Journal of Official Statistics
Vol. 10. No. 2, 1994, pp. 225–228

Book Reviews

Books for review are to be sent to the Book Review Editor Gösta Forsman, Department of Mathematics, University of Linköping, S-581 83 Linköping, Sweden.

Vogt, W.P., Dictionary of Statistics and Methodology. A Nontechnical Guide for the Social Sciences. Sage Publications, Inc., London, U.K., 1993. ISBN 0–8039–5276–7, 0-8039-5277-5 (pbk) 270 pp., cloth £26.95, paper £14.95.

The author's intention, as stated in the preface, is to provide a handy reference book for social science students. Therefore, it is gratifying to find items here such as "Ecological Correlation" and "Ecological Fallacy," exemplified and with acceptable definitions. As regards two further items having to do with social science, "Contextual Effects" and "Snowball Sampling," I am not too happy that the former is given short shrift, while the example chosen to illustrate a possible application of the latter, viz., sample vegetarians, is certainly not the best I could think of.

To turn now to statistical terms in a more narrow sense, I am not too happy about the entry "Nonparametric Statistics." We are told that such techniques are "designed to be used when the data being analyzed depart from the distributions that can be analyzed with parametric statistics." Distribution-free methods, as is the preferred term, are not necessarily subordinate to something called "parametric statistics."

If I compare the present dictionary with what used to be Kendall and Buckland's *A Dictionary of Statistical Terms* (*DST*), now Marrot's 5th edition (1988), not surprisingly, I find the latter much to be preferred by students of statistics. For example, the following three items from DST are conspicuously absent in Vogt's dictionary: "Non-response," "Ridge Regression," and "Randomized Response."

Cross-references are, of course, a good thing in a dictionary of this kind. Sometimes, however, Vogt is too generous with his cross-references. Thus with the entry "Systematic Error," the reader is referred first to "Bias" and "Validity"; then to "Biased Estimates" and "Random Error"; and from there to "Reliability," and so on. By contrast, DST provides an admirably succinct definition of the term in a single entry.

I guess that, with the sub-title "Nontechnical Guide," we have to accept

a certain lack of precision and absence of technical terms. On the whole, however, the dictionary should serve a useful purpose for students of social science.

Sven Berg
Lund University
Lund
Sweden

Hedayat, A.S. and Sinha, B.K., Design and Inference in Finite Population Sampling. John Wiley & Sons, Inc., Somerset, NJ, U.S.A. ISBN 0–471–88073–6, xiii + 377pp, $59.95.

This book provides a mathematical treatment of various aspects of the design of and inferences for samples from finite populations. The book is clearly designed to be a textbook for graduate students. In the preface the authors indicate that "this book is an introduction to design and inference in survey sampling for a two-semester or three-quarter course at the senior or graduate level in statistics. By a judicious selection of topics, one could also teach a one-semester course based on this book." It clearly is appropriate for a course in a mathematical statistics department, as the emphasis is on the mathematics of the techniques, with passing references to practical issues. I would prefer to suggest that the book discusses finite population sampling, as the title indicates, rather than survey sampling, since apart from the material in Chapter 11 there is nothing specific to surveys in the book. I also do not think that this book is a suitable introduction to the area of sampling. It develops the topic from theoretical foundations rather than introducing simple concepts first and building from them. This is illustrated by the fact that the Horvitz-Thompson estimator is introduced in Chapter 3, *followed* in Chapter 4 by simple random sampling. Given the limitations of most graduate programs for course time, I would think a logical approach would be to offer an introductory course using a book such as Cochran (1977), and then use this book as the basis for a more advanced one-semester course.

The book consists of twelve chapters. Chapter 1 introduces concepts and notation for sample designs using a very general approach. First and second order inclusion probabilities are defined. This is complemented in Chapter 2 by the introduction of concepts and notation for estimation and inference. Throughout the book the discussion concerns the estimation of population totals and means, a limitation from the point of view of the practitioner, but no doubt necessary to cover the mathematics of the topics covered in a single volume. Chapter 3 discusses in detail the Horvitz-Thompson estimator (introduced in Chapter 2) for general sample designs, with particular emphasis on its variance and unbiased estimators of the variance. At this point the importance of having a sampling design with all pairwise joint inclusion probabilities being non-zero is introduced. This point is emphasized throughout the book, and might perhaps lead to the mistaken impression that such designs are the only ones useful in practice.

