

Preface

What is the “added value” for readers in purchasing this Second Edition? For this Second Edition, we have added a new Part 4, on “Estimation,” which includes two chapters: “The Measurement of Biodiversity” (Chapter 13), and “Detecting Populations and Estimating their Size” (Chapter 14). These two new chapters describe measurements and methods that are beginning to be widely used and that address questions that are central to the disciplines of community ecology and demography. Some of the methods we describe have been developed only in the last 10 years, and many of them continue to evolve rapidly.

The title of this new section, “Estimation,” reflects an ongoing shift in ecology from testing hypotheses to estimating parameters. Although Chapter 3 describes summary measures of location and spread, most of the focus of the First Edition—and of many ecological publications—is on hypothesis testing (Chapter 4). In the two new chapters, we are more concerned with the process of estimation itself, although the resulting estimators of biodiversity and population size certainly can be used in conventional statistical tests once they are calculated. It is useful to think of biodiversity and population sizes as latent or unknown “state variables.” We want to measure these things, but it is impossible to find and tag all of the trout in a stream or discover all of the species of ants that lurk in a bog. Instead, we work only with small samples of diversity, or limited collections of marked animals, and we use these samples to estimate the “true” value of the underlying state variables. The two new chapters are united not only by their focus on estimation, but also by their common underlying statistical framework: methods for asymptotic estimators of species richness discussed in Chapter 13 were derived from methods developed for mark-recapture studies discussed in Chapter 14. We hope the methods described in these two chapters will provide some fresh statistical tools for ecologists to complement the more classical topics we cover in Chapters 1–12.

Chapter 13 begins with what seems like a relatively simple question: how many species are there in an assemblage? The central problem is that nature is

(still) vast, our sample of its diversity is tiny, and there are always many rare and undetected species. Moreover, because the number of species we count is heavily influenced by the number of individuals and samples we have collected, the data must be standardized for meaningful comparisons. We describe methods for both interpolation (rarefaction) and extrapolation (asymptotic species richness estimators) to effectively standardize biodiversity comparisons. We also advocate the use of Hill numbers as a class of diversity indices that have useful statistical properties, although they too are subject to sampling effects. We thank Anne Chao, Rob Colwell, and Lou Jost for their ongoing collaboration, as well as their insights, extended correspondence, and important contributions to the literature on biodiversity estimation.

In Chapter 14, we delve deeper into the problem of incomplete detection, addressing two main questions. First, how do we estimate the probability that we can detect a species, given that it is actually present at a site? Second, for a species that is present at a site, how large is its population? Because we have a limited sample of individuals that we can see and count, we must first estimate the detection probability of an individual species before we can estimate the size of its population. Estimates of occupancy probability and population size are used in quantitative models for the management of populations of both rare and exploited species. And, just as we use species richness estimators to extrapolate beyond our sample, we use estimates from occupancy models and mark-recapture studies to extrapolate and forecast what will happen to these populations as, for example, habitats are fragmented, fishing pressure increases, or the climate changes. We thank Elizabeth Crone, Dave Orwig, Evan Preisser, and Rui Zhang for letting us use some of their unpublished data in this chapter and for discussing their analyses with us. Bob Dorazio and Andy Royle discussed these models with us; they, along with Evan Cooch, have developed state-of-the-art open-source software for occupancy modeling and mark-recapture analysis.

In this edition, we have also corrected numerous minor errors that astute readers have pointed out to us over the past 8 years. We are especially grateful for the careful reading of Victor Lemes Landeiro, who, together with his colleagues Fabricio Beggiano Baccaro, Helder Mateus Viana Espirito Santo, Miriam Plaza Pinto, and Murilo Sversut Dias, translated the entire book into Portuguese. Specific contributions of these individuals, and all the others who sent us comments and pointed out errors, are noted in the Errata section of the book's website (harvardforest.fas.harvard.edu/ellison/publications/primer/errata). Although we have corrected all errors identified in the first edition, others undoubtedly remain (another problem in detection probability); please contact us if you find any.

All code posted on the Data and Code section of the book's website (harvardforest.fas.harvard.edu/ellison/publications/primer/datafiles) that we used

for the analyses in the book has been updated and reworked in the R programming language (r-project.org). In the time since we wrote the first edition, R has become the de facto standard software for statistical analysis and is used by most ecologists and statisticians for their day-to-day analytical work. We have provided some R scripts for our figures and analyses, but they have minimal annotation, and there are already plenty of excellent resources available for ecological modeling and graphics with R.

Last but not least, we thank Andy Sinauer, Azelie Aquadro, Joan Gemme, Chris Small, Randy Burgess, and the entire staff at Sinauer Associates for carefully editing our manuscript pages and turning them into an attractive book.¹

Preface to the First Edition

Why another book on statistics? The field is already crowded with texts on biometry, probability theory, experimental design, and analysis of variance. The short answer is that we have yet to find a single text that meets the two specific needs of ecologists and environmental scientists. The first need is a general introduction to probability theory including the assumptions of statistical inference and hypothesis testing. The second need is a detailed discussion of specific designs and analyses that are typically encountered in ecology.

