Contents

1 Introduction 1
- 1.1 Longitudinal studies 1
- 1.2 Examples 3
- 1.3 Notation 15
- 1.4 Merits of longitudinal studies 16
- 1.5 Approaches to longitudinal data analysis 17
- 1.6 Organization of subsequent chapters 20

2 Design considerations 22
- 2.1 Introduction 22
- 2.2 Bias 22
- 2.3 Efficiency 24
- 2.4 Sample size calculations 26
  - 2.4.1 Continuous responses 28
  - 2.4.2 Binary responses 30
- 2.5 Further reading 31

3 Exploring longitudinal data 33
- 3.1 Introduction 33
- 3.2 Graphical presentation of longitudinal data 34
- 3.3 Fitting smooth curves to longitudinal data 41
- 3.4 Exploring correlation structure 46
- 3.5 Exploring association amongst categorical responses 52
- 3.6 Further reading 53

4 General linear models for longitudinal data 54
- 4.1 Motivation 54
- 4.2 The general linear model with correlated errors 55
  - 4.2.1 The uniform correlation model 55
  - 4.2.2 The exponential correlation model 56
  - 4.2.3 Two-stage least-squares estimation and random effects models 57
- 4.3 Weighted least-squares estimation 59
- 4.4 Maximum likelihood estimation under Gaussian assumptions 64
- 4.5 Restricted maximum likelihood estimation 66
- 4.6 Robust estimation of standard errors 70

5 Parametric models for covariance structure 81
5.1 Introduction 81
5.2 Models 82
5.2.1 Pure serial correlation 84
5.2.2 Serial correlation plus measurement error 89
5.2.3 Random intercept plus serial correlation plus measurement error 90
5.2.4 Random effects plus measurement error 91
5.3 Model-fitting 93
5.3.1 Formulation 94
5.3.2 Estimation 95
5.3.3 Inference 97
5.3.4 Diagnostics 98
5.4 Examples 99
5.5 Estimation of individual trajectories 110
5.6 Further reading 113

6 Analysis of variance methods 114
6.1 Preliminaries 114
6.2 Time-by-time ANOVA 115
6.3 Derived variables 116
6.4 Repeated measures 123
6.5 Conclusions 125

7 Generalized linear models for longitudinal data 126
7.1 Marginal models 126
7.2 Random effects models 128
7.3 Transition (Markov) models 130
7.4 Contrasting approaches 131
7.5 Inferences 137

8 Marginal models 141
8.1 Introduction 141
8.2 Binary responses 142
8.2.1 The log-linear model 142
8.2.2 Log-linear models for marginal means 143
8.2.3 Generalized estimating equations 146
8.3 Examples 148
8.4 Counted responses 160
8.4.1 Parametric modelling for count data 160
8.4.2 Generalized estimating equation approach 162
8.5 Sample size calculations revisited 165
8.6 Further reading 167

9 Random effects models 169
9.1 Introduction 169
9.2 Estimation for generalized linear mixed models 171
9.2.1 Conditional likelihood 171
9.2.2 Maximum likelihood estimation 172
9.3 Logistic regression for binary responses 175
9.3.1 Conditional likelihood approach 175
9.3.2 Random effects models for binary data 178
9.3.3 Examples of logistic models with Gaussian random effects 180
9.4 Counted responses 184
9.4.1 Conditional likelihood method 184
9.4.2 Random effects models for counts 186
9.4.3 Poisson–Gaussian random effects models 188
9.5 Further reading 189

10 Transition models 190
10.1 General 190
10.2 Fitting transition models 192
10.3 Transition models for categorical data 194
10.3.1 Indonesian children's study example 197
10.3.2 Ordered categorical data 201
10.4 Log-linear transition models for count data 204
10.5 Further reading 206

11 Likelihood-based methods for categorical data 208
11.1 Introduction 208
11.1.1 Notation and definitions 209
11.2 Generalized linear mixed models 209
11.2.1 Maximum likelihood algorithms 212
11.2.2 Bayesian methods 214
11.3 Marginalized models 216
11.3.1 An example using the Gaussian linear model 218
11.3.2 Marginalized log-linear models 220
11.3.3 Marginalized latent variable models 222
11.3.4 Marginalized transition models 225
11.3.5 Summary 231
11.4 Examples 231
11.4.1 Crossover data 231
11.4.2 Madras schizophrenia data 234
11.5 Summary and further reading 243

12 Time-dependent covariates 245
12.1 Introduction 245
12.2 An example: the MSCM study 247

12.3 Stochastic covariates 253
12.3.1 Estimation issues with cross-sectional models 254
12.3.2 A simulation illustration 256
12.3.3 MSCM data and cross-sectional analysis 257
12.3.4 Summary 258
12.4 Lagged covariates 259
12.4.1 A single lagged covariate 259
12.4.2 Multiple lagged covariates 260
12.4.3 MSCM data and lagged covariates 261
12.4.4 Summary 265
12.5 Time-dependent confounders 265
12.5.1 Feedback: response is an intermediate and a confounder 266
12.5.2 MSCM data and endogeneity 268
12.5.3 Targets of inference 269
12.5.4 Estimation using g-computation 273
12.5.5 MSCM data and g-computation 275
12.5.6 Estimation using inverse probability of treatment weights (IPTW) 276
12.5.7 MSCM data and marginal structural models using IPTW 279
12.5.8 Summary 280
12.6 Summary and further reading 280

13 Missing values in longitudinal data 282
13.1 Introduction 282
13.2 Classification of missing value mechanisms 283
13.3 Intermittent missing values and dropouts 284
13.4 Simple solutions and their limitations 287
13.4.1 Last observation carried forward 287
13.4.2 Complete case analysis 288
13.5 Testing for completely random dropouts 288
13.6 Generalized estimating equations under a random missingness mechanism 293
13.7 Modelling the dropout process 295
13.7.1 Selection models 295
13.7.2 Pattern mixture models 299
13.7.3 Random effect models 301
13.7.4 Contrasting assumptions: a graphical representation 303
13.8 A longitudinal trial of drug therapies for schizophrenia 305
13.9 Discussion 316

14 Additional topics 319
14.1 Non-parametric modelling of the mean response 319
14.1.1 Further reading 326
14.2 Non-linear regression modelling 326
14.2.1 Correlated errors 328
14.2.2 Non-linear random effects 329
14.3 Joint modelling of longitudinal measurements and recurrent events 329
14.4 Multivariate longitudinal data 332

Appendix Statistical background 337
A.1 Introduction 337
A.2 The linear model and the method of least squares 337
A.3 Multivariate Gaussian theory 339
A.4 Likelihood inference 340
A.5 Generalized linear models 343
A.5.1 Logistic regression 343
A.5.2 Poisson regression 344
A.5.3 The general class 345
A.6 Quasi-likelihood 346

Bibliography 349

Index 369