Book and Software Reviews

Books for review are to be sent to the Book Review Editor Jaki Stanley McCarthy, USDA/NASS, Research Division, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A.

A.C. Davison and D.V. Hinkley. *Bootstrap Methods and Their Application.* Cambridge, UK.; Cambridge University Press, 1997. ISBN 0-521-57391 (hard). ISBN 0-521-574711 (paper) x+ 582 pp. + disk. £27.95.

By now most, if not all, of us are aware of the importance of the bootstrap as a statistical tool of great utility. Many texts include a brief introduction to the bootstrap but there are few comprehensive treatments. *Bootstrap Methods and Their Application* is not only comprehensive but also lucid and well organized.

I will concentrate in this review on topics likely to be of particular interest to *JOS* readers. Before I do, a brief overview is in order. (Not all the topics in the book are mentioned.) Chapter 1 is the ''Introduction.'' Chapter 2 is on ''The Basic Bootstraps,'' parametric and nonparametric, including simple confidence intervals and variance estimation. There is some discussion of the jackknife and half-sampling methods (balanced repeated replication). Chapter 3, ''Further Ideas,'' includes censoring, missing data, and finite population sampling. Chapter 4 is on ''Tests,'' and Chapter 5 is on ''Confidence Intervals.'' ''Linear Regression,'' including robust regression, is the topic of Chapter 6. ''Further Topics in Regression,'' such as nonlinear models, regression involving survival data, and nonparametric regression, are treated in Chapter 7. Chapter 8, ''Complex Dependence,'' is on times series and point processes. Balanced bootstraps and saddlepoint approximation are treated in Chapter 9, ''Improved Calculation.'' ''Semiparametric Likelihood Inference'' is the topic of Chapter 10. It includes a section on Bayesian bootstraps. The final chapter, Chapter 11, discusses ''Computer Implementation.'' Each of the ten sections of this chapter treats the computer implementation of the methods discussed in the corresponding previous chapter. There is a short appendix on cumulant calculations.

The book has an extensive set of problems that expand on the textual material and should not be overlooked. The book also has ''practicals,'' data-oriented problems that can be worked out using S-Plus programs. To facilitate this, a library of pertinent S-Plus functions developed by Angelo J. Canty is included with the book on diskette.

The library can also be downloaded at

                    http://dmawww.epfl.ch/davison.mosaic/BMA/library.html

where a Unix version and two Windows versions (one for S-Plus Version 3.2 or later and one efor Version 3.1) are provided. Although the practicals enhance the book, the reader not conversant with S-Plus would still be able to benefit. The ambitious reader could even work out the practicals in another programming language.

   The bibliographic notes near the end of Chapters 2 through 10 will be helpful to those wishing to study a particular topic in greater detail. There is an extensive set of references at the end of the book.

   Considering the nature of the topic, the book is remarkably free of typographical errors. Even so, an errata sheet can be downloaded at

                    http://dmawww.epfl.ch/davison.mosaic/BMA/

I am a survey sampler, so one of the first things I checked when I received the book was the coverage of finite population sampling. I was not disappointed. First of all, this topic is treated early in the book rather than being relegated to a miscellaneous topic section at the end. Although there is a concise but elegant discussion of half sampling near the end of Chapter 2, Section 3.7 on ''Finite Population Sampling,'' contains most of the material. The choice of the bootstrap sample size is discussed, as is the mirror-match bootstrap. The population and superpopulation bootstraps, each of which involves forming bootstrap *populations* (not just samples), are introduced and compared. A numerical example based on city population data is presented and used to illustrate the methods. The section concludes with a discussion of stratified sampling. More discussion of unequal probability sampling would have been welcome, but it may have been wise to wait until this area is better developed than it is now.

   I often have to deal with hierarchical data (e.g., data on teachers within schools), so I was pleased to see that this is the topic of Section 3.8. Two simple nonparametric strategies for bootstrapping appropriate to such data are presented. I also frequently encounter the problem of missing data, and there is a nice discussion of this in Section 3.6.

   Time series are important in official statistics applications, and the use of the bootstrap for time series occupies most of Chapter 8. The key idea is to bootstrap in *blocks*. Examples are included that use data on the water height of a river, the price of wool, and the occurrence of sunspots.

   Censored data and survival data are common in the health sciences. In Section 3.5, censored data are introduced and the estimation of the survivor function (essentially one minus the distribution function) and the cumulative hazard function is treated. Section 7.3 discusses situations in which explanatory variables are employed to study survival including uses of the proportional hazards model.

   The Bayesian bootstrap is treated in Section 10.5 (in the chapter on ''Semiparametric Likelihood Inference''). The Bayesian bootstrap provides one way of doing multiple imputation and is of interest for this reason. I would have liked even more coverage (it has about two pages), but the discussion is a good starting point.