This book reflects several collective decades of teaching statistics to undergraduate and graduate students at the University of Oklahoma, the University of Vermont, Mount Holyoke College, and the Harvard Forest. The book represents our personal perspective on what is important at the interface of ecology and statistics and the most effective way to teach it. It is the book we both wish we had in our hands when we started studying ecology in college.

A “primer” suggests a short text, with simple messages and safe recommendations for users. If only statistics were so simple! The subject is vast, and new methods and software tools for ecologists appear almost monthly. In spite of the book’s length (which stunned even us), we still view it as a distillation of statistical knowledge that is important for ecologists. We have included both classic expositions as well as discussions of more recent methods and techniques. We hope there will be material that will appeal to both the neophyte and the experienced researcher. As in art and music, there is a strong element of style and personal preference in the choice and application of statistics. We have explained our own preferences for certain kinds of analyses and designs, but readers and users of statistics will ultimately have to choose their own set of statistical tools.

¹And for dealing with all of the footnotes!

How To Use This Book

In contrast to many statistics books aimed at biologists or ecologists—"biometry" texts—this is not a book of "recipes," nor is it a set of exercises or problems tied to a particular software package that is probably already out of date. Although the book contains equations and derivations, it is not a formal statistical text with detailed mathematical proofs. Instead, it is an exposition of basic and advanced topics in statistics that are important specifically for ecologists and environmental scientists. It is a book that is meant to be read and used, perhaps as a supplement to a more traditional text, or as a stand-alone text for students who have had at least a minimal introduction to statistics. We hope this book will also find a place on the shelf (or floor) of environmental professionals who need to use and interpret statistics daily but who may not have had formal training in statistics or have ready access to helpful statisticians.

Throughout the book we make extensive use of footnotes. Footnotes give us a chance to greatly expand the material and to talk about more complex issues that would bog down the flow of the main text. Some footnotes are purely historical, others cover mathematical and statistical proofs or details, and others are brief essays on topics in the ecological literature. As undergraduates, we both were independently enamored of Hutchinson's classic *An Introduction to Population Biology* (Yale University Press 1977). Hutchinson's liberal use of footnotes and his frequent forays into history and philosophy served as our stylistic model.

We have tried to strike a balance between being concise and being thorough. Many topics recur in different chapters, although sometimes in slightly different contexts. Figure and table legends are somewhat expansive because they are meant to be readable without reference to the text. Because Chapter 12 requires matrix algebra, we provide a brief Appendix that covers the essentials of matrix notation and manipulation. Finally, we have included a comprehensive glossary of terms used in statistics and probability theory. This glossary may be useful not only for reading this book, but also for deciphering statistical methods presented in journal articles.

Mathematical Content

Although we have not shied away from equations when they are needed, there is considerably less math here than in many intermediate level texts. We also try to illustrate all methods with empirical analyses. In almost all cases, we have used data from our own studies so that we can illustrate the progression from raw data to statistical output.

Although the chapters are meant to stand alone, there is considerable cross-referencing. Chapters 9–12 cover the traditional topics of many biometry texts and contain the heaviest concentration of formulas and equations. Chapter 12,

on multivariate methods, is the most advanced in the book. Although we have tried to explain multivariate methods in plain English (no easy task!), there is no way to sugar-coat the heavy dose of matrix algebra and new vocabulary necessary for multivariate analysis.

Coverage and Topics

Statistics is an evolving field, and new methods are always being developed to answer ecological questions. This text, however, covers core material that we believe is foundational for any statistical analysis. The book is organized around design-based statistics—statistics that are to be used with designed observational and experimental studies. An alternative framework is model-based statistics, including model selection criteria, likelihood analysis, inverse modeling, and other such methods.

What characterizes the difference between these two approaches? In brief, design-based analyses primarily address the problem of $P(\text{data} \mid \text{model})$: assessing the probability of the data given a model that we have specified. An important element of design-based analysis is being explicit about the underlying model. In contrast, model-based statistics address the problem of $P(\text{model} \mid \text{data})$ —that is, assessing the probability of a model given the data that are present. Model-based methods are often most appropriate for large datasets that may not have been collected according to an explicit sampling strategy.

We also discuss Bayesian methods, which somewhat straddle the line between these two approaches. Bayesian methods address $P(\text{model} \mid \text{data})$, but can be used with data from designed experiments and observational studies.

The book is divided into three parts. Part I discusses the fundamentals of probability and statistical thinking. It introduces the logic and language of probability (Chapter 1), explains common statistical distributions used in ecology (Chapter 2), and introduces important measures of location and spread (Chapter 3). Chapter 4 is more of a philosophical discussion about hypothesis testing, with careful explanations of Type I and Type II errors and of the ubiquitous statement that “ $P < 0.05$.” Chapter 5 closes this section by introducing the three major paradigms of statistical analysis (frequentist, Monte Carlo, and Bayesian), illustrating their use through the analysis of a single dataset.

