

A Primer of *Ecological Statistics*

Second Edition



Nicholas J. Gotelli *University of Vermont*

Aaron M. Ellison *Harvard Forest*



Sinauer Associates, Inc. Publishers
Sunderland, Massachusetts U.S.A.

Brief Contents

PART I

FUNDAMENTALS OF PROBABILITY AND STATISTICAL THINKING

- 1 *An Introduction to Probability* 3
- 2 *Random Variables and Probability Distributions* 25
- 3 *Summary Statistics: Measures of Location and Spread* 57
- 4 *Framing and Testing Hypotheses* 79
- 5 *Three Frameworks for Statistical Analysis* 107

PART II

DESIGNING EXPERIMENTS

- 6 *Designing Successful Field Studies* 137
- 7 *A Bestiary of Experimental and Sampling Designs* 163
- 8 *Managing and Curating Data* 207

PART III

DATA ANALYSIS

- 9 *Regression* 239
- 10 *The Analysis of Variance* 289
- 11 *The Analysis of Categorical Data* 349
- 12 *The Analysis of Multivariate Data* 383

PART IV

ESTIMATION

- 13 *The Measurement of Biodiversity* 449
 - 14 *Detecting Populations and Estimating their Size* 483
- Appendix** *Matrix Algebra for Ecologists* 523

Contents

PART I

Fundamentals of Probability and Statistical Thinking



CHAPTER 1

An Introduction to Probability 3

What Is Probability? 4

Measuring Probability 4

- The Probability of a Single Event:
Prey Capture by Carnivorous Plants 4
- Estimating Probabilities by Sampling 7

Problems in the Definition of Probability 9

The Mathematics of Probability 11

- Defining the Sample Space 11
- Complex and Shared Events: Combining Simple Probabilities 13
- Probability Calculations: Milkweeds and Caterpillars 15
- Complex and Shared Events: Rules for Combining Sets 18
- Conditional Probabilities 21

Bayes' Theorem 22

Summary 24

CHAPTER 2

Random Variables and Probability Distributions 25

Discrete Random Variables 26

- Bernoulli Random Variables 26
- An Example of a Bernoulli Trial 27
- Many Bernoulli Trials = A Binomial Random Variable 28
- The Binomial Distribution 31
- Poisson Random Variables 34
- An Example of a Poisson Random Variable:
Distribution of a Rare Plant 36

The Expected Value of a Discrete Random Variable	39
The Variance of a Discrete Random Variable	39
Continuous Random Variables	41
Uniform Random Variables	42
The Expected Value of a Continuous Random Variable	45
Normal Random Variables	46
Useful Properties of the Normal Distribution	48
Other Continuous Random Variables	50
The Central Limit Theorem	53
Summary	54

CHAPTER 3

Summary Statistics: Measures of Location and Spread 57

Measures of Location	58
The Arithmetic Mean	58
Other Means	60
Other Measures of Location: The Median and the Mode	64
When to Use Each Measure of Location	65
Measures of Spread	66
The Variance and the Standard Deviation	66
The Standard Error of the Mean	67
Skewness, Kurtosis, and Central Moments	69
Quantiles	71
Using Measures of Spread	72
Some Philosophical Issues Surrounding Summary Statistics	73
Confidence Intervals	74
Generalized Confidence Intervals	76
Summary	78

CHAPTER 4

Framing and Testing Hypotheses 79

Scientific Methods	80
Deduction and Induction	81
Modern-Day Induction: Bayesian Inference	84
The Hypothetico-Deductive Method	87
Testing Statistical Hypotheses	90
Statistical Hypotheses versus Scientific Hypotheses	90
Statistical Significance and P -Values	91
Errors in Hypothesis Testing	100
Parameter Estimation and Prediction	104
Summary	105

CHAPTER 5

Three Frameworks for Statistical Analysis 107

Sample Problem	107
Monte Carlo Analysis	109
Step 1: Specifying the Test Statistic	111
Step 2: Creating the Null Distribution	111
Step 3: Deciding on a One- or Two-Tailed Test	112
Step 4: Calculating the Tail Probability	114
Assumptions of the Monte Carlo Method	115
Advantages and Disadvantages of the Monte Carlo Method	115
Parametric Analysis	117
Step 1: Specifying the Test Statistic	117
Step 2: Specifying the Null Distribution	119
Step 3: Calculating the Tail Probability	119
Assumptions of the Parametric Method	120
Advantages and Disadvantages of the Parametric Method	121

