Contents

List of contributors xiii

Introduction *Vinicio J. Sosa, Simoneta Negrete-Yankelevich, and Gordon A. Fox* 1

Why another book on statistics for ecologists? 1
Relating ecological questions to statistics 5
A conceptual foundation: the statistical linear model 7
What we need readers to know 12
How to get the most out of this book 13

1 Approaches to statistical inference *Michael A. McCarthy* 15

1.1 Introduction to statistical inference 15
1.2 A short overview of some probability and sampling theory 16
1.3 Approaches to statistical inference 19
1.3.1 Sample statistics and confidence intervals 20
1.3.2 Null hypothesis significance testing 21
1.3.3 Likelihood 27
1.3.4 Information-theoretic methods 30
1.3.5 Bayesian methods 33
1.3.6 Non-parametric methods 39
1.4 Appropriate use of statistical methods 39

2 Having the right stuff: the effects of data constraints on ecological data analysis *Earl D. McCoy* 44

2.1 Introduction to data constraints 44
2.2 Ecological data constraints 45
2.2.1 Values and biases 45
2.2.2 Biased behaviors in ecological research 47
2.3 Potential effects of ecological data constraints 48
2.3.1 Methodological underdetermination and cognitive biases 48
2.3.2 Cognitive biases in ecological research? 49
2.4 Ecological complexity, data constraints, flawed conclusions 50
2.4.1 Patterns and processes at different scales 51
2.4.2 Discrete and continuous patterns and processes 52
2.4.3 Patterns and processes at different hierarchical levels 54
2.5 Conclusions and suggestions 56

3 Likelihood and model selection *Shane A. Richards* 58

3.1 Introduction to likelihood and model selection 58
3.2 Likelihood functions 59

3.2.1 Incorporating mechanism into models 61
3.2.2 Random effects 63
3.3 Multiple hypotheses 65
3.3.1 Approaches to model selection 67
3.3.2 Null hypothesis testing 68
3.3.3 An information-theoretic approach 70
3.3.4 Using AIC to select models 73
3.3.5 Extending the AIC approach 74
3.3.6 A worked example 76
3.4 Discussion 78

4 Missing data: mechanisms, methods, and messages
Shinichi Nakagawa 81
4.1 Introduction to dealing with missing data 81
4.2 Mechanisms of missing data 83
4.2.1 Missing data theory, mechanisms, and patterns 83
4.2.2 Informal definitions of missing data mechanisms 83
4.2.3 Formal definitions of missing data mechanisms 84
4.2.4 Consequences of missing data mechanisms: an example 86
4.3 Diagnostics and prevention 88
4.3.1 Diagnosing missing data mechanisms 88
4.3.2 How to prevent MNAR missingness 90
4.4 Methods for missing data 92
4.4.1 Data deletion, imputation, and augmentation 92
4.4.2 Data deletion 92
4.4.3 Single imputation 92
4.4.4 Multiple imputation techniques 94
4.4.5 Multiple imputation steps 95
4.4.6 Multiple imputation with multilevel data 98
4.4.7 Data augmentation 101
4.4.8 Non-ignorable missing data and sensitivity analysis 101
4.5 Discussion 102
4.5.1 Practical issues 102
4.5.2 Reporting guidelines 103
4.5.3 Missing data in other contexts 104
4.5.4 Final messages 105

5 What you don't know can hurt you: censored and truncated data in ecological research *Gordon A. Fox* 106
5.1 Censored data 106
5.1.1 Basic concepts 106
5.1.2 Some common methods you should not use 107
5.1.3 Types of censored data 109
5.1.4 Censoring in study designs 111
5.1.5 Format of data 113
5.1.6 Estimating means with censored data 113
5.1.7 Regression for censored data 116

5.2 Truncated data 124
5.2.1 Introduction to truncated data 124
5.2.2 Sweeping the issue under the rug 125
5.2.3 Estimation 125
5.2.4 Regression for truncated data 127
5.3 Discussion 129

6 Generalized linear models *Yvonne M. Buckley* 131
6.1 Introduction to generalized linear models 131
6.2 Structure of a GLM 135
6.2.1 The linear predictor 135
6.2.2 The error structure 136
6.2.3 The link function 136
6.3 Which error distribution and link function are suitable for my data? 137
6.3.1 Binomial distribution 138
6.3.2 Poisson distribution 141
6.3.3 Overdispersion 143
6.4 Model fit and inference 145
6.5 Computational methods and convergence 146
6.6 Discussion 147

7 A statistical symphony: instrumental variables reveal causality and control measurement error *Bruce E. Kendall* 149
7.1 Introduction to instrumental variables 149
7.2 Endogeneity and its consequences 151
7.2.1 Sources of endogeneity 152
7.2.2 Effects of endogeneity propagate to other variables 154
7.3 The solution: instrumental variable regression 154
7.3.1 Simultaneous equation models 158
7.4 Life-history trade-offs in Florida scrub-jays 158
7.5 Other issues with instrumental variable regression 161
7.6 Deciding to use instrumental variable regression 163
7.7 Choosing instrumental variables 165
7.8 Conclusion 167

8 Structural equation modeling: building and evaluating causal models *James B. Grace, Samuel M. Scheiner, and Donald R. Schoolmaster, Jr.* 168
8.1 Introduction to causal hypotheses 168
8.1.1 The need for SEM 168
8.1.2 An ecological example 169
8.1.3 A structural equation modeling perspective 171
8.2 Background to structural equation modeling 173
8.2.1 Causal modeling and causal hypotheses 173

