Contents

Preface *page* xiii
List of Acronyms xviii

1 Introduction 1
1.1 Almost Every Biological Theory or Hypothesis Is a Model of How Nature Works 1
1.2 We Use Data to Separate Wrong Models from (Possibly) Correct Ones (or Bad Models from Less Bad Ones) 2
1.2.1 What Kind of Data Do We Need? 2
1.2.2 The Signal and the Noise 2
1.2.3 Random and Representative 2
1.2.4 How Do We Decide If a Model Fits Well? What's the "Best" Model? 3
1.2.5 What Do I Do Next? 3
1.3 There Are Many Ways to Reach Wrong Conclusions! 3
1.4 There Are Also Many Ways to be Right 4
1.5 Our Philosophy in This Book 4
1.5.1 Think Clearly 4
1.5.2 Think in Advance 4
1.5.3 Think before You Analyze 5
1.6 How This Book Is Structured 5
1.7 A Bit of Housekeeping 6

2 Things to Know before Proceeding 8
2.1 Samples, Populations, and Statistical Inference 8
2.2 Probability 9
2.3 Probability Distributions 9
2.3.1 Distributions for Variables 10
2.3.2 Distributions for Statistics 11
2.4 Frequentist ("Classical") Estimation 11
2.4.1 Methods for Estimation 12
2.4.2 Simple Parameters and Statistics 13
2.4.3 Sampling Distribution of the Mean 15
2.4.4 Standard Error of the Sample Mean 15
2.4.5 Confidence Intervals for Population Mean 16
2.4.6 Standard Errors and Confidence Intervals for Other Statistics 16
2.4.7 Resampling Methods for Frequentist Estimation 17
2.5 Hypothesis Testing 18
2.5.1 Frequentist Statistical Hypothesis Testing 18
2.5.2 Decision Errors 20
2.5.3 One- and Two-Tailed Tests 21
2.5.4 Multiple Hypothesis Testing 21

2.5.5 Testing Hypotheses about Means and Variances for Two Populations 23
2.5.6 Parametric Tests and Their Assumptions 23
2.6 Comments on Frequentist Inference 24
2.7 Bayesian Statistical Inference 25
2.7.1 Prior Knowledge and Probability 27
2.7.2 Likelihood Function 27
2.7.3 Posterior Probability 27
2.7.4 Model Comparison and Bayes Factors 28
2.7.5 Final Comments 28

3 Sampling and Experimental Design 30
3.1 Sampling Design 30
3.1.1 Probability Sampling 30
3.1.2 Sample Size for Random Sampling 32
3.1.3 Nonprobability Sampling 33
3.2 Experimental Design 33
3.2.1 Replication of Experimental Units 34
3.2.2 Controls 36
3.2.3 Randomization 36
3.2.4 Independence 38
3.2.5 Reducing Unexplained Variance 38
3.2.6 Limitations of Manipulative Experiments 38
3.3 Sample Size for Detecting Differences: Power Analysis 39
*3.3.1 Using Power to Plan Experiments (*a priori *Power Analysis)* 39
3.3.2 Post Hoc *Power Calculation* 42
3.3.3 Effect Size 42
3.3.4 Using Power Analyses 43
3.4 Key Points 44

4 Introduction to Linear Models 45
4.1 What Is a Linear Model? 45
4.2 Components of Linear Models 46
4.2.1 Types of Response Variables 46
4.2.2 Types of Predictor Variables 48
4.3 Assembling our Linear Model 48
4.4 Estimation for Linear Models 49
4.4.1 Ordinary Least Squares 51
4.4.2 Maximum Likelihood 53
4.4.3 Robust Estimation Methods for Linear Models 54
4.5 How Well Does a Model Fit? 54
4.5.1 OLS Measures of Fit: ANOVA 54
4.5.2 ML Measures of Fit: Log-Likelihood and Deviance 55
4.5.3 Information Criteria 55
4.6 Assumptions for Linear Model Inference 55
4.6.1 Assumptions for OLS 55
4.6.2 Assumptions for ML 57
4.6.3 Model Diagnostics 57
4.7 Types of Linear Models 57
4.7.1 General Linear Models 57
4.7.2 Generalized Linear Models 59