Non-Parametric Analysis: A Special Case of Monte Carlo Analysis 121

Bayesian Analysis 122

Step 1: Specifying the Hypothesis 122

Step 2: Specifying Parameters as Random Variables 125

Step 3: Specifying the Prior Probability Distribution 125

Step 4: Calculating the Likelihood 129

Step 5: Calculating the Posterior Probability Distribution 129

Step 6: Interpreting the Results 130

Assumptions of Bayesian Analysis 132

Advantages and Disadvantages of Bayesian Analysis 133

Summary 133

PART II

Designing Experiments

CHAPTER 6

Designing Successful Field Studies 137

What Is the Point of the Study? 137

Are There Spatial or Temporal Differences in Variable Y ? 137

What Is the Effect of Factor X on Variable Y ? 138

Are the Measurements of Variable Y Consistent with the Predictions of Hypothesis H ? 138

Using the Measurements of Variable Y , What Is the Best Estimate of Parameter θ in Model Z ? 139

Manipulative Experiments 139

Natural Experiments 141

Snapshot versus Trajectory Experiments 143

The Problem of Temporal Dependence 144

Press versus Pulse Experiments 146

Replication 148

How Much Replication? 148

How Many Total Replicates Are Affordable? 149

The Rule of 10 150

Large-Scale Studies and Environmental Impacts 150

Ensuring Independence 151

Avoiding Confounding Factors 153

Replication and Randomization 154

Designing Effective Field Experiments and Sampling Studies 158

Are the Plots or Enclosures Large Enough to Ensure Realistic Results? 158

What Is the Grain and Extent of the Study? 158

Does the Range of Treatments or Census Categories Bracket or Span the Range of Possible Environmental Conditions? 159

Have Appropriate Controls Been Established to Ensure that Results Reflect Variation Only in the Factor of Interest? 160



Have All Replicates Been Manipulated in the Same Way Except for the Intended Treatment Application? 160

Have Appropriate Covariates Been Measured in Each Replicate? 161

Summary 161

CHAPTER 7

A Bestiary of Experimental and Sampling Designs 163

Categorical versus Continuous Variables 164

Dependent and Independent Variables 165

Four Classes of Experimental Design 165

Regression Designs 166

ANOVA Designs 171

Alternatives to ANOVA: Experimental Regression 197

Tabular Designs 200

Alternatives to Tabular Designs: Proportional Designs 203

Summary 204

PART III

Data Analysis

CHAPTER 9

Regression 239

Defining the Straight Line and Its Two Parameters 239

Fitting Data to a Linear Model 241

Variances and Covariances 244

CHAPTER 8

Managing and Curating Data 207

The First Step: Managing Raw Data 208

Spreadsheets 208

Metadata 209

The Second Step: Storing and Curating the Data 210

Storage: Temporary and Archival 210

Curating the Data 211

The Third Step: Checking the Data 212

The Importance of Outliers 212

Errors 214

Missing Data 215

Detecting Outliers and Errors 215

Creating an Audit Trail 223

The Final Step: Transforming the Data 223

Data Transformations as a Cognitive Tool 224

Data Transformations because the Statistics Demand It 229

Reporting Results: Transformed or Not? 233

The Audit Trail Redux 233

Summary: The Data Management Flow Chart 235



Least-Squares Parameter Estimates	246
Variance Components and the Coefficient of Determination	248
Hypothesis Tests with Regression	250
The Anatomy of an ANOVA Table	251
Other Tests and Confidence Intervals	253
Assumptions of Regression	257
Diagnostic Tests For Regression	259
Plotting Residuals	259
Other Diagnostic Plots	262
The Influence Function	262
Monte Carlo and Bayesian Analyses	264
Linear Regression Using Monte Carlo Methods	264
Linear Regression Using Bayesian Methods	266
Other Kinds of Regression Analyses	268
Robust Regression	268
Quantile Regression	271
Logistic Regression	273
Non-Linear Regression	275
Multiple Regression	275
Path Analysis	279
Model Selection Criteria	282
Model Selection Methods for Multiple Regression	283
Model Selection Methods in Path Analysis	284
Bayesian Model Selection	285
Summary	287
 CHAPTER 10	
<hr/> <i>The Analysis of Variance</i> 289	
Symbols and Labels in ANOVA	290
ANOVA and Partitioning of the Sum of Squares	290
The Assumptions of ANOVA	295
Hypothesis Tests with ANOVA	296
Constructing F-Ratios	298
A Bestiary of ANOVA Tables	300
Randomized Block	300
Nested ANOVA	302
Two-Way ANOVA	304
ANOVA for Three-Way and n -Way Designs	308
Split-Plot ANOVA	308
Repeated Measures ANOVA	309
ANCOVA	314
Random versus Fixed Factors in ANOVA	317
Partitioning the Variance in ANOVA	322
After ANOVA: Plotting and Understanding Interaction Terms	325
Plotting Results from One-Way ANOVAs	325
Plotting Results from Two-Way ANOVAs	327
Understanding the Interaction Term	331
Plotting Results from ANCOVAs	333
Comparing Means	335
A Posteriori Comparisons	337
A Priori Contrasts	339
Bonferroni Corrections and the Problem of Multiple Tests	345
Summary	348
 CHAPTER 11	
<hr/> <i>The Analysis of Categorical Data</i> 349	
Two-Way Contingency Tables	350
Organizing the Data	350
Are the Variables Independent?	352
Testing the Hypothesis: Pearson's Chi-square Test	354
An Alternative to Pearson's Chi-Square: The G -Test	358

