Ecology, 86(3), 2005, pp. 810–811

ECOLOGICAL STATISTICS

Gotelli, Nicholas J., and Aaron M. Ellison. 2004. **A primer of ecological statistics.** Sinauer, Sunderland, Massachusetts. xviii + 510 p. $34.95, ISBN: 0-87893-269-0.

Key Words: Bayesian inference; experimental design; hypothesis; regression; statistics.

The task of teaching statistics to novice scientists is a daunting one. It is caught in the circularity that students without data in hand find little use for a tedious formality that seemingly only accidentally coincides with the closing stages of a research project, which results in students with data in hand either scratching their heads to recall long forgotten formulas or (worse) digging through an impenetrable literature to find a technique for salvaging data collected under less than ideal circumstances: low replication, unbalanced experimental designs, collinearity, heteroscedasticity, autocorrelation, and bias. Perhaps this new *Primer of ecological statistics* will convince students of the first sort to take seriously the task of collecting scientific data before the experimental or observational projects are begun; perhaps for the second sort it will rescue some term papers or theses. Perhaps for professors of environmental statistics it will render the task of teaching a little less daunting.

This book is divided into three parts. Part I introduces the basic ideas of probability and distributions of random variables (Chapters 1 and 2) and statistical concepts like location and spread (Chapter 3). Chapter 4 is standard fare on frequentist hypothesis testing with a too-hasty introduction to Bayesian inference thrown in for balance. The unfortunate effect is to put the reader slightly off-balance. It concludes with the authors' account of ''three frameworks'' for thinking about hypothesis testing: randomization tests, parametric hypothesis tests, and (again) Bayesian inference. Part II discusses experimental design and data management, emphasizing field experiments. This discussion is straightforward, but useful. Finally, Part III introduces four families of statistical tests for data analysis: regression; ANOVA; categorical data analyses; and multivariate analysis using MANOVA, ordination and classification techniques, and multiple regression. While Part III is a pretty standard catalog of parametric methodology, the first two parts especially are a welcome addition to the literature.

Because methodology exhibits an inherent gradient of sophistication, books on methodology must focus on a narrow audience, which always comprises just a slice of the possible readership. Books are written differently for novices and students, intermediate users, and the various grades of experts. Mixing audiences is disastrous, and it is attentive and devoted authors who know their audience well. The usual error is that some passages will bore the reader while others are tantalizingly elusive. Statistics is no exception to this rule, and neither is this primer. Nonetheless, its authors are to be commended for having produced overall an extremely readable and useful introductory textbook.

The gems are the chapters entitled ''Managing and curating data'' (Chapter 8) and ''Regression'' (Chapter 9). The first of these is brief, valuable, and unique. Though a discussion is taking place at higher levels about the practical aspects of data archiving and ethical discussions about data-sharing, data-mining, and data-dredging, students are rarely encouraged to think very deeply about the tasks of data management. Unfortunately, this practice results in a good deal of potentially valuable data being lost to the scientific community once it has been summarized and published; additionally, results from pilot experiments and student projects rarely become a part of the publicly accessible store of information.

The chapter on regression goes beyond the simple linear model. Regression is introduced in this way (as it should be—the simple linear model is the bread and butter of hypothesis testing), but then after the usual topics have been covered (estimation, variance components, hypothesis tests, assumptions, diagnostics) the chapter proceeds to introduce robust regression, quantile regression, logistic regression, nonlinear regression, multiple regression, and path analysis. The result is just what the instructor wishes to achieve: the reader is left with the impression that facility with regression is a useful and achievable skill (because of the detail in the first part) and not left with the naïve impression that simple regression exhausts the capabilities of the technique, but rather that a raft of other methods exists for the specialized situations in which the ecologist is commonly placed. Unfortunately, nonlinear relationships and non-constant variance seem more like the rule than the exception in ecology. A further noteworthy contribution of this chapter is the discussion of model selection, which, because it is presented in a general way, encourages the student to appreciate the ubiquity of the problem in data analysis. Model selection is not a specialized topic of simple regression, but is a general feature of the statistical task. The student comes away knowing this. The final great feature of this book is the emphasis throughout on the increasing usefulness of non-parametric statistics, particularly randomization techniques, which are presently underappreciated and consequently underused.

