2 residuals	315
17.3	MCMC estimation for non-independent level 2 residuals	317
	17.3.1 Sampling the level 2 covariance matrix	318
	17.3.2 Sampling the level 2 residuals	319
17.4	Adaptive proposal distributions in MCMC estimation	319
17.5	MCMC estimation for non-independent level 1 residuals	320
	17.5.1 A 2-level model formulated as a single level model with non-independent residuals	321
17.6	Modelling the level 1 variance as a function of explanatory variables with random effects	321
17.7	Discrete responses with correlated random effects	322
	17.7.1 The probit ordered response model where the variance is a function of explanatory variables with random effects	323
17.8	Calculating the DIC statistic	324
17.9	A growth dataset	325
17.10	Conclusions	326
18	Software for multilevel modelling	329
18.1	Software packages	329
	References	333
	Author index	347
	Subject index	351