- The Chi-square Test and the G-Test for $R \times C$ Tables 359
- Which Test To Choose? 363
- Multi-Way Contingency Tables 364**
- Organizing the Data 364
- On to Multi-Way Tables! 368
- Bayesian Approaches to Contingency Tables 375
- Tests for Goodness-of-Fit 376**
- Goodness-of-Fit Tests for Discrete Distributions 376
- Testing Goodness-of-Fit for Continuous Distributions: The Kolmogorov-Smirnov Test 380
- Summary 382**
- CHAPTER 12**
-
- The Analysis of Multivariate Data 383*
- Approaching Multivariate Data 383**
- The Need for Matrix Algebra 384
- Comparing Multivariate Means 387**
- Comparing Multivariate Means of Two Samples: Hotelling's T^2 Test 387
- Comparing Multivariate Means of More Than Two Samples: A Simple MANOVA 390
- The Multivariate Normal Distribution 394**
- Testing for Multivariate Normality 396
- Measurements of Multivariate Distance 398**
- Measuring Distances between Two Individuals 398
- Measuring Distances between Two Groups 402
- Other Measurements of Distance 402
- Ordination 406**
- Principal Component Analysis 406
- Factor Analysis 415
- Principal Coordinates Analysis 418
- Correspondence Analysis 421
- Non-Metric Multidimensional Scaling 425
- Advantages and Disadvantages of Ordination 427
- Classification 429**
- Cluster Analysis 429
- Choosing a Clustering Method 430
- Discriminant Analysis 433
- Advantages and Disadvantages of Classification 437
- Multivariate Multiple Regression 438**
- Redundancy Analysis 438
- Summary 444**

PART IV

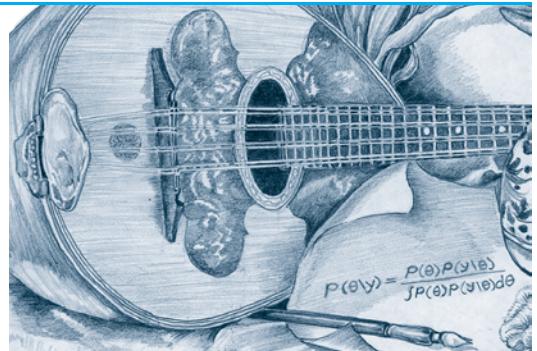
Estimation

CHAPTER 13

The Measurement of Biodiversity 449

Estimating Species Richness 450

- Standardizing Diversity Comparisons through Random Subsampling 453



Rarefaction Curves: Interpolating Species Richness 455	Occupancy of More than One Species 493
The Expectation of the Individual-Based Rarefaction Curve 459	A Hierarchical Model for Parameter Estimation and Modeling 495
Sample-Based Rarefaction Curves: Massachusetts Ants 461	Occupancy Models for Open Populations 501
Species Richness versus Species Density 465	Dynamic Occupancy of the Adelgid in Massachusetts 505
The Statistical Comparison of Rarefaction Curves 466	Estimating Population Size 506
Assumptions of Rarefaction 467	Mark-Recapture: The Basic Model 507
Asymptotic Estimators: Extrapolating Species Richness 470	Mark-Recapture Models for Open Populations 516
Rarefaction Curves Redux: Extrapolation and Interpolation 476	Occupancy Modeling and Mark-Recapture: Yet More Models 518
Estimating Species Diversity and Evenness 476	Sampling for Occupancy and Abundance 519
Hill Numbers 479	Software for Estimating Occupancy and Abundance 521
Software for Estimation of Species Diversity 481	Summary 522
Summary 482	APPENDIX
CHAPTER 14	<hr/>
<i>Detecting Populations and Estimating their Size 483</i>	<i>Matrix Algebra for Ecologists 523</i>
Occupancy 485	<i>Glossary 535</i>
The Basic Model: One Species, One Season, Two Samples at a Range of Sites 487	<i>Literature Cited 565</i>
	<i>Index 583</i>