RHODORA, Vol. 106, No. 928, pp. 378–382, 2004

BOOK REVIEW

A Primer of Ecological Statistics by Nicholas J. Gotelli and Aaron M. Ellison. 2004. xviii + 510 pp. 103 illus. ISBN 0-87893-269-3 $34.95 (paperbound). Sinauer Assiocates, Inc., Sunderland, MA.

In the Preface, the authors state that their goal was to provide a single text that met the needs of ecologists and environmental scientists by providing a general introduction to probability theory, statistical inference, and hypothesis testing, plus a detailed discussion of experimental design and analyses typically encountered in ecological studies. It is intended 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" and the authors hope that it will "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."

The book is divided into three parts. Part I includes chapters on probability theory, random variables, probability distributions, summary statistics, framing and testing hypotheses, and a comparison of three "frameworks" for statistical analyses. Part II is devoted primarily to designing field studies and includes a fairly comprehensive set of experimental designs common to ecological research. It also includes a chapter on "Managing and Curating Data." Part III deals with the nuts and bolts of statistical analyses and includes chapters on regression, analysis of variance, analysis of categorical data (i.e., contingency tables) and multivariate analysis. The Appendix includes a section on "Matrix Algebra for Ecologists" plus a glossary that includes chapter references. The book is generously footnoted with a variety of explanatory and historical information. Figures are clear and include legends that are meant to be readable without reference to the text.

The authors acknowledge that while the term "primer" might suggest a short, simply explained introduction to the subject, the book is not short (510 pp.), nor is all of the covered material simple. The chapters are intended to stand alone and if you are looking for an easily readable section to warm up on, you might rather start with Part II, especially if you are less than fond of mathematical notation. If you choose to start with Part I, Chapter 1 jumps right into the mathematics of probability, incorporating set theory, set notation, Venn diagrams, and a number of definitions. The chapter concludes with a section on Bayes' Theorem explained primarily using Bayes' Formula. Much of the mathematic

notation in the chapter is not used extensively throughout the book. Considering the intended audience and in the spirit of a "primer," the material might better have been covered conceptually. If you are looking for more of a "how-to" approach to ecological statistics, you can safely skip most of this chapter. Chapter 2 introduces the idea of random variables and develops, conceptually and mathematically, the binomial, Poisson, and normal probability distributions. While these three distributions are the underpinnings of many statistical methods, the F-distribution and Chi^2 distributions, which are used extensively in the remainder of the book, are not covered. The chapter ends with an explanation of the Central Limit Theorem, "a cornerstone of probability and statistical analysis," but is missing an introduction to the distribution of sample means and its relationship to a population distribution. Chapter 3 is a fairly conventional treatment of measures of central tendency (arithmetic mean, mode, median, geometric mean) and dispersion (variance, standard deviation, standard error). The explanation of degrees of freedom is better than in most introductory texts and the discussion of skewness and kurtosis is excellent. Some less familiar statistics such as coefficient of dispersion are covered as well. There is a good discussion on when to use which statistic. Confidence intervals are covered with a clear explanation of what they mean, plus a comparison with credibility intervals used in Bayesian statistics. Chapter 4 is a formal discussion of the scientific method, deductive and inductive reasoning, and hypothesis development and testing. Helpful distinctions are made between statistical hypotheses and scientific hypotheses. Interpretation of p-values, the concepts of Type I and Type II errors, and statistical power are presented well. The chapter devotes several pages to the distinction between Bayesian Inference and Hypothetico-Deductive Methods which is interesting, but not essential to an understanding of the remainder of the book. The final chapter in Part I, titled "Three Frameworks for Statistical Analysis," works through a single sample problem using familiar parametric analysis (ANOVA in this case), Monte Carlo analysis, and Bayesian analysis. Similar comparisons are made in less detail later in the book. While the designs and analyses in Parts II and III are based primarily on parametric ("frequentist") methods, the book does a good job of exposing the reader to the concepts of Monte Carlo and Bayesian methods, which have become more prevalent with the availability of faster computers.

Part II, Designing Experiments, is by far the most readable part of the book. It is written in a more advisory style. This part alone may justify purchasing the book. Chapter 6 examines design considerations relevant