Having developed the characteristics of the Horvitz-Thompson estimator, the next two chapters cover sample designs. Chapter 4 gives a detailed treatment of simple random samples, and then Chapter 5 is a major chapter (58 pages) on the use of auxiliary information for developing various probability proportional to size sampling schemes. The chapter does well at providing a general approach to the topic, in addition to examining the properties of

some well known methods (Brewer-Durbin-Rao; Sunter; Lahiri-Midzuno; Sampford; Rao-Hartley-Cochran, etc.). Considerable attention is devoted to the Hedayat-Lin method, which the authors describe as a universal method, encompassing all sampling designs. It is based on the idea of emptying balls from boxes. I did not find the approach described especially insightful into the properties of various methods, but that is perhaps just personal taste; in any case the inclusion of this discussion is surely a unique feature of this book.

Chapter 6 covers the use of auxiliary data in estimation, examining the properties of ratio and regression estimation. Although both topics are covered with balance and depth, I consider that emphasis on the definition of sampling schemes and conditions under which these methods are unbiased to be of less practical importance than would be a consideration of the question of the relative size of the bias with more standard designs, and the effect of the bias on mean squared error and inference. The discussion of regression estimators is then used to introduce the topics of estimation in the case of double sampling and sampling on successive occasions.

Chapter 7 covers cluster sampling and two stage designs. Again the approach is logical and unified. Nevertheless, I had two quibbles. I would have liked to see more material on three stage designs. These designs raise more issues than just an extension of those for two stage designs, since the choice of sample design at each stage, and the questions of optimization across stages and the use of auxiliary data, become more complex. In discussing the choice of optimum allocation between stages in a two stage design, although using a classical cost model, the authors treat it as a discrete problem, rather than approximating a continuous solution. Although this is technically more sound, it does not provide insight into the population variance and cost structure factors that determine a desirable degree of clustering. In practice with large samples the optimum is flat, and the cost and variance components are specified with considerable error, so that using a continuous approximation to a discrete problem is more than adequate in most cases.

Chapter 8 covers systematic sampling. Of course the problem of having zero pairwise inclusion probabilities is prominent. It is interesting to contrast the approach adopted here with a book such as Wolter (1985, ch. 7) where the attention is on finding, for a given population, a systematic sampling scheme and a variance estimator such that the latter has small bias despite these zero joint probabilities. In this book the emphasis is on developing variations of systematic sampling that do not have this problem.

Chapter 9 covers stratified designs, which are relatively straightforward in comparison to the earlier material. This is especially so since only limited mention is made of the topics of poststratification, the number and formation of strata, and multiple variable issues. Furthermore, the question of when to use stratum-by-stratum ratio or regression estimation versus cross-stratum estimation is not covered.

Chapter 10 provides a nice introduction to the superpopulation approach to inference. Some different models are introduced and the main results on inference with respect to the superpopulation model are presented. The difference in objectives between finite population inference using the traditional randomization approach, compared to the superpopulation approach, is clearly indicated.

At first glance Chapter 11 looks out of place in a text such as this, for it covers

the topic of randomized response, used to obtain data on sensitive characteristics in surveys of persons. However, it is nice to see a fairly comprehensive treatment of the mathematics of the various forms of this technique. Introducing mathematical statisticians to issues in survey research, over and above sampling issues, seems like a good idea as well.

Chapter 12 covers some miscellaneous special topics. I found this to be the least satisfactory chapter. The section on small area estimation really deals with the topic of domain estimation, with references to the literature on small area estimation. The section on nonresponse problems mainly consists of a detailed treatment on the mathematics of a two phase approach, where a random subsample of the initial nonrespondents is surveyed (ah, that it were that easy!). The final section on replication techniques gives only a brief introduction to the methods and little insight into their use in practice – the novice reader could be excused for thinking that the main application of such techniques is for bootstrapping the variance of the mean from a simple random sample. Unlike the case for the other chapters, the references in Chapter 12, although quite extensive, are not comprehensive and generally are not those that come to my mind when thinking of useful and established references on these topics.

As indicated, the references in the book (apart from in Chapter 12) are a strength, being numerous, and providing further reading for both the topics covered in depth and those just alluded to. As a potential textbook, the exercises are an invaluable part of the book. There are many (227 in all), they complement the text material very effectively, and they are challenging while being relevant. Throughout the book there are consistently many examples. These are small, artificial examples designed to illustrate concepts and points of theory, rather than to demonstrate the application of methods in real practice. Considering the large number of complex formulas used, and the fact that this is a first edition, there are very few typographical errors.

This book does a strong, consistent job of presenting the mathematics of finite population sampling and inference, developed from fundamental principles. As such it fulfills a need satisfied by few, if any, other texts. I recommend the book for its intended purpose as a textbook for a graduate course in statistics.

References

Cochran, W.G. (1977). Sampling Techniques. Third Edition. Wiley: New York.

Wolter, K.M. (1985). Introduction to Variance Estimation. Springer-Verlag: New York.

Keith Rust
Westat,
Rockville, MD, U.S.A.

Printed in August