4.7.3 Linear Mixed Models (General and Generalized) 60
4.8 Key Points 61

5 Exploratory Data Analysis 62
5.1 Basic Graphical Tools 62
5.1.1 Some Common Basic Graphs 62
5.1.2 Smoothing 66
5.1.3 Residual Plots 68
5.2 Outliers 69
5.3 Am I Fitting the Right Model? 70
5.3.1 The Underlying Probability Distribution 70
5.3.2 Homogeneity of Variances 70
5.3.3 Is My Linear Model "Linear"? 71
5.4 Is It Normal to Transform Data? 71
5.4.1 Transformations and Distributional Assumptions 72
5.4.2 Transformations and Linearity 72
5.4.3 Transformations and Additivity 73
5.4.4 Do We Really Need a Data Transformation? 73
5.5 Standardizations 73
5.6 Missing Data 73
5.6.1 Missing Data Mechanisms 74
5.6.2 Detecting Missing Data 74
5.6.3 Methods for Missing Data 74
5.7 Key Points 75

6 Simple Linear Models with One Predictor 76
6.1 Linear Model for a Single Continuous Predictor: Linear Regression 76
6.1.1 Linear Model for a Continuous Predictor (Linear Regression Model) 76
6.1.2 Model Parameters 81
6.1.3 Inference for Parameters 83
6.1.4 Inference for Predicted Values 83
6.1.5 Model Comparison and the Analysis of Variance 83
6.1.6 Regression Through the Origin 85
6.1.7 Regression with X Random 85
6.2 Linear Model for a Single Categorical Predictor (Factor) 88
6.2.1 Experimental vs. Observational Studies 88
6.2.2 Linear Model for a Categorical Predictor 88
6.2.3 Model Parameters 90
6.2.4 Inference for Parameters 93
6.2.5 Model Comparison and the Analysis of Variance 95
6.2.6 Unequal Sample Sizes (Unbalanced Designs) 96
6.2.7 Specific Comparisons of Group Means 97
6.3 Predictor Effects 101
6.3.1 Continuous Predictor (Regression) Models 101
6.3.2 Categorical Predictor Models 101
6.4 Assumptions 103
6.4.1 Zero Conditional Mean 103
6.4.2 Independence 103
6.4.3 Variance Homogeneity 103
6.4.4 Normality 103

6.5 Model Diagnostics 103
6.5.1 *Residuals* 103
6.5.2 *Leverage* 104
6.5.3 *Influence Measures* 105
6.5.4 *Diagnostic Plots* 105
6.5.5 *Transformations* 105
6.6 Robust Linear Models 107
6.6.1 *Rank-Based ("Nonparametric") Methods* 108
6.6.2 *Generalized (Weighted) Least Squares* 109
6.6.3 *Other Robust Methods* 109
6.6.4 *Resampling and Permutation Methods* 110
6.7 Power of Single-Predictor Linear Models 110
6.7.1 *Regression Models* 111
6.7.2 *Categorical Predictor Models* 111
6.8 Key Points 112

7 Linear Models for Crossed (Factorial) Designs 115
7.1 Two-Factor Fully Crossed (Factorial) Designs 115
7.1.1 *Completely Randomized (Experimental) Designs* 115
7.1.2 *Observational (Nonexperimental) Designs* 115
7.1.3 *Designs That Combine Completely Randomized Factors with Nonrandomized (Observational) Factors* 118
7.1.4 *The Factorial Linear Effects Model* 118
7.1.5 *Model Parameters* 121
7.1.6 *Inference for Parameters* 121
7.1.7 *Model Comparison and Analysis of Variance* 123
7.1.8 *More on Main Effects and Interactions* 126
7.1.9 *Interactions and Transformations* 127
7.1.10 *Specific Comparisons of Marginal Means* 127
7.1.11 *Interpreting Interactions* 128
7.1.12 *Predictor Effects* 128
7.1.13 *Assumptions* 129
7.1.14 *Robust Factorial ANOVAs* 129
7.2 Complex Factorial Designs 129
7.2.1 *Missing Cells* 130
7.2.2 *Fractional Factorial Designs* 134
7.3 Power and Sample Size in Factorial Designs 135
7.4 Key Points 136

8 Multiple Regression Models 137
8.1 Linear Model for Multiple Continuous Predictors: Multiple Regression 137
8.1.1 *The Multiple Linear Regression Model* 137
8.1.2 *Model Parameters* 139
8.1.3 *Inference for Parameters* 144
8.1.4 *Model Comparison and Analysis of Variance* 145
8.1.5 *Assumptions of Multiple Linear Regression Models* 147
8.1.6 *Model Diagnostics* 147
8.1.7 *Diagnostic Graphics* 148
8.1.8 *Transformations* 148
8.1.9 *Collinearity* 148

