Contents

Preface vii

PART I **The Conceptual Basis for Fitting Statistical Models**

1 General Introduction 3
- 1.1 The purpose of statistics 3
- 1.2 Statistics in a schizophrenic state? 4
- 1.3 How is this book organized? 4
- 1.4 How to use this book 6

References 7

2 Statistical Modeling: A short historical background 9
- 2.1 What is a statistical model? 9
- 2.2 What is this thing called probability? 10
- 2.3 Linking probability with statistics 13
- 2.4 The early Bayesian demise during the 1930s 15

References 17

3 Estimating Parameters: The main purpose of statistical inference 19
- 3.1 Introduction 19
- 3.2 Least squares: A theory of errors and the normal distribution 20
- 3.3 Maximum likelihood 20
  - 3.3.1 The basic concepts 20
  - 3.3.2 Obtaining maximum likelihood estimates 22
  - 3.3.3 Using maximum likelihood estimates in statistical inference 27
- 3.4 Bayesian parameter estimation: The basics 29
- 3.5 Bayesian methods: Markov chain Monte Carlo to the rescue 36
- 3.6 Quality control for the algorithms of Bayesian methods 45
- 3.7 More general MCMC variations: Metropolis–Hastings and Gibbs algorithms 46
- 3.8 Recent advances in Bayesian methods: Hamiltonian Monte Carlo 48
- 3.9 Bayesian hypothesis tests 50
- 3.10 Summary of the main differences between maximum likelihood and Bayesian methods 51

References 54

PART II **Applying the Generalized Linear Model to Varied Data Types**

4 The General Linear Model I: Numerical explanatory variables 59
4.1 Introduction 59
4.2 The lognormal distribution and its relation to the general linear model 60
4.3 Simple linear regression: One continuous explanatory variable 61
4.4 Simple linear regression: Frequentist fitting 64
4.5 Tools for model validation in frequentist statistics 66
4.6 Simple linear regression: Bayesian fitting 69
4.7 Tools for model validation in Bayesian statistics 78
4.8 Multiple linear regression: More than one numerical explanatory variable 80
4.9 Multiple linear regression: Frequentist fitting 84
4.10 The importance of standardizing explanatory variables 86
4.11 Polynomial regression 90
4.12 Multiple linear regression: Bayesian fitting 91
4.13 Problems 98
References 98

5 The General Linear Model II: Categorical explanatory variables 101
5.1 Introduction 101
5.2 Student's *t* test: One categorical explanatory variable with two groups 101
5.3 The *t*-test: Frequentist fitting 106
5.4 The *t*-test: Bayesian fitting 110
5.5 Viewing one-way analysis of variance as a multiple regression 115
5.6 One-way analysis of variance: Frequentist fitting 120
5.7 One-way analysis of variance: Bayesian fitting 124
5.8 A posteriori tests in frequentist models 130
5.9 A posteriori tests in Bayesian models? 134
5.10 Problems 136
References 137

6 The General Linear Model III: Interactions between explanatory variables 139
6.1 Introduction 139
6.2 Factorial analysis of variance 139
6.3 Factorial analysis of variance: Frequentist fitting 145
6.4 Factorial analysis of variance: Bayesian fitting 149

6.5 Analysis of covariance: Mixing continuous and categorical explanatory variables 156
6.6 Analysis of covariance: Frequentist fitting 158
6.7 Analysis of covariance: Bayesian fitting 162
6.8 Problems 168
References 168

7 Model Selection: One, two, and more models fitted to the data 169
7.1 Introduction 169
7.2 The problem of model selection: Parsimony in statistics 170
7.3 Model selection criteria in the frequentist framework: AIC 172
7.4 Model selection criteria in the Bayesian framework: DIC and WAIC 176
7.5 The posterior predictive distribution and posterior predictive checks 178
7.6 Now back to the WAIC and LOO-CV 182
7.7 Prior predictive distributions: A relatively "new" kid on the block 185
References 186

8 The Generalized Linear Model 189
8.1 Introduction 189
8.2 What are GLMs made of? 189
8.3 Fitting GLMs 193
8.4 Goodness of fit in GLMs 194
8.5 Statistical significance of GLM 198
References 198

