Contents

Preface xv

1 Introduction 1
- 1.1 Background 1
- 1.2 Scope 1
- 1.3 Notation 6
- 1.4 Distributions related to the Normal distribution 8
  - 1.4.1 Normal distributions 8
  - 1.4.2 Chi-squared distribution 9
  - 1.4.3 t-distribution 10
  - 1.4.4 F-distribution 10
  - 1.4.5 Some relationships between distributions 11
- 1.5 Quadratic forms 11
- 1.6 Estimation 13
  - 1.6.1 Maximum likelihood estimation 13
  - 1.6.2 Example: Poisson distribution 15
  - 1.6.3 Least squares estimation 15
  - 1.6.4 Comments on estimation 16
  - 1.6.5 Example: Tropical cyclones 17
- 1.7 Exercises 17

2 Model Fitting 21
- 2.1 Introduction 21
- 2.2 Examples 21
  - 2.2.1 Chronic medical conditions 21
  - 2.2.2 Example: Birthweight and gestational age 25
- 2.3 Some principles of statistical modelling 35
  - 2.3.1 Exploratory data analysis 35
  - 2.3.2 Model formulation 36
  - 2.3.3 Parameter estimation 36

2.3.4 Residuals and model checking 36
2.3.5 Inference and interpretation 39
2.3.6 Further reading 40
2.4 Notation and coding for explanatory variables 40
2.4.1 Example: Means for two groups 41
2.4.2 Example: Simple linear regression for two groups 42
2.4.3 Example: Alternative formulations for comparing the means of two groups 42
2.4.4 Example: Ordinal explanatory variables 43
2.5 Exercises 44

3 Exponential Family and Generalized Linear Models 49
3.1 Introduction 49
3.2 Exponential family of distributions 50
3.2.1 Poisson distribution 51
3.2.2 Normal distribution 52
3.2.3 Binomial distribution 52
3.3 Properties of distributions in the exponential family 53
3.4 Generalized linear models 56
3.5 Examples 58
3.5.1 Normal linear model 58
3.5.2 Historical linguistics 58
3.5.3 Mortality rates 59
3.6 Exercises 61

4 Estimation 65
4.1 Introduction 65
4.2 Example: Failure times for pressure vessels 65
4.3 Maximum likelihood estimation 70
4.4 Poisson regression example 73
4.5 Exercises 76

5 Inference 79
5.1 Introduction 79
5.2 Sampling distribution for score statistics 81
5.2.1 Example: Score statistic for the Normal distribution 82
5.2.2 Example: Score statistic for the Binomial distribution 82
5.3 Taylor series approximations 83
5.4 Sampling distribution for maximum likelihood estimators 84

5.4.1 Example: Maximum likelihood estimators for the Normal linear model 85
5.5 Log-likelihood ratio statistic 86
5.6 Sampling distribution for the deviance 87
5.6.1 Example: Deviance for a Binomial model 88
5.6.2 Example: Deviance for a Normal linear model 89
5.6.3 Example: Deviance for a Poisson model 91
5.7 Hypothesis testing 92
5.7.1 Example: Hypothesis testing for a Normal linear model 94
5.8 Exercises 95

6 Normal Linear Models 97
6.1 Introduction 97
6.2 Basic results 98
6.2.1 Maximum likelihood estimation 98
6.2.2 Least squares estimation 98
6.2.3 Deviance 99
6.2.4 Hypothesis testing 99
6.2.5 Orthogonality 100
6.2.6 Residuals 101
6.2.7 Other diagnostics 102
6.3 Multiple linear regression 104
6.3.1 Example: Carbohydrate diet 104
6.3.2 Coefficient of determination, R^2 108
6.3.3 Model selection 111
6.3.4 Collinearity 118
6.4 Analysis of variance 119
6.4.1 One-factor analysis of variance 119
6.4.2 Two-factor analysis of variance 126
6.5 Analysis of covariance 132
6.6 General linear models 135
6.7 Non-linear associations 137
6.7.1 *PLOS Medicine* journal data 138
6.8 Fractional polynomials 141
6.9 Exercises 143

7 Binary Variables and Logistic Regression **149**
7.1 Probability distributions 149
7.2 Generalized linear models 150
7.3 Dose response models 151
7.3.1 Example: Beetle mortality 154
7.4 General logistic regression model 158
7.4.1 Example: Embryogenic anthers 159
7.5 Goodness of fit statistics 162
7.6 Residuals 166
7.7 Other diagnostics 167
7.8 Example: Senility and WAIS 168
7.9 Odds ratios and prevalence ratios 171
7.10 Exercises 174