8.1.10 Interactions in Multiple Regression 151
8.1.11 Regression Models with Polynomial Terms 154
8.1.12 Other Issues in Multiple Linear Regression 156
8.1.13 Categorical Predictors in Multiple Regression Models 158
8.2 Analysis of Covariance 158
8.2.1 Linear Models for Simple Analyses of Covariance 160
8.2.2 Model Comparison and the Analysis of (Co)variance 165
8.2.3 Assumptions of ANCOVA Models 166
8.2.4 Homogeneous Within-Group Regression Slopes 167
8.2.5 Robust ANCOVA 170
8.2.6 Unequal Sample Sizes (Unbalanced Designs) 170
8.2.7 Specific Comparisons of Adjusted Means 170
8.2.8 Factorial Designs 170
8.2.9 Designs with Two or More Covariates 172
8.3 Key Points 173

9 Predictor Importance and Model Selection in Multiple Regression Models 174
9.1 Relative Predictor Importance 174
9.1.1 Single Model Methods 174
9.1.2 Multiple Model Methods 181
9.1.3 Recommendations of Relative Importance 182
9.2 Model Selection 182
9.2.1 Model Selection Criteria 183
9.2.2 Traditional Stepwise Selection 184
9.2.3 All Subsets and Information Criteria 184
9.2.4 Model Averaging 185
9.2.5 Model Validation 186
9.3 Regression Trees 187
9.3.1 Standard Regression Trees 187
9.3.2 Bagging and Boosted Regression Trees 189
9.4 Key Points 193

10 Random Factors in Factorial and Nested Designs 194
10.1 Fixed vs. Random Effects and Mixed Models 194
10.1.1 Designs Applicable to Mixed Models 194
10.2 Fitting Linear Models with Fixed and Random Factors 199
10.2.1 Traditional OLS "ANOVA" Models Approach 200
10.2.2 Linear Mixed Effect (or Multilevel) Models Approach 201
10.2.3 Modeling Strategies 204
10.3 Simple Random Factor Designs 205
10.3.1 Traditional OLS Approach 206
10.3.2 Linear Mixed Effects (Multilevel) Models 207
10.4 Multilevel Regressions 207
10.5 Nested (Hierarchical) Designs 209
10.5.1 Two-Level Nested Designs 210
10.5.2 More Complex Nested Designs 214
10.5.3 Sample Size and Nested Designs 215
10.6 Crossed (Factorial) Mixed Designs 215
10.6.1 Types of Factorial Mixed Designs 216
10.6.2 Analysis of Crossed Designs with One Fixed and One Random Factor 217

10.6.3 Crossed Designs with Two or More Fixed Factors and One Random Factor 223
10.6.4 Design and Analysis Issues with Crossed Mixed Designs and Their Models 223
10.7 Key Points 227

11 Split-Plot (Split-Unit) Designs: Partly Nested Models 228
11.1 Simple Split-Plot Designs 228
11.1.1 Analysis for Simple Split-Plot Designs 230
11.1.2 Assumptions 235
11.1.3 Unbalanced Split-Plot Designs 236
11.1.4 Model Building 236
11.2 More Complex Designs 236
11.2.1 Additional Between-Plots Factors 236
11.2.2 Additional Within-Plots Factors 239
11.2.3 Including Continuous Covariates 239
11.3 Key Points 241

12 Repeated Measures Designs 242
12.1 Simple Repeated Measures Designs 243
12.1.1 Analysis of Simple Repeated Measures Designs 243
12.1.2 Assumptions for Simple Repeated Measures Models 247
12.1.3 Missing Observations 249
12.2 More Complex Repeated Measures Designs 249
12.2.1 One Between-Subjects Factor 249
12.2.2 Two or More Between-Subjects Factors 250
12.2.3 Two or More Within-Subjects Factors 256
12.2.4 Model Selection in Complex Repeated Measures Designs 256
12.3 Key Points 256

13 Generalized Linear Models for Categorical Responses 257
13.1 Logistic Regression 258
13.1.1 Binary Response with a Single Continuous Predictor: Simple Logistic Regression 258
13.1.2 Categorical Predictors in GLMs 258
13.1.3 Binary Response with Multiple Predictors: Multiple Logistic Regression 260
13.1.4 Nominal and Ordinal Multinomial Response Variables 260
13.1.5 Proportion Response 260
13.2 Count Responses: Poisson Regression 260
13.3 Goodness-of-Fit for GLMs 263
13.4 Inference for Parameters in GLMs 265
13.5 Assumptions and Diagnostics for Binomial and Poisson GLMs 265
13.6 Overdispersed Data 266
13.6.1 Identifying Overdispersion 267
13.6.2 Correcting for Overdispersion 267
13.6.3 Too Many Zeroes (Zero-Inflated Data) 268
13.6.4 Zero-Truncated Data 268
13.6.5 Binomial Overdispersion 268
13.7 Contingency Tables 269
13.7.1 Two-Way Tables 269
13.7.2 Three-Way and Higher Tables 274
13.7.3 More Complex Tables 279
13.7.4 Loglinear vs. Logistic Models for Tables 279