   Although the book avoids proofs, it does not shy away from presenting important but tough material. It is less mathematical and proof-oriented than Hall (1992) and Shao

and Tu (1995), but more comprehensive than Efron and Tibshirani (1993). The very recent book by Chernick (1999) appears to be at about the level of Efron and Tibshirani. Someone seeking a really introductory treatment should consider Mooney and Duval (1993).

Bootstrap Methods and Their Application fills an important niche. The book is written with clarity and thoughtfulness. It covers many topics not found in other book-length treatments except the more mathematical Shao and Tu (1995). The book would be a valuable addition to the statistician's bookshelf.

References

Chernick, M.R. (1999). Bootstrap Methods: A Practitioner's Guide. New York: Wiley Interscience.

Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. New York: Chapman and Hall.

Hall, P. (1992). The Bootstrap and Edgeworth Expansion. New York: Springer-Verlag.

Mooney, C.Z. and Duval, R.D. (1993). Bootstrapping: A Nonparametric Approach to Statistical Inference. Newbury Park, California: Sage Publications.

Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap. New York: Springer-Verlag.

Michael P. Cohen
National Center for Education Statistics
555 New Jersey Avenue NW #402
Washington, DC 20208-5654
U.S.A.
Phone: (202) 219-1917
Fax: (202) 219-1736
e-mail: mcohen@inet.ed.gov

Edward Korn and Barry Graubard. *Analysis of Health Surveys*. New York: Wiley and Sons, 1999, 346 pp. ISBN 0471137731, 84.95 USD.

Medical survey research is increasing in both numbers and complexity. Statisticians find that the information they gained during training needs to be updated. Statisticians in the medical field are increasingly being called upon to help plan and analyze these surveys.

This book is excellent for that purpose. According to the authors, the purpose of the book is to describe methods for analyzing health surveys so that a reader who knows statistical methods for analyzing data in the nonsurvey setting can apply these methods appropriately to survey data.

The book contains two sections: Chapters 1–5 describe the theory and methods of analyzing surveys with a focus on health surveys; Chapters 6–9 deal with specific applications.

Chapter 1 is a basic introduction, while Chapter 2 goes into basic survey methods (including some sample plans), linearization, replication methods, and nonsampling errors. Chapter 3 deals with statistical analysis where it differs from nonsurvey analysis. (The following areas are covered: comparison of means, linear regression, ANOVA, logistic regression, survival analysis, predictive margins, and a discussion of subpopulations.)

Chapter 4 discusses sample weights and imputations, while Chapter 5 discusses variance estimation in special cases (such as one sampled unit in a strata, subpopulations, imputed values, and superpopulations). Chapters 6–9 basically take the reader through actual health survey applications to experience the particular issues of cross-sectional analysis, longitudinal analysis, multiple surveys, and case-controlled surveys.

This a well-written and well-planned book. The only criticisms are minor and are no doubt the result of having to balance thoroughness versus number of pages. 1) Terms are sometimes not defined or explained as fully as would be expected in a book for statisticians who are not current in the field of survey research (for example, superpopulation and pseudo-design). 2) Several times in the first five chapters, the authors suggest using simulation to check certain situations. While this may be second nature to some statisticians, others may have limited exposure to it. References would have been nice, but an actual example in the appendix would have been better.

On the positive side, 1) The authors did a good job of trying to make the chapters stand alone units, and they should be commended for planning the book in this manner. 2) Several sections had nice summations AT THE BEGINNING! This makes it easier on the casual reader, or a reader searching specific information. 3) Figures are spread throughout the book and referred to as needed. Locating the appropriate figure on the tables is easy, and a page number is provided along with the usual information. 4) The reference list is extensive and current. 5) The index is more thorough than usual, and the authors should be complemented on that as well.

The book fulfills the author's purpose of providing needed expertise for statisticians who are limited in their knowledge of survey research. There are excellent questions at the end of each chapter so it would make a good textbook for a course in survey analysis. However, no answer key is provided, either in the book or as a supplement. Therefore, the book can only be recommended for a classroom setting where the instructor is very knowledgeable in this field.

Barbara Foster, Dr.
Department of Academic Computing
UTSWMC
5323 Harry Hines Blvd.
Dallas, TX 75235-9066
U.S.A.
Phone: 214-648-3055
Fax: 214-648-7673
e-mail: barbara.foster@swmed.edu

Daniel Doring and Stephen Simpson, eds. *Statistics in Society: The Arithmetic of Politics*. Arnold: London, 1999. ISBN 0-340-71994-X, xxvi+ 484 pp. £16.99.