9 When the Response Variable is Binary 201
9.1 Introduction 201
9.2 Key concepts for binary GLMs: Odds, log odds, and additional link functions 202
9.3 Fitting binary GLMs 203
9.4 Ungrouped binary GLM: Frequentist fitting 207
9.5 Further issues about validating binary GLMs 213
9.6 Ungrouped binary GLMs: Bayesian fitting 216
9.7 Grouped binary GLMs 227
9.8 Problems 233
References 233

10 When the Response Variable is a Count, Often with Many Zeros 235
10.1 Introduction 235
10.2 Over-dispersion: A common problem with many causes and some solutions 237

10.3 Plant species richness and geographical variables 239
10.3.1 Frequentist fitting of the count GLM 242
10.3.2 Bayesian fitting of count GLMs 247
10.4 Modeling of counts with an excess of zeros: Zero-inflated and hurdle models 254
10.4.1 Frequentist fitting of a zero-inflated model 256
10.4.2 Bayesian fitting of a zero-augmented model 261
10.5 Problems 268
References 269

11 Further Issues Involved in the Modeling of Counts 271
11.1 "The more you search, the more you find" 271
11.2 Log-linear models as count GLMs 272
11.3 Frequentist fitting of a log-linear model 275
11.4 Bayesian fitting of a log-linear model 284
11.5 Problems 291
References 292

12 Models for Positive, Real-Valued Response Variables: Proportions and others 293
12.1 Introduction 293
12.2 Modeling proportions 293
12.3 Plant cover, grazing, and productivity 295
12.4 Frequentist fitting of a GLM on proportions 297
12.5 Bayesian fitting of a GLM on proportions 303
12.6 Modeling positive, real-valued response variables 311
12.7 Predicting tree seedling biomass 312
12.8 Frequentist fitting of a gamma GLM 314
12.9 Bayesian fitting of a gamma GLM 317
12.10 Other related yet important cases of positive, real-valued response variables 319
12.11 Problems 320
References 321

Approaches to Defining Priors 323

PART III **Incorporating Experimental and Survey Design Using Mixed Models**

13 Accounting for Structure in Mixed/Hierarchical Models 327
13.1 Introduction 327
13.2 Fixed effects and random effects in the frequentist framework 329
13.3 Defining mixed effects models 332

13.4 Problems and inconsistencies with the definition of random effects 335
13.5 Population-level and group-level effects in Bayesian hierarchical models 336
13.6 Fitting mixed models in the frequentist framework 339
13.7 Statistical significance and model selection in frequentist mixed models 353
13.8 The shrinkage or borrowing strength effect in mixed models 356
13.9 Fitting mixed models in the Bayesian framework 358
13.10 Problems 370
References 371

14 Experimental Design in the Life Sciences: The basics 373
14.1 Introduction 373
14.2 The basic principles of experimental design 374
14.3 Surveys and observational studies 376
14.4 The main types of experimental design used in the life sciences 376
14.4.1 Factorial design 377
14.4.2 Randomized block design 378
14.4.3 Split-plot design 380
14.4.4 Nested design 382
14.4.5 Repeated measures design 384
14.4.6 Crossover design 385
14.5 How many samples should we take? 387
References 390

15 Mixed Hierarchical Models and Experimental Design Data 393
15.1 Introduction 393
15.2 Binary GLMM with a randomized block design 394
15.2.1 Binary GLMM with a randomized block design: Frequentist models 398
15.2.2 Binary GLMM with a randomized block design: Bayesian models 407
15.3 Gaussian GLMM with a repeated measures design 416
15.3.1 Gaussian GLMM with a repeated measures design: Frequentist models 420
15.3.2 Gaussian GLMM with a repeated measures design: Bayesian models 423
15.4 Beta GLMM with a split-plot design 428
15.4.1 Beta GLMM with a split-plot design: Frequentist model 432
15.4.2 Beta GLMM with a split-plot design: Bayesian model 439
15.5 Problems 449
References 449

Afterword 451
Appendix A: List of R Packages Used in This Book 453
Appendix B: Exploring and Describing the Evidence in Graphics
(only available online at www.oup.com/companion/InchaustiSMWR)
Appendix C: Using R and RStudio: The Bare-Bones Basics
(only available online at www.oup.com/companion/InchaustiSMWR)
Index 455