13.8 Generalized Linear Mixed Models 279
13.9 Generalized Additive Models 280
13.10 Key Points 283

14 Introduction to Multivariate Analyses 285
14.1 Distributions and Associations 286
14.2 Linear Combinations, Eigenvectors, and Eigenvalues 287
14.2.1 Linear Combinations of Variables 287
14.2.2 Eigenvalues 287
14.2.3 Eigenvectors 287
14.2.4 Derivation of Components 287
14.3 Multivariate Distance and Dissimilarity Measures 291
14.3.1 Dissimilarity Indices for Continuous and Count Variables 292
14.3.2 Dissimilarity Indices for Dichotomous (Binary) Variables 292
14.3.3 General Dissimilarity Indices for Mixed Variables 294
14.3.4 Choosing Dissimilarity Indices 294
14.4 Data Transformation and Standardization 295
14.5 Standardization, Association, and Dissimilarity 296
14.6 Screening Multivariate Datasets 296
14.7 Introduction to Multivariate Analyses 297
14.8 Key Points 298

15 Multivariate Analyses Based on Eigenanalyses 299
15.1 Principal Components Analysis 299
15.1.1 Deriving Components 299
15.1.2 Interpreting the Components 300
15.1.3 How Many Components to Retain? 305
15.1.4 Which Association Matrix to Use? 305
15.1.5 Simplifying Component Structure 306
15.1.6 PCA Assumptions and "Fit" 306
15.1.7 Ordination and Biplots for PCA 307
15.1.8 Principal Components Regression 308
15.1.9 Factor Analysis 308
15.2 Correspondence Analysis 309
15.2.1 Deriving the Axes 309
15.2.2 Ordination and Biplots for CA 311
15.2.3 Reciprocal Averaging 311
15.3 Use of PCA and CA with Ecological Abundance (Count) Data 311
15.4 Constrained (Canonical) Multivariate Analysis 312
15.4.1 Redundancy Analysis 312
15.4.2 Canonical Correspondence Analysis 315
15.5 Linear Discriminant Function Analysis 316
15.5.1 Deriving Discriminant Functions 316
15.5.2 Classification and Prediction 319
15.5.3 Assumptions of Discriminant Function Analysis 320
15.5.4 Multivariate Analysis of Variance 320
15.6 Key Points 320

16 Multivariate Analyses Based on (Dis)similarities or Distances 322
16.1 Multidimensional Scaling or Ordination 322
16.1.1 Metric (Classical) Scaling: Principal Coordinates Analysis 323
16.1.2 Nonmetric (Enhanced) Multidimensional Scaling 323
16.2 Cluster Analysis 329
16.2.1 Agglomerative Hierarchical Clustering 329
16.2.2 Divisive Hierarchical Clustering 330
16.2.3 Nonhierarchical Clustering 330
16.3 Analyses Based on Dissimilarities 331
16.3.1 Contributions of Original Variables to Ordination 331
16.3.2 Relating Dissimilarities to Other Variables 331
16.3.3 Multivariate Linear Models 334
16.4 Key Points 335

17 Telling Stories with Data 337
17.1 Research Doesn't Exist Until You Tell Someone 337
17.1.1 Telling Better Stories: The Importance of Narrative 337
17.2 Summarizing Data Analyses 338
17.2.1 Linear Models 338
17.2.2 Other Analyses 341
17.3 Visualizing Data 341
17.3.1 Just Show Us the Numbers! 343
17.3.2 Tables 343
17.4 Graphical Summaries of the Data 343
17.4.1 Some Basic Principles for Visualizing Data 344
17.4.2 An Appropriate Visual Display 350
17.5 Error Bars: Visualizing Variation and Precision 359
17.5.1 Possible Solutions 360
17.6 Horses for Courses: What You Present Depends on Who's Listening 360
17.6.1 Know Your Audience as Well as Possible 363
17.7 Software and Other Sources 364
17.8 Key Points 364

Glossary 365
References 367
Index 383

Color plates can be found between pages 364 and 365.