Contents

Preface ix

1 Statistical modelling: An overview **1**

1.1 Introduction 1
1.2 Why model? 1
1.3 The general linear statistical model 3
1.4 Model development 6
1.5 Model selection by computer 8
1.6 Statistical software 9

2 Research designs and data **11**

2.1 Units of analysis 11
2.2 Research designs 12
2.3 Sampling, surveys and weighting 12
2.4 Variables and their measurement 14
2.5 Missing data 17
2.6 Data structures 20

3 Statistical preliminaries **21**

3.1 Odds and odds ratios 21
3.2 Logarithms 24
3.3 z, t and chi-squared (χ^2) 25
3.4 Tests of significance 30
3.5 Categorical explanatory variables; dummy variables; reference or base category 31
3.6 Multicollinearity 34
3.7 Interaction terms 35
3.8 Incorporating nonlinear relationships 37
3.9 Residuals 37
3.10 Unobserved variables and unobserved heterogeneity 39

4 Multiple regression for continuous response variables **41**

4.1 Introduction 41
4.2 Example: Determinants of pay 41
4.3 Fitting a multiple regression model in SPSS 43
4.4 How good is the regression model? 45
4.5 Expanding the model 48
4.6 Adding interaction terms 52
4.7 Automated model selection procedures 54
4.8 Diagnostics: Analysis of residuals 55
4.9 Fitting the regression model in Stata 60
4.10 Tobit model 61

5 Logistic regression for binary response variables **63**

5.1 Introduction 63
5.2 Example: Victim of car crime 64
5.3 Undertaking logistic regression in SPSS 65
5.4 Expanding the model 69
5.5 Automated model selection procedures 73
5.6 Diagnostics: Analysis of residuals 74
5.7 Fitting the logistic regression model in Stata 75
5.8 Probit model 75

6 Multinomial logistic regression for multinomial response variables **77**

6.1 Background 77
6.2 Fitting the multinomial in SPSS to data on sentences awarded by the court 78
6.3 Diagnostics: Analysis of residuals 82
6.4 Expanding the model 84
6.5 Relationship between SPSS logistic and SPSS multinomial 87
6.6 Fitting the multinomial in Stata to voting data, incorporating complex design factors 87

7 Loglinear models **93**

7.1 Introduction 93
7.2 Three-factor model 94
7.3 Expanding the model 95
7.4 Loglinear models specifying a response variable 96

8 Ordinal logistic regression for ordered categorical response variables **98**

8.1 Introduction 98
8.2 Example: Ability to influence local political decisions 98

8.3 Fitting the ordinal logistic regression model in SPSS 101
8.4 Fitting the ordinal logistic regression model in Stata 104

9 Multilevel modelling 107

9.1 Introduction 107
9.2 Methods of analysis and their deficiencies 108
9.3 Multilevel models 111
9.4 Statistical software 116
9.5 Fitting models in MLwiN to employment data 117

10 Latent variables and factor analysis 125

10.1 Introduction 125
10.2 How are latent variables constructed? 126
10.3 Factor analysis 127
10.4 Example: Fear of crime 130
10.5 Interpreting factors and rotation 136
10.6 Exploratory and confirmatory factor analysis 139
10.7 Other latent variable models 139

11 Causal modelling: simultaneous equation models 141

11.1 Introduction 141
11.2 Recursive models: Path analysis, determinants of pay 141
11.3 Non-recursive models: Deterrent effect of solving crime 147
11.4 Structural equation modelling (SEM) 154

12 Longitudinal data analysis 156

12.1 Introduction 156
12.2 Time series analysis 158
12.3 Repeated cross-sectional surveys 159
12.4 Panel and cohort studies 160
12.5 Statistical models for panel data 161
12.6 Fitting panel data models in Stata 165
12.7 Example from the British Household Panel Study, time spent on housework 166
12.8 Relationship between panel models and multilevel models 171

13 Event history models 173

13.1 Time to an event: Survival models 173
13.2 Cox's proportional hazards model 177
13.3 Comparing survival times between two or more groups 179
13.4 Fitting Cox's proportional hazards regression model to Training for work data 182

13.5 Discrete time models 184
13.6 Data structures 184
13.7 Time to a single event 189
13.8 Competing risk models 190
13.9 Repeated events or multiple spells 191
13.10 Multiple states 192
13.11 Guidance on restructuring data to fit survival and discrete time models 193

References 197
Appendix 1 The generalised linear model 201
Appendix 2 Handling tabular data 203
Index 207