Part II discusses how to successfully design and execute observational studies and field experiments. These chapters contain advice and information that is to be used before any data are collected in the field. Chapter 6 addresses the practical problems of articulating the reason for the study, setting the sample size, and dealing with independence, replication, randomization, impact studies, and environmental heterogeneity. Chapter 7 is a bestiary of experimental designs. However, in contrast to almost all other texts, there are almost no equations in this

chapter. We discuss the strengths and weaknesses of different designs without getting bogged down in equations or analysis of variance (ANOVA) tables, which have mostly been quarantined in Part III. Chapter 8 addresses the problem of how to curate and manage the data once it is collected. Transformations are introduced in this chapter as a way to help screen for outliers and errors in data transcription. We also stress the importance of creating metadata—documentation that accompanies the raw data into long-term archival storage.

Part III discusses specific analyses and covers the material that is the main core of most statistics texts. These chapters carefully develop many basic models, but also try to introduce some more current statistical tools that ecologists and environmental scientists find useful. Chapter 9 develops the basic linear regression model and also introduces more advanced topics such as non-linear, quantile, and robust regression, and path analysis. Chapter 10 discusses ANOVA (analysis of variance) and ANCOVA (analysis of covariance) models, highlighting which ANOVA model is appropriate for particular sampling or experimental designs. This chapter also explains how to use *a priori* contrasts and emphasizes plotting ANOVA results in a meaningful way to understand main effects and interaction terms. Chapter 11 discusses categorical data analysis, with an emphasis on chi-square analysis of one- and two-way contingency tables and log-linear modeling of multi-way contingency tables. The chapter also introduces goodness-of-fit tests for discrete and continuous variables. Chapter 12 introduces a variety of multivariate methods for both ordination and classification. The chapter closes with an introduction to redundancy analysis, a multivariate analog of multiple regression.

About the Cover

The cover is an original still life by the ecologist Elizabeth Farnsworth, inspired by seventeenth-century Dutch still-life paintings. Netherlands painters, including De Heem, Claesz, and Kalf, displayed their skill by arranging and painting a wide range of natural and manmade objects, often with rich allusions and hidden symbolism. Elizabeth's beautiful and biologically accurate renderings of plants and animals follow a long tradition of biological illustrators such as Dürer and Escher. In Elizabeth's drawing, the centerpiece is a scrolled pen-and-paper illustration of Bayes' Theorem, an eighteenth-century mathematical exposition that lies at the core of probability theory. Although modern Bayesian methods are computationally intensive, an understanding of Bayes' Theorem still requires pen-and-paper contemplation.

Dutch still lifes often included musical instruments, perhaps symbolizing the Pythagorean beauty of music and numbers. Elizabeth's drawing incorporates Aaron's bouzouki, as the writing of this book was interwoven with many sessions of acoustic music with Aaron and Elizabeth.

The human skull is the traditional symbol of mortality in still life paintings. Elizabeth has substituted a small brain coral, in honor of Nick's early dissertation work on the ecology of subtropical gorgonians. In light of the current worldwide collapse of coral reefs, the brain coral seems an appropriate symbol of mortality. In the foreground, an *Atta* ant carries a scrap of text with a summation sign, reminding us of the collective power of individual ant workers whose summed activities benefit the entire colony. As in early medieval paintings, the ant is drawn disproportionately larger than the other objects, reflecting its dominance and importance in terrestrial ecosystems. In the background is a specimen of *Sarracenia minor*, a carnivorous plant from the southeastern U.S. that we grow in the greenhouse. Although *Sarracenia minor* is too rare for experimental manipulations in the field, our ongoing studies of the more common *Sarracenia purpurea* in the northeastern U.S. may help us to develop effective conservation strategies for all of the species in this genus.

The final element in this drawing is a traditional Italian Renaissance chicken carafe. In the summer of 2001, we were surveying pitcher plants in remote bogs of the Adirondack Mountains. At the end of a long field day, we found ourselves in a small Italian restaurant in southern Vermont. Somewhat intemperately, we ordered and consumed an entire liter-sized chicken carafe of Chianti. By the end of the evening, we had committed ourselves to writing this book, and "The Chicken"—as we have always called it—was officially hatched.

Acknowledgments

We thank several colleagues for their detailed suggestions on different chapters: Marti Anderson (11, 12), Jim Brunt (8), George Cobb (9, 10), Elizabeth Farnsworth (1–5), Brian Inouye (6, 7), Pedro Peres-Neto (11, 12), Catherine Potvin (9, 10), Robert Rockwell (1–5), Derek Roff (original book proposal), David Skelly (original book proposal), and Steve Tilley (1–5). Laboratory discussion groups at the Harvard Forest and University of Vermont also gave us feedback on many parts of the book. Special kudos go to Henry Horn, who read and thoroughly critiqued the entire manuscript. We thank Andy Sinauer, Chris Small, Carol Wigg, Bobbie Lewis, and Susan McGlew of Sinauer Associates, and, especially, Michele Ruschhaupt of The Format Group for transforming pages of crude word-processed text into an elegant and pleasing book.

The National Science Foundation supports our research on pitcher plants, ants, null models, and mangroves, all of which appear in many of the examples in this book. The University of Vermont and the Harvard Forest, our respective home institutions, provide supportive environments for book writing.