By a happy coincidence, this reviewer was also the author (*Journal of the American Statistical Association, 75* (1980), pp. 748–750) of a review of a much earlier companion book, *Demystifying Social Statistics*, edited by John Irvine, Ian Miles, and Jeff Evans. Both books are products of the Radical Statistics group of ''statisticians and others who share a common concern about the political assumptions implicit in the process of compiling and

using statistics, and in awareness of the actual and potential misuse of statistics and its techniques'' (Doring and Simpson, 1999, p. xxiv). The group was founded in 1975 and evolved from pre-World War II concern for the role of statistics as a tool for social change that gave rise in particular to the founding of the British Society for Social Responsibility in 1975. The current volume presents a substantial elaboration of ideas presented in the 1979 *Demystifying Social Statistics* book, and illuminates the myriad dimensions of the societal applications of statistics. *Statistics in Society* demonstrates the near-universal relevance of statistics to creating, explaining, addressing, and resolving societal realities. This is praxis – the unity of theory and practice – at its ultimate best, as demonstrated by the prefatory statement by David Godron and Alison Macfarlane: ''Only rational, democratic and progressive planning can tackle the manifest injustices of our present society and help the least 'powerful' groups to realize their full potential. Meaningful statistics are needed for this process. To paraphrase the old Marxist adage, the purpose of statistics in general and Radical Statistics in particular is not only to describe the world but also to change it'' (p. xxvi). The connection between official statistics and these laudable goals is very direct. As stated by Dr. John Martyn, of the Roehampton Institute, London: ''reliable official statistics are a cornerstone of democracy and are essential to good government and proper accountability'' (p. xxvi).

The scope of statistical praxis is both broad and deep, as demonstrated by an overview of the topics addressed in *Statistics in Society*. There are eight broadly defined sections to the 47 essays included in the book: (1) collecting statistics, (2) models and theory, (3) classifying people, (4) counting poverty, (5) valuing health, (6) assessing education, (7) measuring employment, and (8) economics and politics. The individual essays come to grips with more specific topics, including the census, government household surveys, confidentiality of official statistics, sources of funding for statistical research, government dissemination of official statistics, eugenics, statistical applications in psychology, ''barefoot statisticians,'' making official statistics more reachable by laypersons, the inherently subjective nature of social research, obfuscation of cultural differences, the social construction of reality, gender issues, legitimating the privatization of pensions, misleading categorizations of ethnicity and religion, eating habits, international migration, poverty and disabled children, income deprivation, homelessness, mental health, crime and the fear of crime, measuring health and health inequality, poverty and health, the hidden privatization of health and social care services, industrial injuries, comparative educational standards, performance indicators in education, trends in reading standards, schools inspection, special education, labor market participation, unemployment and employment, employment forecasting, historical statistics on poverty and economic distress, measuring the national economy, household projections, the militarism, information industry, electoral systems, meaningful electoral graphics, and the social construction of social statistics. Throughout all of these topical chapters, there is a consistent theme of demonstrating how statistics are used and have been used to foster one set of interests, and what might be done differently or better to serve broader, more humane purposes.

An appendix contains a list of Radical Statistics publications, broken down into books and pamphlets, broadsheets, journal articles, a health report, and the journal *Radical Statistics*. There is an impressively extensive 40-page references list and separate author and subject indexes, essential items that often are missing from other edited volumes.

The extensive coverage and the diversity of topics addressed in the *Statistics in Society*

volume preclude detailed examination here, but certain themes recur in multiple essays and in different contexts. First and foremost is the appeal for a redesigned social policy. This is demonstrated in sets of articles addressing the topics of homelessness, public health, education, and other pressing societal needs. The predominant message is of a (British) society that generally is not served well by methods for statistical data collection and analysis that tend to preserve the interests of society's rulers. A natural concomitant of this realization that the primary latent function of statistical procedures is the preservation of the societal status quo is the call to apply statistics in ways that might make the discipline much more societally relevant and more friendly to the needs of those whose interests are not well served by the existing controllers of the commanding heights of the economy and the political superstructure. The goal, after all, must be to change the world, not just to understand it. As noted by Stephen Simpson and Daniel Dorling in the concluding chapter: ''In a simple sense, statistics are a social product simply because they are produced by people. But they are also firmly located in the aims and tensions of the society that produces them – whether expressed by organisations or government, trade, or campaigns. In literature it is said that every text has a context. With statistics it is not just that what is discovered depends on the society from which those numbers are drawn. The methods and data that are used themselves are shaped by their context'' (p. 414). This reality is elaborated by the process by which statistics in society are ''paid for'': the data must have a purpose, must be assembled, require interpretation, and must be described and disseminated.

Inevitably, not all major societal applications of statistical methods can be addressed in a single compilation, no matter how many essays it contains nor how comprehensive it aspires to be. The military and its social effects is given fairly short shrift emerging research areas such as biotechnology or bioinfomatics are omitted; the political agendas underlying current statistical practices frequently are implicit rather than clearly drawn. This book would work quite well as a supplemental text in social science methods courses but would have somewhat limited utility in a more formal class offered in a statistics program. Nonetheless, this should be must reading for anyone who may yet adhere to the simplistic positivistic vision of a value-free science of society, and it provides a powerful support for those who aspire to make society into a better version of reality.

Mack Shelley
Department of Statistics
323 Snedecor Hall
Iowa State University
Ames, IA 50011-1210
U.S.A.
Phone: (515) 294-8346
Fax: (515) 294-4040
email: mshelley@iastate.edu