8 Nominal and Ordinal Logistic Regression **179**
8.1 Introduction 179
8.2 Multinomial distribution 180
8.3 Nominal logistic regression 181
8.3.1 Example: Car preferences 183
8.4 Ordinal logistic regression 188
8.4.1 Cumulative logit model 189
8.4.2 Proportional odds model 189
8.4.3 Adjacent categories logit model 190
8.4.4 Continuation ratio logit model 191
8.4.5 Comments 192
8.4.6 Example: Car preferences 192
8.5 General comments 193
8.6 Exercises 194

9 Poisson Regression and Log-Linear Models **197**
9.1 Introduction 197
9.2 Poisson regression 198
9.2.1 Example of Poisson regression: British doctors' smoking and coronary death 201
9.3 Examples of contingency tables 204
9.3.1 Example: Cross-sectional study of malignant melanoma 205
9.3.2 Example: Randomized controlled trial of influenza vaccine 206

9.3.3 Example: Case–control study of gastric and duodenal ulcers and aspirin use 207
9.4 Probability models for contingency tables 209
9.4.1 Poisson model 209
9.4.2 Multinomial model 209
9.4.3 Product multinomial models 210
9.5 Log-linear models 210
9.6 Inference for log-linear models 212
9.7 Numerical examples 212
9.7.1 Cross-sectional study of malignant melanoma 212
9.7.2 Case–control study of gastric and duodenal ulcer and aspirin use 215
9.8 Remarks 216
9.9 Exercises 217

10 Survival Analysis **223**
10.1 Introduction 223
10.2 Survivor functions and hazard functions 225
10.2.1 Exponential distribution 226
10.2.2 Proportional hazards models 227
10.2.3 Weibull distribution 228
10.3 Empirical survivor function 230
10.3.1 Example: Remission times 231
10.4 Estimation 233
10.4.1 Example: Exponential model 234
10.4.2 Example: Weibull model 235
10.5 Inference 236
10.6 Model checking 236
10.7 Example: Remission times 238
10.8 Exercises 240

11 Clustered and Longitudinal Data **245**
11.1 Introduction 245
11.2 Example: Recovery from stroke 247
11.3 Repeated measures models for Normal data 253
11.4 Repeated measures models for non-Normal data 257
11.5 Multilevel models 259
11.6 Stroke example continued 262
11.7 Comments 265
11.8 Exercises 266

12 Bayesian Analysis **271**
12.1 Frequentist and Bayesian paradigms 271
12.1.1 Alternative definitions of p-values and confidence intervals 271
12.1.2 Bayes' equation 272
12.1.3 Parameter space 273
12.1.4 Example: Schistosoma japonicum 273
12.2 Priors 275
12.2.1 Informative priors 276
12.2.2 Example: Sceptical prior 276
12.2.3 Example: Overdoses amongst released prisoners 279
12.3 Distributions and hierarchies in Bayesian analysis 281
12.4 WinBUGS software for Bayesian analysis 281
12.5 Exercises 284

13 Markov Chain Monte Carlo Methods **287**
13.1 Why standard inference fails 287
13.2 Monte Carlo integration 287
13.3 Markov chains 289
13.3.1 The Metropolis–Hastings sampler 291
13.3.2 The Gibbs sampler 293
13.3.3 Comparing a Markov chain to classical maximum likelihood estimation 295
13.3.4 Importance of parameterization 299
13.4 Bayesian inference 300
13.5 Diagnostics of chain convergence 302
13.5.1 Chain history 302
13.5.2 Chain autocorrelation 304
13.5.3 Multiple chains 305
13.6 Bayesian model fit: the deviance information criterion 306
13.7 Exercises 308

14 Example Bayesian Analyses **315**
14.1 Introduction 315
14.2 Binary variables and logistic regression 316
14.2.1 Prevalence ratios for logistic regression 319
14.3 Nominal logistic regression 322
14.4 Latent variable model 324
14.5 Survival analysis 326
14.6 Random effects 328

14.7 Longitudinal data analysis 331
14.8 Bayesian model averaging 338
14.8.1 Example: Stroke recovery 340
14.8.2 Example: *PLOS Medicine* journal data 340
14.9 Some practical tips for WinBUGS 342
14.10 Exercises 344

Postface **347**

Appendix **355**

Software **357**

References **359**

Index **371**