8.2.2 Mediators, indirect effects, and conditional independence 174
8.2.3 A key causal assumption: lack of confounding 175
8.2.4 Statistical specifications 175
8.2.5 Estimation options: global and local approaches 176
8.2.6 Model evaluation, comparison, and selection 178
8.3 Illustration of structural equation modeling 179
8.3.1 Overview of the modeling process 179
8.3.2 Conceptual models and causal diagrams 180
8.3.3 Classic global-estimation modeling 181
8.3.4 A graph-theoretic approach using local-estimation methods 186
8.3.5 Making informed choices about model form and estimation method 190
8.3.6 Computing queries and making interpretations 193
8.3.7 Reporting results 196
8.4 Discussion 197

9 Research synthesis methods in ecology *Jessica Gurevitch and Shinichi Nakagawa* 200

9.1 Introduction to research synthesis 200
9.1.1 Generalizing from results 200
9.1.2 What is research synthesis? 201
9.1.3 What have ecologists investigated using research syntheses? 201
9.1.4 Introduction to worked examples 202
9.2 Systematic reviews: making reviewing a scientific process 203
9.2.1 Defining a research question 204
9.2.2 Identifying and selecting papers 204
9.3 Initial steps for meta-analysis in ecology 204
9.3.1 What not to do 205
9.3.2 Data: What do you need, and how do you get it? 205
9.3.3 Software for meta-analysis 207
9.3.4 Exploratory data analysis 207
9.4 Conceptual and computational tools for meta-analysis 210
9.4.1 Effect size metrics 210
9.4.2 Fixed, random and mixed models 210
9.4.3 Heterogeneity 211
9.4.4 Meta-regression 213
9.4.5 Statistical inference 213
9.5 Applying our tools: statistical analysis of data 214
9.5.1 Plant responses to elevated CO_2 214
9.5.2 Plant growth responses to ectomycorrhizal (ECM) interactions 220
9.5.3 Is there publication bias, and how much does it affect the results? 221
9.5.4 Other sensitivity analyses 222
9.5.5 Reporting results of a meta-analysis 223
9.6 Discussion 224
9.6.1 Objections to meta-analysis 224
9.6.2 Limitations to current practice in ecological meta-analysis 226
9.6.3 More advanced issues and approaches 226

10 Spatial variation and linear modeling of ecological data *Simoneta Negrete-Yankelevich and Gordon A. Fox* 228
10.1 Introduction to spatial variation in ecological data 228
10.2 Background 232
10.2.1 Spatially explicit data 232
10.2.2 Spatial structure 232
10.2.3 Scales of ecological processes and scales of studies 236
10.3 Case study: spatial structure of soil properties in a *milpa* plot 237
10.4 Spatial exploratory data analysis 238
10.5 Measures and models of spatial autocorrelation 239
10.5.1 Moran's *I* and correlograms 240
10.5.2 Semi-variance and the variogram 242
10.6 Adding spatial structures to linear models 246
10.6.1 Generalized least squares models 247
10.6.2 Spatial autoregressive models 250
10.7 Discussion 259

11 Statistical approaches to the problem of phylogenetically correlated data *Marc J. Lajeunesse and Gordon A. Fox* 261
11.1 Introduction to phylogenetically correlated data 261
11.2 Statistical assumptions and the comparative phylogenetic method 262
11.2.1 The assumptions of conventional linear regression 263
11.2.2 The assumption of independence and phylogenetic correlations 265
11.2.3 What are phylogenetic correlations and how do they affect data? 266
11.2.4 Why are phylogenetic correlations important for regression? 272
11.2.5 The assumption of homoscedasticity and evolutionary models 278
11.2.6 What happens when the incorrect model of evolution is assumed? 280
11.3 Establishing confidence with the comparative phylogenetic method 281
11.4 Conclusions 283

12 Mixture models for overdispersed data *Jonathan R. Rhodes* 284
12.1 Introduction to mixture models for overdispersed data 284
12.2 Overdispersion 286
12.2.1 What is overdispersion and what causes it? 286
12.2.2 Detecting overdispersion 288
12.3 Mixture models 289
12.3.1 What is a mixture model? 289
12.3.2 Mixture models used in ecology 292
12.4 Empirical examples 293
12.4.1 Using binomial mixtures to model dung decay 293
12.4.2 Using Poisson mixtures to model lemur abundance 299
12.5 Discussion 306

13 Linear and generalized linear mixed models *Benjamin M. Bolker* 309
13.1 Introduction to generalized linear mixed models 309
13.2 Running examples 310

13.3 Concepts 311
13.3.1 Model definition 311
13.3.2 Conditional, marginal, and restricted likelihood 319
13.4 Setting up a GLMM: practical considerations 322
13.4.1 Response distribution 322
13.4.2 Link function 323
13.4.3 Number and type of random effects 323
13.5 Estimation 323
13.5.1 Avoiding mixed models 324
13.5.2 Method of moments 324
13.5.3 Deterministic/frequentist algorithms 324
13.5.4 Stochastic/Bayesian algorithms 325
13.5.5 Model diagnostics and troubleshooting 326
13.5.6 Examples 327
13.6 Inference 328
13.6.1 Approximations for inference 328
13.6.2 Methods of inference 329
13.6.3 Reporting the GLMM results 331
13.7 Conclusions 333

Appendix 335
Glossary 345
References 354
Index 379