to ecological studies. It contrasts manipulative versus natural experiments, snapshot versus trajectory studies, and press versus pulse treatments. The authors make suggestions on replication including their "rule of 10" and tradeoffs with their "rule of 5." Chapter 7 is a collection of experimental and sampling designs that is initially broken down by the variable types (categorical versus continuous) of the independent and dependent variables. It differs from chapters in other Biostatistics books in that the mechanics of the analyses associated with each design are reserved for Part III. This makes the chapter readable and the concepts clear. General considerations are included for regression, ANOVA, logistic regression, and contingency table analysis. Designs discussed include one-way, RCB, nested, factorial, split-plot, and repeated measures designs. Special considerations for competitive experiments are discussed including additive, substitutive, and response surface approaches. Before-After, Control-Impact (BACI) designs are described for environmental impact studies. The authors suggest that in some cases, ANOVA may be an "intellectual straightjacket" causing scientists to neglect other useful designs. They suggest several Experimental Regression designs as preferable for certain applications. There is a general discussion of tabular designs and contingency tables. Chapter 8 provides useful general advice on managing and curating data. Much of this is common sense, but they make a strong case for keeping your data in good order. They stress the importance of transcribing field and lab data into spreadsheets immediately, constructing meta-data files to record critical information about the data (who, when, where it was collected, etc.). There is a section in the chapter on handling "outliers" in data sets, and some good ideas on using graphical exploratory data analysis for outlier detection. Data transformations are covered in this chapter as well.

Part III, Data Analysis, covers the mechanics of statistical analysis along with the underlying assumptions. Analyses are exemplified with data from some of the authors' studies. Chapter 9 is devoted to regression analysis. It starts with simple linear regression and provides a conceptual introduction plus methods and formulas for fitting the least-square regression line and calculating sums of squares and an F-ratio for hypothesis testing. There is a thorough section on confidence bands and prediction bands, plus an outstanding section on residual analysis. Conceptual coverage of Monte Carlo and Bayesian regression analyses is included. Cursory introductions to robust regression, logistic regression, multiple regression, and path analysis are included. Model selection methods for multiple regressions are covered including forward selection,

backward elimination, and stepwise methods. Chapter 10 covers ANOVA computational methods for the designs covered in Part II, Chapter 7. Although in practice these calculations are generally done using commercial statistical software, the authors stress the importance of understanding how the analyses work. For many designs, the default setting in statistical packages will not produce the correct analysis. This is especially true for some nested designs, split-plots, and mixed models including a combination of fixed and random treatments. There is a lot of information in this chapter but a few complete examples of some of the more complex designs would be helpful. Planned comparisons (orthogonal contrasts) and multiple comparison tests (Tukey's HSD) for pairwise and groupwise testing of means are explained well, although a number of alternative *a posteriori* tests are not covered. There is a spirited discussion in which the authors argue convincingly against the use of Bonferroni and Dunn-Sidak corrections for multiple analyses within a study. Advice is given on representing ANOVA and multiple comparison results for publication. It is not certain that the authors' suggestion to plot interaction between categorical treatments as line graphs will be universally embraced by journal editors. Most of us were taught that connecting means with a line implies a continuous treatment variable. Chapter 11 is devoted primarily to the analysis of counts and categorical data that are normally represented in contingency tables. Tests covered include Chi^2 Test of Independence, the G-Test, and Fisher's Exact Test. Criteria for when to use each test are discussed. Methods are supplied for calculating expected values for multi-way tables. Conceptual explanations for analyzing tables via Monte Carlo and Bayesian methods are given. A Goodness of Fit test is demonstrated using the fairness of a tossed Belgian Euro as an example. The Kolmogorov-Smirnov Goodness of Fit Test is demonstrated for testing for normality (or any other defined distribution). Chapter 12 provides an introduction to multivariate statistics. Explanations rely heavily on matrix algebra, and for those who need a refresher course in basic matrix operations, it is included in the Appendix. The chapter starts with Hotelling's T^2 Test, which is the multivariate equivalent of a T-Test. The authors walk the reader through a MANOVA, including calculations of common associated test statistics: Wilke's Lambda, Pillai's Trace, Hotelling-Lawley Trace, and Roy's Greatest Root. A short discussion of multiple comparisons is included. Ordination via PCA (Principal Components Analysis) is illustrated using one of the author's Cobra Lily morphological data sets. Factor Analysis, Principal Coordinate Analysis (PCoA), Correspondence Analysis (CA), and Multi-Dimensional Scaling are

briefly covered and guidance is given on how to choose between them. Where the goal of Ordination is to separate observations along the fewest axes, the goal of Classification is to group similar objects into identifiable groups that can be distinguished from neighboring classes. Agglomerative and divisive Cluster Analyses are compared. Hierarchical and non-hierarchical methods are discussed with K-Means analysis as an example of the latter. The section on Discriminant Analysis is good, but the snail data set used was not especially well suited. Two types of Multivariate Multiple Regression are discussed that are appropriate for examining associations between multivariate biological data and environmental data. Canonical Correspondence Analysis (CCA) is briefly explained, while Redundancy Analysis (RDA) is demonstrated using the less than ideal snail data set.

Overall, the book achieves the goals of the authors. It is a useful volume for graduate students and scientists involved in ecological field research. Part II is recommended for anyone planning a field study. Part III provides good explanations of statistical analyses that are regularly encountered in ecological work and journal articles.

—Christopher D. Neefus, Department of Plant Biology and Office of Biometrics, University of New Hampshire, Spaulding Hall, 38 College Rd., Durham, NH 03824.