very low level (no mention of prediction intervals), methods based on k-nearest neighbors, classification and regression trees (six pages), and neural networks (17 pages). The mechanics of the latter are adequately described, but there is too much emphasis on arithmetic, and little effort is made to intuitively justify the prediction process. At the end of the chapter, there is one paragraph on multiple regression analysis and one sentence on logistic regression. A major disappointment to me was the almost exclusive reliance in the examples on a rather old automobile fuel efficiency dataset (there is one observation for a Datsun 1200 vehicle). I had hoped to see some real business applications.

After a four page "Deployment" chapter, the book ends with a "Conclusions" chapter containing one large-scale example involving data on the incidence of diabetes among Pima Indians. Here we find histograms, box plots, a two-sample t test, some derived associative rules which I did not find overly insightful, and a brief summary of prediction results via neural networks. I think students coming out of an undergraduate regression course could do a fine job of analyzing these data without resorting to the bells and whistles of data mining. And once again, where are the business applications?

Although the text does give a brief snapshot of the subject, it is lacking in detail, applications, and opportunities for practice. Someone considering becoming involved in a data mining project or teaching an introductory course in the subject would be advised to learn much more than what *MSD* offers. Good information sources are the much more ambitious books by Hastie, Tibshirani, and Friedman (2001) (the best-selling Springer statistics book ever, thanks to purchases by those outside our discipline) and Bishop (2006).

Jay DEVORE
California Polytechnic State University

REFERENCES

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, New York: Springer.

Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning*, New York: Springer.

Introduction to Computer-Intensive Methods of Data Analysis in Biology.

Derek A. ROFF. New York, NY: Cambridge University Press, 2006, vii+368 pp., $135.00 (H), ISBN 0-521-84628-5; $65.00 (P), ISBN: 0-521-60865-1.

In many disciplines, scientists are interested in estimating characteristics of the underlying population whose sampling distributions are unknown. In these situations, inferences have traditionally made use of elaborate assumptions in order to make analytical approaches tractable. In these nonstandard situations, more desirable analytical approaches involve such computationally intensive methods as bootstrapping, Monte Carlo methods, and Bayesian tools. Despite the many advances in statistical software, prevalence of these computationally intensive methods is still limited. The presentation of these computationally intensive methods is sparse at best outside of statistics graduate courses. Unfortunately, these methods are often absent in the classrooms where they may be most needed: the graduate service courses to nonstatistics majors.

This text attempts to make computationally intensive methods accessible to graduate students and researchers across the biological sciences. Although an audience with a biological background will be most interested in this material, due to the presented examples, this text is a valuable resource for anyone interested in computationally intensive statistics. The author's presentation of the material is meticulous in terms of organization and the use of well-defined notation. Nearly every method presented in the text has accompanying code in S-PLUS. Examples are interesting and most always involve real data. The topics are presented in delightfully simple and intuitive fashion. The author avoids getting bogged down in theoretical discussions and seems to have a knack for knowing when to provide formulas and when not to. In many other books where theory is avoided, one is left thinking that the methods are a "black-box." This is not the case here. Quite possibly one of the most appealing features of this text is the extensive collection (nearly 65 pages) of well-documented S-PLUS programs which exists in one of the appendixes.

There are six major topics in this text. The first chapter begins with an introduction to computer-intensive methods. The motivation for the need of these methods is nicely done. Chapter 2 provides a wonderful, applied coverage of point and interval estimation within the context of maximum likelihood. After laying the foundation for computational methods in Chapters 1 and 2, the text turns its attention to specific computational methods, including the jackknife (Chapter 3), the bootstrap (Chapter 4), randomization (permutation) and Monte Carlo methods (Chapter 5), nonparametric regression (LOESS, splines, generalized additive models and tree-based models) (Chapter 6), and finally, a chapter on Bayesian methods. My only wish is that the author would have provided more detail on LOESS, splines, and generalized additive models. The coverage of these regression topics was truly done at an "exposure" level.

The reader of this text needs a solid background in applied statistics (a rigorous one semester, graduate-level statistical methods course and a one-semester graduate course in regression). They should also feel comfortable with software packages that are not windows-based (I believe that R and S-PLUS require a more mature audience than, say, MINITAB, SAS JMP, or SYSTAT). If you have these tools and are a scientist or student who is interested in computationally intensive methods, this text will be a terrific resource. This text has inspired me to incorporate some computational methods into my own service courses.

Timothy J. ROBINSON
University of Wyoming

Dyadic Data Analysis.

David A. KENNY, Deborah A. KASHY, and William L. COOK. New York, NY: Guilford Publications, 2006, xix+458 pp., $52.00 (H), ISBN: 1-57230-986-5.

It is always interesting to read a statistical text written by and for practitioners of other disciplines. *Dyadic Data Analysis* is authored by professors of psychology and psychiatry and is aimed at social and behavioral researchers. As the title suggests, the book investigates the causes of paired data and the associated consequences of the resulting correlation. However, since the authors are not statisticians, their perspective is different as are their vocabulary and notational conventions. For instance, they use "nonindependence" instead of "dependence" to describe association between observations and denote model parameters by western letters instead of the Greek characters more familiar to statisticians.

Although the intended audience is not statisticians, it is clear that the authors expect a certain level of statistical sophistication from their readership. For example, they illustrate computation of Cohen's kappa (Cohen 1960), noting "[f]or any given category the expected number of agreements is the product of the number of times each judge chose the category divided by the total sample size" (p. 40). Although technically correct, this statement presupposes an understanding of joint probability, parameter estimation, and hypothesis testing which some readers may lack.

The book's scope is extensive with topics including hierarchical models (Chapter 4), social networks (Chapter 11) and longitudinal data (Chapters 13 and 14) to name a few. Yet, because the text addresses so many topics, its treatment of each is necessarily limited. Consequently, the authors often omit mathematical details from their exposition, opting instead for a high-level interpretation. Of particular value are their summary sections at the end of each chapter and their periodic digressions including "what not to do" and the consequences of various analytic missteps. On the whole, this decision benefits the practitioner seeking guidance, but it has its drawbacks. For example, seemingly inconsequential typographical errors (e.g., the use of a lower-case "ℓ" instead of the number "1" in the formula on page 29) may go unnoticed, potentially causing confusion or improper implementation.

Another example of the book's "big picture" philosophy is its discussion on testing equality of variances. Rather than test the hypothesis directly, the authors suggest a transformation from (X_1, X_2) to $(X_1 - X_2, X_1 + X_2)$ of which they remark "[a]lthough it is not obvious, the test of whether this correlation coefficient [of the transformed variables] equals zero evaluates a difference in the variances for X_1 and X_2" (p. 122). Though clever, this is stated without proof or intuition, merely a citation. While the book is to-the-point, its brevity, at times, makes it difficult to grasp the concepts behind formulas and procedures. Without this, the reader has very little chance of internalizing and assimilating the techniques or rederiving them from first principles.

Dyadic Data Analysis would not be my first choice of a companion text for a statistical methods course. Those looking for a focused, comprehensive development of a few topics will not find it here. Instead, they will find cursory treatment of many. And while reasonable people may disagree about the value of such an approach, applied researchers in need of a statistical users' manual of sorts may find the book's style and breadth of coverage useful and appealing.

David H. ANNIS
Wachovia Bank

REFERENCES

Cohen, J. (1960), "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, 20, 37–46.