Some disappointments: I think that students are often trained to think of statistics as being a collection of idiosyncratic and conceptually disconnected techniques (e.g., regression, correlation, ANOVA, split plot design). If one's ideas about what information is to be collected happen to fit one of these molds then so much the better, because a more compelling story can be pieced together. But the student's general impression is that there is no coherent theory of statistics. This is backwards, and this book does little to correct this impression. Experimental design or data collection should begin with a thoughtful reflection on the research question to be answered (this part is emphasized in this book) and then a statistical procedure for inference should be derived to answer this question (this part is not emphasized). In general, data shouldn't be shoe-horned into existing structures just because the models are simple or because a particular piece of computer software will perform the analysis auto-

matically. Experiments should be constrained by the amount of data required to answer a question and not by a paucity of ways to represent a hypothesis. The reason that there is a standard set of parametric techniques is because there are deep theoretical relationships among the kinds of data that can be collected and equally deep connections among the different kinds of models that might be used to represent relationships among factors and between factors and observations. As we seek and teach unifying concepts in our science, so also we should seek and teach unifying concepts in our procedures for inference. The generalized linear model is one such unifying concept. Nonparametric regression is another. While the generalized linear model is referred to offhandedly, neither of these unifying concepts is discussed in detail. This is this book's greatest deficiency.

In conclusion, the conspicuous appearance of Bayesian inference throughout this book requires a comment. It is first introduced as an alternative ''paradigm'' for thinking about statistics and discussed in greater detail as one of the ''three frameworks.'' It comes up again in the chapter on regression and in the discussion of contingency tables. But in all of these places, the idea of prior and posterior probabilities versus frequencies is more confusing than enlightening. Indeed, in no case do the authors present the reader with enough information to implement the Bayesian methodology and the mystique surrounding Bayesian statistics is confirmed, rather than dispelled. The authors would do well to eliminate the references to Bayesian statistics in this volume entirely and instead write an additional primer. If that book should be as readable an introduction as this one, it would indeed be a valuable contribution.

JOHN M. DRAKE

National Center for Ecological Analysis and Synthesis
735 State Street, Suite 300
Santa Barbara, California 93101
E-mail: drake@nceas.ucsb.edu

Submit books and monographs for review to the Book Review Editor, Janet Lanza, Biology Department, University of Arkansas at Little Rock, Little Rock, Arkansas 72204 (telephone (501) 569-3500).

We welcome offers to review books for *Ecology*, but we cannot accept an offer to review a *specific* book. Anyone who wishes to review books for *Ecology* should send a current *curriculum vitae*, a description of competencies, and a statement of reviewing interests to the Book Review Editor. Authors of reviews must verify that they have no conflict of interest that might interfere with their objectivity, and that they have not offered (and will not offer) a review of the same book to another journal.

Spotlight

RECENT PUBLICATIONS OF PARTICULAR INTEREST

Darby, John, R. Ewan Fordyce, Alan Mark, Keith Probert, AND Colin Townsend, editors. 2003. **The natural history of southern New Zealand.** University of Otago Press, Dunedin, New Zealand. 387 p. $120.00, ISBN: 877133-51-5.

Young, David. 2004. **Our islands, our selves: a history of conservation in New Zealand.** University of Otago Press, Dunedin, New Zealand. 298 p. $59.95, ISBN: 1-877276-94-4.

Going to New Zealand? You should probably look at this pair of books. Never thought about going to New Zealand. This pair of books might tempt you. Both books have substantial text accompanied by intriguing photographs and drawings. The book by Darby et al. includes information on geology, climate, and the biota. The book by Young traces the history of land from the Maori to the modern society with its ecological problems and the way New Zealanders have developed their approach to conservation.